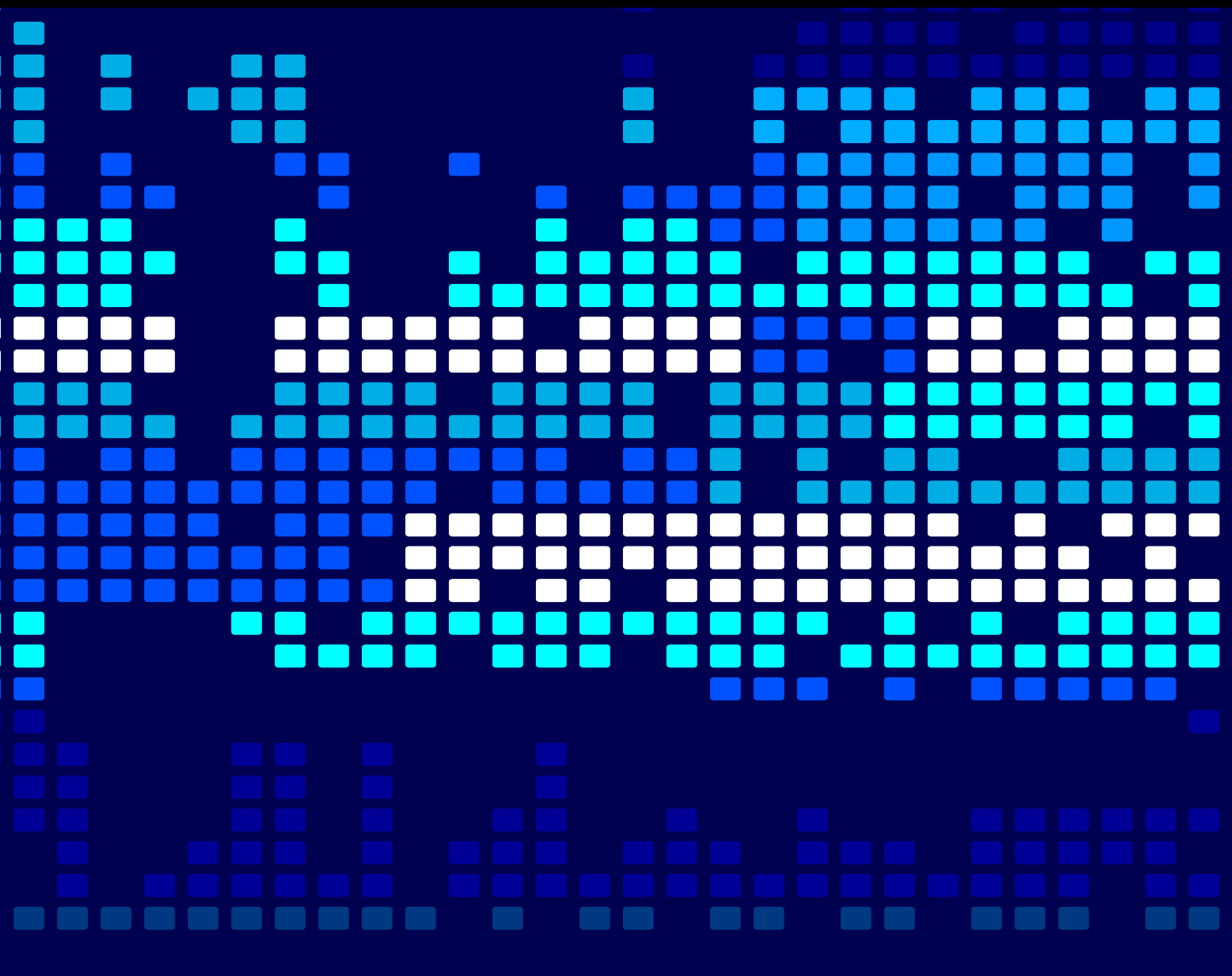
Scientific Programming

Optimization Models and Algorithms for Operation and Control with Advanced Information Technologies

Guest Editors: Junhu Ruan, Xuping Wang, Chengyan Yue, Guo Chen, and Minsoo Kim



Optimization Models and Algorithms for Operation and Control with Advanced Information Technologies

Scientific Programming

Optimization Models and Algorithms for Operation and Control with Advanced Information Technologies

Guest Editors: Junhu Ruan, Xuping Wang, Chengyan Yue, Guo Chen, and Minsoo Kim

Editorial Board

Contents

Optimization Models and Algorithms for Operation and Control with Advanced Information Technologies
Junhu Ruan, Xuping Wang, Chengyan Yue, Guo Chen, and Minsoo Kim
Volume 2017, Article ID 4706714, 2 pages

Assessing Partnership Alternatives in an IT Network Employing Analytical Methods
Vahid Reza Salamat, Alireza Aliahmadi, Mir Saman Pishvaee, and Khalid Hafeez
Volume 2016, Article ID 1319249, 18 pages

Modeling and Optimization of the Drug Extraction Production Process
Dakuo He, Zhengsong Wang, Le Yang, Tongshan Liu, Yao Yao, and Zhizhong Mao
Volume 2016, Article ID 3279423, 15 pages

Optimal Scheduling of Doctors Outpatient Departments Based on Patients' Behavior
Zongwei Ren, Haini Guo, Shizhen Bai, and Peng Li
Volume 2016, Article ID 6123103, 9 pages

Impact of Personnel Flexibility on Job Shop Scheduling
Ren Lin, Guohua Zhou, Aijun Liu, Hui Lu, and Tonglei Li
Volume 2016, Article ID 3403058, 8 pages

Exploring an Interactive Value-Adding Data-Driven Model of Consumer Electronics Supply Chain Based on Least Squares Support Vector Machine
Xiao-le Wan, Zhen Zhang, Xiao-xia Rong, and Qing-chun Meng
Volume 2016, Article ID 3717650, 13 pages

A Recovery Model for Production Scheduling: Combination of Disruption Management and Internet of Things
Yang Jiang, Qiulei Ding, and Xuhui Wang
Volume 2016, Article ID 8264879, 9 pages

A Novel Evaluation Approach for Tourist Choice of Destination Based on Grey Relation Analysis
Xudong Guo and Zilai Sun
Volume 2016, Article ID 1812094, 10 pages

Collaborative Product Design for Tasks Sorting Based on Shortest Delivery
Xuedong Liang, Wenting Zhou, Aijun Liu, and Hongmei Xue
Volume 2016, Article ID 9613293, 6 pages

Research on Healthy Anomaly Detection Model Based on Deep Learning from Multiple Time-Series Physiological Signals
Kai Wang, Youjin Zhao, Qingyu Xiong, Min Fan, Guotan Sun, Longkun Ma, and Tong Liu
Volume 2016, Article ID 5642856, 9 pages

Optimal Control for Bufferbloat Queue Management Using Indirect Method with Parametric Optimization
Amr Radwan, Hoang-Linh To, and Won-Joo Hwang
Volume 2016, Article ID 4180817, 10 pages

An Optimal Decision Assessment Model Based on the Acceptable Maximum LGD of Commercial Banks and Its Application
Baofeng Shi, Bin Meng, and Jing Wang
Volume 2016, Article ID 9751243, 9 pages

A Multiple Classifier Fusion Algorithm Using Weighted Decision Templates
Aizhong Mi, Lei Wang, and Junyan Qi
Volume 2016, Article ID 3943859, 10 pages

Intelligence in Ecology: How Internet of Things Expands Insights into the Missing CO_2 Sink
Wenfeng Wang, Xi Chen, Hongwei Zheng, Zhihan Lv, Zhengjia Liu, Jing Qian, and Ping Hu
Volume 2016, Article ID 4589723, 8 pages

Two-Phase Algorithm for Optimal Camera Placement
Jun-Woo Ahn, Tai-Woo Chang, Sung-Hee Lee, and Yong Won Seo
Volume 2016, Article ID 4801784, 16 pages

Event-Triggered Faults Tolerant Control for Stochastic Systems with Time Delays
Ling Huang, Xuhuan Xie, and Wenbo Xie
Volume 2016, Article ID 5606234, 13 pages

Identifying and Analyzing Strong Components of an Industrial Network Based on Cycle Degree
Zhiying Zhang, Xiaozhen Chen, Wenwen Xiao, and Guijie Qi
Volume 2016, Article ID 7340859, 11 pages

Acquisition Pricing and Inventory Decisions on Dual-Source Spare-Part System with Final Production and Remanufacturing
Yancong Zhou, Xudong Guo, and Xiaochen Sun
Volume 2016, Article ID 8038045, 10 pages

A Heuristics-Based Parthenogenetic Algorithm for the VRP with Potential Demands and Time Windows
Chenghua Shi, Tonglei Li, Yu Bai, and Fei Zhao
Volume 2016, Article ID 8461857, 12 pages

Hindawi Publishing Corporation
Scientific Programming
Volume 2017, Article ID 4706714, 2 pages
http://dx.doi.org/10.1155/2017/4706714

Editorial

Optimization Models and Algorithms for Operation and Control with Advanced Information Technologies

Junhu Ruan,[1,2] Xuping Wang,[2] Chengyan Yue,[3] Guo Chen,[4] and Minsoo Kim[5]

[1]*College of Economics and Management, Northwest A&F University, Yangling, China*
[2]*Faculty of Management and Economics, Dalian University of Technology, Dalian, China*
[3]*Department of Applied Economics, University of Minnesota, Minneapolis, MN, USA*
[4]*School of Electrical and Information Engineering, University of Sydney, Sydney, NSW, Australia*
[5]*Department of Systems Management and Engineering, Pukyong National University, Busan, Republic of Korea*

Correspondence should be addressed to Junhu Ruan; rjh@nwsuaf.edu.cn

Received 12 December 2016; Accepted 12 December 2016; Published 25 January 2017

In the literature, lots of optimization models and algorithms have been presented to deal with the operation and control issues in various domains such as intelligent manufacturing, production scheduling, transportation routing, service arrangement, and space exploration. Extant related works have promoted the development of conventional optimization and control theories, and quite a few of them have been widely used in the real world.

With the emergence of recent information technologies including Wireless Sensing Networks, Internet of Things (IOT), Cloud Computing, and Distributed Computing, people can obtain and process far more than what they could, generally within shorter time and with fewer costs. In addition, the application of the information technologies has been bringing about new challenges. One growing challenge is the Big Data problem. Although the Big Data problems and the like may provide chances for people to explore new rules and laws of the universe, they will become catastrophes if not well dealt with.

Thus, people should find new ways and tools in various domains to cope with the emerging challenges. Without exception, extant optimization models and algorithms need to be extended or reformulated for dealing with the reshaped operation and control issues by new information technologies.

This special issue is intended to present optimization models and algorithms for dealing with operation and control issues based on the emerging information technologies. It also aims at promoting the exchanges and interactions between investigators across different fields. We solicit high quality, original research articles as well as review articles focused on models and algorithms for operation and control issues with advanced information technologies.

We received 67 manuscripts in total, made strict preliminary and peer reviews for all the works, and finally accepted 18 papers for the special issue. The included papers can be grouped into five specific categories, that is, (1) models and algorithms with advanced information technologies, (2) healthcare related models and algorithms, (3) models and algorithms of production or delivery, (4) evaluation and classification models, and (5) control models and algorithms.

The first category (models and algorithms with advanced information technologies) includes four papers. The paper "Assessing Partnership Alternatives in an IT Network Employing Analytical Methods" by V. R. Salamat et al. presents an optimization model for the collaborative network building process in the IT industry. In the paper "Exploring an Interactive Value-Adding Data-Driven Model of Consumer Electronics Supply Chain Based on Least Squares Support Vector Machine," X. Wan et al. use the least squares support vector machine to develop a data-driven model of consumer electronics supply chains. The paper "Intelligence in Ecology: How Internet of Things Expands Insights into the Missing CO_2 Sink" by W. Wang et al. analyzes the role of the

Internet of Things (IoT) in monitoring the soil abiotic CO_2 uptake and develops an IoT monitoring system. In the work "Two-Phase Algorithm for Optimal Camera Placement," J.-W. Ahn et al. develop a two-phase algorithm for the camera placement problem in visual sensor networks.

The second category (healthcare related models and algorithms) includes three papers. The paper "Modeling and Optimization of the Drug Extraction Production Process" by D. He et al. contributes a data-driven iterative learning control method to the drug extraction production process. In the paper "Optimal Scheduling of Doctors Outpatient Departments Based on Patients' Behavior," Z. Ren et al. use the prospect theory to model the patient satisfaction and present a doctor outpatient scheduling model and its plant growth simulation algorithm. The paper "Research on Healthy Anomaly Detection Model Based on Deep Learning from Multiple Time-Series Physiological Signals" by K. Wang et al. uses the convolutional neural network to develop a healthy anomaly detection model.

The third category (models and algorithms of production or delivery) includes six papers. The paper "Impact of Personnel Flexibility on Job Shop Scheduling" by R. Lin et al. contributes a measurement of personnel flexibility as well as a job shop scheduling model with personnel flexibility. The paper "Collaborative Product Design for Tasks Sorting Based on Shortest Delivery" by X. Liang et al. presents a collaborative design task scheduling model. In the work "Identifying and Analyzing Strong Components of an Industrial Network Based on Cycle Degree," Z. Zhang et al. relied on the cycle degree of industrial networks to formulate the concept of circular centrality consisting of the lengths and the numbers of cycles through a vertex. In the paper "Acquisition Pricing and Inventory Decisions on Dual-Source Spare-Part System with Final Production and Remanufacturing," Y. Zhou et al. apply stochastic dynamic programming and recursion algorithm to solve a spare-part inventory problem with two replenishment strategies. The paper "A Recovery Model for Production Scheduling: Combination of Disruption Management and Internet of Things" by Y. Jiang et al. presents a recovery model of production scheduling with the integration of disruption management and Internet of Things. In the paper "A Heuristics-Based Parthenogenetic Algorithm for the VRP with Potential Demands and Time Windows," C. Shi et al. extend the classical vehicle routing problem (VRP) to a VRP with potential demands and time windows (VRP-PDTW) and develop a heuristics-based parthenogenetic algorithm for solving the extended problem.

The fourth category (evaluation and classification models) includes three papers. In the paper "A Novel Evaluation Approach for Tourist Choice of Destination Based on Grey Relation Analysis," X. Guo and Z. Sun apply grey relation analysis into the tourist destination evaluation and show the advantage by the comparison with the Technique for Order Preference by Similarity to Ideal Solution (TOPSIS). The paper "An Optimal Decision Assessment Model Based on the Acceptable Maximum LGD of Commercial Banks and Its Application" by B. Shi et al. presents an optimal credit rating model for customers' credit risk evaluation and credit decision. The paper "A Multiple Classifier Fusion Algorithm Using Weighted Decision Templates" by A. Mi et al. presents a multiple classifier fusion algorithm using weighted decision templates and shows the advantage by an experimental comparison.

The fifth category (control models and algorithms) includes two papers. In the paper "Optimal Control for Bufferbloat Queue Management Using Indirect Method with Parametric Optimization," A. Radwan et al. integrate an indirect approach with parametric optimization to develop a forward-backward optimal control queue algorithm. The paper "Event-Triggered Faults Tolerant Control for Stochastic Systems with Time Delays" by L. Huang et al. presents an event-triggered scheme of state-feedback controller design for stochastic networked control systems.

Acknowledgments

We would like to thank all the authors for contributing their works to the special issue and thank all the reviewers for their supporting efforts. Junhu Ruan would like to acknowledge partial supports from China Ministry of Education Social Sciences and Humanities Research Youth Fund Project (no. 16YJC630102); China Postdoctoral Science Foundation (no. 2016M600209); Science and Technology Plan Projects of Yangling Demonstration Zone (no. 2016RKX-04).

Junhu Ruan
Xuping Wang
Chengyan Yue
Guo Chen
Minsoo Kim

Hindawi Publishing Corporation
Scientific Programming
Volume 2016, Article ID 1319249, 18 pages
http://dx.doi.org/10.1155/2016/1319249

Research Article

Assessing Partnership Alternatives in an IT Network Employing Analytical Methods

Vahid Reza Salamat,[1] Alireza Aliahmadi,[1] Mir Saman Pishvaee,[1] and Khalid Hafeez[2]

[1] *Socio-Economic Systems Engineering Department, Iran University of Science and Technology, Tehran, Iran*
[2] *Business & Management Division, Dundee Business School, University of Abertay, Bell Street, Dundee DD1 1HG, UK*

Correspondence should be addressed to Alireza Aliahmadi; alireza@iust.ac.ir

Received 13 June 2016; Revised 1 September 2016; Accepted 15 September 2016

Academic Editor: Junhu Ruan

One of the main critical success factors for the companies is their ability to build and maintain an effective collaborative network. This is more critical in the IT industry where the development of sustainable competitive advantage requires an integration of various resources, platforms, and capabilities provided by various actors. Employing such a collaborative network will dramatically change the operations management and promote flexibility and agility. Despite its importance, there is a lack of an analytical tool on collaborative network building process. In this paper, we propose an optimization model employing AHP and multiobjective programming for collaborative network building process based on two interorganizational relationships' theories, namely, (i) transaction cost theory and (ii) resource-based view, which are representative of short-term and long-term considerations. The five different methods were employed to solve the formulation and their performances were compared. The model is implemented in an IT company who was in process of developing a large-scale enterprise resource planning (ERP) system. The results show that the collaborative network formed through this selection process was more efficient in terms of cost, time, and development speed. The framework offers novel theoretical underpinning and analytical solutions and can be used as an effective tool in selecting network alternatives.

1. Introduction

Business networks have become an essential part of conducting business. This is because the nature of business environment has changed and in many industries strategic business networks are replacing market structure theories [1, 2]. As a consequence of increased global competition, technological complexities, and environmental uncertainties, the firms could not internally perform all of the value creation activities, and instead they create strategic collaborations and interfirm ties to focus on their core competencies by externalizing some of the value creation activities to other network actors [3]. The significance of the concept leads to many efforts from the authors in various disciplines and provides a huge body of knowledge in different branches of science like management, economy, computer science, industrial engineering, sociology, law, and so forth, with various keywords like "collaborative network" [4], "collaborative networked organization" [5], "strategic networks" [1], "relational view" [6], "business networks" [7], "network approach to strategic management" [2], and "industry network approach" [8, 9]. This diversity of literature convinced the authors [10] to provide a metareview in different types of interorganizational relationships. This diversity of literature is because we are in the preparadigmatic phase, as Kuhn named it, of interpreting this phenomenon which is sometimes called a new discipline [4]. There are some debates on the economic reason of this paradigm shift; [11] named the innovation economy and growth of information technology as the reason behind decentralization of firms and emergence of business ecosystems.

Firms exploit interorganizational relationships for a number of reasons. According to [12] there are four motives to these: (1) acquiring resources, (2) reducing uncertainties, (3) enhancing legitimacy, and (4) attaining collective goals. Gulati et al. [1] argue that a firm's network could be a source

of competitive advantages, because the firm's network is idiosyncratic, inimitable, and nonsubstitutable, and it could be a source of sustainable competitive advantage based on resource-based view of firms. In addition, these networks could provide some benefits for firms like learning, achieving economies of scale and scope, and providing access to information, resources, markets, and technologies. Moreover, firms can share their risks through these relations and [12] and increase the agility. On the other hand, the networks may have a negative impact on the firm's performance and could lead to some constraints like "lock-in and lock-out" effect which occurs because of resource constraints, and also loyalty expectation of partners. As a result, a specific relationship may preclude organizations from relating to some other networks [1, 13].

In addition, strategic networks may have major influence on customers and/or end users. A network may combine resources and capabilities to deliver unique products or services to customers, and this is the outcome of a complex value activity integration process. Similarly, any failure or ineffectiveness would be transferred to customers, thereby progressively ruining the competitive advantage of the entire network. Moreover, the formation of a collaborative network is a long-term decision and any structural changes in the network may lead to significant pay-offs. Therefore, the decision on how to form collaborative network requires sufficient managerial attention and organizational resources.

In this paper, the problem of how to develop an effective "collaborative network" (CN) is considered in the context of IT industry. An Iranian large-scale IT firm wants to establish a goal-oriented network as a kind of "collaborative networked organization" (CNO) to develop an "enterprise resource planning" (ERP) for universities, called "integrated university management system" (IUMS). The assumed CN consists of a focal firm and members and is adapted to CNs' definition as described in the next section. The focal firm, as the manager of the network, selects some firms as the network's members and each member undertakes some parts of the project and possesses a portion of the final product's share. Obviously, the final products belong to all members of the network and the network is responsible for updating and supporting the product. In other words, the network has two forms in two levels of business; at first, the network is responsible for product development and after that the network changes its role to a service provisioning network in which it is responsible for supporting and upgrading the software and selling it in "software as a service" or any other form. In this paper, the focus is on the first level in which the product in under development by the network.

To address the problem, an integrated framework is proposed that involves both qualitative and quantitative methodologies to solve this problem. In the framework, two contemporary theories of interorganizational relationship, resource-based view and transaction cost theory, are combined; and a more comprehensive and applicable theory is provided. The justifications and advantages are discussed in the paper. Furthermore, a structured framework is proposed which includes a hybrid AHP and multiobjective programming and tries to provide a holistic view of the problem. At the end, the framework is implemented in the real case study and applicability of the framework is discussed.

In the subsequent sections, the proposed framework is described in detail. In Section 2, the theoretical background is discussed and the key concepts about collaborative networks, network formation, and rationale of interorganizational relationships are provided. In Section 3, the proposed framework is explained in detail. In Section 4, the implementation reports and the corresponding results are presented. In Section 5, the managerial implications, limitations, and the relevant future work are discussed.

2. Theoretical Background

In this section, theoretical backgrounds and the essential materials are provided. Related literature about collaborative networks, network formation, and rationale of interorganizational relationships are reviewed. These concepts form the underlying basis of the integrated framework proposed in this paper.

2.1. Collaborative Networked Organization (CNO). As mentioned by Camarinha-Matos and Afsarmanesh [14], the literature of interorganizational network is "a large but disjoint body of literature"; thus, we try to review different parts and provide a general big picture. The first part is "collaborative networks" which are defined by Camarinha-Matos and Afsarmanesh [4] as "an alliance constituted by a variety of entities (e.g., organizations and people) that are largely autonomous, geographically distributed, and heterogeneous in terms of their operating environment, culture, social capital, and goals." They classified the collaborative networks as (1) "collaborative networked organization" and (2) "ad hoc collaboration," where ad hoc collaboration can be accomplished in virtual communities and in not-business-oriented issues and most kinds of collaborative networks which act as an organization can be classified as the first class. They also classified the CNOs as two classes of (1) "long-term strategic networks" and (2) "goal-oriented networks." The first class tries to prepare the conditions and proper environment for fast and fluid configuration of the networks, when faced with opportunities. The second type–goal-oriented networks–derives from a specific goal like production or service providing activities [14].

Another stream is "strategic networks" which are defined as the firm's set of relationships, which are enduring and have strategic importance for firms entering them [1]. Brass et al. [12] define the strategic networks as a set of ties and nodes, where ties represent the relationships and nodes represent the actors including individuals, work units, and organizations. Gulati et al. [1] state that these strategic networks include interorganizational ties like strategic alliances, joint ventures, long-term buyer-supplier partnerships, and other similar ties. Jarillo [15] tried to develop the concept of strategic networks and discuss them in order to show their role as a mode of organization, also to analyze the economic conditions of strategic networks' existence. Stuart and Sorenson [16] discuss strategic networks and their role in entrepreneurial ventures. They present a well-organized literature review on

the strategic network in the context of entrepreneurial issues and provide a research agenda in five major areas for a better understanding of the strategic networks. In this perspective, actors are embedded in greater networks of relationships and their conduct and performance should be analyzed with relevance to these networks which provide opportunities as well as constraints for the actors. Also, Kohtamäki et al. [17] discuss the strategic networks and focus on the effect of organizational identity on strategic network behavior and performance.

"Business networks" is another important stream which is known by some works like [3, 7, 18–21]. "Network organization" and "network of organizations" are two important concepts in this area. "Network of organizations" refers to "any group of organizations or actors that are interconnected with direct or indirect exchange relationships." Any market and the entire composition of the economy can be described in this manner as a network of ties and actors. But "network organizations" refers to the networks that define value system and the actor's roles and responsibilities [3]. Ritter et al. [13] propagate the concepts of deliberate and emergent networks or intentional and unintentional networks. Jarillo [15] believed that networks are intentional organization modes built to enhance firm's competitive positioning. Möller and Rajala [20] suggest that networks are emerging systems that individual firms are restricted by or dependent on. Other researchers provide different classifications; they distinguish between hub-driven networks and strategic multiactor networks [22, 23]. Also, Pulles et al. [24] provide an empirical study and discuss the essential characteristics of an actor for being an innovative contributor in the business network.

Based on the extant literature, the interorganizational collaborations can be analyzed at three different levels. The first level discusses dyadic relationships and effect of relationships in accessing and controlling resources. The development, management, and termination of relationships are discussed at this level. The interactions between two actors are the main issue of this level that is identified by many researchers [25–32]. The second level discusses the portfolio of relations that a focal firm maintained, sometimes called "portfolio of alliances" or "strategic nets." The questions are how to create these structures, how they affect the firm's performance, and how to manage this portfolio. See, for example, [20, 33–41]. The third level is the broadest one and concentrates on networks. Usually, there is not an individual or focal firm but the whole structure of industry or market will be considered. The question is how network structures evolve and how to explain their dynamics. In addition, this level examines the effect of networks on firms and vice versa [7, 23, 42–46].

In this paper, the problem was defined in the class of a "goal-oriented network" in collaborative networks context where a network formed to accomplish a specific goal as an organization. In the "business network" context, the problem can be classified as an intentional, deliberate, and manageable network and also a hub-driven one which consists of a hub as coordinator and some members as value creators. Moreover, the problem was defined to tackle the portfolio level in interorganizational relationships and not in dyadic or industry network level.

2.2. Fundamental Theories on Interorganizational Relationships and Networks. There are a number of theories that describe the rationale behind the formation of interorganizational relationships. For example, [47] mentioned six theories about the formation of interorganizational relationships, including (1) transaction cost economics, (2) resource dependence, (3) strategic choice, (4) stakeholder theory of the firm, (5) organizational learning, and (6) institutional theory. Moreover, Parmigiani and Rivera-Santos [10] explore the key theories for explaining interorganizational relationships and mentioned theories like transaction cost economics, the resource-based view, knowledge-based view, dynamic capabilities, agency theory, resource dependence, stakeholder theory, institutional theory, and social networks. In this paper, two dominant theories, (1) transaction cost theory and (2) the resource-based view, are selected for consideration and the framework was based on them. This is because these two theories are the most general ones which other theories can be justified based on them and other theories can be seen as a branch of them or a combination of them [10]. Also, they are completely dominant in the literature in comparison with other theories and can perfectly cover the different aspects of the problem as described in the subsequent sections. These two theories are discussed in the following sections.

2.2.1. Transaction Cost Theory (TCT). Transaction cost theory has been widely used in literature as a theory of interorganizational collaboration [48–50]. The theory proposes that firms should determine their boundaries in order to minimize the sum of production cost and transaction cost. Indeed, this theory is developed to answer to the popular "make or buy?" question and describes that each value creation activity should be internalized or externalized in a way that minimize production plus transaction costs. In this theory, interorganizational relationships lead to reducing the uncertainties induced by market failure in addition to reducing associated costs [47]. The transaction cost refers to cost associated with exchange activities like writing and enforcing contracts and production cost associated with internal activities like producing, organizing, and managing.

As mentioned, TCT tries to address the make or buy decision. The firm should decide to produce using its own resource or purchase from the market. In addition, a firm could produce jointly through an alliance or interorganizational relationships. According to TCT, it is better to internalize when there are high transaction costs and low production costs. Transaction costs increased when the firm is encountered with a high rate of opportunistic behavior and high asset specificity. Merger and acquisitions could be seen as ways to internalization. It is better to externalize and use market transactions when production costs are high and transaction costs are low. Moreover, the firm could choose to ally when there are medium transaction and production costs and transaction costs are not high enough to justify internalization. Indeed, alliances could be seen as semi-internalization, and firms use it when internalization cost is more than market transaction but some limitations prevent full internalization [51].

There are some authors who criticize this theory in its ability to describe the formation of interorganizational relationships and generalization of this theory. There is a view that this theory is limited to economical and cost related issues, but some relations formed for some other reasons like learning, and these reasons could not be included in TCT [47]. However, [52] provide a critique that, according to TCT, all firms with similar transaction and production costs should make similar decisions about internalization and externalization, while this statement is untenable, and it could easily be seen that internalization/externalization decision dramatically differs along with a distinctive industry and in similar firms. Indeed, it should be argued that some complex factors (apart from economic and cost-based factors) should be considered in alliances, and using cost-based factors limits the quality of decision-making [53]. Although the simple logic of TCT provides useful and intuitive insights and causes widespread use of it, this simplicity causes some weakness that limits its effectiveness. Therefore, we suggest combining this theory with some other concept to improve its suability. In this paper, it is argued that TCT could be seen as complementary with resource-based view (RBV) and a combination of these two theories could be more advantageous.

2.2.2. Resource-Based View (RBV). Resource-based view of firm has been originated from the work of [54] and expanded in works by Wernerfelt [55] and Barney [56]. Resource-based view is seen as an inward-looking approach and in contrast with porter's competitive force. Teece et al. [57] mentioned four approaches in business strategy: (1) competitive force, (2) strategic conflict, (3) resourced-based perspective, and (4) dynamic capabilities. They introduce two former approaches as outward looking that emphasize on market power exploitation and the two latter as inward looking that emphasize on efficiency. They state that, according to RBV, firms are heterogeneous with their resources, and there is some degree of stickiness between firms and these resources and firms are forced to live with what they have and what they lack. This is because some resources and assets like reputation and tacit knowledge could not be easily tradable. Also, business development is complex and could not be easily achieved through the market.

The RBV emerged as an alternative approach for understanding the firm's competitive advantages. In this view, firms can be seen as a complex bundle of tangible and intangible resources. As mentioned before, RBV is in the contrast to the porter's competitive forces. Still, RBV emphasizes on the firm's resources and provides an inward-looking approach. RBV interprets the origin of competitive advantage in possessing valuable resources. Indeed, firms are heterogeneous with their resources and resources often are imitable, immobile, and firm-specific, and these resources could generate competitive advantage that could lead to economic rents or returns [58, 59]. What a firm can perform effectively is a function of what it has possessed and so a firm's competitive advantages are defined by its resources [51]. One of the most important concepts of RBV is named as VRIO/N framework. According to this framework, not all types of resources could lead to sustained competitive advantages, but firm's resources that have four attributes could lead to sustained competitive advantages: (1) resources must be valuable (V), (2) rare among the firm's competitors (R), (3) imperfectly imitable (I), (4) have potential to be used by organizational processes (O), or in another variant be nonsubstitutable (N). These four attributes expose the heterogeneity, immobility, and usefulness of the firm's resources. According to this framework, if resources are not valuable, then they decrease firm's revenue and could be classified as weaknesses and named as competitive disadvantages. If resources are valuable but not rare, exploiting them will generate a competitive parity. If resources are valuable and rare, but easily imitable, they will generate temporary competitive advantages. And finally, if resources are valuable, rare, and costly to imitate and could be exploited by the organization then these resources could lead to sustained competitive advantages [56, 60].

The TCT and RBV are dominant theories in justifying the rationale of interorganizational relationships. As previously mentioned, resources that are valuable, rare, costly to imitate, and nonsubstitutable can make competitive advantages and these types of resources could not be purchased from the market; thus, firms try to achieve them through interorganizational relationships, sometimes called intermediate market. Firms rationale for creating relationships are to exchange, share, and combine valuable resources with other firms. In summary, there are two general motives for firms to ally: (1) to achieve the other's resources and (2) to maintain and develop its own resource in combination with other's resource [51]. Some authors [51, 61] compare the TCT and RBV logic in interpreting firm's relationships. TCT tries to economize transaction and production costs and presents proper governance structure, while RBV tries to maximize profits in the long-term by exploiting and developing resources. Moreover, TCT proposes to ally when transaction costs are not so high and take an intermediate place, while RBV proposes to ally when critical resources are needed by the focal firm owned by other parties and these resources are inseparable from other resources of the owner.

In summary, despite being criticized, these two theories have been widely used over a period of time by many authors. TCT concentrates on minimizing costs whereas RBV concentrates on maximizing value obtained from the relations. These two approaches can be integrated and used in a complementary way, to increase the potential relative strength of the two theories [53, 61, 62]. TCT only attends to cost-based issues and RBV could compensate this weakness. In addition, TCT could consider the opportunistic behavior of firms which occurs in specific circumstances and it proposes useful statements that could not be found in RBV. Also, TCT assumes that all firms can produce or offer services equally effective and this weakness could be compensated by incorporating RBV, because in RBV each firm's performance is determined by its complex bundle of resources and RBV explicitly differentiates firms in their performance [61]. Furthermore, RBV and TCT theories have been employed in justification of issues in the various fields of study; for example, [63] discussed the portfolio selection problems from TCT viewpoint and [64] discussed alliance

formation justification by RBV theory. In this paper, these two theories are integrated to maximize their potential and implemented employing analytical tools. We propose that the combination of these two could interpret the reality better and lead to the better results.

2.3. Network Formation. There are a handful of papers which discuss the "network formation" problem and they are from different scientific streams. Thus, in this section, the papers related to "network formation" are identified and briefly reviewed. Some authors explain the formation of relations at the dyadic level. David et al. [65] explain the stages of dyadic relationship building in five phases, namely, (1) the prerelationships stage, (2) the early stage, (3) the development stage, (4) the long-term stage, and (5) the final stage. After that, Dwyer et al. [66] provide a framework for developing buyer-seller relationships. They proposed a formation process that incorporates five phases: (1) awareness, (2) exploration, (3) expansion, (4) commitment, and (5) dissolution. Also, Larson [67] discussed the three stages of relationship making, (1) preconditions for exchange, (2) conditions for building, and (3) integration and control. Wilson [68] tried to combine the empirical knowledge about the successful relationship in an integrated model and provide a conceptual process for developing buyer-seller relationships. The proposed model consists of five stages including (1) partners search and selection, (2) defining the purpose, (3) setting relationship boundaries, (4) creating relationship value, and (5) relationship maintenance. In recent papers, Abosag and Naudé [69] focus on the development of a special form of B2B relationship and discuss the role of interpersonal liking and trust in the relationship development. They provide a conceptual model and use the SEM methodology. Gumerov et al. [70] discuss the features of forming a dyadic relationship in interaction network environment and entrepreneurial structure. Lacoste and Johnsen [71] discuss the dynamics and balance of power in relationships between supplier and customer and describe their actions to improve countervailing power. Mandják et al. [72] focus on the question of "how a business relationship emerge?" and discuss the first stage: "birth of a new relationship." They provide a conceptual framework and discuss the different trust building scenarios. Then, they worked on trigger issues in initial stage which boost an emerging relationship and provide a logical framework [73]. Finally, Hennelly and Wong [74] explore the formation of new relationship in nascent industries with high uncertain environment by longitudinal case studies. They discussed three concepts: compatibility, complementarity, and relationship life cycle in the relationships formation.

At the network level, Ebers [75] explain the network formation and provide useful insights and they discuss dynamics and costs of interorganizational networks. David and Keely [76] discuss collaborative network formation in multiple funding agencies. Hinterhuber [77] explains the concept of virtual value chain as an emergent phenomenon and provides a process model for forming such a network. The proposed model consists of six phases: (1) analysis of firm value chain, (2) analysis of the flow of goods and total value created, (3) identification of ways to increase the amount of value created by the extended value chain, (4) configuration of the network around value creation opportunities, (5) identification of ways to capture value created (alliances, joint ventures, and acquisitions), and (6) managing the cross-industry value chains. Ozcan and Eisenhardt [34] develop a process view for building portfolio of relations in the context of entrepreneurial firms. Their process framework consists of three strategies for building portfolio of interorganizational relations: (1) to advocate unique industry architecture that proactively shapes the industry, (2) to take long jumps to exploit opportunities to coordinate unconnected firms, and (3) to defend against emerging industry uncertainties when they occur. Partanen and Möller [23] develop a process model for the strategic network formation problem based on value creation system approach. Their process consists of 8 stages: (1) determination of value creation activities, (2) determination of the value creation system, (3) determination of objectives and analysis of target activities, (4) comparing resources and activities with target activities, (5) analysis of delegated activities, (6) conducting preliminary partner assessment, (7) negotiating with alternative partners, and (8) launching interfirm collaborations. They implemented their proposed process model in the ICT sector and discussed the implementation points and managerial implications. Gaspar et al. [78] organize a descriptive-qualitative study and determine the factors which influence, promote, or restrict the formation of a cooperative network. In another work, Hale and White [79] develop a framework for the establishment of a collaborative networks. They explore four case studies in aerospace and renewable industry and propose a framework consists of eight elements: (1) purpose definition, (2) funding mechanism, (3) identification and selection of partners, (4) defining collaborative objectives, (5) network structure, (6) governance processes, (7) defining risks, and (8) requirement for success. They claim that the process of member identification and selection in the networks is one the key gaps. Jussila et al. [80] conduct a qualitative case study and focus on strategic network formation between second-tier actors in the field of construction megaprojects. They determined six strategic networks in the case and discuss their types and roles and also uncertainties regarding network formation. Zhang and Pezeshkan [81] discuss the alliance formation of a foreign firm in the host country and explore the role of firm's position in the network. They integrate theories of resource dependence and social network to provide a better understanding of alliance formation phenomenon. Moreover, there are a lot of works which discuss the formation of virtual organizations and enterprises as a kind of collaborative networks, like [82–85].

To sum up this section, it should be said that the literature on interorganizational relationships, as described by Parmigiani and Rivera-Santos [10], is very expanded and complex. Thus, we try to classify it and provide a big picture on the field of interorganizational relationships formation. Therefore, Table 1 categorizes the reviewed literature on the formation phase of interorganizational relationships.

Table 1: Summary of reviewed papers in formation phase of interorganizational relationships.

Paper	Level			Theoretic	Descriptive/prescriptive	Type of achievement
	Dyadic	Portfolio of relations	Networks			
[66]	√				Descriptive	Framework
[67]	√				Descriptive	Framework
[68]	√				Descriptive	Framework
[75]			√		Descriptive	Conceptual model
[48]		√		√	Descriptive	Theoretical propositions
[65]	√				Descriptive	Framework
[49]		√		√	Descriptive	Theoretical propositions
[51]		√		√	Descriptive	Theoretical propositions
[61]		√		√	Descriptive	Theoretical propositions
[47]		√		√	Descriptive	Theoretical propositions
[58]		√		√	Descriptive	Conceptual model
[77]		√			Prescriptive	Framework
[76]			√		Descriptive	Conceptual model
[34]		√			Prescriptive	Framework
[23]		√			Prescriptive	Framework
[69]	√				Descriptive	Statistical propositions
[78]			√		Descriptive	Affecting factors
[79]			√		Descriptive	Framework
[70]	√				Descriptive	Mathematical analysis
[71]	√				Descriptive	Conceptual model
[72]	√				Descriptive	Conceptual framework
[73]	√				Descriptive	Logical framework
[74]	√				Descriptive	Theoretic implications
[80]			√		Descriptive	Conceptual model
[81]		√		√	Descriptive	Theoretic implications

Based on Table 1 and reviewed literature, the gaps in the literature which are in the focus of this paper are listed below as follows:

(i) There is a significant lack of prescriptive solutions for the formation problem in the level of networks and portfolio of relationships. The prescriptive solutions, which generally include mathematical models, are common and tackle the formation problem in the dyadic level.

(ii) There is a weak relationship between fundamental theories and applied models, and applied models generally are constructed by case study's requirements and not by considering fundamental theories.

(ii) The complementarity of RBV and TCT theories are mentioned by some authors, as described earlier, but there is a lack of solutions for conceptualization and employment of it in a hybrid theory.

Owing to these points, the aim of this paper is to fill the mentioned gaps by providing a practical, prescriptive framework based on the fundamental theories. A hybrid of RBV and TCT theories was considered as fundamental theory and a practical framework and analytic solution were developed based on it.

3. The Proposed Framework for Collaborative Network Formation

As stated before, the framework aims at addressing network formation problem where a focal firm as a hub wants to establish a portfolio of relations for developing a specific business or product. In this problem, the focal firm needs to determine the scope to be internalized and those that should be assigned to the others, also choosing among available alternatives for externalized activities. The set of these actors, as a portfolio of relations, makes the network a synergic network that provides the predefined goals in a highly flexible, agile, and responsive to environmental changes way. It should be mentioned that the process of establishing a portfolio of alliances, as mentioned by Wilson [68], is a difficult and time-consuming process and sometimes needs to do iterative actions of gathering information, planning, acting, and reforming but may provide superior results.

The framework proposed in the following subsections consists of five stages. The framework involves both qualitative and quantitative aspects of the problem definition and solution. The framework includes a multiobjective mathematical model that formulates RBV and TCT theories. Indeed, the proposed framework integrates two important problems, namely, determination of externalization activities

and choosing the best partner for them; this integration could lead to better answers and will result in an effective network.

3.1. Stage 1: Determining the Value Creation Activities Map (VCAM). Firstly, the hub firm assesses its value creation activities to develop a map that identifies high-level value creation activities and their relations. Indeed, VCAM shows the activities that should be accomplished through the network. The VCAM is an important perquisite of the framework, because it should identify the areas of responsibilities among network members. Some models like "value chain model" [86] and "value creation system" [87] can be useful for determining value creation activities. Adopting Porter's concept, the value creation system divides the activities into two main categories, realization and support activities. Realization activities are those that create goods or services, and support activities are those that do not directly participate in creating products, but help the other activities in terms of their efficiency. It is noticeable that value creation activities are not necessarily organizational units but those which provide value to end users. Customers' needs and desires are important factors for this map. The resulted map should properly reflect these needs and illustrate the activities that lead to them.

3.2. Stage 2: Determining the Most Important Activities That Should Be Internalized. In this stage, the most important activities that should be accomplished by the hub firm will be identified. The activities that are included in VCAM and the network will be formed based on them. The hub firm wants to assign activities in VCAM to the network members and perform key activities by itself in order to maintain its place as a hub. The firm focuses on the activities that are in conformation with its capabilities. For this purpose, the hub firm capabilities will be classified, and the most important activities from VCAM that are in conformity with these capabilities will be selected for internalization. Researchers have addressed the problem of identifying and classifying the firm's capabilities [58, 59, 88]. For this part, we adopt and extend the work of [89]. According to resource-based view of the firm, they state that each firm has some resources; resources could potentially build capabilities and some capabilities with special features from firm's competencies. Any capabilities that have two important criteria, "uniqueness" and "collectiveness," could be classified as firm's competencies. As illustrated in Figure 1, each capability that has a high score in uniqueness and collectiveness could be seen as firm's competencies.

Based on the above-mentioned concepts, the hub firm capabilities will be classified and the most important activities that are in conformity with these capabilities will be selected for internalization. This classification is resolved by using the criteria of "uniqueness" and "collectiveness." Uniqueness is assessed by three measures, namely, rareness, Inimitability, and nonsubstitutability. Likewise, collectiveness is assessed by three measures, namely, Across-product, Across-function, and Across-business. The capabilities are classified under three categories, namely, (1) core competences, (2) competences, and (3) capabilities. Core competences are the most important capabilities and activities related to this category of capabilities selected for internalization.

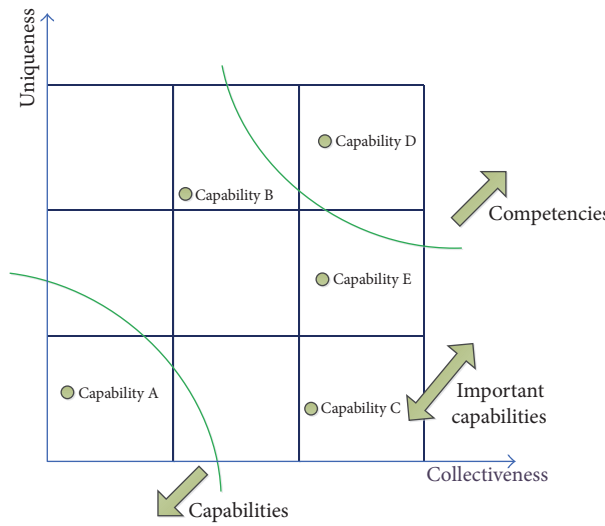


Figure 1: Capability classification scheme.

3.3. Stage 3: Determining Potential Partners for Activities and Their Preliminary Assessment. In this stage, efforts concentrate on identifying potential partners for each remaining activity from the previous stage (that still need to be selected). Subsequently, a preliminary assessment of these is conducted at this stage. Each remaining activity from the previous stage could be performed by hub firm or by one of the potential partners identified in this stage, and, obviously, each potential partner could be considered for more than one activity. The alternatives which will be chosen are those that possess necessary resources and capabilities for accomplishing the activities. Potential partner selection will be done by evaluating industry network. Industry network assessment could provide useful information by analyzing relationships, communities, and network properties by tools like "social network analysis" (SNA) and tools introduced for network economy analysis like [90]. Also, alternatives could be chosen by considering previous alliances and collaborators that have valuable, rare, and noninimitable resources.

Following this, initial assessment of potential partners will be conducted. The first prerequisite at this stage is gathering enough information about each alternative. Some alternatives' attributes could be more important at this stage, like information about their previous experience in alliances, level of trust and commitment, and their willingness to share tacit knowledge. Also, some other information like strategic goals and objectives, market position, organizational culture, structure, and processes could be useful at this stage.

3.4. Stage 4: Eliminating Unqualified Alternatives through Assessment of Noncompensatory Criteria. In this stage, some unqualified alternatives would be eliminated from the decision-making procedure. For this purpose, a selective

Figure 2: Noncompensatory criteria conceptualization.

literature review was conducted to determine the important criteria. Subsequently, the efforts are to be concentrated on constructing a model that can conceptualize the determination criteria. In this stage, the focus is on noncompensatory criteria, where alternatives must score more than a specific threshold level, allowing to discard any activates below the threshold. In other words, each potential partner that scores below the acceptable level would be eliminated. Indeed, the criteria mentioned in this stage are intrinsically different with the next stages, because in this stage the criteria will be used as noncompensatory and for elimination of alternatives and in the next stages the aim is to rank the alternatives with compensatory criteria. Thus, it is possible to find similarities between this stage's criteria with the next stages. These noncompensatory criteria were conceptualized and illustrated in Figure 2. The criteria used in this stage consist of three categories: (1) organizational fit, (2) strategic fit, and (3) reputation. "Organizational fit" is related to fitness between hub firm and its potential partners. This fitness should occur in three dimensions like cultural, process, and also structure similarity. "Strategic fit" is related to conformity of strategic issues between hub firm and its potential partners. These issues are adaptation in goals, objectives and values, adaptation in motives for entering into an alliance, and also the similarity of their market position. The third category is named "reputation" and is related to some social attribute of the potential partner firm, like anticipating trust and commitment, his willingness to share valuable expertise and tacit/explicit knowledge, and also record of his behavior in previous alliances.

The aim at this stage is that the managers and decision-makers (DMs) are assured that selected alternatives meet the minimum requirements. In the final stage, the best partner alternatives will be chosen based on some compensatory criteria.

3.5. Stage 5: Final Selection of Network Members. In this stage, network members will be selected and activities of VCAM assigned to them. In other words, the hub firm's portfolio of alliances will be determined. For this purpose, a mathematical model is proposed that leads to hub firm's network. The proposed model complements the previous stages. The model is a multiobjective decision-making model one that exploits two objectives, (1) maximizing the total value gained from the portfolio of relationships (RBV objective) and (2) minimizing total transaction and production cost (TCT objective). Table 1 describes the linear programming model parameters. There are some works in the literature which have similar aim, like [91].

Parameters of a Linear Programming Model for Network Partner Selection

Indices

i: activities $i \in I = \{1, 2, \ldots, n\}$.

j: candidates for each activity $j \in J = \{1, 2, \ldots, m_i\}$.

H: set of related activities $\{(s \cdot k) \in H$. activity s is related to activity $k\}$.

Input Parameters

Un_{ij}: relative uniqueness achieved through relationship with candidate j in activity i.

Co_{ij}: relative collectiveness achieved through relationship with candidate j in activity i.

Pr_{ij}: relative production cost of candidate j in activity i.

Tr_{ij}: relative transaction cost of candidate j in activity i.

C_{ij}: total budget needed by candidate j to accomplish activity i.

T_{ij}: time needed for candidate j to accomplish activity i.

PDD: planned due date of the project.

β: total accessible budget.

Problem Variables

DUE(i): the planned finish time of activity i.

$X_{ij} = 1$. If activity i allocated to candidate j; otherwise $X_{ij} = 0$.

One key point in this formulation is that the hub firm could be a potential alternative for performing activities, and, in such case, the first counter of set J (set of potential candidates for each activity) will dedicate to hub firm. The proposed model consists of two objectives and three sets of constraints. The mathematical formulation consists of objectives and constraints defined as follows:

$$\max \quad Z_1 = f_1\left(X_{ij}\right) = \sum_{i=1}^{n}\sum_{j=1}^{m_i}\left(W_1 \mathrm{Un}_{ij} \cdot X_{ij} + W_2 \mathrm{Co}_{ij} X_{ij}\right) \quad \text{objective (1)} \tag{1}$$

$$\min \quad Z_2 = f_2\left(X_{ij}\right) = \sum_{i=1}^{n}\sum_{j=1}^{m_i}\left(W_1' \mathrm{Pr}_{ij} \cdot X_{ij} + W_2' \mathrm{Tr}_{ij} X_{ij}\right) \quad \text{objective (2)} \tag{2}$$

$$\text{Subject to:} \quad \sum_{j=1}^{m_i} X_{ij} = 1. \quad \forall i \in I \tag{3}$$

$$\mathrm{DUE}(i) \leq \mathrm{DUE}(k) - \sum_{j=1}^{m_i} T_{kj} \cdot X_{kj} \quad \forall (i \cdot k) \in H \tag{4}$$

$$\mathrm{DUE}(n) = \mathrm{PDD}; \tag{5}$$

$$\sum_{i=1}^{n}\sum_{j=1}^{m_i} C_{ij} \cdot X_{ij} \leq \beta \tag{6}$$

contingent constraints related to portfolio effect of alternatives (7)

The above model is described detailing required objectives and constraints.

Objective (1) (Maximizing Total Value Gained from the Portfolio of Relationships (1)). This objective aims to maximize total gained value from the portfolio of relationships. In the formulation of this objective, two parameters of uniqueness and collectiveness are considered as introduced by Hafeez et al. [89]. They state that the resources with high degree of these two characteristics could qualify as firm's competencies and could lead to competitive advantages. Uniqueness shows the degree at which the firm's resources are solitary in type or characteristics in the market place, and collectiveness depicts that these capabilities are wide spread in the organization and could be found in any product and services and constitute a new pattern of competition for the firm. Uniqueness and collectiveness were measured by AHP method as shown in Figure 3. The AHP criteria were taken directly from the work by Hafeez et al. [89] and the weights will be determined by pairwise comparison of alternatives and criteria in conventional AHP manner, and final alternative weights will be used in objective function.

Objective (2) (Minimizing Total Production and Transaction Costs (2)). This objective function tries to minimize total production and transaction costs. According to TCT theory, total transaction costs and production costs must be minimized by answering the internalization/externalization question. This objective function utilizes two parameters of production cost and transaction cost. Production and transaction costs were measured by AHP method as shown in Figure 4 and final weights will be used in the objective (2) as a parameter. The criteria of production and transaction cost are inspired from the works by Dyer [48], Madhok and Tallman [49], and Tsang [61] and the weights will be determined by pairwise comparison, in conventional AHP way, and final alternative weights will be used in objective function. It is noteworthy to say that objective (2) (TCT objective) describes the short-term benefits of the hub firm and objective (1) (RBV objective) relies on the long-term benefits. Thus, the decision-maker could trade-off between these two objectives by a balancing coefficient.

Equation (3) (Activity Allocation Constraint). The constraint guarantees that each activity will be allocated to at least one candidate. It should be noticed that the hub firm could be one of the candidates and the activity may be allocated to the hub firm.

Equations (4), (5), and (6) (Total Time and Cost Constraint). As network will be formed for a specific software development project and the network is responsible for updating, maintenance, marketing, sales, and so on, so each activity should be accomplished in a time window and the total project must be finished before the predetermined due date. In this situation, perquisite relations between activities cause a formulation as suggested in (4). It is noticeable that a dummy start and end activity should be added to activity map so that the start activity depicts the start time of the project and the end activity depicts the final project due date. In addition, there is a cost constraint. The focal firm as network manager should control the total cost of the product.

Constraints (7) (Portfolio Effect). Some constraints arise according to the portfolio effect that different potential partners imposed to the hub firm. For example, it is possible that the hub firm does not want to have two conflicting partners in its portfolio simultaneously.

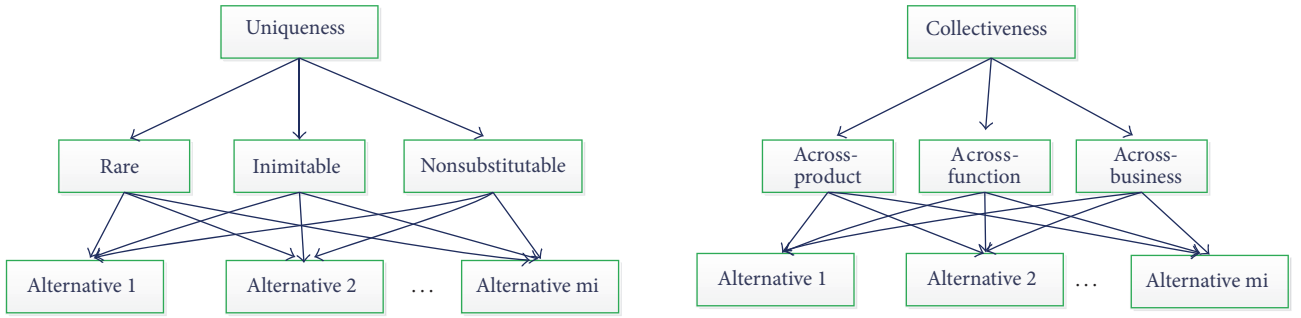


Figure 3: AHP hierarchies for assessment of uniqueness and collectiveness.

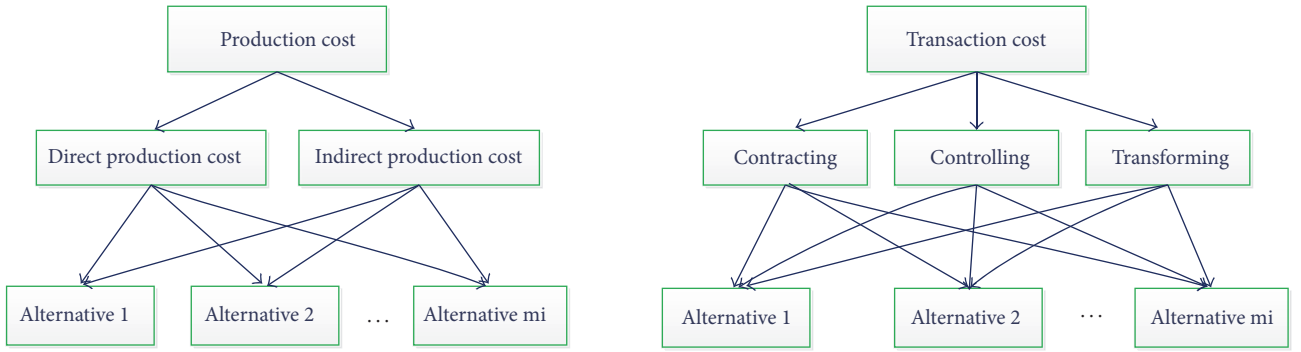


Figure 4: AHP hierarchies for assessment of production and transaction cost.

4. Implementation

The proposed framework was implemented in a project established for development of a total solution for universities, named as integrated university management system (IUMS). The IUMS project was planned in three scenarios, namely, plans A, B, and C. In plan A, the project will be performed completely by the project team. In plan B, the project team will be engaged in the development of the system in the modules that are completely related to university functions. Finally, in plan C, almost all of system modules will be performed by partner firms and the project team will be focused on the integration of modules and control of transactions between modules. Plan A has the highest cost and duration, whereas plan C has the lowest cost. Thus the framework was implemented in order to determine the optimal solution. The components and users of IUMS system were shown in Figure 5. The system was designed based on service-oriented architecture (SOA), where the IPIMS works as integration module and relates other modules to each other based on SOA. Also, the system should serve some users like students, lecturers, professors, staff, partners, and governments as shown in the figure. The value creation activity map was shown in Figure 6. This map illustrates the activities that should be done by the network, and each activity creates a specific value for the end user.

In the next stage, the aim is to determine the most important activities for the hub firm to internalize. A methodology was proposed for this stage in the previous section. The methodology was based on analysis of hub firm's capabilities and two criteria of "uniqueness" and "collectiveness." Thus, the hub firm's capabilities were analyzed and classified based on the two aforementioned criteria and by AHP technique, as described in stage 2 of the framework, and the result was shown in Figure 7.

Then, the potential partners for each activity were identified. Thus, after interviews with experts and a brief analysis of the results, an initial number of alternatives were determined. These alternatives were investigated precisely and unqualified ones were omitted from the procedure.

In stage 4, qualitative analysis of alternatives was conducted, and unqualified alternatives were excluded from the rest of procedure. The main criteria, as described earlier, are reputation, organizational fit, and strategic fit. In the IUMS case, reputation was the most important criterion.

In stage 5, the final selection of network members will be done. The most important parameters that are utilized in the mathematical model of this stage are Un, Co, Pr, and Tr. Also, there are four weight parameter in the objectives as shown in formulas (1) and (2). The weights are set as 0.65 for uniqueness and 0.35 for collectiveness and 0.65 for production cost and 0.35 for transaction cost.

Subsequently, the mathematical model was implemented in GAMS and the obtained results were discussed. As the model was a hybrid of AHP and multiobjective mathematical programming, the appropriate techniques were employed, namely, (1) weighting method, (2) method of global criterion, (3) lexicographic method, (4) the method of satisfactory goals

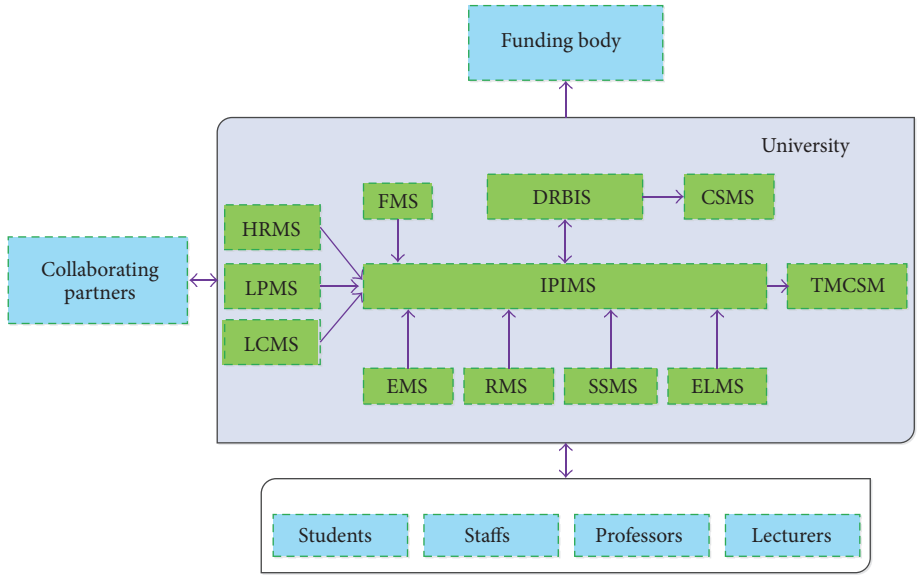


Figure 5: The IUMS components.

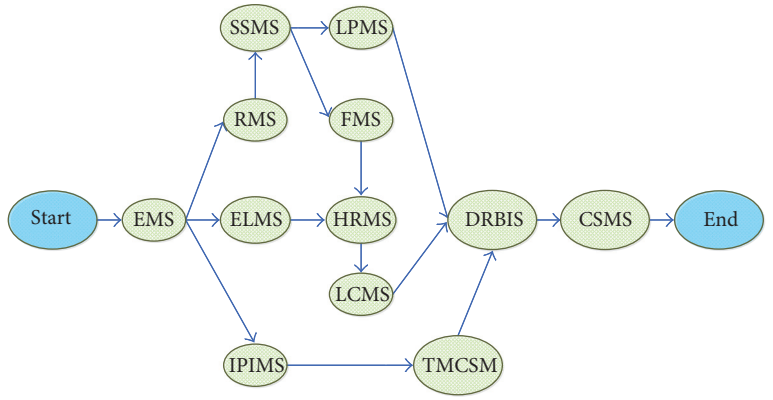


Figure 6: IUMS value creation activities map.

Table 2: Classification of the optimization methods employed.

Classification	Method name
Basic methods	(1) Weighting method
No preference	(2) Method of global criterion
Priori preference	(3) Lexicographic method
Posteriori preference or interactive methods	(4) The method of satisfactory goals
	(5) SIMOLP

[92, 93], and (5) simplified interactive multiple objective linear programming (SIMOLP) [94]. The employed techniques were illustrated in Table 2. The techniques were selected based on their adaptation with problem structure. Also, it was tried to select at least one technique in each class of techniques.

The weighting method, as a simplistic and straightforward method, was selected because it can provide a simple, reasonable, and understandable way of solving the MOADM problem. The method of global criterion was selected from the no preference class, where there was no need to judge the preferences. The lexicographic was selected from the priori preference class in which the decision-makers should determine their preferences before the running of the algorithm. Also, two methods, SIMOLP and the method of satisfactory goals, were selected from the class of posteriori or interactive preferences where decision-makers provide their judgement interactively or after the running of the problem. The obtained results from above-mentioned methods were illustrated in Table 3. In weighting method, the solution achieved by solving this problem

$$\max \quad Z = w_1 f_1 + w_2 f_2 \equiv \max \; Z' = f_1 + \frac{w_2}{w_1} f_2 \qquad \alpha = \frac{w_2}{w_1}. \tag{8}$$

As illustrated in the formulation above, the objectives are sum up with different weights, and it is possible to divide the formulation by one of the weights and solve the problem

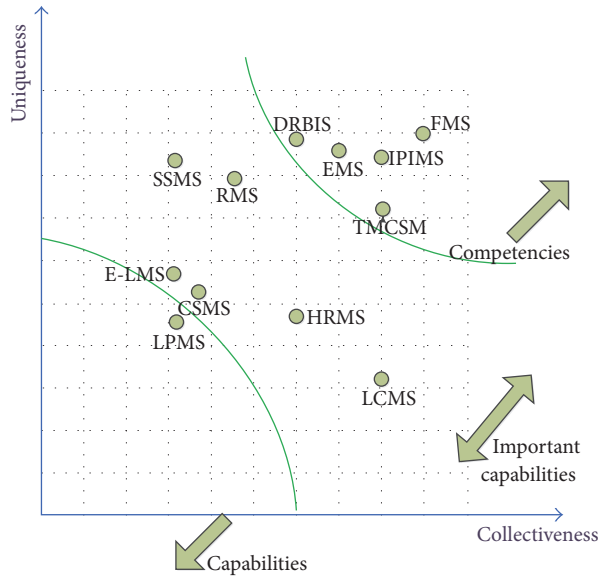


Figure 7: Capabilities classification.

Table 3: The obtained results from the different methods.

Methods	Parameters	Results	
		RBV	TCT
Weighting method	$\alpha < 0.08$	2.320	1.760
	$0.08 \le \alpha < 0.2$	2.274	1.773
	$0.2 \le \alpha < 0.6$	2.261	1.852
	$\alpha \ge 0.6$	1.166	3.258
Global criterion	$P = 1$	2.261	1.852
Lexicographic	Preferred obj. = RBV, RBV ≥ 2.3	2.32	1.76
	RBV ≥ 2.2	2.261	1.852
	RBV ≥ 2.1	2.138	1.988
	RBV ≥ 2.0	2.032	2.155
	RBV ≥ 1.9	1.972	2.247
	RBV ≥ 1.8	1.804	2.396
	Preferred obj. = TCT; TCT ≥ 3.2	1.166	3.258
	TCT ≥ 3.0	1.166	3.258
	TCT ≥ 2.8	1.454	2.863
	TCT ≥ 2.6	1.684	2.603
	TCT ≥ 2.4	1.743	2.511
	TCT ≥ 2.2	1.972	2.247
	TCT ≥ 2.0	1.986	2.168
	TCT ≥ 1.8	2.261	1.852
Satisfactory goals	TCT ≥ 1.8; RBV ≥ 2.0	2.261	1.852
SIMOLP	—	2.261	1.852

with one parameter named alpha, which can vary from zero to infinity. The problem was solved with different alpha by steps equal to 0.1; after numerous run, the four solutions were obtained. In the second method, the global criterion method, the efforts are on the minimization of the gap between objectives and their optimum value as shown in formula (9) as follows:

$$\min \quad Z = \sum_{i=1}^{k} \left(\frac{f_i(x^*) - f_i(x)}{f_i(x^*)} \right)^p. \tag{9}$$

The problem was solved with this method and p set to 1. Hence, the problem was transformed into a nonlinear problem. The problem was solved with "MINLP" solver of GAMS and the results were reported. In the lexicographic method, the objectives were ranked based on their importance. The solution was obtained by optimizing the most important objectives and the procedure will continue by objectives orders. The first-order objective sets as equal to its optimum value and was placed in constraints, as shown in formula (10).

$$\begin{aligned} \max \quad & f_2(x) \\ \text{s.t.} \quad & g(x) \le 0 \\ & f_1(x) = f_1^*. \end{aligned} \tag{10}$$

In the formula, objective 1 is preferred to objective 2. After solving the problem with this objective, the preferred objective with the optimal value was set as a constraint and the problem was solved with the second-order objective. In an extension of the lexicographic method, it is possible to let the first-order objective tolerate with level as shown in formula (11).

$$\begin{aligned} \max \quad & f_2(x) \\ \text{s.t.} \quad & g(x) \le 0 \\ & f_1(x) \ge f_1^* - \delta_1. \end{aligned} \tag{11}$$

The lexicographic method was implemented with different tolerate level and with different preferences about the objectives, and the results reflected in Table 3. Since it was difficult to select one objective as the dominant, the problem was solved with the two alternative manners. In the satisfactory goal method, the decision-maker determines a set of acceptable goals for each constraint, after that the problem was solved with the objective which has the least satisfactory level and the rest of objectives are set as constraints. The problem was solved in different situations until reaching the solution that fulfills satisfactory levels of all objectives. Since there are only two objectives, the decision-makers could find the most satisfactory solution easily. At the end, the SIMOLP technique was run. In this technique, the problem was solved with weights that were achieved by a specific procedure. In the first, the problem was solved with each objective separately; then the weights were attained by solving the system of equation as depicted in formula (12).

$$\sum_{j=1}^{j} w_j f_j^l - w_{k+1} = 0 \quad \forall l \in H. \tag{12}$$

After that the problem was solved with achieved weights; if the obtained solution was satisfactory, the procedure was

Table 4: The favorable (or optimum) solution.

RBV	TCT	Variables	Alt. 1	Alt. 2	Alt. 3
2.261	1.852	RMS			1
		SSMS			1
		LPMS		1	
		HRMS	1		
		LCMS		1	
		CSMS	1		

terminated; otherwise, the new solution was replaced with one of the previous and the procedure was repeated until a satisfactory solution would be obtained. In this problem, there are two objectives and a system of two equations. It is noticeable that the system of equations does not depend on w_{k+1} and this parameter could be determined arbitrarily. In this problem, after two runs, the procedure leads to duplicated solution and was terminated automatically. But, the first solution after solving the system of equations was the most favorable and selected by the decision-makers.

The final solution was selected by comparing different solutions. Finally, the intersection of all methods' solutions was selected as reported in Table 4. It is noticeable that the selected solution was repeated in all methods as shown in Table 5.

5. Methods Comparison and Discussions

A firm's collaborative network, that consists of a focal firm and its relatives, has a significant role in the success of the members and also is an outstanding source of competitive advantage for the firms. In summary, the formation of a collaborative network is a long-term decision and therefore wants a careful selection process. Thus, the decision should be made with enough allocated time and budget to lead to good results.

The proposed framework was implemented in IUMS project and the results show that by consideration of such a network, the costs and project time can be reduced significantly. In addition, by this plan, the managerial load was decreased and the flexibility, in addressing the environmental and technological changes, was increased. The aim of the present work is to extend the current knowledge of collaborative network building. Previous works based on problem scope and employed tools could be classified into two major categories: (1) the works which concentrated on partner or alliance selection and often provide mathematical models and (2) the works which focused on network level but often they are descriptive and qualitative, and there are a handful of works which provide step wise processes for network building. Thereupon, according to lack of academic efforts which concentrate on network level and provide mathematical tools, this paper tried to (1) focus on network level, (2) investigate qualitative aspects of problem and provide systematic managerial tool, and (3) provide a mathematical decision-making tool which can assess all possible solutions and lead to precise and reliable solution. These three points are the main difference points which crucially distinguished this work from the previous efforts.

The solutions attained from different methods were shown in Figure 8. Indeed, the "weighting method" is easy to use and produce all possible optimal solutions. "Lexicographic" has a high rate of flexibility and can incorporate detailed opinion of DMs through its flexible parameters. Thus, "lexicographic" and "weighting" methods produce very wide range of Pareto optimal solutions and decision on this wide range is difficult. The "global criterion" has only one parameter to set and it is easy to use, but DMs cannot reflect their judgement through this one parameter, and, also, the results of this method are very dependent on the parameter "P," whereas different "P" parameter provide totally different and wide solutions. The best methods are "SIMOLP" and "Satisfactory goals" because they are easy to use and provide high-quality solutions in this problem. They also appropriately consider RBV objective as a long-term issue. SIMOLP is an iterative method and some iterations are needed to achieve the solution, whereas "satisfactory goals" reach the solution in one step. Indeed, these two methods are interactive methods which incorporate the DMs preferences interactively, and also, the problem is based on DMs trade-offs. Thus, the suitable trade-off can be reached through interactive parameter setting. In this problem, the results produced by SIMOLP and "satisfactory goals" was in accordance with the DMs final trade-off, and as, the satisfactory goal was more easy to use and reach the solution in one step it was found as most suitable method for this problem in the case study. In other words, the "satisfactory goal" method is in compliance with nature of the problem and DMs have a suitable sense about its way of achieving solutions. The "satisfactory goal" in combination with one of "weighting" or "lexicographic" methods leads to much appropriate method, because the "weighting" or "lexicographic" method provides a big picture of all solutions and clears the problem space and this knowledge helps the DMs to easily set the parameters of "satisfactory goal" method and find the appropriate solution.

Noticeably, the framework entails a learning process that occurs for the participants in focus groups. Indeed, the decision-maker's points of view were evolved during the process. Moreover, the framework tried to structure the criteria in managers' mind and clarify the ambiguity of the problem. At the early stages of the process, the managers tend to build the network with possible lowest cost and implicitly ignore the long-terms issues but, at the end, their view was moderated and they compromised between two objectives. The cost-based issues that were revealed in TCT theory are more important in temporal networks which were created for performing temporary tasks, whereas long-term issues and RBV objective are more suitable for permanent networks in which the hub tends to form the network based on members' resources and capabilities. In such a network, time limits are not important and the networks formed over times. In this problem, the network is not temporal and also there are time and cost limits and the hub should accomplish the task in a predefined time window. Thus, it is essential to compromise between the two objectives.

Table 5: Selected solution in all methods.

Weighting method	Global criterion	Lexicographic	Satisfactory goals	SIMOLP
$0.2 \leq \alpha < 0.6$	$P = 1$	Preferred obj. = RBV, RBV ≥ 2.2	TCT ≥ 1.8	First iteration solution

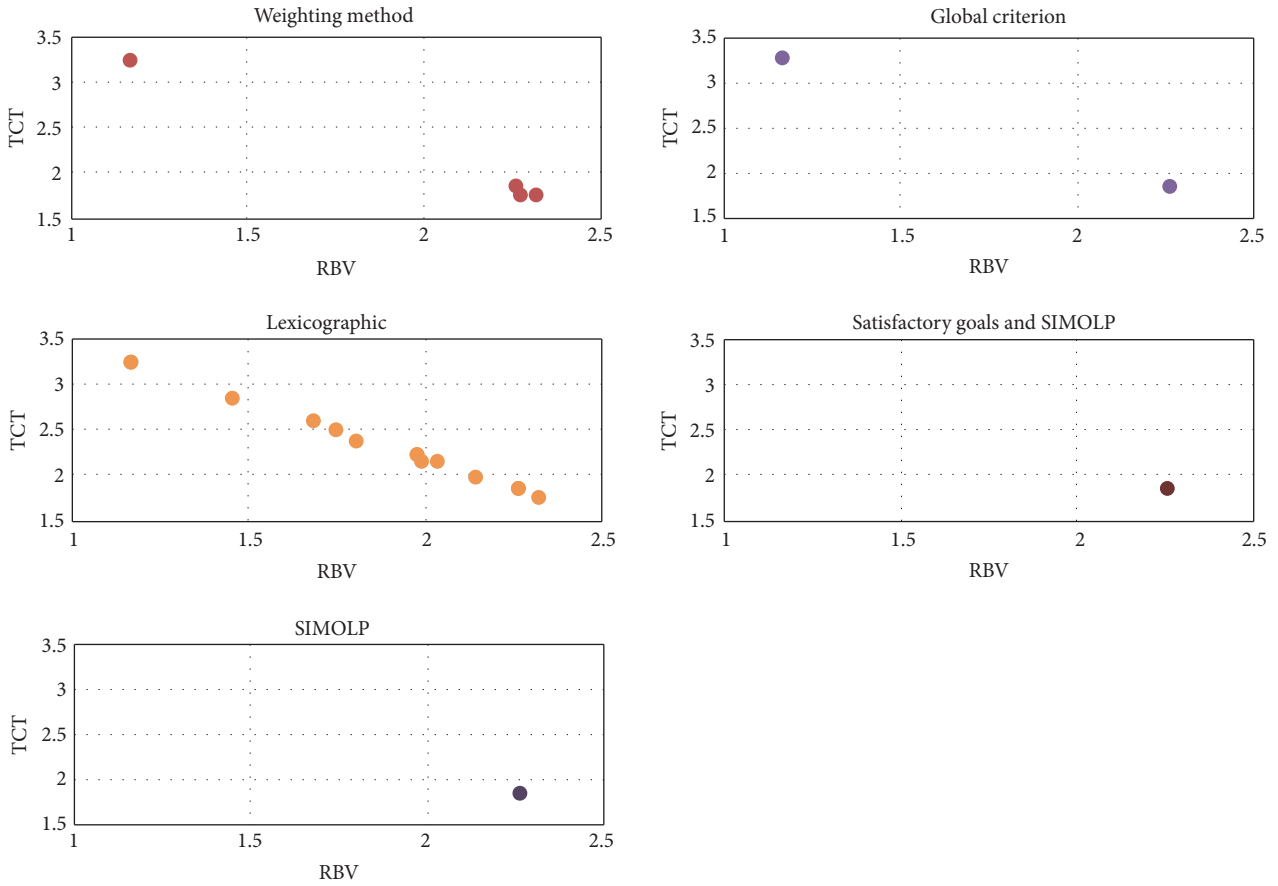


Figure 8: Comparison of solutions attained from different methods.

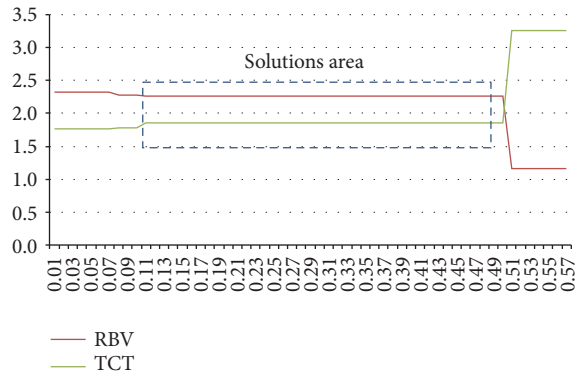


Figure 9: RBV and TCT objective values based on different weighting method.

The solutions obtained from the weighting method are shown in Figure 9, and selected solution based on the decision-makers trade-off was illustrated in it. The horizontal axis is the weight (Alpha) and the vertical is the objectives' values. The increase in horizontal axis is equal to increase in alpha and it means that the DMs prefer the TCT objective. In other words, the horizontal axis shows a spectrum in which at one hand the total preference is dedicated to TCT and on the other hand to RBV. It means that the solutions showed on the right-hand side of the figure are equivalent to the solution that is obtained by solving a one objective problem with TCT objective, and the solution illustrated in the left-hand side of the figure is equivalent to solving the problem singly with RBV objective. The selected solution depicts the compromise between two objectives and is strictly different from the solutions related to one objective problem. The selected solution's cost is the second desirable, but its RBV value is much closer to the optimum value. So the DMs accept higher costs to achieve a higher value of RBV objective and this is another justification of the selected solution. The selected solution is achieved by weights between 0.2 and 0.6.

6. Conclusions

This paper provides an analytic method that consists of a hybrid AHP and multiobjective programing and

a managerial framework for collaborative network building process by structuring the various aspects of the problem in a step by step procedure. In addition, five various multiobjective methods applied to the problem and their performances were compared. The research findings provide a more robust solution for selecting network alternatives on the basis of cost and time and speed of product development. As declared in Discussion, the cost of project reduced significantly in comparison with the plan A of the project (in which the project was conducted internally). This research has made the contributions into two areas, namely, (1) conceptual contribution and (2) methodological contribution.

Conceptual Contribution. This research provides a combinatorial theory of interorganizational relationships. The two contemporary theories of interorganizational relationships were combined. The first is the transaction cost theory which stands for short-term and financial benefits of firm, and the second is resource-based theory that stands for the long-term and nonfinancial benefits of firms. On one hand, the managers tend to accomplish their tasks with lowest possible cost and on the other hand they prefer to maintain their long-term advantages by developing sustainable capabilities which are in contrast to low-cost decisions. The outcome of the combination is more robust as it complements the weaknesses of the individual theories.

Methodological Contribution. The research proposes a mathematical model combining AHP and multiobjective programming with the aim to structure the various aspects of the problem and to model it in a systematic way. One important advantage of the framework is that it tries to provide a holistic solution for a firm's portfolio of relations, instead of optimizing it at a dyadic level. This is important because a piecemeal approach to solving each independent relation would restrict optimum solution. Also, a holistic solution overcomes the phenomenon of lock-in and lock-out effect that arises due to the presence of one or more dyadic relationships in the system. Secondly, the mathematical model is a hybrid AHP and multiobjective programming model which considers time, cost, and the internal relation of activities. The model made it possible to consider all possible solutions systematically and to be assured that selected solution is the best one based on available knowledge.

6.1. Limitations and Future Research Directions. Although this research contributes to the network formation literature, it has some limitations. First, the most important limitation of this research is about generalization. The proposed framework was implemented in a case study in IT industry, but it needs to be strengthened in other longitudinal studies, other case studies, and in other cross-industry studies. Second, the framework needs a huge volume of data that should be provided by participant experts. Although the network formation problem is a critical and long-term decision and this volume of data is completely logical, it causes boredom of participants and consequently decreases the data accuracy. Third, although a hybrid theory of RBV and TCT was employed in the framework, more efforts are needed to establish a new, comprehensive, hybrid theory; and there are many justifications and implications that should be discussed in the new hybrid theory. Forth, although the effects of the relationships on each other were considered on the model, the more dynamic analysis can be done in the future works. One of the research directions for the future works is to employ fuzzy variables in the model to control the ambiguity of the data. Another research direction is to do deep longitudinal and multiple case studies to strength it and discuss the generalizability of the model. Furthermore, the hybrid theory of TCT and RBV can be employed for governance of the interorganizational relationships and not just for formation phase.

Competing Interests

The authors declare that there are no competing interests regarding the publication of this paper.

References

[1] R. Gulati, N. Nohria, and A. Zaheer, "Strategic networks," *Strategic Management Journal*, vol. 21, no. 3, pp. 203–215, 2000.

[2] J. Tikkanen and A. Halinen, *Network Approach to Strategic Management—Exploration to the Emerging Perspective*, 2003.

[3] K. Möller, A. Rajala, and S. Svahn, "Strategic business nets—their type and management," *Journal of Business Research*, vol. 58, no. 9, pp. 1274–1284, 2005.

[4] L. M. Camarinha-Matos and H. Afsarmanesh, "Collaborative networks: a new scientific discipline," *Journal of Intelligent Manufacturing*, vol. 16, no. 4-5, pp. 439–452, 2005.

[5] L. M. Camarinha-Matos, H. Afsarmanesh, N. Galeano, and A. Molina, "Collaborative networked organizations—concepts and practice in manufacturing enterprises," *Computers & Industrial Engineering*, vol. 57, no. 1, pp. 46–60, 2009.

[6] J. H. Dyer and H. Singh, "The relational view: cooperative strategy and sources of interorganizational competitive advantage," *Academy of Management Review*, vol. 23, no. 4, pp. 660–679, 1998.

[7] K. K. Möller and A. Halinen, "Business relationships and networks: managerial challenge of network era," *Industrial Marketing Management*, vol. 28, no. 5, pp. 413–427, 1999.

[8] B. Axelsson and G. Easton, *Industrial Networks: A New View of Reality*, Routledge, London, UK, 1992.

[9] I. Snehota and H. Hakansson, *Developing Relationships in Business Networks*, Routledge Londres, 1995.

[10] A. Parmigiani and M. Rivera-Santos, "Clearing a path through the forest: a meta-review of interorganizational relationships," *Journal of Management*, vol. 37, no. 4, pp. 1108–1136, 2011.

[11] T. H. Davenport, M. Leibold, and S. C. Voelpel, *Strategic Management in the Innovation Economy: Strategic Approaches and Tools for Dynamic Innovation Capabilities*, John Wiley & Sons, New York, NY, USA, 2007.

[12] D. J. Brass, J. Galaskiewicz, H. R. Greve, and W. Tsai, "Taking stock of networks and organizations: a multilevel perspective," *Academy of Management Journal*, vol. 47, no. 6, pp. 795–817, 2004.

[13] T. Ritter, I. F. Wilkinson, and W. J. Johnston, "Managing in complex business networks," *Industrial Marketing Management*, vol. 33, no. 3, pp. 175–183, 2004.

[14] L. M. Camarinha-Matos and H. Afsarmanesh, "Classes of collaborative networks," in *Encyclopedia of Networked and Virtual Organizations*, Information Science Reference, 2008.

[15] J. C. Jarillo, "On strategic networks," *Strategic Management Journal*, vol. 9, no. 1, pp. 31–41, 1988.

[16] T. E. Stuart and O. Sorenson, "Strategic networks and entrepreneurial ventures," *Strategic Entrepreneurship Journal*, vol. 1, no. 3-4, pp. 211–227, 2007.

[17] M. Kohtamäki, S. Thorgren, and J. Wincent, "Organizational identity and behaviors in strategic networks," *Journal of Business & Industrial Marketing*, vol. 31, no. 1, pp. 36–46, 2016.

[18] B. Tan, "Modelling and analysis of a network organization for cooperation of manufacturers on production capacity," *Mathematical Problems in Engineering*, vol. 2006, Article ID 85103, 24 pages, 2006.

[19] S. C. Henneberg, S. Mouzas, and P. Naudé, "Network pictures: concepts and representations," *European Journal of Marketing*, vol. 40, no. 3-4, pp. 408–429, 2006.

[20] K. Möller and A. Rajala, "Rise of strategic nets—new modes of value creation," *Industrial Marketing Management*, vol. 36, no. 7, pp. 895–908, 2007.

[21] S. Mouzas, S. Henneberg, and P. Naudé, "Developing network insight," *Industrial Marketing Management*, vol. 37, no. 2, pp. 167–180, 2008.

[22] F. Cruijssen, P. Borm, W. Dullaert, and H. Hamers, "A versatile framework for cooperative hub network development," *European Journal of Industrial Engineering*, vol. 4, no. 2, pp. 210–227, 2010.

[23] J. Partanen and K. Möller, "How to build a strategic network: a practitioner-oriented process model for the ICT sector," *Industrial Marketing Management*, vol. 41, no. 3, pp. 481–494, 2012.

[24] N. J. Pulles, J. Veldman, and H. Schiele, "Identifying innovative suppliers in business networks: An Empirical Study," *Industrial Marketing Management*, vol. 43, no. 3, pp. 409–418, 2014.

[25] M. Nippa, S. Beechler, and A. Klossek, "Success factors for managing international joint ventures: a review and an integrative framework," *Management and Organization Review*, vol. 3, no. 2, pp. 277–310, 2007.

[26] W. Shi, J. Sun, and J. E. Prescott, "A temporal perspective of merger and acquisition and strategic alliance initiatives: review and future direction," *Journal of Management*, vol. 38, no. 1, pp. 164–209, 2012.

[27] J. Christoffersen, "A review of antecedents of international strategic alliance performance: synthesized evidence and new directions for core constructs," *International Journal of Management Reviews*, vol. 15, no. 1, pp. 66–85, 2013.

[28] J. R. Brown, A. S. Krishen, and C. S. Dev, "The role of ownership in managing interfirm opportunism: a dyadic study," *Journal of Marketing Channels*, vol. 21, no. 1, pp. 31–42, 2014.

[29] E. Gomes, B. R. Barnes, and T. Mahmood, "A 22 year review of strategic alliance research in the leading management journals," *International Business Review*, vol. 25, no. 1, pp. 15–27, 2016.

[30] P. C. van Fenema and C. Loebbecke, "Towards a framework for managing strategic tensions in dyadic interorganizational relationships," *Scandinavian Journal of Management*, vol. 30, no. 4, pp. 516–524, 2014.

[31] J. M. Whipple, R. Wiedmer, and K. K. Boyer, "A dyadic investigation of collaborative competence, social capital, and performance in buyer–supplier relationships," *Journal of Supply Chain Management*, vol. 51, no. 2, pp. 3–21, 2015.

[32] B.-G. Son, C. Kocabasoglu-Hillmer, and S. Roden, "A dyadic perspective on retailer–supplier relationships through the lens of social capital," *International Journal of Production Economics*, vol. 178, pp. 120–131, 2016.

[33] D. Lavie, "Alliance portfolios and firm performance: a study of value creation and appropriation in the U.S. software industry," *Strategic Management Journal*, vol. 28, no. 12, pp. 1187–1212, 2007.

[34] P. Ozcan and K. M. Eisenhardt, "Origin of alliance portfolios: entrepreneurs, network strategies, and firm performance," *Academy of Management Journal*, vol. 52, no. 2, pp. 246–279, 2009.

[35] T. J. Vapola, M. Paukku, and M. Gabrielsson, "Portfolio management of strategic alliances: an international business perspective," *International Business Review*, vol. 19, no. 3, pp. 247–260, 2010.

[36] U. Wassmer, "Alliance portfolios: a review and research agenda," *Journal of Management*, vol. 36, no. 1, pp. 141–171, 2010.

[37] I. Castro, C. Casanueva, and J. L. Galán, "Dynamic evolution of alliance portfolios," *European Management Journal*, vol. 32, no. 3, pp. 423–433, 2014.

[38] J. Schmutzler, R. Gutiérrez, E. Reficco, and P. Márquez, "Evolution of an alliance portfolio to develop an inclusive business," in *Social Partnerships and Responsible Business: A Research Handbook*, pp. 143–163, 2014.

[39] R. Gutiérrez, P. Márquez, and E. Reficco, "Configuration and development of alliance portfolios: a comparison of same-sector and cross-sector partnerships," *Journal of Business Ethics*, vol. 135, no. 1, pp. 55–69, 2016.

[40] S. Haider and F. Mariotti, "The orchestration of alliance portfolios: the role of alliance portfolio capability," *Scandinavian Journal of Management*, vol. 32, no. 3, pp. 127–141, 2016.

[41] U. Lichtenthaler, "Alliance portfolio capability: a conceptual framework for the role of exploration or exploitation alliances," *Journal of Strategy and Management*, vol. 9, no. 3, pp. 281–301, 2016.

[42] H. Håkansson and I. Snehota, "No business is an island: the network concept of business strategy," *Scandinavian Journal of Management*, vol. 5, no. 3, pp. 187–200, 1989.

[43] K. Möller, "Theory map of business marketing: relationships and networks perspectives," *Industrial Marketing Management*, vol. 42, no. 3, pp. 324–335, 2013.

[44] Y. Bramoullé, R. Kranton, and M. D'Amours, "Strategic interaction and networks," *The American Economic Review*, vol. 104, no. 3, pp. 898–930, 2014.

[45] J. Popp, G. MacKean, A. Casebeer, H. Milward, and R. Lindstrom, *Inter-Organizational Networks: A Review of the Literature to Inform Practice*, Collaborating Across Boundaries Series, IBM Center for the Business of Government, Washington, DC, USA, 2014.

[46] H. Hoang and A. Yi, "Network-based research in entrepreneurship: a decade in review," *Foundations and Trends in Entrepreneurship*, vol. 11, no. 1, pp. 1–54, 2015.

[47] B. R. Barringer and J. S. Harrison, "Walking a tightrope: creating value through interorganizational relationships," *Journal of Management*, vol. 26, no. 3, pp. 367–403, 2000.

[48] J. H. Dyer, "Effective interfirm collaboration: how firms minimize transaction costs and maximize transaction value," *Strategic Management Journal*, vol. 18, no. 7, pp. 535–556, 1997.

[49] A. Madhok and S. B. Tallman, "Resources, transactions and rents: managing value through interfirm collaborative relationships," *Organization Science*, vol. 9, no. 3, pp. 326–339, 1998.

[50] O. E. Williamson, "Comparative economic organization: the analysis of discrete structural alternatives," *Administrative Science Quarterly*, vol. 36, no. 2, pp. 269–296, 1991.

[51] T. K. Das and B.-S. Teng, "A resource-based theory of strategic alliances," *Journal of Management*, vol. 26, no. 1, pp. 31–61, 2000.

[52] M. J. Leiblein and D. J. Miller, "An empirical examination of transaction-and firm-level influences on the vertical boundaries of the firm," *Strategic Management Journal*, vol. 24, no. 9, pp. 839–859, 2003.

[53] T. R. Holcomb and M. A. Hitt, "Toward a model of strategic outsourcing," *Journal of Operations Management*, vol. 25, no. 2, pp. 464–481, 2007.

[54] E. T. Penrose, *The Theory oí the Growth oí the Firm*, John Wiley & Sons, New York, NY, USA, 1959.

[55] B. Wernerfelt, "A resource-based view of the firm," *Strategic Management Journal*, vol. 5, no. 2, pp. 171–180, 1984.

[56] J. Barney, "Firm resources and sustained competitive advantage," *Journal of Management*, vol. 17, no. 1, pp. 99–120, 1991.

[57] D. J. Teece, G. Pisano, and A. Shuen, "Dynamic capabilities and strategic management," *Strategic Management Journal*, vol. 18, no. 7, pp. 509–533, 1997.

[58] K. Hafeez, Y. Zhang, and N. Malak, "Determining key capabilities of a firm using analytic hierarchy process," *International Journal of Production Economics*, vol. 76, no. 1, pp. 39–51, 2002.

[59] K. Hafeez, Y. Zhang, and N. Malak, "Core competence for sustainable competitive advantage: a structured methodology for identifying core competence," *IEEE Transactions on Engineering Management*, vol. 49, no. 1, pp. 28–35, 2002.

[60] J. Kraaijenbrink, J.-C. Spender, and A. J. Groen, "The resource-based view: a review and assessment of its critiques," *Journal of Management*, vol. 36, no. 1, pp. 349–372, 2010.

[61] E. W. K. Tsang, "Transaction cost and resource-based explanations of joint ventures: a comparison and synthesis," *Organization Studies*, vol. 21, no. 1, pp. 215–242, 2000.

[62] O. E. Williamson, "Strategy research: governance and competence perspectives," *Strategic Management Journal*, vol. 20, no. 12, pp. 1087–1108, 1999.

[63] W. Yue, Y. Wang, and C. Dai, "An evolutionary algorithm for multiobjective fuzzy portfolio selection models with transaction cost and liquidity," *Mathematical Problems in Engineering*, vol. 2015, Article ID 569415, 15 pages, 2015.

[64] K. M. Eisenhardt and C. B. Schoonhoven, "Resource-based view of strategic alliance formation: strategic and social effects in entrepreneurial firms," *Organization Science*, vol. 7, no. 2, pp. 136–150, 1996.

[65] F. David, L.-E. Gadde, and S. Ivan, *Managing Business Relationship*, John Wiley & Sons, West Sussex, UK, 1998.

[66] F. R. Dwyer, P. H. Schurr, and S. Oh, "Developing buyer-seller relationships," *Journal of Marketing*, vol. 51, no. 2, pp. 11–27, 1987.

[67] A. Larson, "Network dyads in entrepreneurial settings: a study of the governance of exchange relationships," *Administrative Science Quarterly*, vol. 37, no. 1, pp. 76–104, 1992.

[68] D. T. Wilson, "An integrated model of buyer-seller relationships," *Journal of the Academy of Marketing Science*, vol. 23, no. 4, pp. 335–345, 1995.

[69] I. Abosag and P. Naudé, "Development of special forms of B2B relationships: examining the role of interpersonal liking in developing Guanxi and Et-Moone relationships," *Industrial Marketing Management*, vol. 43, no. 6, pp. 887–896, 2014.

[70] A. V. Gumerov, L. E. Fatikhova, K. O. Kolpakov et al., "Features of forming an interaction network of an entrepreneurial structure," *Review of European Studies*, vol. 7, no. 4, pp. 209–215, 2015.

[71] S. Lacoste and R. E. Johnsen, "Supplier–customer relationships: a case study of power dynamics," *Journal of Purchasing and Supply Management*, vol. 21, no. 4, pp. 229–240, 2015.

[72] T. Mandják, Z. Szalkai, E. Neumann-Bódi, M. Magyar, and J. Simon, "Emerging relationships: how are they born?" *Industrial Marketing Management*, vol. 49, pp. 32–41, 2015.

[73] T. Mandják, Z. Szalkai, E. Neumann-Bódi, M. Magyar, and J. Simon, "Trigger issues in emerging relationships," *Industrial Marketing Management*, vol. 58, pp. 137–147, 2016.

[74] P. Hennelly and C. Y. Wong, "The formation of new inter-firm relationships: a UK offshore wind sector analysis," *International Journal of Energy Sector Management*, vol. 10, no. 2, pp. 172–190, 2016.

[75] M. Ebers, "Explaining inter-organizational network formation," *The Formation of Inter-Organizational Networks*, vol. 1, pp. 3–40, 1997.

[76] P. A. David and L. C. Keely, "The economics of scientific research coalitions: collaborative network formation in the presence of multiple funding agencies," in *Science and Innovation: Rethinking the Rationales for Funding and Governance*, pp. 251–308, Edward Elgar Publishing, Cheltenham, UK, 2003.

[77] A. Hinterhuber, "Value chain orchestration in action and the case of the global agrochemical industry," *Long Range Planning*, vol. 35, no. 6, pp. 615–635, 2002.

[78] M. A. Gaspar, R. S. Silva, M. C. Farina, and J. P. L. de Siqueira, "Inter-organizational relationships: promoters and restrictive factors in the formation of cooperation network," *Revista Eletrônica de Estratégia & Negócios*, vol. 7, no. 1, pp. 213–241, 2014.

[79] A. Hale and L. White, "Developing a framework to establish collaborative enterprise networks," in *Proceedings of the 8th Annual IEEE International Systems Conference (SysCon '14)*, pp. 431–438, IEEE, Ottawa, Canada, April 2014.

[80] A. Jussila, T. Mainela, and S. Nätti, "Formation of strategic networks under high uncertainty of a megaproject," *Journal of Business & Industrial Marketing*, vol. 31, no. 5, pp. 575–586, 2016.

[81] J. Zhang and A. Pezeshkan, "Host country network, industry experience, and international alliance formation: evidence from the venture capital industry," *Journal of World Business*, vol. 51, no. 2, pp. 264–277, 2016.

[82] K. Keahey, I. Foster, T. Freeman, and X. Zhang, "Virtual workspaces: achieving quality of service and quality of life in the Grid," *Scientific Programming*, vol. 13, no. 4, pp. 265–275, 2005.

[83] M. H. Danesh, B. Raahemi, S. A. Kamali, and G. Richards, "A framework for process and performance management in service oriented virtual organisations," *International Journal of Computer Information Systems and Industrial Management Applications*, vol. 5, pp. 203–215, 2013.

[84] J. Crispim, N. Rego, and J. Pinho de Sousa, "Stochastic partner selection for virtual enterprises: a chance-constrained approach," *International Journal of Production Research*, vol. 53, no. 12, pp. 3661–3677, 2015.

[85] W. Lin, W. Wu, and J. Z. Wang, "A heuristic task scheduling algorithm for heterogeneous virtual clusters," *Scientific Programming*, vol. 2016, Article ID 7040276, 10 pages, 2016.

[86] M. E. Porter, "Competitive strategy," *Measuring Business Excellence*, vol. 1, no. 2, pp. 12–17, 1997.

[87] C. Parolini, *The Value Net: A Tool for Competitive Strategy*, Wiley, Chichester, UK, 1999.

[88] K. Hafeez, Y. Zhang, and N. Malak, "Identifying core competence," *IEEE Potentials*, vol. 21, no. 2, pp. 2–7, 2002.

[89] K. Hafeez, N. Malak, and Y. B. Zhang, "Outsourcing non-core assets and competences of a firm using analytic hierarchy process," *Computers and Operations Research*, vol. 34, no. 12, pp. 3592–3608, 2007.

[90] P. Fiala, "An ANP/DNP analysis of economic elements in today's world network economy," *Journal of Systems Science and Systems Engineering*, vol. 15, no. 2, pp. 131–140, 2006.

[91] E. Deelman, G. Singh, M.-H. Su et al., "Pegasus: a framework for mapping complex scientific workflows onto distributed systems," *Scientific Programming*, vol. 13, no. 3, pp. 219–237, 2005.

[92] C.-L. Hwang and A. S. M. Masud, in *Multiple Objective Decision Making—Methods and Applications*, Springer, Berlin, Germany, 1979.

[93] J. Branke, K. Deb, K. Miettinen, and R. Slowiński, Eds., *Multi-objective Optimization: Interactive and Evolutionary Approaches*, vol. 5252 of *Theoretical Computer Science and General Issues*, Springer Science & Business Media, 2008.

[94] G. R. Reeves and L. S. Franz, "A simplified interactive multiple objective linear programming procedure," *Computers & Operations Research*, vol. 12, no. 6, pp. 589–601, 1985.

Hindawi Publishing Corporation
Scientific Programming
Volume 2016, Article ID 3279423, 15 pages
http://dx.doi.org/10.1155/2016/3279423

Research Article

Modeling and Optimization of the Drug Extraction Production Process

Dakuo He,[1,2] Zhengsong Wang,[1] Le Yang,[1] Tongshan Liu,[1] Yao Yao,[1] and Zhizhong Mao[1,2]

[1]*College of Information Science and Engineering, Northeastern University, Shenyang, Liaoning 110004, China*
[2]*State Key Laboratory of Synthetical Automation for Process Industries, Northeastern University, Shenyang 110004, China*

Correspondence should be addressed to Dakuo He; hedakuo@ise.neu.edu.cn

Received 17 June 2016; Revised 31 August 2016; Accepted 14 September 2016

Academic Editor: Chengyan Yue

Optimized control of the drug extraction production process (DEPP) aims to reduce production costs and improve economic benefit while meeting quality requirements. However, optimization of DEPP is hampered by model uncertainty. Thus, in this paper, a strategy that considers model uncertainty is proposed. Mechanistic modeling of DEPP is first discussed in the context of previous work. The predictive model used for optimization is then developed by simplifying the mechanism. Optimization for a single extraction process is first implemented, but this is found to lead to serious wastage of herbs. Hence, the optimization of a multiextraction process is then conducted. To manage the uncertainty in the model, a data-driven iterative learning control method is introduced to improve the economic benefit by adjusting the operating variables. Finally, fuzzy parameter adjustment is adopted to enhance the convergence rate of the algorithm. The effectiveness of the proposed modeling and optimization strategy is validated through a series of simulations.

1. Introduction

Drug extraction, or drug leaching, is one of the most significant operations in the pharmaceutical industry. As the basic and primary process of drug production, it has been widely applied to many medicinal plants [1–7]. In the drug extraction production process (DEPP), a solvent or chemical reagent that offers high solubility for the effective constituent (EC) in the herbs and poor solubility for constituents that need not be extracted is applied to solid herbs. The EC then dissolves out of the herb organization and into the solvent or chemical reagent [8]. The extracted EC is finally used in various types of drugs, such as granules, tablets, and capsules, in subsequent pharmaceutical processes.

The disadvantages of this production process include low extraction yield, wastage of materials, and high energy consumption. Modeling and process optimization, which aim to increase the extraction yield and minimize production costs while meeting specific quality requirements, are therefore particularly significant in both theoretical research and practical applications.

The optimization of drug extraction process has received increasing attention. For example, Alam et al. [9] studied the optimization of the extraction parameters of embelin from *Embelia ribes* through ultrasound-assisted extraction with a Box–Behnken design. Bochi et al. [10] optimized the extraction conditions and anthocyanin yields in experiments with high proportions of water. Chen et al. [11] applied an orthogonal experiment to optimize the extraction conditions of polysaccharides from *Ornithogalum caudatum* Ait. Chen et al. [12] used the response surface methodology to optimize the experimental conditions for ultrasonic-assisted extraction of functional components from sugar beet molasses. Bae et al. [13] successfully developed and validated a simple qualitative and quantitative method for the simultaneous determination of 15 phenolic compounds and caffeine from teas and then optimized the extraction process using the response surface methodology based on a central composite design. In addition to these innovations, many other scholars have made tremendous contributions to the theoretical and practical optimization of DEPP [3, 14–16]. However, many of these reports focus on optimizing the extraction conditions

using experimental design or simulation techniques. To the best of the authors' knowledge, the optimization of DEPP in terms of improving the economic benefit of the whole production process has rarely been reported in the literature. The present work develops a mechanism model for DEPP based on some previous efforts and then proposes an optimization strategy to resolve practical problems such as the production efficiency and production costs.

To implement this optimization, a mechanism model of DEPP should be developed. Such mechanistic modeling plays several significant roles: first, the modeling process will analyze the mechanism of the drug extraction process, confirm its operating variables and quality indicators, and analyze the relationship between them; second, an appropriate model can be used to simulate the actual production process to generate the required data and analyze the optimization results. As an actual simulator, the mechanism model should be able to link the extraction efficiency of the EC and volatile oil—the quality indicators—with the operating variables (steam flow at the bottom and side of the extraction tank and the extraction time).

The mechanistic modeling of DEPP mainly consists of modeling the following components: extraction tank, mass transfer process of EC, volatile oil recycling, and efficiency of the oil-water separator. A mathematical model of the extraction tank that describes the changes in temperature and liquid level over time has been developed [17]. As the principle and mechanism of the mass transfer process in drug extraction are very similar to those of metal leaching, mechanistic models of metal leaching [3, 18–22] can be used as references in developing a mass transfer model for the EC. Additionally, Han et al. derived a diffusion rate equation for volatile oils, allowing a mechanistic model of volatile oil recovery to be developed [23]. Finally, the efficiency of the oil-water separator is introduced to describe the separation efficiency of oil droplets with a certain diameter [24]. The above mechanistic models are combined to simulate the whole DEPP, and the rationality and validity of this approach are verified through a series of simulations. The innovation and contribution of the DEPP model presented in this paper can be summarized as follows. First, we extend a modeling approach for the leaching process to model the mass transfer process of ECs. Second, we present a combined mechanistic modeling framework for the overall DEPP by integrating four discrete components described in previous studies.

In addition to a mechanistic model, a predictive model is also necessary for such optimization. In this work, the established mechanism is simplified to serve as a predictive model. The economic benefit per unit time is considered as the optimization objective, while the steam flow at the bottom and side of the extraction tank and the extraction time are treated as decision variables. The optimization model is then established under certain quality constraints. Several classical optimization algorithms, such as particle swarm optimization (PSO) [25, 26], differential evolution (DE) [27], and artificial bee colony (ABC) [28], are adopted to solve this optimization problem. Comparing these algorithms, it is concluded that PSO achieves the best performance in determining the optimal solution. Thus, a PSO algorithm is used to solve this optimization problem and obtain the optimal economic benefit and corresponding parameters, including optimal quality indicators and operating variables. The present work first considers the optimization of a single extraction process. However, the extraction efficiency of the EC in this case is very low, leading to serious wastage of herb materials. The optimization control for multiextraction is therefore investigated. In the multiextraction process, the extraction frequency is defined as the number of times that the herbs are subjected to the extraction process. However, the optimal extraction frequency in terms of economic benefit is unknown. Thus, the extraction frequency is treated as one of the decision variables when establishing the optimization problem. Note that the extraction frequency is an integer in this problem.

The optimal economic benefit of the multiextraction process is based on the predictive model, whereas the actual optimal economic benefit can be calculated by plugging the best operating variables into the mechanism model. There are some discrepancies between the predicted and actual values of the optimal economic benefit. In practice, however, it is unlikely that the actual production process can be modeled accurately (or even approximately) with a process model [29, 30]. Process optimization is hampered by such model uncertainty, and so the "optimal" value given by the model may not necessarily mean "optimal for the process" [31].

Iterative learning control (ILC) is highly effective in controlling systems with repetitive operations that precisely follow a desired target trajectory. It has been widely applied in repetitive industrial processes because of its perfect learning ability from the repetitive tracking task [32–34]. Recently, data-driven ILC methods have been proposed to deal with complicated practical systems [35–37]. The control scheme is data-driven because there is no explicit model information or training process, and only the measured input and output data are used for the controller design, analysis, and implementation [36]. As one such method, data-driven optimal terminal ILC (DDOTILC) works under the principle that the control law is only updated using the terminal output tracking error [35]. In this work, DDOTILC is applied to implement iterative optimization control for DEPP in order to overcome the model uncertainty via the online adjustment of the operating variables. The fuzzy adaptive adjustment of parameters in DDOTILC is then implemented to enhance the convergence rate. Using this approach, the model uncertainty can be reduced and the economic benefit improved by adjusting the operating variables.

2. Principle and Production Technology of the Drug Extraction Process

2.1. Principle of the Drug Extraction Process

2.1.1. Mass Transfer Theory of Effective Constituent. Drug extraction is one kind of solid-liquid extraction. Thus, the mass transfer principle and computing method follow the solid-liquid extraction system model shown in Figure 1. It is assumed that the herb particles are composed of solute

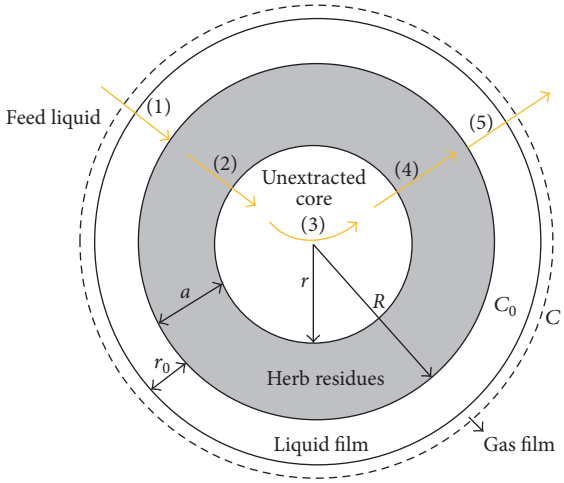


Figure 1: Unextracted core shrinking model of effective component.

(EC) and inert carriers (herb residues), so the solid-liquid extraction system is composed of solute, solvent, and inert solid. Additionally, there will be a gas-liquid film at the interface between the solid particles and liquid phase. It is generally acknowledged that this mass transfer process can be divided into the five steps described in Figure 1 [3, 8]:

(1) The diffusion of solvent into the herb particle surface

(2) The permeation of solvent from the herb surface to the interior

(3) The dissolution of the EC inside the herb particles

(4) The diffusion of EC from inside the herb particles to the surface (internal diffusion)

(5) The diffusion of EC from the herb particle surface to the solvent across the gas-liquid film (external diffusion).

2.1.2. Recycling Theory of Volatile Oil. Volatile oil (VO), also named essential oil, is a generic term for the volatile and oily components in plants. VOs are insoluble in water and can be distilled out along with the water vapor. During the extraction process, the recycling of VO is divided into the following three steps (see Figure 2):

(1) Internal diffusion: with increasing temperature, VOs diffuse into the outer surface of the solid herb particle from the interior

(2) External diffusion: VOs diffuse into the water vapor from the surface of the solid particle by threading the gas-liquid film

(3) Gas-liquid transition: VOs and water vapor enter the condenser and are condensed into the liquid phase, after which the VOs are separated from water though the oil-water separator.

2.2. Production Technology of the Drug Extraction Process. Generally, DEPP includes the following three procedures.

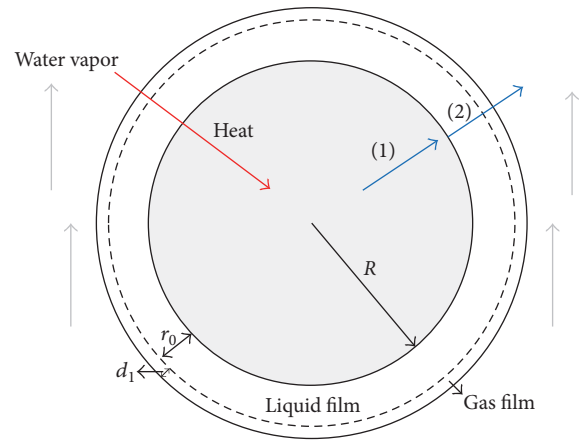


Figure 2: Schematic diagram of volatile oil recycling.

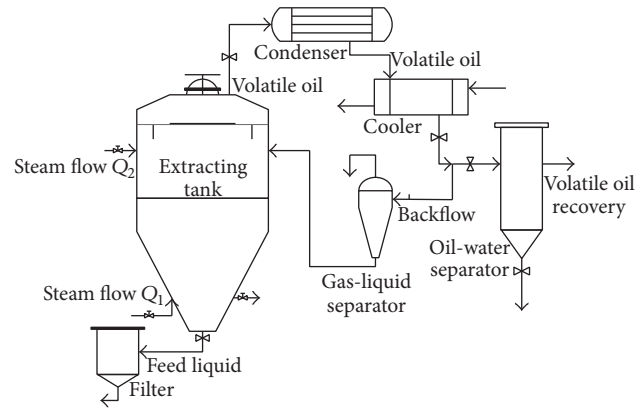


Figure 3: Diagram of equipment used in the extraction procedure.

(1) Preprocessing of Herb Materials. In this procedure, the herb materials are smashed and soaked in water for some time to inflate the herb organization. This preprocessing accelerates the dissolution and diffusion of EC during extraction.

(2) Extraction. As the most important process of DEPP, this procedure involves various pieces of equipment including the extraction tank, condenser, gas-liquid separator, oil-water separator, and filter (see Figure 3). In this process, the preprocessed herb materials are placed into the extraction tank, and steam is added through both the bottom and side of the tank. As the temperature increases, the valuable components of the herb particles, such as EC and VO, will continually diffuse into the extraction solution and water vapor, respectively. For a batch of herb materials, it is extremely difficult to achieve acceptable yields of EC and VO with a single extraction process. Therefore, multiple extractions are required to prevent the wastage of raw materials.

(3) Separation. In this procedure, the feed liquid containing EC and VO are separated from the herb residues and steam, respectively, using the filter and oil-water separator. The feed liquid and VO are then sent to the next phase for further processing.

The extraction efficiencies of EC and VO are the main quality indicators of DEPP. There are numerous factors influencing these indicators, such as the radius of the herb material particles, extraction temperature, extraction time, concentration difference, and solid-liquid contact condition. For the purpose of DEPP optimization, the present study explores the impact on the extraction efficiency of the steam flow at both the bottom and side of the extraction tank and the extraction time.

3. Mechanistic Modeling of Drug Extraction Production Process

DEPP is a complex heat and mass transfer process, and its mechanistic model mainly consists of the following components: mathematical model of drug extraction tank, mass transfer model of EC, mechanistic modeling of VO recycling, and an efficiency model of the oil-water separator. The mechanistic modeling work for DEPP is implemented based on the following assumptions:

(1) The herb particle has a spherical shape after crushing.

(2) The EC of herb materials in the solvent is distributed uniformly under the condition of full stirring.

(3) The EC in the outer layer of the herb particles is first spread to the solvent, after which the EC in the inner layer spreads to the outer layer and then to the solvent.

3.1. Mathematical Modeling of Drug Extraction Tank. A mechanistic model of the extraction tank has been developed according to the principles of mass and energy conservation [17]. In this model, the liquid level H and feed liquid temperature T in the extraction tank can be calculated using the steam flow at the bottom and side of the tank, Q_1, Q_2, and the extraction time t. The variation of liquid level and temperature with extraction time t can be expressed as

$$\frac{dH}{dt} = \frac{(Q_1 - Q_3) P_1}{A}, \tag{1}$$

$$\begin{aligned}\frac{dT}{dt} = {} & \frac{1}{AS_1H}\left[\left(P_2 - S_1TP_1 + S_2T_1 - S_2T\right)Q_1\right. \\ & \left. + S_2Q_2\left(T_1 - T\right) + \left(S_1TP_1 - P_2\right)Q_3\right].\end{aligned} \tag{2}$$

The mixture flow of VO and steam Q_3 is assumed to have a nonlinear relationship with T:

$$Q_3 = K_1 e^{K_2 T}. \tag{3}$$

The symbols used in this model are as follows: Q_1, Q_2: the steam flow at the bottom and side of the extraction tank, respectively; Q_3: mixture flow of VO and steam; T_1: steam temperature before entering the extraction tank; S_1: specific heat capacity of feed liquid; S_2: specific heat capacity of water; P_1: volume ratio of water (liquefied from steam) and steam; P_2: standard heat of liquefaction of vapor under standard atmospheric pressure; A: cross-sectional area of the extraction tank.

3.2. Mathematical Modeling of the Mass Transfer Process of Effective Constituent. The mass transfer process for EC was described in detail in Section 2, where it was noted that the transfer could be divided into five steps. In the usual extraction process, the penetration of solvent and dissolution of solute are not the controlling factors and can thus be neglected. The diffusion processes undoubtedly have a decisive effect on the speed of mass transfer. Hence, external diffusion and internal diffusion are primarily considered in the modeling of the mass transfer process.

External diffusion, also called liquid film diffusion, is the main resistance to mass transfer. The gas film will not hamper the mass transfer process, so it is neglected in the modeling process, as illustrated by the dashed line in Figure 1. Internal diffusion is the shrinking core process of particles containing EC, as shown in Figure 1. In this study, using the unreacted shrinking core theory from the metal ore leaching process, the drug extraction process can be approximately regarded as the leaching of metallic minerals, albeit under different reactions. When the EC in the outer layer has diffused into the extraction solvent, the insoluble components are left on the surface of the solid particles. Thus, the solid particles are surrounded by a grey layer composed of insoluble components, as shown in Figure 1. The inner EC then diffuses into the surface of the solid particles through internal diffusion, after which this part of the EC in the solid particle surface diffuses into the solvent once again. This procedure is repeated, and the unextracted core will gradually shrink.

The diffusion rate of the extraction process can be expressed as [3, 18–20]

$$\frac{dG}{dt} = JS, \tag{4}$$

where G is the mass of EC extracted from solid herb particles, J is the mass of EC diffused from solid herb particles to solvent per unit time, S expresses the area of the interface between the solid and liquid phase, and t is the extraction time.

It is assumed that internal diffusion and external diffusion occur in series, so the law of additive resistance can be applied as follows [3, 21, 22]:

$$J = \frac{C_0 - C}{1/\beta_1 + 1/\beta_2} = \frac{C_0 - C}{(R - r)/D_s + r_0/D}, \tag{5}$$

where C_0 is the solubility of EC in the liquid film on the surface of solid particles; C is the instantaneous concentration of EC in extraction solvent; D and D_s express the diffusion coefficients of EC passing through the liquid film and the inside of solid particles, respectively; R and r denote the radius of the original solid particle and the solid particle as the core gradually shrinks, respectively; and r_0 expresses the thickness of the liquid film (see Figure 1). In addition, β_1 is the coefficient of internal diffusion ($1/\beta_1$ is the internal diffusion resistance) and β_2 is the coefficient of external diffusion ($1/\beta_2$

is the external diffusion resistance). The parameters β_1 and β_2 are given by

$$\beta_1 = \frac{D_s}{R-r},$$
$$\beta_2 = \frac{D}{r_0}. \tag{6}$$

Combined with the model of the extraction tank, it can be concluded that

$$\frac{dG}{dt} = \frac{d(VC)}{dt} = \frac{d(AHC)}{dt} = AH\frac{dC}{dt} + AC\frac{dH}{dt}. \tag{7}$$

Based on (1), (4), (5), and (7), we have that

$$\frac{dC}{dt} = \frac{1}{AH}\left[\frac{(C_0 - C)S}{(R-r)/D_s + r_0/D} - CP_1(Q_1 - Q_3)\right]. \tag{8}$$

The diffusion coefficients can then be determined as follows.

(1) For the solution with macromolecules of EC, the external diffusion coefficient is obtained by

$$D = BT\frac{\nu_A}{f_A}. \tag{9}$$

The resistance of EC sustained in solution f_A can be calculated by Stokes' equation:

$$f_A = 6\pi R\mu_B\nu_A. \tag{10}$$

Combining (9) and (10), we obtain

$$D = \frac{BT}{6\pi R\mu_B}, \tag{11}$$

where B is the Boltzmann constant, ν_A is the kinematic velocity of EC molecules, and μ_B is the viscosity of the solvent.

(2) For the dilute solution with micromolecules of EC, the external diffusion coefficient is generally expressed by Wilke's equation [38]:

$$D = 7.4\times10^{-12}(\psi M_B)^{0.5}\frac{T}{\mu_B V_A^{0.6}}, \tag{12}$$

where ψ is the associating parameter of the solvent, M_B is the molecular weight of the solvent, and V_A is the molecular volume of EC. In this work, (12) is used to calculate the diffusion coefficient D.

(3) The internal diffusion coefficient of EC inside the solid particle D_s is calculated by [39]

$$D_s = \varepsilon\frac{D}{\tau}, \tag{13}$$

where ε and τ are the porosity and sinuosity of holes inside the solid particles, respectively.

Another equation connects the radius of solid particles containing EC with the internal diffusion coefficient D_s, the concentration of EC in the liquid film C_0, and the instantaneous concentration of EC in extraction solvent C [39]:

$$r = R - \left(\frac{kD_sCt}{C_0}\right)^{1/2}. \tag{14}$$

3.3. Mechanistic Modeling of Volatile Oil Recycling. The recycling principle of VO was described in Section 2. Because the internal diffusion occurs in the same phase, it can be neglected. However, the external diffusion process, in which VO diffuses into the gas phase (water vapor) from the surface of the solid phase (solid particle) across the gas-liquid film, is interphase diffusion. This external diffusion is the main factor controlling the diffusion velocity. Because the VO is insoluble in water, its explicit diffusion in the liquid phase can be ignored, as denoted by the dashed line in Figure 2. Consequently, the resistance to external diffusion mainly occurs in the gas film [40]. The diffusion velocity is expressed as [40]

$$\frac{dq}{dt} = K_GS(C_l - C_i), \tag{15}$$

where q is the concentration of VO in the solid herb particles, K_G is the gas phase mass transfer film coefficient, C_l is the concentration of VO in the gas phase, and C_i is the concentration of VO in the gas film. The following assumptions are made:

(1) Because the VO in the gas phase will constantly enter the condenser with water vapor, $C_l \approx 0$.

(2) There is a linear counterbalance between the concentrations of VO in the solid phase and gas film; that is, $C_i = Kq$ [23], where K is a constant coefficient.

Equation (15) can then be expressed as

$$\frac{dq}{dt} = -K_GSKq = -\frac{e^{k_0D}SK}{d_1}q, \tag{16}$$

where

$$K_G = \frac{e^{k_0D}}{d_1}, \tag{17}$$

in which k_0 is a constant coefficient and d_1 is the thickness of the gas film.

As for the liquid density, the density of VO will decrease as the feed liquid temperature T increases. The VO density can be calculated as

$$\rho_0 = A_1 + A_2T + A_3T^2 + A_4T^3 + A_5T^4, \tag{18}$$

where ρ_0 is the density of VO and A_1–A_5 are constant coefficients.

3.4. Mechanistic Modeling of the Efficiency of the Oil-Water Separator. Under the effect of gravity, the oil-water separator achieves the separation of oil and water in the static or flowing state by taking advantage of the density difference between oil and water. The kinetic state of a droplet in the disperse phase can be described by Stokes' law [24]:

$$\nu_{ow} = \frac{(\rho_w - \rho_0)gd_2^2}{18\mu_0}, \tag{19}$$

where ν_{ow} is the lifting velocity of the oil droplet, ρ_w is the density of water, g is acceleration due to gravity, d_2 is the diameter of the oil droplet, and μ_0 is the viscosity of water.

The separation efficiency of oil droplets of a certain diameter can be determined according to shallow pool theory [41] as

$$\eta_0 = \frac{S_0 v_{ow}}{Q_0} = \frac{S_0 g}{18\mu_0 Q_0}\left(\rho_w - \rho_0\right) d_2^2, \tag{20}$$

where η_0 is the lifting efficiency of oil droplets of diameter d_2, S_0 is the lifting area of the oil layer, and Q_0 is the flow of the oil-water mixture.

The separation efficiency η_s is a key performance indicator in any oil-water separator and can be expressed as follows:

$$\eta_s = 1 - \frac{C_d}{C_b}, \tag{21}$$

where C_d and C_b denote the oil content at the exit and entrance of the separator, respectively.

3.5. Processing the Mechanistic Model of the Drug Extraction Production Process. For the convenience of calculations, differential equations in the models are discretized to give a set of algebraic equations. Taking the model for the liquid level in the extraction tank, we have

$$\frac{H(k+1) - H(k)}{T_S} = \frac{\left(Q_1(k) - Q_3(k)\right) P_1}{A}, \tag{22}$$

where T_S is the sampling period.

Thus, the liquid level at sampling time $k+1$ is

$$H(k+1) = \frac{\left(Q_1(k) - Q_3(k)\right) P_1}{A} T_S + H(k), \tag{23}$$

where $Q_3(k) = K_1 e^{K_2 T(k)}$.

Similarly, the discretized model for the feed liquid temperature is

$$\begin{aligned} T(k+1) &= \frac{T_S}{AS_1 H(k)}\left[\left(P_2 - S_1 T(k) P_1 + S_2 T_1 - S_2 T(k)\right)\right. \\ &\cdot Q_1(k) + S_2 Q_2(k)\left(T_1 - T(k)\right) \\ &\left. + \left(S_1 T(k) P_1 - P_2\right) Q_3(k)\right] + T(k). \end{aligned} \tag{24}$$

The discretized model for the concentration of EC is

$$\begin{aligned} C(k+1) = \frac{T_S}{AH(k)}\left[\frac{\left(C_0 - C(k)\right) S}{\left(R - r(k)\right)/D_S(k) + r_0/D(k)}\right. \\ \left. - C(k) P_1\left(Q_1(k) - Q_3(k)\right)\right] + C(k), \end{aligned} \tag{25}$$

where

$$\begin{aligned} D(k) &= 7.4 \times 10^{-12}\left(\Phi M_B\right)^{0.5} \frac{T(k)}{\mu_B(k) V_A^{0.6}}, \\ D_S(k) &= \varepsilon \frac{D(k)}{\tau}, \\ r(k) &= R - \left(\frac{k D_S(k) C(k) t(k)}{C_0}\right)^{1/2}, \end{aligned} \tag{26}$$

and the discretized model for the concentration of VO is

$$q(k+1) = \left(1 - \frac{e^{k_0 D(k)} S K T_S}{d_1}\right) q(k). \tag{27}$$

It is assumed that the extraction time can be divided into six equal intervals, so the entire drug extraction process has 13 input variables: six steam flows at the bottom of the extraction tank ($Q_{11}, Q_{12}, Q_{13}, Q_{14}, Q_{15}, Q_{16}$), six steam flows at the side of the extraction tank ($Q_{21}, Q_{22}, Q_{23}, Q_{24}, Q_{25}, Q_{26}$), and the extraction time t. The output variables are the extraction yields of EC and VO (F_1, F_2, resp.), their extraction efficiencies (η_1, η_2), and the total consumption of steam F_3.

The extraction yield of EC F_1 and its extraction efficiency η_1 can be expressed as

$$\begin{aligned} F_1 &= f_1(k) = V(k) C(k) = AH(k) C(k), \\ \eta_1 &= \frac{F_1}{F_{1\,\text{total}}}, \end{aligned} \tag{28}$$

where $F_{1\,\text{total}}$ is the total mass of EC in a batch of herbs, which we consider to be 25 kg in this study.

The decrement of VO in the extraction tank between adjacent sampling periods is

$$\begin{aligned} \Delta m(k) &= m(k-1) - m(k) \\ &= \left(q(k-1) - q(k)\right) F_{2\,\text{total}}. \end{aligned} \tag{29}$$

Thus, the recycle yield of VO F_2 and its extraction efficiency η_2 are computed by the following equations:

$$\begin{aligned} F_2 &= f_2(k) = f_2(k-1) + \Delta m(k) \eta_s(k), \\ \eta_2 &= \frac{F_2}{F_{2\,\text{total}}}, \end{aligned} \tag{30}$$

where $F_{2\,\text{total}}$ is the total mass of VO in a batch of herbs, assumed to be 36 kg, and

$$\begin{aligned} \eta_s(k) &= 1 - \frac{C_d(k)}{C_b(k)} = \frac{\left(T_S P_1 Q_3(k) - \Delta m(k)\right) \eta_0(k)}{T_S P_1 Q_3(k) - \Delta m(k) \eta_0(k)}, \\ \eta_0(k) &= \frac{S_0 g}{18\mu_0 Q_0(k)}\left(\rho_w - \rho_0\right) d_2^2 \\ &= \frac{S_0 g}{18\mu_0 P_1 Q_3(k)}\left(\rho_w - \rho_0\right) d_2^2. \end{aligned} \tag{31}$$

The total consumption of steam F_3 is given by

$$F_3 = \sum_{i=1}^{6} \frac{\left(Q_{1i} + Q_{2i}\right) t}{6}. \tag{32}$$

In conclusion, the input and output variables of the mechanistic model for DEPP can be expressed as

$$\begin{aligned} u &= \left(Q_{11}, Q_{12}, Q_{13}, Q_{14}, Q_{15}, Q_{16}, Q_{21}, Q_{22}, Q_{23}, Q_{24}, Q_{25},\right. \\ &\left. Q_{26}, t\right), \\ y &= \left(F_1, F_2, F_3, \eta_1, \eta_2\right). \end{aligned} \tag{33}$$

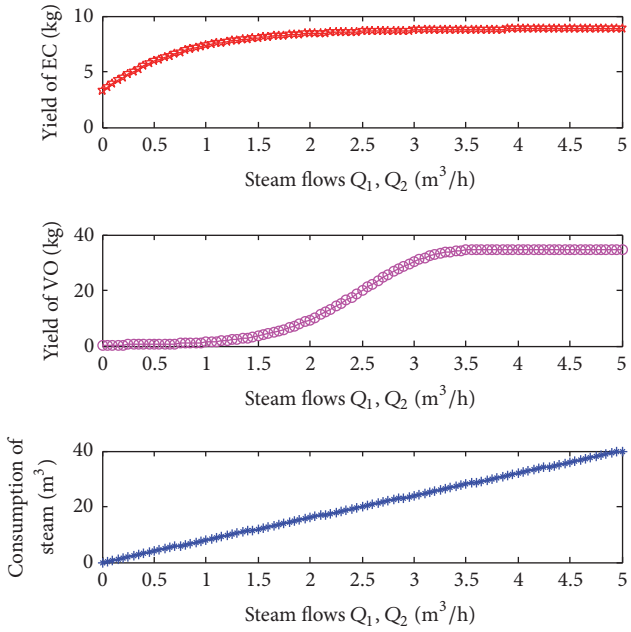


Figure 4: Relationship between model outputs and steam flow in the extraction process.

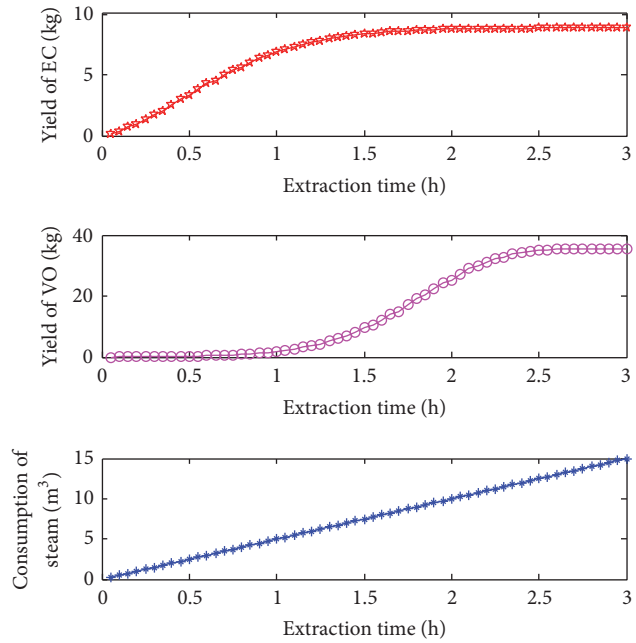


Figure 5: Relationship between model outputs and extraction time in the extraction process.

To verify the rationality and effectiveness of the established mechanistic model, a number of simulations were conducted. Figure 4 shows the effect of steam flows on the model outputs. In this simulation, all the steam flows were identical and the other operating or state conditions remained unaltered. For example, $t = 3$ h. The parameters used in the mechanistic model are listed in Table 1.

From Figure 4, it is clear that the extraction yields of EC and VO increase with the steam flow until reaching their respective equilibrium states, and the consumption of steam exhibits a gradual, linear rise. First, with other parameters fixed, the increased steam flow causes the feed liquid temperature to rise, and this rising temperature results in an increment in the diffusion coefficients, which will accelerate the extraction of EC. However, when the temperature reaches a stable value, owing to the characteristics of the feed liquid, the extraction of EC will gradually slow down and eventually stop once the concentration of EC in the solvent is equal to its solubility at this temperature. The recycling of VO has the similar tendency to that of EC, but the reason for such an equilibrium state is that the VO has been fully extracted at this temperature.

The effect of the extraction time on the model outputs is plotted in Figure 5. In this simulation, it was assumed that all other operating or state conditions remained unaltered and the steam flows were set as $Q_{ij} = 2.5$, $i = 1, 2$ and $j = 1, 2, \ldots, 6$. It can be seen that the extraction yields of EC and VO increase with the extraction time until they reach the equilibrium state. As the extraction time increases, the heat absorbed by the feed liquid will continue to rise, which leads to the increment in the feed liquid temperature. The analysis is then similar to that for the effect of the various steam flows.

4. Optimization of the Drug Extraction Production Process

4.1. Identification of Predictive Model Parameters. To carry out such optimization, a predictive model of DEPP has been developed. This model links the operating variables (steam flow at both the bottom and side of the extraction tank and the extraction time) with the extraction efficiencies of EC and VO. In the present work, the established mechanistic model was simplified to give a predictive model. The following simplifications were made.

(1) In modeling the extraction tank, the nonlinear relationship between the mixed steam flow Q_3 and temperature T was linearized; that is, $Q_3 = k_1 T + k_2$, where k_1, k_2 are the parameters to be identified.

(2) As internal diffusion has less effect on the mass transfer than external diffusion, the internal diffusion was neglected to simplify the mechanistic model.

Hence, the differences between the predictive model f_m and mechanistic model f_p can be described as

$$Q_3 = k_1 T + k_2, \tag{34}$$

$$\frac{dC}{dt} = \frac{1}{AH}\left[\frac{SD}{r_0}(C_0 - C) - CP_1(Q_1 - Q_3)\right], \tag{35}$$

and the other parts of the mechanistic model were retained in the predictive model.

To predict the extraction efficiencies under different operating conditions, the parameters k_1, k_2 must be identified. A best fit approach was adopted in which k_1, k_2 in (34) are the values that best predict the actual DEPP data [42]. The extraction efficiencies of EC and VO were used as criteria

Table 1: Parameters used in mechanistic model.

Parameter	Description	Value
A	Cross-sectional area of extraction tank (m^2)	3
T_1	Steam temperature before entering extraction tank (°C)	125
P_1	Volume ratio of water and steam	1/80
P_2	Standard heat of liquefaction of vapor (kJ/mol)	40.68
S_1	Specific heat capacity of feed liquid (kJ/(kg × °C))	4
S_2	Specific heat capacity of water (kJ/(kg × °C))	4.2
K_1	Coefficient in Q_3	4.4
K_2	Coefficient in Q_3	0.001
S_0	Lifting area of oil layer (m^2)	2.8
ρ_w	Density of water (g/cm^3)	1
ρ_0	Density of volatile oil (g/cm^3)	0.72
g	Acceleration of gravity (m/s^2)	9.8
d_2	Diameter of oil droplet (m)	$0.3912E-5$
q_0	Initial concentration of volatile oil in herbs	0.3
R	Radius of solid herb particle (m)	$0.8202E-3$
r_0	Thickness of liquid film (m)	$0.222E-3$
d_1	Thickness of gas film (m)	$0.6302E-3$
τ	Sinuosity of holes inside gray layer in solid particle	1.2
ε	Porosity of gray layer in solid particle	0.98
C_0	Solubility of EC in liquid film	0.85
ψ	Associating parameter of solvent	2.65
M_B	Molecular weight of solvent (D)	100
V_A	Molecular volume of EC (m^3)	$30.62 * 10^{(-8)}$
K	Linear equilibrium proportional constant	0.15
H_0	Initial liquid level in extraction tank (m)	3.5
T_0	Initial temperature of feed liquid in extraction tank (°C)	20
C_{in}	Initial concentration of EC in solvent	0
T_S	Sampling period (s)	0.05

for the parameter identification. The identification algorithm used for the predictive model is as follows.

Step 1. n sets of process data (u_i, η_i) are produced by mechanistic model f_p, in which u_i denotes the vector of operating variables and η_i is the vector of extraction efficiencies of EC and VO; that is, $\eta_i = (\eta_{1i}, \eta_{2i}), i = 1, 2, \ldots, n$.

Step 2. Insert u_i into predictive model f_m to obtain the predicted extraction efficiencies as follows:

$$\widehat{\eta}_i = \left(\widehat{\eta}_{1i}, \widehat{\eta}_{2i}\right) = f_m\left(u_i\right). \tag{36}$$

Step 3. Parameters k_1, k_2 are determined by minimizing the following objective function of the sum of square errors (SSE) of the extraction efficiencies through an optimization algorithm:

$$\min_{k_1,k_2} \left\{ \text{SSE} = \sum_{i=1}^{n} \left(\eta_{1i} - \widehat{\eta}_{1i}\right)^2 + \sum_{i=1}^{n} \left(\eta_{2i} - \widehat{\eta}_{2i}\right)^2 \right\}. \tag{37}$$

Table 2: Parameter values and optimization results of each algorithm.

n	w	Algorithm	k_1	k_2	SSE
30	10	PSO	0.0044	4.4223	$4.646E-4$
		DE	0.0049	4.3734	$4.971E-4$
		ABC	0.0023	4.6205	$6.532E-4$

We solved (37) using PSO, DE, and ABC and compared the performances of the algorithms with one another. The algorithms were executed w times and the results averaged. The optimization results are summarized in Table 2, from which we can see that PSO achieved the best performance in solving the optimization problem.

4.2. Optimization of Single Extraction. The economic benefit per unit time is considered as the optimization objective, and the operating variables are treated as decision variables. Taking economic income and production consumption into consideration (including the economic income of EC and VO

Table 3: Parameter values in optimization model.

Parameters	a (USD/Kg)	b (USD/Kg)	c (USD/m^3)	d (USD/Kg)	F_4 (Kg)
Value	152.7	122.17	0.031	1.53	120

Table 4: Constraints in optimization model.

Constrained variables	Quality indexes		Operating variables	
	η_1	η_2	Q_{ij} (m^3/h)	t (h)
Range	[0.8, 1]	[0.8, 1]	[0, 5]	[1, 4]

Table 5: Optimal economic benefits obtained by each algorithm.

Algorithm	PSO	DE	ABC
Economic benefit (USD/h)	2293	2240	2252

and the consumption cost of herb materials and steam), the following optimization model was established based on the predictive model:

$$\begin{aligned} \max \quad & J = \frac{a \times F_1 + b \times F_2 - c \times F_3 - d \times F_4}{t} \\ \text{s.t.} \quad & (F_1, F_2, F_3, \eta_1, \eta_2) = f(u); \\ & \eta_{1\min} \le \eta_1 \le \eta_{1\max}; \\ & \eta_{2\min} \le \eta_2 \le \eta_{2\max}; \\ & Q_{\min} \le Q_{ij} \le Q_{\max}; \\ & t_{\min} \le t \le t_{\max}; \\ & i = 1, 2; \\ & j = 1, 2, \ldots, 6, \end{aligned} \tag{38}$$

where a, b, c, and d are the prices of EC, VO, steam, and herb materials and F_4 is the weight of a batch of herb materials. The parameters and constraints in this model are listed in Tables 3 and 4, respectively. The PSO, DE, and ABC algorithms were executed 10 times to determine the optimal solution to (38), and the results are presented in Table 5. The optimization results once again validate the superiority of PSO for the problem considered in this study.

The optimal values obtained by PSO for the operating variables are listed in Table 6. The optimal operating conditions were applied to the single extraction process, and the output indexes are given in Table 7. Additionally, the variation in the extraction yield of EC and VO and the consumption of steam are plotted in Figure 6.

From this figure, we can see that the mass of EC increases with extraction time in the initial stage of the single extraction, but the rate of increase gradually slows as we reach the equilibrium state. The reason for this is that the concentration of EC in the extracted solution reaches its saturated solubility, and so no more EC will be extracted from the herb materials, even if we continue the extraction process.

There is also a comparatively small recycle yield of VO in the initial 0.75 h, because the temperature in the extraction

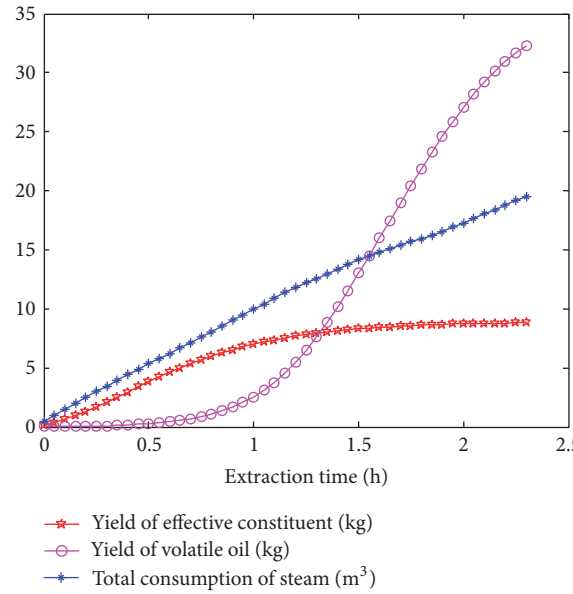


Figure 6: Variation trend of various output indexes in single extraction.

tank is too low to boil VO and little of it can be volatilized and brought out together with vapor in the initial stages of extraction. When the temperature reaches the boiling point of VO, its recycle yield grows exponentially. As the extraction progresses, there is less VO remaining in the herbs, and the increasing tendency of the recycle yield becomes slower until all VO has been extracted.

The actual results in Table 7, however, suggest that the extraction efficiency of EC is only 35.28%, and so most of the EC has not been extracted. Thus, the quality constraints cannot be satisfied, although there is a significant economic benefit. According to national production management regulations, such a huge wastage of herb materials is forbidden, and a multiextraction process is needed to avoid wasting raw materials.

4.3. Optimization of Multiextraction. As the extraction process progresses, the extraction temperature does not always rise with an increase in the operating variables. This is because of the self-regulating characteristics of the feed liquid. As described above, the concentration of EC in the feed liquid will reach a maximum, and only VO will be recycled along with the extraction. This causes the low extraction efficiency of EC in the single extraction case. Therefore, to avoid such a situation, the first extraction should be terminated and the feed liquid containing EC should be separated from the solid herb via a filter. After this, a certain amount of new solvent can be added to the extraction tank and a second extraction can be performed. Depending on the total content of EC

Table 6: Best settings of operating variables produced by PSO.

Steam flow Q_1 (m^3/h)		Steam flow Q_2 (m^3/h)		Extraction time (h)	
Q_{11}	4.9554	Q_{21}	4.9545		
Q_{12}	4.7633	Q_{22}	4.3348	t	2.3
Q_{13}	4.5968	Q_{23}	4.8391		
Q_{14}	3.6422	Q_{24}	4.3850	Optimal economic benefit (USD/h)	
Q_{15}	3.5999	Q_{25}	2.3083	J	2293
Q_{16}	3.1419	Q_{26}	4.2665		

Table 7: Actual optimal output indexes and economic benefit in single extraction.

Single extraction	F_1 (Kg)	F_2 (Kg)	F_3 (m^3)	η_1	η_2	J (USD/h)
Actual value	8.82	32.24	19.51	35.28%	89.56%	2218

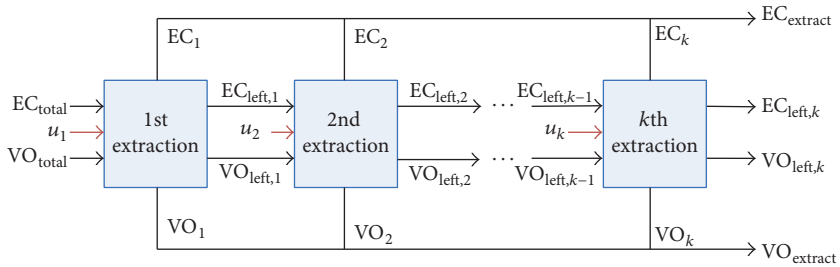


Figure 7: Schematic diagram of multiextraction.

in the herb and its solubility in the solvent, at least three extractions are needed to remove all of the EC. A schematic diagram of the multiextraction process is shown in Figure 7, where each extraction represents a single extraction. Based on the principle of multiextraction, any surplus in the quality indexes after the $(i-1)$th extraction, such as the total residual mass of VO and radius of unextracted core of EC, is regarded as the initial input conditions of the ith extraction, where $i = 2, 3, \ldots, k$. According to this principle, a predictive model for the multiextraction process can be developed.

In developing a predictive model for the multiextraction scenario, the extraction frequency k is treated as one of the operating variables ($k = 3, 4, 5$). In this study, predictive models were established for three, four, and five extractions. We now describe the three-time extraction model; the modeling approach for four and five extractions is similar. The three-extraction case is composed of three single extractions, each of which has the 13 input variables described previously. Thus, there are 39 input variables in the three-extraction predictive model:

$$U = (u_1, u_2, u_3), \tag{39}$$

where u_i is the vector of operating variables in the ith single extraction and

$$u_i = (Q_{11,i}, Q_{12,i}, Q_{13,i}, Q_{14,i}, Q_{15,i}, Q_{16,i}, Q_{21,i}, Q_{22,i}, Q_{23,i}, Q_{24,i}, Q_{25,i}, Q_{26,i}, t_i), \quad i = 1, 2, 3. \tag{40}$$

The predictive outputs for the three extractions, including the total extraction yield of EC and VO, consumption of steam, and extraction efficiencies, can be calculated as

$$\begin{aligned} F_1 &= \sum_{i=1}^{3} F_{1,i}, \\ F_2 &= \sum_{i=1}^{3} F_{2,i}, \\ F_3 &= \sum_{i=1}^{3} \left(\sum_{j=1}^{6} \left(\frac{\left(Q_{1j,i} + Q_{2j,i}\right) t_i}{6} \right) \right), \\ \eta_1 &= \frac{F_1}{F_{1\,\text{total}}}, \\ \eta_2 &= \frac{F_2}{F_{2\,\text{total}}}. \end{aligned} \tag{41}$$

Therefore, the predictive model of three-time extraction can be summarized as

$$(F_1, F_2, F_3, \eta_1, \eta_2) = F_{m,3}(U). \tag{42}$$

After developing predictive models for $k = 3, 4$, and 5, the optimization model for multiextraction was derived

by treating the extraction frequency as one of the decision variables:

$$\begin{aligned}
\max \quad & J = \frac{a \times F_1 + b \times F_2 - c \times F_3 - d \times F_4}{\sum_{i=1}^{k} t_i} \\
\text{s.t.} \quad & (F_1, F_2, F_3, \eta_1, \eta_2) = F_{m,k}(U); \\
& \eta_{1\min} \le \eta_1 \le \eta_{1\max}; \\
& \eta_{2\min} \le \eta_2 \le \eta_{2\max}; \\
& Q_{\min} \le Q_{1j,k} \le Q_{\max}; \\
& Q_{\min} \le Q_{2j,k} \le Q_{\max}; \\
& t_{\min} \le t_i \le t_{\max}; \\
& k = 3, 4, 5; \\
& j = 1, 2, \ldots, 6.
\end{aligned} \tag{43}$$

In this optimization model, the extraction frequency k is treated as one of the decision variables ($k = 3, 4, 5$). The decision variables must be carefully handled in solving this optimization problem. For example, when $k = 3$, the decision variable vector U_d can be expressed as

$$U_d = (k, u_1, u_2, u_3, u_4, u_5), \tag{44}$$

where $u_4 = u_5 = \mathbf{0}$. Similarly, $u_5 = \mathbf{0}$ when $k = 4$.

The advantages of PSO in solving such optimization problems have been verified many times and are not covered here. The parameters used in the PSO algorithm are presented in Table 8, and the optimal multiextraction solution obtained by the PSO method with 10 runs is given in Table 9.

From Table 9, it is clear that the optimal extraction frequency is 3, so the operating variables in the fourth and fifth extractions are all zero. These optimal operating variables were used in the mechanistic model of the multiextraction process to calculate the actual output indexes and economic benefit. The results, together with their predicted values, are given in Table 10. However, owing to the inherent errors of predictive models, it is difficult to achieve the optimized economic benefit in an actual process. The results in Table 10 indicate that a discrepancy of 6.4% exists between the predicted and actual economic benefit, which means that "optimal in the model" may not translate to "optimal for the process," as mentioned before. This model uncertainty hampers the optimization of the multiextraction process. Hence, as discussed in Section 5, the ILC method was adopted to overcome the problem of model uncertainty.

5. Use of Iterative Learning Control to Overcome Model Uncertainty

To eliminate the impact of the model uncertainty, the iterative nature of numerical optimization and certain repetitive properties can be utilized. The idea of ILC is introduced to improve the economic benefit. In this study, we used DDOTILC with the control law (see [35] for details):

Table 8: Parameter settings of PSO.

Swarm size	Iterations	Inertia weight	Learning factor
100	100	Linear decrease: 0.9~0.2	c_1: 2.5~1.0 c_2: 1.0~2.5

$$\begin{aligned}
\widehat{\varphi}_k &= \widehat{\varphi}_{k-1} + \frac{\xi \left(\Delta y_{k-1}(N) - \widehat{\varphi}_{k-1} \Delta u_{k-1}\right) \Delta u_{k-1}^T}{\mu + \left\|\Delta u_{k-1}\right\|^2} \\
u_k &= u_{k-1} + \frac{\rho \widehat{\varphi}_k^T}{\lambda + \left\|\widehat{\varphi}_k\right\|^2} e_{k-1}(N),
\end{aligned} \tag{45}$$

where ρ, λ, ξ, and μ are positive constants; k is the iteration number of the batch process and N is their terminal moment; $\widehat{\varphi}_k$ is the estimated pseudo-partial-derivative of the kth batch process; and $e_{k-1}(N)$ is the terminal tracking error of the output.

In the application of this approach, the economic benefit is treated as the terminal output and the operating variables are adjusted online to improve the economic benefit. The desired economic benefit J_d is set as the predicted value obtained by solving (43). The parameters used in (45) are listed in Table 11. The blue curves in Figure 8 describe the evolution of the economic benefit under DDOTILC. The results indicate that the actual economic benefit achieves an increment of 1.5% after six batches, which partly improves the economic benefit, although this is still short of the desired economic benefit. In addition, the extraction efficiencies decrease with the increase in economic benefit, as shown in Figure 8. When the extraction efficiencies reach a threshold, the economic benefit will be inversely proportional to the extraction yields of EC and VO, because the costs of extracting the remaining EC and VO will be much greater than the additional income. Thus, the optimal economic benefit may not correspond to adjusting the operating variables to maximize the extraction of EC and VO.

The parameter settings will affect the convergence rate of ILC, and the best parameter settings should have the ability to change adaptively with the tracking error and variations in system output. In this study, the parameter ρ in (45) was varied though fuzzy adaptive adjustment over the range $[0, \alpha]$. The convergence rate of economic benefit can be changed by adjusting ρ. When the absolute values of the tracking error of economic benefit and its variation are high, the value of ρ should also be high. Moreover, the value of ρ should decrease as the absolute values of tracking error and its variation decrease. According to this principle, the fuzzy adjustment rule was formulated as shown in Table 12, where E_J and ΔE_J represent the absolute values of the tracking error of economic benefit and its variation, respectively. The membership functions of E_J, ΔE_J, and ρ are shown in Figure 9. The first red curve in Figure 8 shows the trajectory of economic benefit based on the fuzzy adjustment of parameter ρ,

Table 9: Best settings of operating variables.

		1st extraction	2nd extraction	3rd extraction	4th extraction	5th extraction
Steam flow Q_1 (m^3/h)	Q_{11}	4.5538	3.6784	4.2887	0.0000	0.0000
	Q_{12}	4.6789	2.7554	0.9895	0.0000	0.0000
	Q_{13}	4.4342	3.2861	4.1627	0.0000	0.0000
	Q_{14}	1.5664	4.9991	2.2495	0.0000	0.0000
	Q_{15}	3.7135	4.9160	0.9995	0.0000	0.0000
	Q_{16}	0.0477	1.6587	4.2214	0.0000	0.0000
Steam flow Q_2 (m^3/h)	Q_{21}	4.9544	0.5090	3.6287	0.0000	0.0000
	Q_{22}	4.3760	3.2264	3.4749	0.0000	0.0000
	Q_{23}	3.7216	4.4979	3.2713	0.0000	0.0000
	Q_{24}	0.2749	1.1098	3.9415	0.0000	0.0000
	Q_{25}	4.0857	3.6746	4.5692	0.0000	0.0000
	Q_{26}	1.0008	2.8459	4.1023	0.0000	0.0000
Extraction time (h)	t	1.6628	1.0992	3.0475	0.0000	0.0000

Table 10: Output indexes and economic benefit in multiextraction.

Multiextraction	F_1 (Kg)	F_2 (Kg)	F_3 (m^3)	η_1	η_2	J (USD/h)
Predicted value	24.39	34.46	37.59	97.57%	95.73%	1333
Actual value	21.09	34.40	37.59	84.36%	95.50%	1247.3

Table 11: Parameters of DDOTILC.

Parameter	ρ	μ	ξ	λ
Value	0.003	1	0.1	0.01

Table 12: Fuzzy adjustment rule.

ρ	ΔE_J						
E_J	NB	NM	NS	Z	PS	PM	PB
NB	Z	Z	Z	PS	PS	PM	PM
NM	Z	Z	Z	PS	PM	PM	PM
NS	Z	PS	PS	PM	PM	PM	PM
Z	Z	PS	PM	PM	PM	PM	PM
PS	PS	PS	PM	PM	PM	PM	PM
PM	PS	PM	PM	PM	PM	PM	PM
PB	PM	PM	PM	PM	PM	PM	PM

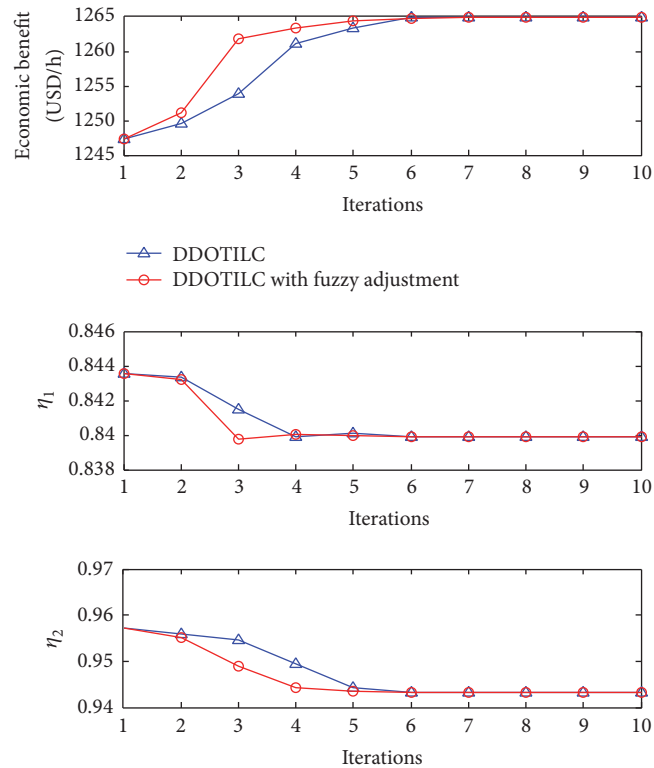


Figure 8: Comparison of economic benefit between original DDOTILC and DDOTILC with fuzzy adjustment.

with which the advantage of the fuzzy-adjusted DDOTILC method is validated. In this simulation, $\alpha = 0.005$.

6. Conclusion

This paper proposed a mechanistic modeling and optimization control strategy for DEPP. First, a mechanistic model of the drug extraction process was developed based on previous efforts to simulate the actual production process and produce process data. A simulation was conducted to verify the rationality and effectiveness of this mechanistic model. A predictive model was then developed by simplifying some of the processes in the mechanistic model. Simulation results demonstrated the effectiveness of this predictive model. Process optimization was first implemented for single extraction, and the results indicated an extremely low extraction yield of EC, which indicates serious wastage of herbs, although

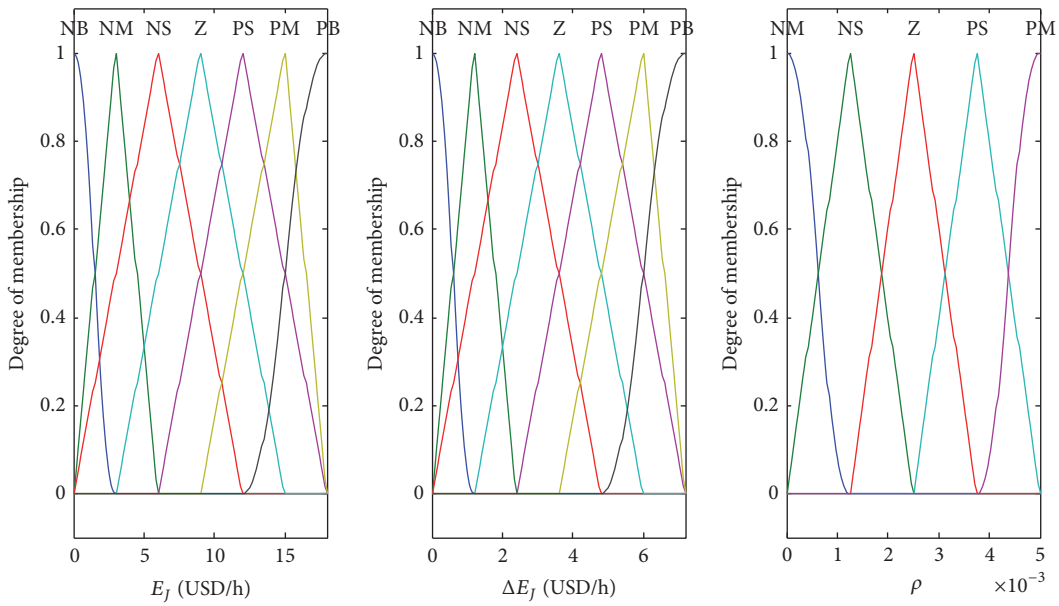


Figure 9: Membership functions of E_J, ΔE_J, and ρ.

this gave a preferable economic benefit. Such wastage of raw materials is forbidden according to national production management regulations. A multiextraction DEPP is therefore necessary, and so process optimization for multiextraction was investigated. However, such optimization is hampered by model uncertainty. Therefore, the DDOTILC method was applied to partly overcome this model uncertainty. The simulation results indicate that the economic benefit can be improved by 1.5%. Finally, the idea of fuzzy adjustment was introduced to adaptively adjust the parameter ρ in DDOTILC, which increases the convergence rate of economic benefit. The simulation results indicate that the proposed modeling and optimization control method can effectively solve the practical problems encountered in DEPP.

The proposed modeling and optimization control strategy is easy to implement in practical applications because it requires lesser accuracy of predictive models. The research in this paper is based on a mechanistic model developed to simulate actual DEPP and produce process data. However, there is no empirical research from pharmaceutical enterprises to test the validity of the model and method. All parameters used for the mechanistic modeling, such as the device parameters for the extraction tank and state parameters of DEPP, correspond to the actual production situation, so the model proposed in this work has the ability to simulate the actual DEPP, and it is reasonable to test the validity of the proposed method through simulations. The second limitation of this study is that the fuzzy rules should be based on the actual situation. For example, the initial range of ρ must be set according to the actual DEPP scenario when applying DDOTILC.

Competing Interests

The authors declare that they have no competing interests.

Acknowledgments

The authors would like to thank National Natural Science Foundation of China under Grant nos. 61374147, 61004083, and 61203103 and Fundamental Research Funds for the Central Universities under Grant nos. N150402001 and N120404014, for supporting this research work.

References

[1] B. E. Richter, B. A. Jones, J. L. Ezzell, N. L. Porter, N. Avdalovic, and C. Pohl, "Accelerated solvent extraction: a technique for sample preparation," *Analytical Chemistry*, vol. 68, no. 6, pp. 1033–1039, 1996.

[2] B. Benthin, H. Danz, and M. Hamburger, "Pressurized liquid extraction of medicinal plants," *Journal of Chromatography A*, vol. 837, no. 1-2, pp. 211–219, 1999.

[3] K. F. Hou, Q. Zheng, Y. Li, J. Shen, and S. Hu, "Modeling and optimization of herb leaching processes," *Computers & Chemical Engineering*, vol. 24, no. 2–7, pp. 1343–1348, 2000.

[4] R. M. Smith, "Extractions with superheated water," *Journal of Chromatography A*, vol. 975, no. 1, pp. 31–46, 2002.

[5] P. Li, S. P. Li, C. M. Fu et al., "Pressurized solvent extraction in quality control of Chinese herb," *China Journal of Chinese Materia Medica*, vol. 29, no. 8, pp. 723–726, 2004.

[6] Y. Chen, J. K. Tian, and Y. Y. Cheng, "A novel technique of extracting essential oil from Chinese herb and medicine–subcritical water extraction," *Chemical Engineering*, vol. 34, no. 8, pp. 59–62, 2006.

[7] Y. T. Lau, K. M. Ng, D. T. W. Lau, and C. Wibowo, "Quality assurance of Chinese herbal medicines: procedure for single-herb extraction," *AIChE Journal*, vol. 59, no. 11, pp. 4241–4254, 2013.

[8] G. C. Rodríguez-Jimenes, A. Vargas-Garcia, D. J. Espinoza-Pérez, M. A. Salgado-Cervantes, V. J. Robles-Olvera, and M. A. García-Alvarado, "Mass transfer during vanilla pods solid liquid

extraction: effect of extraction method," *Food and Bioprocess Technology*, vol. 6, no. 10, pp. 2640–2650, 2013.

[9] M. S. Alam, Z. A. Damanhouri1, A. Ahmad et al., "Development of response surface methodology for optimization of extraction parameters and quantitative estimation of embelin from Embelia ribes Burm by high performance liquid chromatography," *Pharmacognosy Magazine*, vol. 11, supplement 1, pp. S166–S172, 2015.

[10] V. C. Bochi, M. T. Barcia, D. Rodrigues, C. S. Speroni, M. M. Giusti, and H. T. Godoy, "Polyphenol extraction optimisation from Ceylon gooseberry (*Dovyalis hebecarpa*) pulp," *Food Chemistry*, vol. 164, pp. 347–354, 2014.

[11] R. Z. Chen, Y. Li, H. Dong et al., "Optimization of ultrasonic extraction process of polysaccharides from *Ornithogalum Caudatum Ait* and evaluation of its biological activities," *Ultrasonics Sonochemistry*, vol. 19, no. 6, pp. 1160–1168, 2012.

[12] M. Chen, Y. Zhao, and S. Yu, "Optimisation of ultrasonic-assisted extraction of phenolic compounds, antioxidants, and anthocyanins from sugar beet molasses," *Food Chemistry*, vol. 172, pp. 543–550, 2015.

[13] I. K. Bae, H. M. Ham, M. H. Jeong, D. H. Kim, and H. J. Kim, "Simultaneous determination of 15 phenolic compounds and caffeine in teas and mate using RP-HPLC/UV detection: method development and optimization of extraction process," *Food Chemistry*, vol. 172, pp. 469–475, 2015.

[14] M. Bravi, R. Bubbico, F. Manna, and N. Verdone, "Process optimisation in sunflower oil extraction by supercritical CO_2," *Chemical Engineering Science*, vol. 57, no. 14, pp. 2753–2764, 2002.

[15] Y.-J. Choi, T.-I. Kwon, and Y.-K. Yeo, "Optimization of the sulfolane extraction plant based on modeling and simulation," *Korean Journal of Chemical Engineering*, vol. 17, no. 6, pp. 712–718, 2000.

[16] X. C. Gong, Y. Zhang, J. Y. Pan, H. B. Qu, and Q. Zhang, "Optimization of the ethanol recycling reflux extraction process for saponins using a design space approach," *PLoS ONE*, vol. 9, no. 12, Article ID e114300, 2014.

[17] Z. X. Huang, *A Study on Strategy of Optimal Control in Procedure of Chinese Traditional Medicine*, Central South University Press, Changsha, China, 2006.

[18] G. H. Hu, Z. Z. Mao, D. K. He, and F. Yang, "Hybrid modeling for the prediction of leaching rate in leaching process based on negative correlation learning bagging ensemble algorithm," *Computers and Chemical Engineering*, vol. 35, no. 12, pp. 2611–2617, 2011.

[19] Z. Tong, "Leaching kinetics of calcium aluminate slag," *The Chinese Journal of Process Engineering*, vol. 5, no. 4, pp. 399–402, 2005.

[20] S. Zhan, "Study on leaching kinetics of pyrite cinder," *Chemical Engineering*, vol. 34, no. 11, pp. 36–39, 2006.

[21] J. M. Del Valle and J. C. De La Fuente, "Supercritical CO_2 extraction of oilseeds: review of kinetic and equilibrium models," *Critical Reviews in Food Science and Nutrition*, vol. 46, no. 2, pp. 131–160, 2006.

[22] M. Perrut, J. Y. Clavier, M. Poletto, and E. Reverchon, "Mathematical modeling of sunflower seed extraction by supercritical CO_2," *Industrial and Engineering Chemistry Research*, vol. 36, no. 2, pp. 430–435, 1997.

[23] Y. P. Han, Y. C. Xiang, S. B. Wang et al., "Relationship between the yield and extraction time of volatile oil," *Chinese Traditional Patent Medicine*, vol. 23, no. 1, pp. 11–13, 2001.

[24] L. C. Wang, *The Indoor Experiment Study of Crown-Type Oil-Water Separator*, Northeast Petroleum University Press, Daqing, China, 2011.

[25] J. Kennedy and R. Eberhart, "Particle swarm optimization," *IEEE International Conference on Neural Networks*, vol. 4, no. 8, pp. 1942–1948, 1995.

[26] T. C. Carneiro, S. P. Melo, P. C. M. Carvalho, and A. P. D. S. Braga, "Particle swarm optimization method for estimation of weibull parameters: a case study for the Brazilian northeast region," *Renewable Energy*, vol. 86, pp. 751–759, 2016.

[27] K. Rajesh, A. Bhuvanesh, S. Kannan, and C. Thangaraj, "Least cost generation expansion planning with solar power plant using Differential Evolution algorithm," *Renewable Energy*, vol. 85, pp. 677–686, 2016.

[28] W. F. Gao, L. L. Huang, S. Y. Liu, and C. Dai, "Artificial bee colony algorithm based on information learning," *IEEE Transactions on Cybernetics*, vol. 45, no. 12, pp. 2827–2839, 2015.

[29] H. Fei, X. Jinwu, L. Min, and Y. Jianhong, "Product quality modelling and prediction based on wavelet relevance vector machines," *Chemometrics and Intelligent Laboratory Systems*, vol. 121, pp. 33–41, 2013.

[30] K. Kim, J.-M. Lee, and I.-B. Lee, "A novel multivariate regression approach based on kernel partial least squares with orthogonal signal correction," *Chemometrics and Intelligent Laboratory Systems*, vol. 79, no. 1-2, pp. 22–30, 2005.

[31] Z. H. Xiong and J. Zhang, "A batch-to-batch iterative optimal control strategy based on recurrent neural network models," *Journal of Process Control*, vol. 15, no. 1, pp. 11–21, 2005.

[32] N. Sanzida and Z. K. Nagy, "Iterative learning control for the systematic design of supersaturation controlled batch cooling crystallisation processes," *Computers and Chemical Engineering*, vol. 59, pp. 111–121, 2013.

[33] X. Jin and J.-X. Xu, "Iterative learning control for output-constrained systems with both parametric and nonparametric uncertainties," *Automatica*, vol. 49, no. 8, pp. 2508–2516, 2013.

[34] X. H. Bu, F. H. Yu, Z. S. Hou, and F. Z. Wang, "Iterative learning control for a class of nonlinear systems with random packet losses," *Nonlinear Analysis: Real World Applications*, vol. 14, no. 1, pp. 567–580, 2013.

[35] R. H. Chi, D. W. Wang, Z. S. Hou, and S. T. Jin, "Data-driven optimal terminal iterative learning control," *Journal of Process Control*, vol. 22, no. 10, pp. 2026–2037, 2012.

[36] R. H. Chi, Z. S. Hou, and S. T. Jin, "A data-driven adaptive ILC for a class of nonlinear discrete-time systems with random initial states and iteration-varying target trajectory," *Journal of the Franklin Institute*, vol. 352, no. 6, pp. 2407–2424, 2015.

[37] S.-T. Jin, Z.-S. Hou, R.-H. Chi, and X.-B. Liu, "Data-driven model-free adaptive iterative learning control for a class of discrete-time nonlinear systems," *Control Theory and Applications*, vol. 29, no. 8, pp. 1001–1009, 2012.

[38] C. R. Wilke and P. Chang, "Correlation of diffusion coefficients in dilute solutions," *AIChE Journal*, vol. 1, no. 2, pp. 264–270, 1955.

[39] E. L. Cussler, *Diffusion: Mass Transfer in Fluid Systems*, Cambridge University Press, New York, NY, USA, 2002.

[40] M. Y. Wu, J. H. Zhao, H. Liang, H. Kang, and S. Xu, "Kinetic model for extraction process of volatile oil from traditional Chinese medicine," *Journal of Chemical Industry and Engineering*, vol. 59, no. 12, pp. 2990–2995, 2008.

[41] Y. B. Xu, S. X. Wang, and X. W. Yang, "Applying shallow pool theory and coalescence technology to improve oily water

treatment system in oilfield," *Technology of Water Treatment*, vol. 36, no. 10, pp. 115–118, 2010.

[42] H. Liu and M. Li, "Population balance modelling and multi-stage optimal control of a pulsed spray fluidized bed granulation," *International Journal of Pharmaceutics*, vol. 468, no. 1-2, pp. 223–233, 2014.

Hindawi Publishing Corporation
Scientific Programming
Volume 2016, Article ID 6123103, 9 pages
http://dx.doi.org/10.1155/2016/6123103

Research Article

Optimal Scheduling of Doctors Outpatient Departments Based on Patients' Behavior

Zongwei Ren, Haini Guo, Shizhen Bai, and Peng Li

School of Management, Harbin University of Commerce, No. 1, Xuehai Street, Harbin 150028, China

Correspondence should be addressed to Zongwei Ren; rzw79@163.com

Received 16 June 2016; Revised 11 September 2016; Accepted 21 September 2016

Academic Editor: Junhu Ruan

The low operational efficiency in the field of medical and health care has become a serious problem in China; the long time that the patients have to wait for is the main phenomenon of difficult medical service. Medical industry is service-oriented and its main purpose is to make profits, so the benefits and competitiveness of a hospital depend on patient satisfaction. This paper makes a survey on a large hospital in Harbin of China and collects relevant data and then uses the prospect theory to analyze patients' and doctors' behavioral characteristics and the model of patient satisfaction is established based on fuzzy theory with a triplet $\alpha/\beta/\gamma$. The optimal scheduling of clinic is described as a problem with the rule of first come, first served which maximizes patient satisfaction for the main goal and minimizes operating costs for the secondary goal. And the corresponding mathematical model is established. Finally, a solution method named plant growth simulation algorithm (PGSA) is presented. Then, by means of calculating of the example and comparing with genetic algorithm, the results show that the optimum can be reached; meanwhile the efficiency of the presented algorithm is better than the genetic algorithm.

1. Introduction

With the improvement of living standards, people tend to pursue a higher quality of life, and the limitation of treatment level in secondary hospitals and community hospitals leads to a serious problem of registered queue in high-level hospitals. There are usually three types of registration in hospitals, namely, general-clinic, specialist-clinic, and emergency-clinic. How to allocate the number of doctor outpatient departments for every type of registration is very essential for operational efficiency in hospital and has a close relationship with patient satisfaction at the same time. Because of the different-type registrations, this paper makes optimal scheduling for the number of doctor outpatient departments considering the patient behavioral characteristics in order to minimize patient waiting time and increase their satisfaction and reduce operating costs.

The main goal of current researches in hospital is lowering costs and running better. Many papers achieved these goals from the perspective of the effective utilization of resource. The resource can be divided into human resource and material resource. Researchers used different methods to optimize the material resource such as Operating Room (OR), multihospital network, and limited number of wards [1–4]. All of the methods mentioned above made some assumptions which can simplify the solution of the problems although there were some discrepancies with reality. Vile et al. [5] investigated several interrelated advanced statistical and operational research methods and sought to predict future demands better, recommended minimum staffing requirements, and generated low-cost rosters. This article aims at optimizing the human resource; however, it only considered a single class of customers and its methods were only reliable under limited conditions. Wright and Mahar [6] studied the staff nurse scheduling in order to improve nurse satisfaction. It was also about the use of human resource and there were two limitations in this paper. First, it did not contain other scheduling policies and another limitation was that they only centralized scheduling decisions across three units in their study.

In many literatures, patients' waiting time is a major factor to be considered because it is closely linked with the quality of hospital operation. Now the hospital aims to deliver the highest quality of care, so reducing waiting time

of patient is important [7–9]. These papers gave a function to compute waiting time or gave a new concept called "break-in-moments" (BIMs) to reduce waiting time. However the latter one is only considering the emergency surgery. In the case of BIM, further researches are also needed to investigate whether inserting breaks between surgeries which can increase the stability of the schedule is needed. From another viewpoint, Shuangxi et al. [10] analyzed historical data of patients' medical treatment time in the last three years, concluded its regularity, and then calculated the number of doctors' configuration. But the research focused on children's regularity based on large amounts of data of the children's hospital which are much different from other high-level hospitals.

Queuing theory is always being used to deal with the problem of minimum waiting time. Lakshmi and Iyer [11] reviewed the contributions and applications of queuing theory in the field of health care management problems and proposed a system of classification of health care areas which are examined with the assistance of queuing models. The research's goal is to provide sufficient information to analysts who are interested in using queuing theory to model a health care process and who want to locate the details of relevant models.

Our research applies prospect theory and membership functions of patient satisfaction to build a model, taking into account the patient behavioral characteristics. Xuping et al. [12] made optimal scheduling of emergency supplies considering the public psychological perception of risk and Yang et al. [13] proposed interference management models of single machine scheduling from multiple angles actors. None of these articles used prospect theory into medicinal service.

It is worth noting that the optimization of service industries is different from manufacturing industries; the former one is always accompanied by the participation of people, and usually it will be affected by people's behavior and subjective emotion, so the hospital should focus on patient satisfaction. Combined with prospect theory and fuzzy theory and taking into account the patient behavioral characteristics at the same time, we have the research of the optimal scheduling for doctor outpatient departments in one high-level hospital of Harbin. And the starting point matches the urgent social medical problems in China.

The remainder of this paper is organized as follows. In Section 2, we introduce the queue classification analysis based on satisfaction degree of patients; Section 3 focuses on the description of problem and establishment of model; a genetic algorithm is described in Section 4; a medical case whose data is collected in a Harbin hospital is given in Section 5, together with simulation results. Conclusions are drawn in Section 6.

2. Queue Classification Analysis Based on Patient Satisfaction Degree

2.1. The Hospital Registration Type Analysis. The types of hospital registration always include general-clinic, specialist-clinic, and emergency-clinic. Under normal circumstances, the fee of these three types of registrations increases in turn. In general, the number of emergency-clinics is much less than the former two. The difference between general-clinic and specialist-clinic is that doctors' experience and technology are relatively richer in specialist-clinics and the registration's fee is higher. Usually, if the patient's situation is not too serious, he will choose general-clinic; if the patient needs comprehensive thorough check, he will choose specialist-clinic; if there is urgent situation, the patient will choose emergency-clinic. Patients can select the appropriate type of registrations according to their own conditions and actual demands. Due to the needs of doctor outpatient departments are different in each type of registration; this paper makes optimal scheduling under actual situations. The following is description of registration categories:

(1) general-clinic A: suitable for the milder clinical symptoms of patients;

(2) specialist-clinic B: suitable for the heavy condition and needs a thorough examination of patients;

(3) emergency-clinic C: suitable for the emergency condition of patients.

The emphasis of this article is on the lean staffing of doctor outpatient departments according to different needs of patients under the varied registration. That is, how much doctors should be arranged for each type of registrations in order to minimize patient waiting time and increase their satisfaction and reduce operating costs.

2.2. Patient Behavior Analysis. In general, different patients in a hospital have different expected waiting time when they are queuing and here are several factors affecting it. First, people with different personality have different expectation of waiting time. Character is the factor that influences patients' emotion themselves. The expected waiting time of patients who have urgent character is much shorter than the patients who have slow character. Second, it depends on patients' different conditions. Usually, the expected waiting time of patients who have serious illness is much shorter than the patients whose illness is lighter. But it does not adapt to all patients absolutely and the length of expected waiting time is different from person to person. If patients' actual waiting time exceeds the length of expected waiting time, the patients will produce bad emotion such as anxiety and discontent. So we divide patient satisfaction degree into three levels: satisfaction (the value is equal to 1), dissatisfaction, and very dissatisfaction (the value is equal to 0). We make survey of the patients in a high-level hospital in Harbin and collect data through the way of interview and questionnaire. And then arrange the data to draw the figure of patient waiting time distribution in the same satisfaction degree, as shown in Figures 1–3.

About patients' satisfaction level, the following points should be noted.

(1) The hospital is a special service agency; its service system is different from dining hall or bank and its process is more complicated. Patients can choose any type of registration for treatment, so after receiving

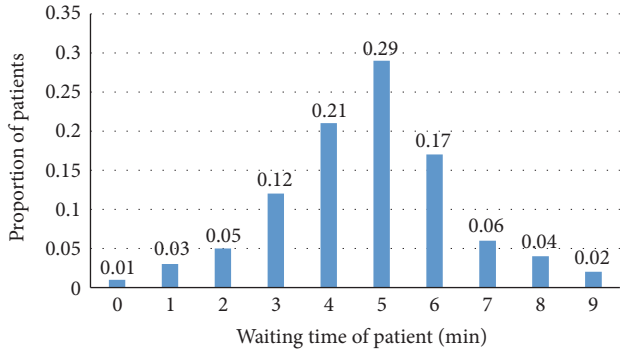


Figure 1: The distribution of waiting time under satisfaction level.

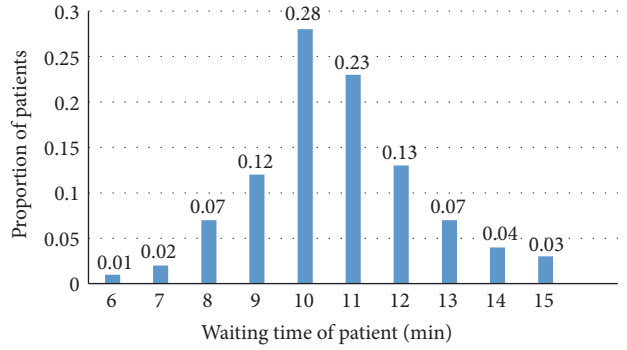


Figure 2: The distribution of waiting time under dissatisfaction level.

the first service from one doctor outpatient department, the patients are likely to go to other service systems such as pharmacy and radiology and, then through the queuing, wait to receive service again and finally they may leave or go back to the first queuing system. In this paper, the satisfaction level of patients we study only refers to the fact that patients produce negative emotions when they get into queuing system for the first time.

(2) In the whole process of treatment, patients not only have to queue for many times, but also through a number of staff positions in hospital. So the factors that influence patient satisfaction degree not only contain the length of waiting time at the first time entering into queuing system, but also will be affected by other aspects of service quality. This paper highlights the number of doctor outpatient departments for every type of registration ignoring other factors and only considers the influence that the length of waiting time brings to patient satisfaction for the first time of queuing.

2.3. Doctor Behavior Analysis. If the number of doctor outpatient departments has irrational allocation for every type of registration, it will not only make patients wait long time, but also affect the doctor work's efficiency and accuracy. On one hand, if more doctor outpatient departments are opened, it will greatly shorten the waiting time of patients, but it increases operating costs of the hospital and causes the waste of doctor resources. On the other hand, if setting up the low cost as the center goal of hospital and reducing the number of doctor outpatient departments, it will lead to serious queuing phenomenon of patients and increase the doctors' workload. As a result, it will let the misdiagnosis rate rise because of no time to rest for doctors and ultimately affect patient satisfaction. What is more, it can make the relationship between doctor and patient become more nervous.

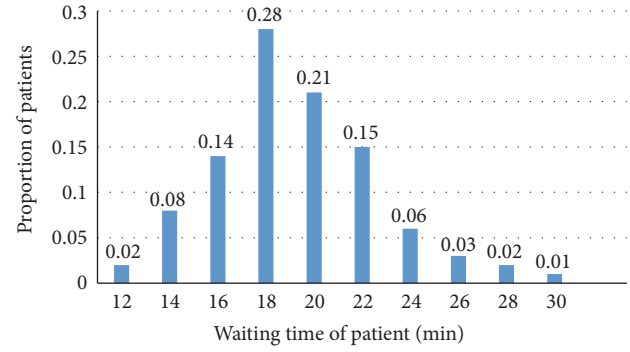


Figure 3: The distribution of waiting time under very dissatisfaction level.

2.4. Patient Satisfaction Model. The major research of customer satisfaction is the relationship between their psychological characteristics and activities; thus human behavior characteristics should be mainly considered to improve the customer satisfaction. There are many valuable researches about human behavior. Zhihong et al. [14] used the similarity of Tanimoto to measure customer behavior and then designed genetic clustering algorithm to divide customer groups and studied different purchasing behaviors of these groups. They designed a feature extraction method of multigroup based on genetic algorithm. We can apply this method into many fields of customers. Chun and Yiping [15] proposed an agent-based modeling and simulation method to analyze customer behavior and the effectiveness of personalized recommendation strategy. This paper only considered the mobile electronic commerce situation. Tingting et al. [16] made a survey of a large supermarket and analyzed customer behavior characteristics based on prospect theory and then built corresponding mathematical model to make optimal scheduling for retail checkout. The situation of this research was similar to hospital. All these articles are involved in customer satisfaction.

The environment of hospitals, staff service attitude, and the length of waiting time can affect patient satisfaction. This paper analyzes patient and doctor behavior characteristics by using the theory of behavior science and then establishes patient satisfaction function of waiting time.

People's perception of behavior and satisfaction is fuzzy and uncertain. The prospect theory has great advantage to describe people behavior characteristics; we can accurately

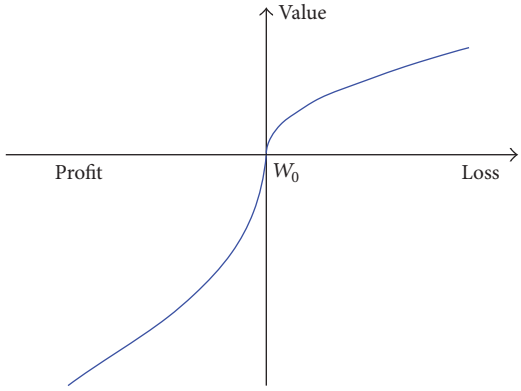


Figure 4: The value function diagram.

build descriptive model through analyzing patient satisfaction degree based on this theory. However, the prospect theory lacks mathematical theory to support. Fuzzy theory has good mathematical theory basis, but it cannot accurately describe patient behavior characteristics. So we combined the two methods effectively to establish patient satisfaction function.

Prospect theory combines psychology with economics effectively and uses the value function to represent policy-makers' subjective value. Its core is the selection of reference point. The evaluation of effectiveness about prospect theory is based on the reference point W_0, and when the result of event is W, we can use amplitude and direction which W deviates W_0 to evaluate effectiveness. The formula is $W - W_0$. We treat the expected waiting time of patient as W_0 and according to the issue we study, we can make the plot describe the problem as shown in Figure 4.

When patient's actual waiting time is longer than expected waiting time (reference point W_0), there will be a loss; otherwise it represents profit.

The value function transports the surface value into decision value. This paper treats the expected waiting time of patient as a reference point because patients pay more attention to the value of difference between expected waiting time and actual waiting time rather than the result itself. According to the prospect theory which Kahneman and Tversky [17] have put forward, we establish the value function of patient j (j means any one of patients):

$$V_j(x) = \begin{cases} x^{\alpha}, & x \geq 0, \\ -\lambda(-x)^{\beta}, & x < 0, \end{cases} \quad j = 1, 2, \ldots, n, \tag{1}$$

where α, β are coefficients of attitude to risk. The bigger α and β are, the more likely decision makers are to pursue risk. λ is the loss aversion coefficient and when $\lambda > 1$, it means that the decision makers are more sensitive to loss. Kahneman and Tversky's study concluded that $\alpha = \beta = 0.88$ and $\lambda = 2.25$ which are consistent with empirical data. There are also other researchers showing that β is bigger than α. Jianmin [18] proposed that $\alpha = 1.21$, $\beta = 1.02$, and $\lambda = 2.25$ under the situation of China and these results based on China actual conditions are different from foreign countries, so they are more valuable.

As the patient's perception of events is fuzzy and uncertain and the prospect theory can only describe this uncertainty qualitatively, it cannot make decision accurately. So this paper uses the membership function of fuzzy theory to obfuscate objective function. The specific steps are as follows.

Step 1. Determine the set of evaluations $T = (T_1, T_2, T_3) = (1, T(t_j), 0)$. T represents the satisfaction degree set; each value in this set means satisfaction, dissatisfaction, and very dissatisfaction. t_j means the actual time that patient j leaves this queuing system; d_j means the expected time that patient j leaves this queuing system; and $M_j = t_j - d_j$ means the lag time of the patient j. When $t_j > d_j$, it means a loss; otherwise there will be profit.

Step 2. Determine the threshold of each satisfaction degree. According to the analysis of survey results, the threshold of satisfaction degree is related to the expected waiting time (d_j) of patient j, coefficient of attitude to risk (β), and loss aversion coefficient (λ). There is a formula $R_j = d_j + (1/\lambda)^{1/\beta}$ describing the relationship between them which is concluded by other researchers. When $t_j \geq R_j$, the value of satisfaction membership degree equals 0 ($T(t_j) = 0$). When $0 \leq t_j < d_j$, the value of satisfaction membership degree equals 1 ($T(t_j) = 1$). When $d_j \leq t_j < R_j$, the people prefer to pursue risk when there is a loss. According to prospect theory, we can make derivation under formula (1):

$$\begin{aligned} T(t_j) &= 1 - \left(-V_j(x)\right) = 1 + V\left(-t_j + d_j\right) \\ &= 1 + \left[-\lambda\left(-\left(-t_j + d_j\right)\right)^{\beta}\right] = 1 - \lambda\left(t_j - d_j\right)^{\beta}. \end{aligned} \tag{2}$$

This paper uses the typical parameters $\alpha = 1.21$, $\beta = 1.02$, and $\lambda = 2.25$.

Step 3. Establish satisfaction function.

In the end, the satisfaction membership function can be established, as in the following formula. Figure 5 can clearly describe trend of the function:

$$T(t_j) = \begin{cases} 1, & 0 \leq t_j < d_j \\ 1 - \lambda\left(t_j - d_j\right)^{\beta}, & d_j \leq t_j < R_j, \\ 0, & t_j \geq R_j \end{cases} \quad j = 1, 2, \ldots, n. \tag{3}$$

According to formula (1) and formula (3) and $R_j = d_j + (1/\lambda)^{1/\beta}$ which reveals the relationship between threshold R_j and expected waiting time d_j of patients, we can deduce that $\lambda(R_j - d_j)^{\beta} = 1$. As the domain of deduction is $d_j \leq t_j < R_j$ and the value of λ is equal to 2.25, the value interval of $\lambda(t_j - d_j)^{\beta}$ is $(0, 1)$. So it verifies correctness of the above model we build.

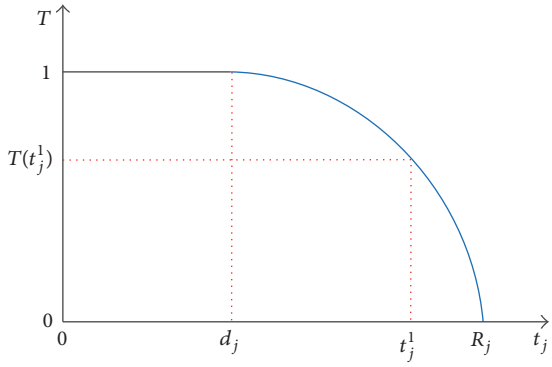


Figure 5: The satisfaction membership function diagram of patient j.

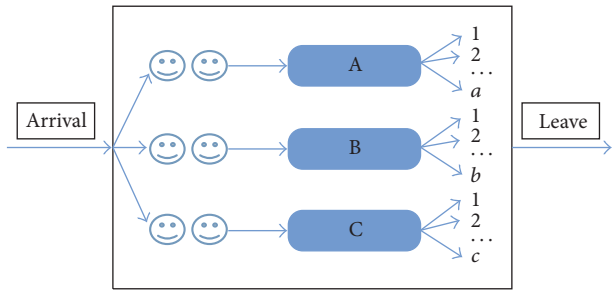


Figure 6: The treatment process diagram of patient j.

3. Description of Problem and Establishment of Model

3.1. Description of Problem. Hospital managers need to adjust the number of doctor outpatient departments according to patients' number and their registration type in order to improve patient satisfaction and reduce operating costs. Hospital belongs to service-oriented industry, and its highest goal is patient satisfaction. Thus, the problem can be described as maximizing patient satisfaction for the main goal and minimizing operating costs for the secondary goal. So it is a multiobjective scheduling problem. We can use a triple $\alpha/\beta/\gamma$ to describe this scheduling problem. The domain of α mainly describes the environment of resources; the domain of β mainly analyzes the constraints; and the domain of γ describes optimal scheduling goals.

3.1.1. Environment of Resources (α Domain). Patients can choose different types of registration according to their own conditions and demands. The service level is different under each type of registration (usually, the doctors of specialist-clinic have higher qualification and rich experience). As this paper focuses on solving the optimal scheduling problem of doctor outpatient departments, we make the following description. n indicates the total number of doctor outpatient departments; a indicates the number of doctor outpatient departments under general-clinic; b indicates the number of doctor outpatient departments under specialist-clinic; and c indicates the number of doctor outpatient departments under emergency-clinic. Figure 6 shows the simple process of patient treatment in hospital.

A, B, and C respectively represent general-clinic, specialist-clinic, and emergency-clinic.

3.1.2. Constraints and Rules (β Domain). Each patient should choose the appropriate registration type according to their own actual situation to accept service in doctor outpatient departments, so there is constraint of using different registration types. Let N_j describe this constraint. Before patient j entering into queuing system, the previous patient $j-1$ has priority to receive service. That means the present patient must have finished accepting service before the next patient receives service. In other words, the rule is first come first served (FCFS).

3.1.3. Optimal Scheduling Target (γ Domain). The hospital's ultimate target is to make profit as a service-oriented institution. For hospitals, the goal is to reduce operating costs as far as possible, but it is not a major one. The service quality is directly related to performance of hospital. If patients produce bad emotion because of long waiting time, it will influence the revenue of hospital. So improving patient satisfaction is the most important. Patient satisfaction is affected by many factors, but this article focuses on the waiting time that patient enters into the queuing system at first time, ignoring the interference of other factors. The first goal is to maximize patient satisfaction and we use the formula $\max \sum_{j=1}^{m} T(t_j)/m$ to express. m indicates the total number of patients. The second goal is to minimize the number of doctors, and we use the formula $\min \sum_{j=1}^{m}(a_{ij} + b_{ij} + c_{ij})$ to express. $\sum_{j=1}^{m} a_{ij}$, $\sum_{j=1}^{m} b_{ij}$, and $\sum_{j=1}^{m} c_{ij}$, respectively, represent the number of doctor outpatient departments under each type of registration.

3.2. Parameter Design. n means the total number of doctor outpatient departments; m means the total number of patients; i means any one of the doctor outpatient departments ($i \in [1, n]$); j means any one of patients ($j \in [1, m]$); η means the number of patients arrived at the hospital in a unit time and we treat it as arrival rate; μ means the number of patients finished accepting service in a unit time; it was treated as service rate and it can be regarded as doctor's work efficiency; c_i means the doctor's work time of any one doctor outpatient department and every doctor has fixed department; p_j means the length of time that patient j received service and it is related to the patients' situation; r_j means the submitted time of report for patient j; t_j means the actual time that patient j leaves this queuing system; d_j means the expected time that patient j leaves this queuing system; and M_j means the lag time of the patient j. The decision variables are as follows:

$$x_j = \begin{cases} 1, & \text{The patient is being served,} \\ 0, & \text{other cases,} \end{cases}$$

$$D_{ij} = \begin{cases} 1, & \text{The type of registration is A } (i \in \text{A}), \\ 2, & \text{The type of registration is B } (i \in \text{B}), \\ 3, & \text{The type of registration is C } (i \in \text{C}). \end{cases} \tag{4}$$

3.3. Basic Assumptions

Assumption 1. Each patient must select only one doctor outpatient department to get service.

Assumption 2. Do not consider the case that the patients go back to the queuing system for the second time.

Assumption 3. The patients will leave when waiting time exceeds a certain range.

Assumption 4. Each type of registration follows the principle of FCFS and the patients cannot change the queue when they selected.

3.4. Model Design. Treating doctor outpatient departments scheduling problem as parallel machine scheduling problem with different speed, we establish the following models:

$$\left(\frac{Q_m}{\text{prec}}, \frac{N_j}{\max \sum_{j=1}^{m} T\left(t_j\right)/m\,(\text{opt})}, \min \sum_{j=1}^{m} \left(a_{ij} + b_{ij} + c_{ij}\right) \right). \tag{5}$$

The corresponding mathematical models are as follows:

$$\max \quad Z = \frac{\sum_{j=1}^{m} T\left(t_j\right)}{m}, \tag{6}$$

$$\min \quad W = \sum_{j=1}^{m} \left(a_{ij} + b_{ij} + c_{ij}\right), \tag{7}$$

$$\text{S.t.} \quad M_j = \max\left(t_j - d_j, 0\right), \tag{8}$$

$$T\left(t_j\right) = \begin{cases} 1, & 0 \le t_j < d_j \\ 1 - \lambda\left(t_j - d_j\right)^{\beta}, & d_j \le t_j < R_j, \\ 0, & t_j \ge R_j \end{cases} \quad j = 1, 2, \ldots, n, \tag{9}$$

$$d_j \ge p_j + r_j, \tag{10}$$

$$c_i \ge \sum_{j=1}^{m} \left(a_{ij} + b_{ij} + c_{ij}\right)\left(p_j + r_j\right), \tag{11}$$

$$\begin{aligned} a_{ij} &= D_{ij} x_j, \\ b_{ij} &= \frac{1}{2} D_{ij} x_j, \\ c_{ij} &= \frac{1}{3} D_{ij} x_j, \end{aligned} \tag{12}$$

$$\sum_{i=1}^{n} \left(a_{ij} + b_{ij} + c_{ij}\right) = 1, \tag{13}$$

$$\sum_{j=1}^{m} \left(a_{ij} + b_{ij} + c_{ij}\right) \le m, \tag{14}$$

$$\begin{aligned} &x_j \in \{0, 1\}, \\ &D_j \in \{0, 1, 2, 3\}. \end{aligned} \tag{15}$$

In this model, formula (6) is the main objective function, formula (7) is the secondary objective function, and the rest of formulas from formula (8) to formula (14) are constraints. Formula (8) measures the delay time compared with expected waiting time of patient j. Formula (9) measures patient satisfaction degree. Formula (10) ensures the rationality of the patients' expected waiting time. Formula (11) ensures each doctor's work time is limited in any one of the doctor outpatient departments. Formula (12) means the registration type which patient j chooses to get service. Formula (13) ensures that each patient just only enters into one doctor outpatient departments and receives service for only once. Formula (14) ensures that the number of opened doctor outpatient departments is less than the total number of doctor outpatient departments. Formula (15) ensures the range of decision variables.

4. Algorithm Design

Since this problem about optimal scheduling of doctor outpatient departments belongs to the multiobjective combinatorial optimization. The number of combinations has an exponential relationship with the number of patients, so using the traversal method to search results is not feasible. In our article, the plant growth simulation algorithm (PGSA) is used to solve this problem. PGSA is the algorithm which applies the rules of plant phototropism into solving optimal scheduling problem. This algorithm compares the feasible region of optimal problem to plant growth environment and it compares global optimum solution to light which the plant requires to grow. So it can simulate the plant phototropism growth rules and then build growth model of tree trunks and branches under the action of different morphactins. In recent years, many scholars put PGSA into different research areas and have received satisfactory results. Such as Xuping et al. [19] in solving the distributive problem of emergency supplies, Qingkui et al. [20] in solving vehicle scheduling problem and Yang et al. [21] in solving scheduling of the bottleneck in auto maintenance shop are all using this algorithm.

4.1. Coding Design. Firstly, we number the patients and then encode them according to the way of real number. An array shows a treatment plan and the length of array is $m + n - 1$ and we use $-1 \sim -(m + 1)$ as identifiers of different clinics. The array of Figure 6 shows the case that there are ten patients and four clinics. We can know which clinic that patients enter according to above number. For example, in Table 1, the patient number 10 is in number 3 clinic.

Table 1: The coding example.

3	4	−1	6	8	−2	7	10	2	5	−3	1	9

4.2. The Probability Model of Plant Growth Simulation Phototropism. We assume the growth process of plant including roots, trunks, and branches. The initial feasible solution x_0 of the problem represents roots and the number of initial growing point is s which are in trunk L. These initial growing points are denoted by $A_{L1}, A_{L2}, \ldots, A_{Ls}$ and the corresponding morphactins are, respectively, denoted by $P_{L1}, P_{L2}, \ldots, P_{Ls}$. The number of initial growing point is r which are in branch H. These initial growing points are denoted by $A_{H1}, A_{H2}, \ldots, A_{Hr}$ and the corresponding morphactins are, respectively, denoted by $P_{H1}, P_{H2}, \ldots, P_{Hr}$. The computational formulas are as follows:

$$P_{Li} = \frac{f(x_0) - f(A_{Li})}{\sum_{i=1}^{s}(f(x_0) - f(A_{Li})) + \sum_{i=1}^{r}(f(x_0) - f(A_{Hi}))},$$
$$P_{Hi} = \frac{f(x_0) - f(A_{Hi})}{\sum_{i=1}^{s}(f(x_0) - f(A_{Li})) + \sum_{i=1}^{r}(f(x_0) - f(A_{Hi}))}. \quad (16)$$

In the formulas, $f(*) = 1/Z$ represents the objective function value of growing point and the goal of this algorithm is to search growing point which minimizes the value of $f(*)$. After determining the morphactin, we can decide whether to produce new branches in existing branches according to the value of morphactin. It means that the higher morphactin concentration of growing point will receive prior opportunity. According to formula (16), we know that $\sum_{i=1}^{s} P_{Li} + \sum_{i=1}^{r} P_{Hi} = 1$. First of all, we use computer generated random number in the domain of $[0, 1]$ and then use the way of throwball to select prior growing point. After generating a new growing point, we should compute its objective function value and update the value of morphactin. Repeat this process until there is no new branch and then we can obtain the optimal solution.

4.3. Algorithm Steps

Step 1. Generate initial feasible solution randomly according to the way of encoding we designed and then calculate the value of $f(x_0)$. Firstly, $x_{\min} = x_0$, $f_{\min}(x_{\min}) = f(x_0)$.

Step 2. Based on the root x_0, it generates 200 new growing points according to the method of two-point switching for improving the efficiency of search. The first switching position generates randomly and the next switching position generates according to the way of increasing four bits. If this position is beyond the individual interval after increasing four bits, the remaining bits should be researched from the initial part of individual again. It is similar to treating individual as a closed loop. If there are the same values in switching interval under adopting such operation, we make second adjustment of bits which have the same value and allow adjusting bits in switching interval in order to eliminate generating individual wrongly. Verify the feasibility of growing points according to constraints.

Step 3. Calculate the objective function value of growing points and then compare them with $f_{\min}$. If the objective function value of growing points is less than $f_{\min}$, then update $x_{\min}$, $f_{\min}(x_{\min})$.

Step 4. Calculate the morphactin value of feasible growing points according to formula (16).

Step 5. Use computer generated random number in the domain of $[0, 1]$ and then use the way of throwball to select the next prior growing point.

Step 6. Repeat Step 2~Step 3 until new growing point stopped generating or reached the number of iterations.

Step 7. Choose the optimal solution which minimizes the number of doctor outpatient departments and then determine the optimal scheduling scheme.

5. Case Study

A survey has been carried out in a high-level hospital in Harbin, China, and the goal is to make the optimal scheduling of doctor outpatient departments for the peak attendance period (8:00–10:00). The total number of doctor outpatient departments is 10, and the general-clinic's number is 4, the specialist-clinic's number is 3, and the emergency-clinic's number is 3. The average service time of general-clinic is 10 minutes and the service rate is 6 persons per hour; the average service time of specialist-clinic is 20 minutes and the service rate is 3 persons per hour and the average service time of emergency-clinic is 8 minutes and the service rate is 7.5 persons per hour. The total number of patients under this period is 112. The arrival rate of patients is $\eta = 0.02$ person/s which obey Poisson distribution. We use MATLAB to solve this scheduling problem. The number of iterations is set 89 times and it runs five times under the Windows 8 platform with RAM of 4 GB and basic frequency is 2.30 GHz. Its average running time is 68 s. In the steady state, the number of general-clinic is 3; the number of specialist-clinic is 3; and the number of emergency-clinic is 1; and the ultimate patient satisfaction degree is 0.92. The optimization results are shown in Table 2.

The row of "State" represents whether the doctor outpatient department is open. "1" indicates that it is open and "0" indicates that it is closed.

In order to analyze the performance of PGSA, we compare results in Table 2 with the results of traditional genetic algorithm (GA). The initialized population of GA is 200, the crossover rate is 0.85, the mutation rate is 0.05, and the number of iterations is set 89 times. The two algorithms run 5 times, respectively, under the same constraints and configuration of the computer. The convergence process of optimal solution is shown in Figure 7. In the figure, x-axis

Table 2: Optimization results.

	Category									
	General-clinic				Specialist-clinic			Emergency-clinic		
Number	1	2	3	4	1	2	3	1	2	3
State	1	1	1	0	1	1	1	1	0	0
The number of patients	21	21	20	0	12	12	13	13	0	0

Table 3: The comparative analysis results of PGSA and GA.

Algorithm	The number of general-clinics	The number of specialist-clinics	The number of emergency-clinics	Satisfaction degree	Running time of CPU
GA	3	2	2	0.88	75
PGSA	3	3	1	0.92	68

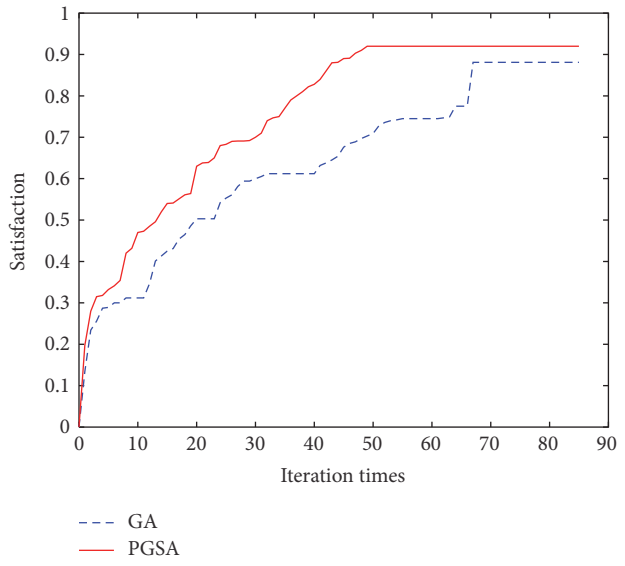


Figure 7: The optimal solution of convergence.

indicates the number of algorithm iterative times; y-axis indicates the average satisfaction degree of patients and we can see that patients' average satisfaction degree is increasing. Compared with GA, PGSA is able to stabilize the value in less iterations and it has characteristics of high convergence speed and efficiency which can meet patients' psychological needs minimizing the waiting time better. The comparative optimal result of PGSA and GA is shown in Table 3.

The results' simulation shows that setting appropriate number of doctor outpatient departments under different types of clinics can improve patient satisfaction effectively and reduce hospital operating costs at the same time and then achieve excellent management. Thus, adjusting the doctor outpatient departments' quantity under different types of clinics according to dynamic periods is very necessary choice in hospital.

6. Conclusion

Patients are main part of hospital, so it is very significant to improve patients' satisfaction. In this paper, an optimal scheduling strategy has been designed for the problem of patients queuing in the hospital. Firstly, the patient behavioral characteristics are analyzed, and then the patient membership functions and basic model are built to make optimal scheduling for doctor outpatient departments. Secondly, PGSA is used to solve this problem. Comparing with traditional algorithm GA, it takes less time to find the optimal solution which has better stability. The PGSA can solve optimal scheduling of doctor outpatient departments effectively in order to increase patient satisfaction and at the same time reduce operating costs in hospital. Finally, the simulation results of an example have proved the correctness and validity of the model and PGSA we designed. The paper focuses on different types of registrations without considering the different types of doctor outpatient departments. In the future, different types of doctor outpatient departments as a variable will be added in order to make the research of this problem close to reality greatly.

Competing Interests

The authors declare that they have no competing interests.

Acknowledgments

This work is supported by the National Nature Science Foundation of China (71371061) and Postdoctoral Foundation of Heilongjiang Province (LBH-Z15108).

References

[1] S. Mahar, K. M. Bretthauer, and P. A. Salzarulo, "Locating specialized service capacity in a multi-hospital network," *European Journal of Operational Research*, vol. 212, no. 3, pp. 596–605, 2011.

[2] N. Meskens, D. Duvivier, and A. Hanset, "Multi-objective operating room scheduling considering desiderata of the surgical team," *Decision Support Systems*, vol. 55, no. 2, pp. 650–659, 2013.

[3] J. R. Villamizar, F. C. Coelli, W. C. A. Pereira, and R. M. V. R. Almeida, "Discrete-event computer simulation methods in the optimisation of a physiotherapy clinic," *Physiotherapy*, vol. 97, no. 1, pp. 71–77, 2011.

[4] P. Landa, R. Aringhieri, P. Soriano, E. Tànfani, and A. Testi, "A hybrid optimization algorithm for surgeries scheduling," *Operations Research for Health Care*, vol. 8, pp. 103–114, 2016.

[5] J. L. Vile, J. W. Gillard, P. R. Harper, and V. A. Knight, "Time-dependent stochastic methods for managing and scheduling Emergency Medical Services," *Operations Research for Health Care*, vol. 8, pp. 42–52, 2016.

[6] P. D. Wright and S. Mahar, "Centralized nurse scheduling to simultaneously improve schedule cost and nurse satisfaction," *Omega*, vol. 41, no. 6, pp. 1042–1052, 2013.

[7] D. A. Stanford, J. M. Lee, N. Chandok, and V. McAlister, "A queuing model to address waiting time inconsistency in solid-organ transplantation," *Operations Research for Health Care*, vol. 3, no. 1, pp. 40–45, 2014.

[8] S. P. J. van Brummelen, W. L. de Kort, and N. M. van Dijk, "Waiting time computation for blood collection sites," *Operations Research for Health Care*, vol. 7, pp. 70–80, 2015.

[9] J. T. van Essen, E. W. Hans, J. L. Hurink, and A. Oversberg, "Minimizing the waiting time for emergencysurgery," *Operations Research for Health Care*, vol. 1, no. 2-3, pp. 34–44, 2012.

[10] L. Shuangxi, S. Yi, and J. Lian, "Establishing an outpatient service management system that suits the time pattern of patients' visits," *Chinese Journal of Hospital Administration*, vol. 17, no. 5, pp. 296–298, 2001.

[11] C. Lakshmi and S. A. Iyer, "Application of queueing theory in health care: a literature review," *Operations Research for Health Care*, vol. 2, no. 1-2, pp. 25–39, 2013.

[12] W. Xuping, M. Chao, and R. Junhu, "Emergency supplies optimal scheduling considering the public's psychological risk perception," *System Engineering Theory and Practice*, vol. 33, no. 7, pp. 1735–1742, 2013.

[13] J. Yang, S. Wei, and D. Qiulei, "Model of disruption management with actors in single machine scheduling," *Journal of Mechanical Engineering*, vol. 49, no. 14, pp. 191–198, 2013.

[14] Z. Zhihong, K. Jisong, and C. Fuzan, "Feature extraction of customer purchase behavior based on genetic algorithm," *Pattern Recognition and Artificial Intelligence*, vol. 23, no. 2, pp. 256–266, 2010.

[15] J. Chun and Z. Yiping, "Agent-based simulation model of customer behavior and personalized recommendation," *Systems Engineering-Theory & Practice*, vol. 33, no. 2, pp. 463–472, 2013.

[16] W. Tingting, Y. Qin, and H. Lin, "Optimization scheduling of retail checkout based on customers' behavior," *Computer Engineering and Applications*, vol. 52, no. 3, pp. 266–270, 2016.

[17] D. Kahneman and A. Tversky, "Prospect theory: an analysis of decision under risk," *Econometrica*, vol. 47, no. 2, pp. 263–291, 1979.

[18] Z. Jianmin, "An experimental test on cumulative prospect theory," *Journal of Jinan University (Natural Science)*, vol. 28, no. 1, pp. 44–48, 2007.

[19] W. Xuping, Z. Nana, and Z. Hongxin, "Emergency material allocation model considering the non-rational psychological comparison of the victims," *Chinese Journal of Management*, vol. 13, no. 7, pp. 1075–1080, 2016.

[20] C. Qingkui, L. Xinyu, and R. Xiangyang, "Vehicle scheduling problem based on plant growth simulation algorithm," *Systems Engineering-Theory Practice*, vol. 35, no. 6, pp. 1449–1456, 2015.

[21] Q. Yang, G.-H. Zhou, and J.-J. Lin, "Scheduling of machines in parallel with different speeds based on plant growth simulation algorithm-taking an example of the bottleneck in 4S auto dealership maintenance shop," *System Engineering Theory and Practice*, vol. 32, no. 11, pp. 2433–2438, 2012.

Hindawi Publishing Corporation
Scientific Programming
Volume 2016, Article ID 3403058, 8 pages
http://dx.doi.org/10.1155/2016/3403058

Research Article

Impact of Personnel Flexibility on Job Shop Scheduling

Ren Lin,[1,2] Guohua Zhou,[2] Aijun Liu,[3] Hui Lu,[4] and Tonglei Li[5]

[1]*School of Economics and Management, Southwest Jiaotong University, Chengdu 610031, China*
[2]*College of Mathematics and Computer Science, Hunan City University, Yiyang 413000, China*
[3]*School of Economics and Management, Xidian University, Xi'an 710071, China*
[4]*Tianhua College, Shanghai Normal University, Shanghai 201815, China*
[5]*Department of Information and technology, Hebei Vocational College for Correctional Police, Hebei, China*

Correspondence should be addressed to Tonglei Li; 19667312@qq.com

Received 17 June 2016; Accepted 7 September 2016

Academic Editor: X. Wang

Considering the lack of the research on the relationship between HR flexibility and scheduling effect, a resource-competency matrix-based method was proposed in order to reveal the quantitative relationship between them. Meanwhile, a job shop scheduling model with HR flexibility was established and the improved genetic algorithm was used to solve the model. A case analysis demonstrated significant impact of HR flexibility on the scheduling effect, which provided valuable guidance for building flexible manufacturing systems.

1. Introduction

In the traditional production scheduling problems, it is usually assumed that the workpiece processing time is a constant and the permanent workers are always available. However, in the actual production activities, the processing time of workpiece is not fixed, which changes with shifts of processing workers. For instance, in a standardized operating environment, the varying technical proficiencies of skilled and new workers will result in differences in processing time. For the problems concerning the optimal scheduling of product design projects, Yingzi et al. [1] proposed evaluation indices for task-personnel-resource matching degree and relevant calculation methods. They introduced the matching degree between the designer, technical resources, and design tasks into the scheduling model, along with the intensity of technical resources. Bixi et al. [2] studied the HR scheduling process in SMEs, where the staff deployment program is adjusted according to the production task requirements and different loads in production units to implement a self-adaptive division of labor model. For the problems concerning the optimal scheduling of production management, Yaling et al. [3] studied the single-piece, small-batch flexible job shop scheduling problem by considering the diversity of operator skills and difference in skill levels for different tasks and calculated the matching degree between task members based on the operators' cumulative operating time on each task. A variety of methods are used to study the task-personnel-resource matching problem. Li et al. [4] acquired the optimal labor allocation scheme for flexible manufacturers with the genetic algorithm and dynamic programming. Smet et al. [5] solved the staffing problem where labor cost was decided by job sequence by employing the dynamic programming. Pan et al. [6] proposed a two-phase heuristic algorithm for integer programming to solve the staff scheduling problem. Personnel arrangements in flow shop scheduling were studied preliminarily in literatures [7, 8]. Di Francesco et al. [9] classified multiskilled assemblers into several categories according to the skill level and created a mixed integer programming model targeting minimizing the number of staff rotations between assembly lines to solve the problem. Corominas et al. [10] divided motorcycle assemblers into two categories, that is, skilled and unskilled ones, and built a staffing model aiming to reduce the tact time and lower the number of workers. Parisio and Neil Jones [11] built an integrated optimization-based framework for workforce planning in the retail market. Sophisticated forecasting methods are integrated with stochastic programming. Miralles et al. [12]

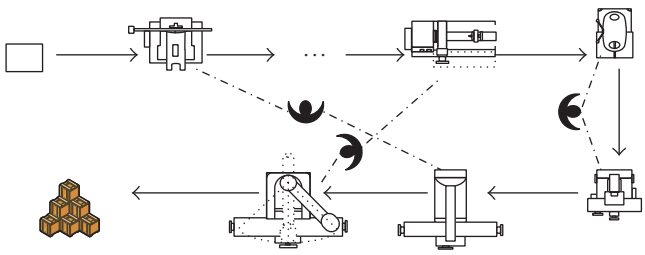


Figure 1: U-shaped production line.

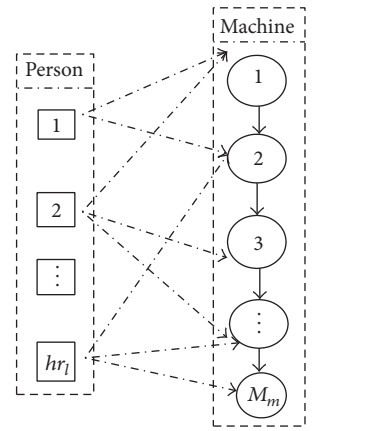


Figure 2: Personnel-machine relationship diagram.

created a staffing model with the maximum productivity and employee satisfaction as the goals and then used branch and bound algorithm to solve the problem. Li and Kelin [13] built a worker-task matching model for the flexible manufacturers with the aim of optimizing labor allocation and job sequencing. Genetic algorithm and dynamic programming method were used for solving the problem. Jun et al. [14] proposed the scheduling model based on the resources optimization disposition, the resource costs, and earliness or tardiness penalty, in which three-layer encoding method was used to realize dynamic scheduling of flexible constraints based on the variable craft route and the man-machine coordination. The nondominated sorting genetic algorithm was also used to optimize production path under the coexistence of multiple production lines, simultaneously, and optimize the wage cost and earliness-tardiness penalty of workpiece. Cao et al. [15] built an assembler scheduling model to maximize the sum of job adaptability. A heuristic algorithm based on the fitness matrix is presented. Mingzhou and Na [16] created a task competence index-based optimized shop staffing model for the goals of job quality and operating time. The hybrid particle swarm optimization (PSO) algorithm was used to solve the model, and the specific solving process was given.

The above representative research findings on scheduling problem under the constraint of personnel flexibility, respectively, indicated the preliminary modeling and relationship between human resources and task matching in the product design; personnel assignment problem under different process routes; and worker-task matching model for production shops. However, these studies were limited to specific cases merely, which lack in-depth exploration on the relationship between personnel flexibility and scheduling effect, particularly the analysis and research into the quantitative relationship between them. As the most important resource in production activities, the study of scheduling problem with personnel flexibility is of great practical significance. Thus, this paper studies the job shop scheduling problem with personnel flexibility with the improved genetic algorithm by referring to the existing research on personnel flexible job scheduling and analyzes the rules and the interaction between personnel flexibility and scheduling effect, striving to provide a theoretical basis for the optimal designing and implementation of job shop scheduling.

2. Mathematical Model of Personnel Flexible Job Scheduling

Suppose the number of workpieces is n, then $J = \{J_1, J_2, \ldots, J_n\}$, $\forall J_i \in J$, which need to be processed through procedures $O_i = \{O_{i,1}, O_{i,2}, \ldots, O_{i,j}, \ldots, O_{i,n_i}\}$. The machine set is $A = \{M_1, M_2, \ldots, M_m\}$. One or more processing machines are available for each procedure and the processing time might vary by different machines. Each worker can operate one or more machines and the worker set is denoted as $H = \{hr_1, \ldots, hr_l\}$.

2.1. Measurement of Personnel Flexibility. Personnel flexibility refers to the corporate personnel's ability to quickly and efficiently handle different tasks with uncertain changes during the production process, which emphasizes the versatility of employees. To study the impact of personnel flexibility on corporate productivity, the personnel flexibility is characterized by the personnel-machine relationship diagram. Assume that there are hr_l workers, who are independent of each other, as well as M_m machines in the production system. The corresponding actual production line and personnel-machine relationship diagram are shown in Figures 1 and 2. Figure 2 demonstrates that different workers have the skills to operate different machines. For example, person one can operate machine one and machine two; however, person two can machine one, machine two, and another machine.

Based on the personnel-machine relationship diagram, a matrix structure can be mapped, which is known as personnel-machine (PM) matrix. PM matrix is a matrix of $hr_l \times M_m$ denoted by PM. PM_{gv} is the element corresponding to the vth column of the gth row.

$$\mathrm{PM}_{gv} = \begin{cases} 1, & \text{Personnel } g \text{ is capable of operating machine } v \\ 0, & \text{Personnel } g \text{ is incapable of operating machine } v. \end{cases} \quad (1)$$

PM matrix can be obtained.

$$\mathrm{PM} = \begin{bmatrix} \mathrm{PM}_{11} & \mathrm{PM}_{12} & \cdots & \mathrm{PM}_{1M_m} \\ \mathrm{PM}_{21} & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ \mathrm{PM}_{hr_l 1} & \cdots & \cdots & \mathrm{PM}_{hr_l M_m} \end{bmatrix}. \quad (2)$$

For PM matrix of any size, the flexibility of production line workers can be measured with the flexibility equation. The relevant calculation equation is shown

$$\mathrm{FI} = \frac{\left(\sum_{g=1}^{hr_l} \sum_{v=1}^{M_m} \mathrm{PM}_{gv}\right)}{hr_l \cdot m}. \quad (3)$$

FI value ranges between $(0, 1]$. The greater the value of FI is, the higher the personnel flexibility of system will be, and vice versa. The greater FI value suggested that there were many workers in the system that can operate multiple machines and several scheduling plans available. Equation (3) effectively expresses the relationship between personnel flexibility and scheduling allocation. The existing literatures pay little attention to this issue.

For (3), three situations would occur during calculation: the number of rows was greater than the number of columns and square matrix; the number of columns was greater than the number of rows. To facilitate the description of subsequent cases, these three types of PM matrices were discussed preliminarily.

(a) When the number of rows was greater than the number of columns, $hr_l > M_m$. The demarcation between personnel flexibility and nonflexibility was the presence of at least one person who was capable of operating more than one device. To simplify the proof process, it was assumed that the machines were operated in ascending order of row numbers in the PM matrix. In this way, there would be $(hr_l - M_m)$ number of idle workers, who were ready for processing tasks. Under such conditions, FI calculation satisfied the following inequality:

$$\mathrm{FI} = \frac{\left(\sum_{g=1}^{hr_l} \sum_{v=1}^{M_m} \mathrm{PM}_{gv}\right)}{hr_l \times m} > \frac{M_m}{hr_l \times M_m} = \frac{1}{hr_l}. \quad (4)$$

(b) When the PM matrix was square, the number of rows was equal to the number of columns. Assuming that all the diagonal elements were 1, the calculation of FI satisfied the following inequality:

$$\mathrm{FI} = \frac{\left(\sum_{g=1}^{hr_l} \sum_{v=1}^{M_m} \mathrm{PM}_{gv}\right)}{hr_l \cdot m} > \frac{M_m\left(hr_l\right)}{hr_l \times M_m} = \frac{1}{hr_l\left(M_m\right)}. \quad (5)$$

(c) When the number of columns was greater than the number of rows, $M_m > hr_l$. To simplify the proof process, it was assumed that the machines were operated in ascending order of row numbers in the PM matrix. Then $(M_m - hr_l)$ number of machines were idle, and thus it was necessary to select $(M_m - hr_l)$ number of employees from hr_l number of them to operate the remaining $(M_m - hr_l)$ number of machines. Workers were already flexible in such an environment and correspondingly FI was calculated as shown in (5).

$$\mathrm{FI} = \frac{\left(\sum_{g=1}^{hr_l} \sum_{v=1}^{M_m} \mathrm{PM}_{gv}\right)}{hr_l \times m} > \frac{hr_l}{hr_l \times M_m} = \frac{1}{M_m}. \quad (6)$$

2.2. Personnel Flexibility Model. Conventional production scheduling problems seldom consider the impact of personnel. But, in the actual production, a machine may be operated by different workers, where the processing time varies with the workers' skills and experience levels. Therefore, production scheduling problems which take personnel flexibility into account are more complex ones. Firstly, arranging suitable machines for a certain procedure is required in the processing of workpiece. Secondly, appropriate personnel needs to be selected from a set of workers who were capable of operating the machine. Only in this way can the processing and sequencing of products actually be completed. $t_{i,j,hr}$ was utilized to represent the processing time of workpiece i on machine j (including the workpiece preparation time) by worker hr; $C(j_i, k, hr)$ denoted the completion time of workpiece j_i on machine k by worker hr; and π represented a sequence of all workpiece; then the completion time of n number of workpieces on m number of machines can be expressed by the following equations:

$$C\left(j_1, 1, hr\right) = t_{j_1,1} \quad (7)$$

$$C\left(j_i, 1, hr\right) = C\left(j_{(i-1)}, 1, hr\right) + t_{j_i,1} \quad i = 2, \ldots, n \quad (8)$$

$$C\left(j_{i_q}, k, hr\right) = \max\left\{C\left(j_i, k-1, hr\right), C\left(j_{(i-1)}, k, hr\right)\right\} + t_{j_i,k} \quad i = 2, \ldots, n;\ k = 2, \ldots, m \quad (9)$$

$$C_{\max}(\pi) = C\left(j_{n_q}, m, hr\right) \quad (10)$$

$$\pi^* = \arg\left\{C_{\max}(\pi) = C\left(j_{n_q}, m, hr\right)\right\} \longrightarrow \min, \quad (11)$$

where (10) was the maximum completion time and (11) denoted the scheduling order corresponding to the minimization of the maximum completion time.

In addition, production scheduling problems considering personnel flexibility also needed to satisfy the following constraints:

(1) One machine can only process one workpiece at a time.

(2) A workpiece can only be processed with a single machine at a time.

(3) No processing procedure can be interrupted once started.

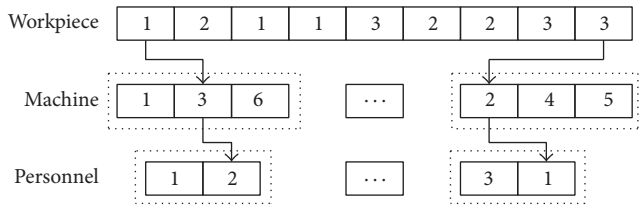


Figure 3: Three-layer encoding.

(4) Different workpiece owns the same priority level.

(5) There are no precedence constraints between procedures of different workpiece, while precedence constraints exist between procedures of the same workpiece.

(6) All workpieces are processable at time zero.

(7) Workers are available at any time as long as there is no conflict.

3. Improved Simulated Annealing Genetic Algorithm (ISAGA)

Genetic algorithm (GA) has fast convergence speed. When an excellent chromosome has a far higher fitness value than the average population in the computation, its probability of being selected increases in case of proportional selection, thereby leading to the "prematurity" phenomenon. The simulated annealing algorithm (SA) has the ability to jump out of local optima, but the principal shortcoming of simulated annealing (SA) is that it takes too much computer time. To solve this premature convergence and time-consuming problem, the paper proposes the improved simulated annealing genetic algorithm (ISAGA), so as to improve the optima searching performance.

3.1. Encoding. Efficient encoding mechanism can help reduce the complexity of computation and avoid repair mechanisms. In this paper, three-layer encoding was adopted. The first layer was procedure-based procedure sequence encoding, where the processing order of various procedures was determined. The second layer was machine-based machine allocation encoding, where the processing machine for each procedure was identified. The third layer was the encoding of workers who operate the machines. Such an encoding approach directly reflected the feasible allocation schemes during scheduling process, where feasible solutions could always be produced. A three-encoding example is illustrated in Figure 3. In this example, there are three workpieces; each workpiece has three operations. For the first operation of workpiece one, three machines (machines one, three, and six) can be used to execute this operation; if the machine three is selected, persons one and two meet the operational requirement; in this example, person two is selected to operate machine three.

Procedure-Based Encoding. Code length was the total number of procedures, while each code was a permutation of all procedures. The workpiece number which appeared at the jth time represented the jth procedure of the workpiece.

Machine-Based Encoding. Its length was the same as the procedure-based encoding. Each encoded bit corresponded to the processing machine selected for each procedure. Each position in the machine codes represented the sequence number of machine selected for the procedure in the set of available machines.

Personnel Encoding. Its length was the same as the procedure-based encoding. Each code corresponded to the processing machine for each procedure. Each position in the personnel codes represented the position of worker selected for operating the machine in the set of available workers.

3.2. Crossover Operator. The purpose of crossover operation is to retain the good information in the parent chromosomes through information exchange between them. In this paper, chromosomes consisted of three parts. The specific crossover process can be represented as follows.

(1) Procedure chromosomes: multiple workpieces were operated in each chromosome using workpiece-based POX crossover, which can well inherit the fine characteristics of parents.

(2) The workpiece set $J = \{J_i\}_{1 \le i \le n}$ was randomly divided into two sets: Jobset1 and Jobset2.

(3) Workpieces included in Jobset1/Jobset2 in the parent chromosome P1/P2 were copied to progeny chromosome C1/C2, while maintaining their locations and sequences.

(4) Workpieces not included in Jobset1/Jobset2 in the parent chromosome P1/P2 were copied to progeny chromosome C1/C2 according to their original sequences.

(5) Crossover of machinery and personnel chromosomes was done by the same method as used for the procedure chromosomes, while ensuring the correspondence between them was unchanged during the crossover process.

3.3. Mutation Operator. In mutation operation, minor disturbances were made on chromosomes by randomly altering certain genes in them to increase population diversity.

For procedure sequencing section, three mutation methods, that is, exchange, insertion, and reverse sequencing, were adopted. Each time, one of these mutation methods was randomly selected for operation.

For processing machinery selection part, a procedure was randomly selected, and then the processing machine currently in the chromosome was replaced with a different machine selected from the set of machines available for the procedure.

For personnel chromosome selection section, a machine was randomly selected from machine codes, and then the current staff in the chromosome was replaced with a different worker who was selected from the set of personnel capable of operating the machine.

3.4. Algorithmic Flow. For job scheduling problem in a personnel flexibility environment, the flow of ISAGA was as follows.

Step 1. Initialize the algorithm parameters (number of population *popsize,* maximum number of iterations $T_{\max}$, initial acceptance probability p_r, crossover probability *pc*, mutation probability *pm*, and annealing rate λ).

Step 2. Randomly generate initial population. Calculate the fitness value of each individual and assign f_{best} to the best solution among the current population, while assigning f_{worst} to the worst solution. Calculate the initial temperature $T_0 = -|f_{\text{best}} - f_{\text{worst}}|/\ln(p_r) = -|\Delta f|/\ln p_r$.

Step 3 (termination condition). The algorithm terminates when the maximum number of iterations N is reached. If the condition is satisfied, turn to Step 7; otherwise, turn to Step 4.

Step 4. Implement genetic operation on the population and calculate the fitness values of new individuals. If the fitness value is better than the optimal individual of the previous generation, replace the parent with the progeny while updating f_{best}; otherwise, retain the optimal individual of the previous generation.

Step 5. Implement ISAGA operation on the current optimal individual in the population, calculate the fitness value of newly generated individual, and compare the variation of fitness $\Delta f = f(k+1) - f(k)$ caused by two locations. If $\Delta f < 0$, accept the new location and if $\exp(-\Delta f/T_{k+1}) > \text{rand}$, also accept the new location; otherwise, retain the old location.

Step 6. $k = k+1$; then perform the annealing operation $T_{k+1} = \lambda T_k$, $\lambda \in (0, 1)$, and return to Step 3.

Step 7. Output the optimal solution obtained in this calculation.

4. Application and Analysis

In this paper, simulation experiment was performed on a computer with Intel Core 2 CPU/2.00 GHz/2.00 GB RAM using Matlab R2009b programming language. Algorithm parameters were set as follows: number of iterations 100; population size 50; crossover probability 0.8; mutation probability 0.1; annealing rate 0.98; initial acceptance probability $p_r = 0.7$.

To verify the impact of different personnel flexibilities on scheduling effect and to analyze the differences in scheduling effect between three situations of PM matrix (number of rows > number of columns; square matrix; number of rows < number of columns) and thereby identify the prominent impact of key personnel on scheduling effect, discussion was made for three situations.

When the number of rows is greater than the number of columns, the number of workers is greater than the number of machines. Relevant processing information is shown in Table 1, and corresponding PM matrix is shown in (12). Scheduling effect is shown in Figure 4.

Table 1: Flexible machine processing information.

Workpiece	Procedure	M_1	M_2	M_3	M_4	M_5
J_1	O_{11}	50	37	40	—	—
	O_{12}	—	30	—	20	20
	O_{13}	115	14	15	—	—
J_2	O_{21}	31	—	35	—	32
	O_{22}	40	30	—	—	60
	O_{23}	—	—	40	—	57
J_3	O_{31}	50	60	—	—	—
	O_{32}	—	40	—	30	50
	O_{33}	—	—	13	—	12
J_4	O_{41}	29	—	27	29	—
	O_{42}	—	26	—	24	—
	O_{43}	10	—	13	—	—
Worker		(1)	(2)	(3)	(4)	(5)

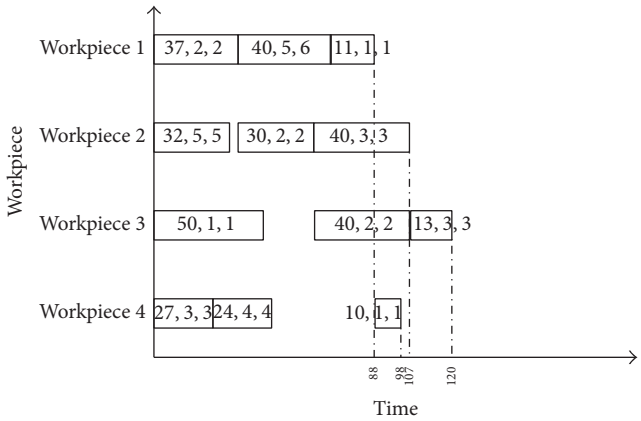


Figure 4: Scheduling effects.

$$\text{PM} = \begin{bmatrix} & M_1 & M_2 & M_3 & M_4 & M_5 \\ hr_1 & 1 & 0 & 0 & 0 & 0 \\ hr_2 & 0 & 1 & 0 & 0 & 0 \\ hr_3 & 0 & 0 & 1 & 0 & 0 \\ hr_4 & 0 & 0 & 0 & 1 & 0 \\ hr_5 & 0 & 0 & 0 & 0 & 1 \\ hr_6 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}. \quad (12)$$

As can be seen from (12), at that time, PM = 0.2 > 1/6.

Figure 4 shows the multiple simulation results of data in Table 1, where the data in block represent the run time, machines, and workers. For instance, 37, 2, and 2 represented that the first procedure of workpiece is processed by worker 2 on machine 2 for 37 time units. In addition, minimized maximum completion time is 120 time units.

To accurately reflect the correlation between scheduling effect and personnel flexibility for the case of number of rows greater than the number of columns, this paper begins the

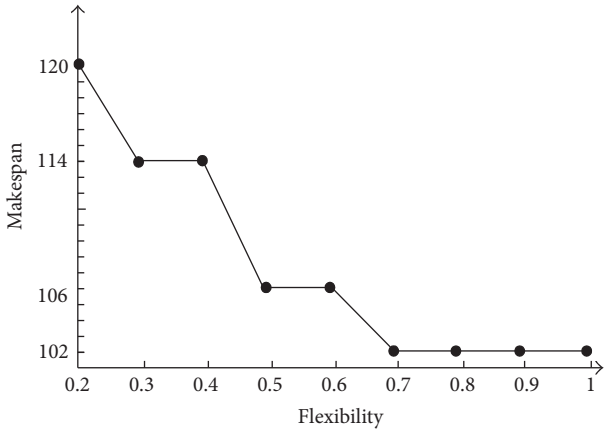


Figure 5: Relationship curve between personnel flexibility and scheduling effect.

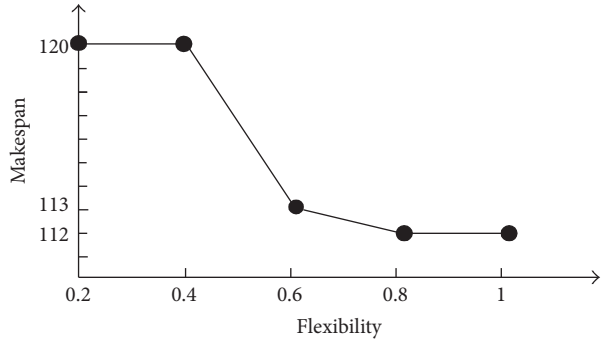


Figure 6: Relationship curve between personnel flexibility and scheduling effect.

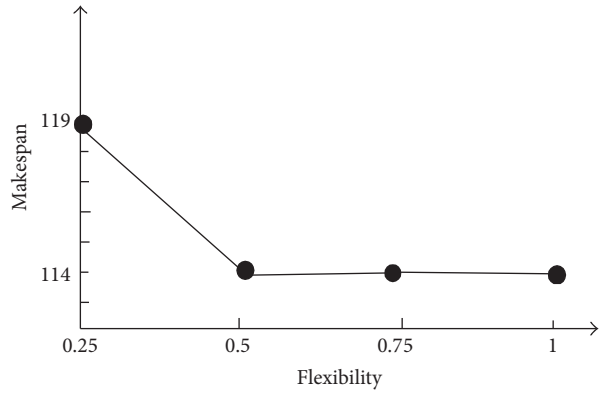


Figure 7: Relationship curve between personnel flexibility and scheduling effect.

study where each of six employees is only able to operate one machine till all employees could operate any machine. First rule of increasing personnel flexibility is to increase the number of operators in descending order of processing time. For instance, processing times of O_{22}, O_{31} on different machines are (M_1: 40, M_2: 30, M_5: 60), (M_1: 50, M_2: 60), and preferred machines for increasing personnel flexibility are M_5 or M_2. Meanwhile, number greater than machine code is filled when increasing the personnel flexibility. If the maximum number of personnel limits has reached, number was added starting from 1. If machine $M\{5\} = (5)$, operator of machine 5 is worker 5 and $M\{5\} = (5, 6)$ is added in the first step. As 6 is the maximum number of workers, $M\{5\} = (5, 6, 1)$ is added in the second step, and so forth. Each type of flexible configuration is run 20 times, and the most frequently recurring schedulability is taken as the scheduling result under respective flexibility configuration. Relationship curve obtained is shown in Figure 5.

As can be seen from Figure 5, when the row number is greater than the column number, that is, when the number of personnel is greater than the number of equipment pieces, minimized maximum completion time decreases gradually with increasing personnel flexibility and eventually stabilized.

The number of rows equaled the number of columns. If only the first five workers are considered, a processing environment comprising five machines and five workers would be formed. For initial production environment, it is assumed that the diagonal element in PM matrix is 1, while the rest is zero. Principle of adding worker flexibility is the same as above. Increasing step size is 5, which meant that five operators are added at a time for one machine. Relationship curve obtained is shown in Figure 6.

As can be seen from Figure 6, when the row number equaled the column number, that is, when the number of personnel equals the number of equipment pieces, minimized maximum completion time decreases gradually with increasing personnel flexibility, which eventually stabilizes. Comparing with Figure 5, it is found that the scheduling result where the number of personnel equaled the number of equipment pieces is inferior to the situation where the number of personnel is greater than the number of equipment pieces. The main reason is that more available personnel corresponded to better processing scheme. Difference in machine operation between workers would inevitably lead to a larger solution space, so more solutions can be searched when there are more personnel.

When the number of rows is less than the number of columns, the number of personnel is less than the number of machines. Assuming that only the first four workers are considered of whom hr_4 can operate machines 4 and 5. Principle of adding worker flexibility is the same as above, with a step size of 5. Relationship curve obtained is shown in Figure 7.

As can be seen from Figure 7, when the row number is less than the column number, that is, when the number of personnel is less than the number of equipment pieces, minimized maximum completion time decreases rapidly with increasing personnel flexibility, which eventually stabilizes. Comparing with Figures 5 and 6, it is found that the scheduling result where the number of personnel is less than the number of equipment pieces is inferior to the former two situations at flexibility of 1. This is mainly because the small number of personnel increases the flexibility of each worker, which however is still unable to completely satisfy enough choice space.

Table 2: Comparison of computational structure between algorithms.

Situation	Algorithm	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
(1)	ISAGA	100	100	100	100	100	97	96	97
	GA	100	100	100	100	99	93	91	90
	SA		100	100	95	92	91	90	90
(2)	ISAGA	100		100		100		100	100
	GA	100		100		100		95	95
	SA	100		100		100		94	95
(3)	Algorithm	0.25	0.5	0.75	1				
	ISAGA	100	100	100	100				
	GA	100	100	100	100				
	SA	100	100	100	100				

To comparatively analyze the performance of the proposed improved algorithm, three situations (1)~(3) are comparatively analyzed utilizing ISAGA, GA, and SA, separately. The algorithms are run 100 times at each degree of flexibility, and the numbers of times they converged to optimal solutions are recorded as shown in Table 2.

After the above analyses, the following conclusions can be drawn.

At a fixed number of equipment pieces, minimized maximum completion time decreases gradually with increasing personnel flexibility, which eventually stabilizes and no longer changes with changing personnel flexibility. This indicates that the unlimited increases in personnel do not necessarily bring high efficiency.

Comparative analysis of the three situations found that the minimized maximum completion time is superior for the situation when the number of personnel is greater than the number of devices to the other two situations at flexibility of 1. Moreover, the results are worst for the situation when the number of personnel is less than the number of devices. As shown in Figure 7, when the number of devices is greater than the number of personnel, the devices are in a starvation state. As shown in Figure 5, personnel flexibility has a theoretical optimal point. When the personnel flexibility is greater than this point, it cannot impact the scheduling result. Figure 6 presents the trend of changes in scheduling results when the number of personnel equaled the number of devices. As can be seen from the figure, the relationship curve changes slowly in such situation, where the scheduling result is not stabilized until the flexibility reached 0.8.

Simulation analysis demonstrates that the ISAGA has certain advantages in solving optimization problems of similar scale, with number of times converging to the optimal solution significantly superior to the other two algorithms. The improved algorithm can be used to solve large-scale optimization problems.

The above findings have important implications for guiding the design and optimization of flexible production lines.

5. Conclusion

In this paper, personnel flexibility scheduling problem aimed at minimizing maximum completion time is studied; ISAGA for solving the problem is proposed to make a classified study based on different personnel flexibilities. The results demonstrate certain interaction between the personnel flexibility and the scheduling effect. These conclusions have important guiding value for the improvement of corporate productivity. In the next step, our team will study the quantitative proof method, hoping to get more general, regular knowledge to better guide the design and optimization of production system flexibility.

Competing Interests

The authors declared no competing interests.

Acknowledgments

The authors are supported by Key Program of Hunan City University Grants (2015XJ02), China Postdoctoral Science Foundation Funded Project (2016M590929), and MOE (Ministry of Education in China) Project of Humanities and Social Sciences (13XJC630011).

References

[1] L. Yingzi, Z. Xiaodong, Z. Shuo et al., "Multi-mode optimal scheduling of product design project based on person-task-resource matching degree," *Industrial Engineering and Management*, vol. 17, no. 3, pp. 74–80, 2012.

[2] Z. Bixi, Z. Wanli, and S. Jing, "Ant colony-based labor allocation method for small-and-medium-size manufacturing workshops," *Industrial Engineering Journal*, vol. 14, no. 4, pp. 119–123, 2011.

[3] Z. Yaling, G. E. Maogen, H. Jing et al., "Personnel scheduling of flexible job-shop problem based on person-task matching degree," *Journal of Hefei University of Technology: Natural Science Edition*, vol. 38, no. 2, pp. 270–273, 2015.

[4] G. Li, X. Kelin, Z. Wei et al., "Staffing model in flexible manufacturing and its scheduling algorithm," *Journal of Harbin Institute of Technology*, vol. 44, no. 5, pp. 144–148, 2012.

[5] P. Smet, P. Brucker, P. De Causmaecker, and G. Vanden Berghe, "Polynomially solvable personnel rostering problems," *European Journal of Operational Research*, vol. 249, no. 1, pp. 67–75, 2016.

[6] Q.-K. Pan, P. N. Suganthan, T. J. Chua, and T. X. Cai, "Solving manpower scheduling problem in manufacturing using mixed-integer programming with a two-stage heuristic algorithm," *The International Journal of Advanced Manufacturing Technology*, vol. 46, no. 9–12, pp. 1229–1237, 2010.

[7] R. Ruiz and J. A. Vázquez-Rodríguez, "The hybrid flow shop scheduling problem," *European Journal of Operational Research*, vol. 205, no. 1, pp. 1–18, 2010.

[8] T. Nishi, Y. Hiranaka, and M. Inuiguchi, "Lagrangian relaxation with cut generation for hybrid flowshop scheduling problems to minimize the total weighted tardiness," *Computers & Operations Research*, vol. 37, no. 1, pp. 189–198, 2010.

[9] M. Di Francesco, N. D. M. Llorente, S. Zanda, and P. Zuddas, "An optimization model for the short-term manpower planning problem in transhipment container terminals," *Computers & Industrial Engineering*, vol. 97, pp. 183–190, 2016.

[10] A. Corominas, R. Pastor, and J. Plans, "Balancing assembly line with skilled and unskilled workers," *Omega*, vol. 36, no. 6, pp. 1126–1132, 2008.

[11] A. Parisio and C. Neil Jones, "A two-stage stochastic programming approach to employee scheduling in retail outlets with uncertain demand," *Omega*, vol. 53, pp. 97–103, 2015.

[12] C. Miralles, J. P. García-Sabater, C. Andrés, and M. Cardós, "Branch and bound procedures for solving the assembly line worker assignment and balancing problem: application to sheltered work centres for disabled," *Discrete Applied Mathematics*, vol. 156, no. 3, pp. 352–367, 2008.

[13] G. Li and X. Kelin, "Staffing model in flexible manufacturing and its scheduling algorithm," *Journal of Harbin Institute of Technology*, vol. 44, no. 5, pp. 144–148, 2012.

[14] L. A. Jun, Y. Yu, Z. M. Hua et al., "Multi-objective dynamic job-shop scheduling based on the optimal disposition of man-machine coordination," *Systems Engineering*, vol. 28, no. 3, pp. 46–52, 2010.

[15] L. Cao, B. Wang, F. Liu, and C.-B. Li, "Mixed-skill worker assignment optimization model for assembly line," *Journal of Chongqing University*, vol. 33, no. 12, pp. 21–26, 2010.

[16] L. Mingzhou and G. Na, "Optimization configuration model of manufacturing plant operators based on competence index," *Journal of Hefei University of Technology*, vol. 34, no. 10, pp. 1466–1469, 2011.

Hindawi Publishing Corporation
Scientific Programming
Volume 2016, Article ID 3717650, 13 pages
http://dx.doi.org/10.1155/2016/3717650

Research Article

Exploring an Interactive Value-Adding Data-Driven Model of Consumer Electronics Supply Chain Based on Least Squares Support Vector Machine

Xiao-le Wan,[1,2] Zhen Zhang,[3] Xiao-xia Rong,[3] and Qing-chun Meng[1,2]

[1]*School of Management, Shandong University, Jinan 250100, China*
[2]*Research Center for Value Co-Creation Network, Shandong University, Jinan 250100, China*
[3]*School of Mathematics, Shandong University, Jinan 250100, China*

Correspondence should be addressed to Qing-chun Meng; meqich@sdu.edu.cn

Received 18 May 2016; Accepted 17 July 2016

Academic Editor: Junhu Ruan

The differences in supply chains and their competitiveness depend on the differences in supply chain value creation systems. On the basis of the theory of value cocreation, this study investigates the interactive value creation of consumer electronics supply chains from the perspective of the interaction and added value created by the main value creation bodies in supply chains. Least squares support vector machine (LS-SVM) is innovatively introduced into the study on consumer electronics supply chains. A data-driven model is also established, the parameters of the method and kernel functions are optimized and selected, and an LS-SVM algorithm of consumer electronics supply chains is proposed to deal with the limited number of samples. Then, an empirical analysis of the top 10 smartphone supply chains in the Chinese market is conducted, and the LS-SVM model and other forecasting methods are compared. Results suggest that the LS-SVM model achieves a good predictive accuracy. This study also analyzes the value-adding structure of supply chains from the perspective of interaction and enriches the theory of value creation among supply chains. This study is conducive to helping consumer electronics enterprises to conduct market analyses and determine value growth points accurately.

1. Introduction

The continuous development of information technologies in recent years has resulted in the constant advancement of the consumer electronics industry under the social background of technological innovation. The consumer electronics industry has become the largest and fastest growing manufacturing industry in the world [1]. After 10 years of rapid development, the Chinese consumer electronics market reached the industrial scale of $4,059.01 billion in 2014; this value equated to 54.1% of the global share [2]. China has become the largest producer and consumer of electronic products. The Chinese consumer electronics industry has maintained its output value in trillion yuan and has become one of the major high-tech industries of the national economy. This industry has also played a positive role in improving the development level of social informatization and in promoting the integration of information and industrialization.

However, the extensive integration of the "Internet+" strategy and advanced information technologies, including mobile Internet, wireless sensing networks, Internet of Things, and cloud computing, has exerted unprecedented effects on consumer behavior and has raised new user requirements, thereby leading to new product types and intense market competition. In addition, high and new technologies promote the rapid emergence of new consumer electronic products. Intelligent production enables online production to keep abreast of offline production, subjects the traditional consumer electronics industry to the catfish effect, and intensifies the global competition in the consumer electronics industry. Therefore, the core competitive advantage of consumer electronics enterprises and the international

competitiveness of the consumer electronics industry must be enhanced.

The competition among consumer electronics enterprises revolves around supply chains. From the perspective of performance, the formation and source of supply chain competitive advantage depend on the operational efficiency and effectiveness of supply chains, the essence of which is the supply chain value creation system. The differences in supply chains and their competitiveness depend on the differences in supply chain value creation systems [3].

Traditional value creation theory contends that value is independently created by enterprises, whereas customers are the consumers of value [4, 5]. However, the theory of value cocreation fundamentally subverts the traditional value creation logic by indicating that a producer is not the sole value creator; that is, consumers can also create value [6, 7]. Consumers can create value in production and consumption [7]. However, the theory of value cocreation constantly expands toward the outer periphery of supply chains in practice. Some main value creation bodies excluded from supply chains tend to participate in supply chain value creation in different forms and exert an unexpected effect; specifically, these value creation bodies participate in open innovation [8], government policies and regulation [9, 10], and guidance of industry associations [11].

From the perspective of the supply chain structure, different value creation bodies represent the contradiction between a traditional value creation idea and the theory of value cocreation either in a broad sense or in a narrow sense. The theory of value cocreation is a redeconstruction of the original supply chain value creation body, which transforms from the original single-value creator enterprise into one in which two value creators, namely, enterprises and consumers, or multiple value creators are involved. Therefore, the changes in value creators inevitably lead to different views on the value creation method. These changes also result in the differences in supply chain value creation systems.

Many scholars have considered consumers as supply chain value creators. Xiao et al. [12] considered that, under the e-commerce situation, the partner-based supply chain should be transformed into a consumer-oriented supply chain; they also suggested that consumers are the future and end of supply chain value creation. Meng et al. [13, 14] adopted a new prosumer model and mathematical models to conduct a series of studies and determined that the alliance of consumers and enterprises can improve supply chain value. Huang et al. [15], Cavalieri and Pezzotta [16], and Zhen [17] introduced consumers into the service-oriented manufacturing supply chain management system and turned them into "coproducers"; specifically, they provided production services and service-oriented production activities to maximize and increase customer value and enterprise value. Shen [3] noted that the overall value of a supply chain is created by the joint efforts of suppliers, manufacturers, distributors, and consumers. In addition, some off-chain value creators (such as an open innovation body) are incorporated into supply chain value creation systems, as described in some preliminary theoretical studies and practices. Niu [18] used fuzzy mathematics theory, established a factor analysis model, and determined that open innovation exerts a crucial effect on the development of cluster supply chains. Sakamoto et al. [8] contended that enterprises should use open innovation as the key factor to enhance supply chain competitiveness and add value. IBM established the first supply chain innovation center in China to integrate and transform the global supply chain capability of the country by assisting global enterprises [19]. The aforementioned studies are based on the maximization of supply chain value, and they include consumers or other relevant participants into the supply chain value cocreation system. However, these studies fail to explain the microscopic mechanism of supply chain value creators after deconstruction.

As a result of the development of advanced information technologies, the integration of intelligence and the interconnection highlights the increasingly prominent role of consumers with regard to consumer electronic products, including mobile Internet, wireless sensing networks, Internet of Things, and cloud computing. An increasing number of consumers involved in the value creation process have become the most important value creators of consumer electronics supply chains. For example, Xiaomi represents the typical value cocreation of a consumer electronics supply chain; this company took only four years to become the fifth largest mobile phone provider in the world (IDC Report, 2015). In addition, open innovation gradually narrows the interaction between an enterprise and external parties; thus, most of the innovative ideas of an enterprise come from outside forces. For example, consumer electronics supply chain brands Haier and Midea have successfully shifted to open innovation. Obviously, the traditional enterprise-oriented supply chain creation mode cannot effectively explain the value creation problems of consumer electronics supply chains. Thus, value cocreation has become a new method for creating supply chain value.

From a deep point of view, the core of value cocreation theory is that the value creator completes value cocreation behavior through heterogeneity interaction; interaction provides the value creator with a means to achieve value cocreation [7, 20]. Technical innovation blurs the boundaries between companies and individuals and provides increasing opportunities for the interaction between enterprises and consumers and for the interaction among consumers [21]. Consumers create value through their consumption practices or through the interaction opportunities provided by enterprises [22]. Interaction is a service experience and user value creation model [23]. It also represents the capacity of enterprises and consumers to obtain useful information and achieve value effectively and constantly [24]. Therefore, interaction can be regarded as a competitive advantage of enterprises and a heterogeneous resource to improve value creation ability [24, 25].

The least squares support vector machine (LS-SVM), which is based on the equality constraint and the least squares loss function, is an extension of SVM [26]. LS-SVM is also used to solve pattern recognition, image classification, and nonlinear function estimation problems [27, 28]. Langone et al. [29] applied the LS-SVM approach with spectral clustering and regression to predict the maintenance of

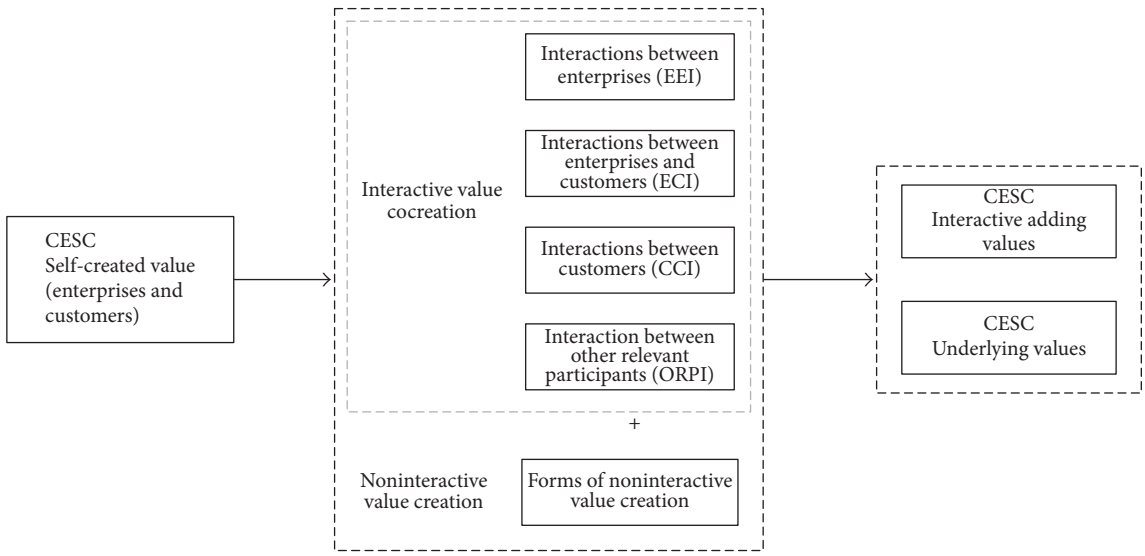


Figure 1: Value creation system of consumer electronics supply chains based on interaction.

industrial machines. Zhang and Shetty [30] adopted the LS-SVM approach to predict the surface roughness in machined surfaces. Yuan et al. [31] built an LS-SVM model to predict short-term wind power. Yu et al. [32] used LS-SVM to predict the temperature in a Chinese solar greenhouse. The LS-SVM method achieves a good prediction performance in regression forecasting. Although the LS-SVM method has been widely used for regression prediction in many disciplines, such as engineering, chemistry, meteorology, and environmental issues, the application of LS-SVM in the prediction of supply chain value is not extensively studied. The supply chain value creation problem exhibits industry-specific features and involves complex nonlinear problems that are jointly constituted by upstream and downstream enterprises, consumers, and other related parties [33]. The problem of a limited sample size (such as consumer electronics supply chains) inevitably emerges when focusing on the value creation issues of a specific industrial supply chain. The LS-SVM method achieves good adaptability in view of these two characteristics: the LS-SVM method has remarkable nonlinear mapping capabilities in prediction [34], and the LS-SVM method usually demonstrates good generalization ability even with a small sample size [35].

In the present study, we first consider the theory of value cocreation to explore the internal value creation mechanism of consumer electronics supply chains. This study also considers consumers, open innovation bodies, and other internal and off-chain bodies included in the value creation system of consumer electronics supply chains. The value creation system of consumer electronics supply chains is established such that it meets the actual development condition and the status of the consumer electronics industry. It innovatively introduces LS-SVM into the study on consumer electronics supply chains, establishes an interactive value-adding data-driven model, and optimizes and selects the parameters and kernel functions of the method to improve its prediction accuracy. Finally, the LS-SVM model and other prediction methods are compared, and the results are combined with the empirical data.

The remaining parts of this paper are arranged as follows. Section 2 establishes the value creation system of consumer electronics supply chains. Section 3 presents an LS-SVM-based data-driven model of consumer electronics supply chains and proposes a relevant algorithm. Section 4 conducts an empirical analysis of the Chinese smartphone supply chain and offers preliminary policy recommendations. Section 5 discusses the comparative study on the four different forecasting models and common kernel functions. Section 6 presents the conclusion.

2. Supply Chain Value Creation System Based on Interaction

According to the perspective of value cocreation, the supply chain value (system output) process is established on the basis of the independent creation of supply chain value creators (system input). This process is jointly achieved by various value creators through heterogeneous interactions and other noninteractive value behaviors (Figure 1).

The supply chain value creation process can be divided into two organic stages. The first stage is the basic value creation process of supply chains. Porter value chain theory explains that enterprises and consumers can create their own value, use this value as an initial supply chain value, input this value into the basic process subsystem of supply chain value creation, adopt a series of noninteractive value creation behaviors, and output the basic value of the supply chain. The second stage is the interactive value-adding supply chain process. Self-created value plays a key role in interactive value creation [36]. The basic value of a supply chain is the input value of the value creation subsystem; the added value of a supply chain is output through the heterogeneous

interactions among various value creators. These two stages occur simultaneously, and they serve as mutually initial values in the continuous and dynamic value creation process. These processes also reflect and enhance each other, and they promote harmony.

In the process of creating added value in the supply chain, interaction effectively increases value and serves as a main carrier of supply chain value-adding factors. Interaction mainly comes from the interaction among supply chain value creators. The theory of value cocreation mainly refers to the value cocreation or interaction between enterprises and consumers. However, Payne et al. [37] considered that the main parts of value cocreation activities are customer value creation, supplier value creation, and interface. Saarijärvi et al. [38] emphasized that "value," "cocreation," and "creation" mean that enterprise value or consumer value is created by different bodies through different forms of activities (B2C, B2B, C2B, or C2C). Moreover, interaction occurs between a consumer and a product (or service) provider and among all the bodies in the value network [39]. However, in practice, interaction occurs in all the cocreation behaviors of the main bodies in the supply chain. Furthermore, cocreation behaviors constitute the interactive value-adding structure of the supply chain.

The present study does not focus on the basic value creation process of supply chains and mainly explores the influence of interactive cocreation behaviors on supply chain value in the value-adding process. The report [40] performs a literature review and qualitative research and starts from the interaction and cocreation among enterprises, consumers, other participants, and main bodies; it also adopts the Delphi-AHP evaluation method to construct an interactive value-adding index system of supply chains (Table 1). This system includes 4 primary indicators, 14 secondary indicators, and 44 tertiary indicators. The index system characterizes the features of consumer electronics supply chains. Any change in an indicator causes a change in the output value of the interactive value-adding system.

3. Data-Driven Model of Consumer Electronics Supply Chains Based on LS-SVM

3.1. Data Collection. According to the interactive value-adding index system of consumer electronics supply chains in Table 1, the senior supply chain experts of the consumer electronics industry are invited to participate. Moreover, a five-point Likert scale is adopted, the vth v ($v = 1, 2, 3$) indicator of N supply chains is quantified and scored, and the input vector of the lth interactive value-adding electronics supply chain system x_l^v is obtained. The actual value of the lth interactive value-adding electronics supply chain system is recorded as y_l. Thus, the original data set of the data-driven model of consumer electronics supply chain $\{x_l^v, y_l\}_{l=1}^N$ is obtained.

3.2. Model Construction. The training set $\{x_k^v, y_k\}$, $k = 1, 2, \ldots, n$, $1 < n \leq N$, is selected; here, $x_k^v \in R^{d^v}$, where x_k^v refers to the variable of the vth indicator that affects the value of consumer electronics products, $v = 1, 2, 3$, and y_k refers to the value of the consumer electronics supply chain. The vth indicator's driving model of the consumer electronics supply chain based on LS-SVM is assumed as follows:

$$y_{\mathrm{CESC}}^v\left(x^v\right) = \left(G^v\right)^T \varphi\left(x^v\right) + b^v, \tag{1}$$

where y_{CESC}^v is the model predictive value of the vth indicator, $(G^v)^T$ is the weight vector, $\varphi(x^v)$ denotes nonlinear mapping from the input space to the high-dimensional feature space, and b^v is a partial vector [41].

According to the LS-SVM method, the error loss function $e_{v_k}^2$ and error penalty factor γ are introduced. The above problem (1) can be transformed into Model (2) as follows:

$$\begin{aligned} \min \quad & J\left(G^v, e^v\right) = \frac{1}{2}\left(G^v\right)^T G^v + \frac{\gamma^v}{2}\sum_{k=1}^{n} e_{v_k}^2 \\ \text{s.t.} \quad & y_k = \left(G^v\right)^T \varphi\left(x_k^v\right) + b^v + e_{v_k}, \\ & \qquad\qquad k = 1, 2, \ldots, n. \end{aligned} \tag{2}$$

The Lagrange function is defined as

$$\begin{aligned} & L\left(G^v, \gamma^v, e^v, a^v, b^v\right) \\ & = \frac{1}{2}\left(G^v\right)^T G^v \\ & \quad + \sum_{k=1}^{n}\left(\frac{\gamma^v}{2} e_{v_k}^2 - a_k^v\left(\left(G^v\right)^T \varphi\left(x_k^v\right) + b^v + e_{v_k} - y_k\right)\right). \end{aligned} \tag{3}$$

According to the KKT conditions of MP, the solution of problem (2) is equivalent to solving the following equations:

$$\begin{aligned} & G^v = \sum_{k=1}^{n} a_k^v \varphi\left(x_k^v\right), \\ & \sum_{k=1}^{n} a_k^v = 0, \\ & \qquad a_k^v = \gamma^v e_{v_k}, \ k = 1, \ldots, n, \\ & G^v \cdot \varphi\left(x_k^v\right) + b^v + e_{v_k} - y_k = 0, \quad k = 1, \ldots, n. \end{aligned} \tag{4}$$

The kernel function is defined as $K(x_i^v, x_j^v) = \varphi(x_i^v) \cdot \varphi(x_j^v)$, and G^v and e_v are eliminated. Problem (4) can be converted into the following equation:

$$\begin{pmatrix} 0 & \left(I_n^v\right)^T \\ 1 & K\left(x_i^v, x_j^v\right) + \dfrac{1}{\gamma^v} \end{pmatrix} \begin{Bmatrix} b^v \\ a^v \end{Bmatrix} = \begin{Bmatrix} 0 \\ y_{\mathrm{CESC}}^v \end{Bmatrix}. \tag{5}$$

The solutions a^v and b^v are calculated. Thus, the data-driven model based on LS-SVM is as follows:

$$y_{\mathrm{CESC}}^v\left(x^v\right) = \sum_{k=1}^{n} a_k^v K\left(x^v, x_k^v\right) + b^v. \tag{6}$$

Table 1: Interactive value-adding index system of consumer electronics supply chains.

Primary indicator B		Secondary indicator C		Tertiary indicator D	
B_1	Interaction between enterprises	C_1	Supply chain relationship quality	D_1	Cooperation
				D_2	Trust
				D_3	Communication
B_1	Interaction between enterprises	C_2	Supply chain learning effect level	D_4	Communication and exchange of knowledge
				D_5	Transition from tacit knowledge into dominant knowledge
				D_6	Absorption of knowledge and construction of new knowledge
B_1	Interaction between enterprises	C_3	Supply chain collaborative effect level	D_7	Common plan
				D_8	Shared decision making
				D_9	Joint implementation
				D_{10}	Long-term collaboration
				D_{11}	Win-win cooperation
B_2	Interaction between enterprise and consumer	C_4	Dialogue level between enterprise and consumer	D_{12}	Continuous, two-way, complete, and real-time information exchange
				D_{13}	Open and good information exchange
				D_{14}	Timely and effective dialogue to resolve disputes
B_2	Interaction between enterprise and consumer	C_5	Trust degree between enterprise and consumer	D_{15}	Not using "asymmetric information" to deceive consumers
				D_{16}	Not using "asymmetric information" to bring potential risks to consumers
B_2	Interaction between enterprise and consumer	C_6	Quality of interactive channel between enterprise and consumer	D_{17}	Easy access to enterprise information platform
				D_{18}	True information provided by the platform
				D_{19}	Can interact with any supplier through effective channels
B_2	Interaction between enterprise and consumer	C_7	Shared risk degree of enterprise and consumer	D_{20}	Informing consumers of all possible risks
				D_{21}	Establishing risk assessment and avoidance mechanism
				D_{22}	Establishing damage compensation system according to risk level
B_3	Interaction between consumers	C_8	Information acquisition level	D_{23}	Timely communicating information
				D_{24}	Inexperienced consumers can seek advice by obtaining information
				D_{25}	The information obtained can avoid potential risks
B_3	Interaction between consumers	C_9	Information sharing level	D_{26}	Sharing product (service) information and experience
				D_{27}	Sharing the psychological feelings of using the product or service
				D_{28}	Information sharing enhances the product's use value (or satisfaction, brand image)
B_3	Interaction between consumers	C_{10}	Mutual assistance degree	D_{29}	Close cooperation with each other
				D_{30}	Establishing special friendship under the same brand
B_3	Interaction between consumers	C_{11}	Innovation level	D_{31}	Consumer community providing or (publishing) innovative information or requirements
				D_{32}	Consumers innovatively using products or services through their own knowledge or skills
B_4	Interaction between other relevant participants	C_{12}	Government regulation level	D_{33}	Healthy and stable supply chain and good environment for enterprise growth (political, economic, social, and cultural aspects)
				D_{34}	Providing guarantees for legal governance
				D_{35}	Encouraging and fulfilling supply chain social responsibility
				D_{36}	Effectively controlling the risks of supply chain emergencies
				D_{37}	Government participates in the design of incentive mechanisms of closed-loop supply chain
B_4	Interaction between other relevant participants	C_{13}	Industry guidance level	D_{38}	Developing technical standards to guide industrial development
				D_{39}	Mitigating the internal interest conflicts of the industry
				D_{40}	Enhancing corporate brand image
				D_{41}	Considering the interests and needs of consumers
B_4	Interaction between other relevant participants	C_{14}	Open innovation activity participation level	D_{42}	Open, innovative participants participate in R&D
				D_{43}	Open, innovative participants participate in product (service) production
				D_{44}	Open, innovative participants participate in corporate governance

In the forecasting model, the commonly used kernel functions include the polynomial kernel function (POLY; $K(x_i, x_j) = (x_i, x_j + b)^d$), linear kernel function (LINEAR; $K(x_i, x_j) = (x_i, x_j)^d$), and sigmoid kernel function ($K(x_i, x_j) = \tanh[c_1(x_i, x_i) + c_2]$). In these kernel functions, the form of the radial basis kernel function (RBF kernel) is as follows:

$$K\left(x_i^v, x_j^v\right) = \exp\left(\frac{\left\|x_i^v - x_j^v\right\|}{2\left(\delta^v\right)^2}\right). \tag{7}$$

b and d in POLY must be adjusted and optimized. Thus, c_1 and c_2 in the sigmoid kernel function are uncertain, and LINEAR is a special case of the RBF kernel. The RBF kernel only comprises one nuclear parameter δ, and such simple structure makes it widely used for research. This study establishes LS-SVM models on the basis of different kernel functions to compare the performance of the RBF kernel and that of the other kernel functions. Moreover, this study performs a comparative analysis of the effect of kernel functions on model accuracy.

3.3. Model Solution

3.3.1. Model Parameter Optimization. The analysis in Section 3.1 indicates that the learning and generalization capabilities of the data-driven model greatly depend on the penalty coefficient γ and kernel function parameter δ. The grid search algorithm (GSA) is a popular method used to obtain the optimal solution of (γ, δ) [42]. Therefore, this study adopts the GSA to determine the two parameters for conducting the k-fold cross-validation of each set of parameters. However, errors are caused by different initial search areas when using the GSA embedded in the MATLAB LS-SVM toolbox (version 1.8). Therefore, this study tests the combinations of primary indicators for m times to ensure that the selected parameter achieves good adaptability. The average relative error is considered as the error of the model at the indicator level. Considering that the index system is constituted by three levels, the final error of the model is defined as the average error of the indicator model at different levels to ensure the consistency of the model.

If h test samples are under the vth level indicator, then the previous section assumes that n samples exist in the training set; thus, C_{h+n}^n combinations exist. In the sth combination, the average relative error of the corresponding training set is

$$e_s^{\text{train}} = \frac{(1/m)\sum_{j=1}^{m}\sum_{v=1}^{3}\sum_{i=1}^{n}\left|y_i - \left(\sum_{k=1}^{n} a_k^v K\left(x_i^v, x_k^v\right) + b^v\right)\right|}{y_i}. \tag{8}$$

The total error evaluation function is

$$e_s = \frac{e_s^{\text{train}}}{3(h+n)}. \tag{9}$$

3.3.2. Comparison and Optimization of Training Set and Test Set. The robustness and generalization ability of the LS-SVM model are affected by the number of training sets. This study produces an optimized design for each ratio of the test set over the training set to determine the optimal number of training sets. First, the ratio-related average relative error is defined as $E^h = (1/C_{h+n}^h)\sum_{s=1}^{C_{h+n}^h} e_s$. Then, the average relative error-standard deviation relationship between the number of test sets h and E^h is observed. The test set sample h^* that produces a small E^h and small standard deviation is selected. Finally, according to Formula (9), all the combinations with the smallest "training set–test set" combinations are chosen. The penalty coefficient of the sample γ^v and the width of the kernel function δ^v are calculated according to corresponding indicators.

3.3.3. LS-SVM Algorithm for Limited Samples. The LS-SVM algorithm of consumer electronics supply chains that is suitable for limited samples is obtained through the preceding analysis.

If the number of test set samples is h, then the number of training set samples is n, and the total number of samples is $N = h + n$. For each h, the total number of training set combinations is $p_h = C_N^h$. m is the number of repeated tests. v ($v = 1, 2, 3$) is the level of indicator.

Step 1. All the vth indicators of N are quantified and scored to derive $\{x_l^v, y_l\}_{l=1}^N$, which is the original data set of the data-driven model of consumer electronics supply chains.

Step 2. As for all $h = 1, \ldots, N-1$, the grid fivefold cross-validation method is adopted to calculate all the corresponding parameter combinations $(\gamma_1, \delta_1), \ldots, (\gamma_{p_h}, \delta_{p_h})$ of combinations $S_1, \ldots, S_{p_h}$. The calculation is repeated for m times. According to Formulas (8)-(9), the total average error of tertiary indicator $e_1, \ldots, e_{s_{p_h}}$ is calculated. Then, $E^h = e_1 + \cdots + e_{s_{p_h}}$, and the total average relative error is obtained.

Step 3. According to the relationship $h - E^h$, the low value of E^h and a number of selected test sets are considered, and the optimal h^* is selected.

Step 4. As for the optimal h^*, all the corresponding parameter combinations $(\gamma_1, \delta_1), \ldots, (\gamma_{p_h}, \delta_{p_h})$ of combinations $S_1, \ldots, S_{p_h}$ are calculated. The calculation is repeated for m times. According to Formulas (8)-(9), the total average error of tertiary indicator $e_1, \ldots, e_{s_{p_h}}$ is calculated, and the smallest error is selected and recorded as e_s^*. The corresponding optimal combination is S_s^*.

Step 5. The grid fivefold cross-test method is adopted to analyze S_s^*. The indicator of each level (γ^*, δ^*) is calculated.

Step 6. Parameter combination (γ^*, δ^*) is substituted into Formula (5). Then, solution (a^v, b^v) is obtained. According to Formula (6), the LS-SVM-based data-driven model of each level indicator can be calculated.

Table 2: Smartphone market in China (quarter 3, 2015).

Brand number	Brand name	Market share (%)
A	Xiaomi	16.4
B	Huawei	16.2
C	Apple	12.1
D	OPPO	10.2
E	Vivo	10.0
F	Samsung	8.8
G	Meizu	6.7
H	Coolpad	4.8
I	Lenovo	4.4
J	ZTE	2.7
	Others	7.6

Source: HIS Technology, 2015 [43].

4. Empirical Analysis of Smartphone Supply Chain

4.1. Data Sources and Preprocessing. This study selects the data on the market share of the top 10 consumer electronic product brands in the Chinese market in the third quarter of 2015; the data are retrieved from the report released by HIS Technology [43] (Table 2).

This study starts from a practical point of view. The market share represents the value creation level of the supply chain. Therefore, the real value of the value creation system of the consumer electronics supply chain may be assumed to be approximately equal to its market share and is represented by $\vec{Y}$. Then, $\vec{Y} = (y_1, y_2, \ldots, y_{10})^\tau = (0.164, 0.162, 0.121, 0.102, 0.1, 0.088, 0.067, 0.048, 0.044, 0.027, 0.076)$. The senior management representatives of the aforementioned top 10 brands were invited to rate the interactive value-adding index system of the supply chain using a five-point Likert scale. The scores constitute the system's input vector x_k^v, $v = 1, 2, 3$.

This study implements the z-score standardization of all the independent variables and dependent variables of the samples to achieve a good model fit (the data on each dimension are changed into the data with a mean value of 0 and a variance of 1).

4.2. Optimal Ratio of Training Set over Test Set. The mean value of all the possible combinations $e_i(\gamma, \delta)$ of each n is calculated according to Formulas (8)-(9) to determine the optimal number of training sets n ($n = 2, 3, \ldots, 9$). The $h - E^h$ relationship diagram of the indicator at each level is obtained.

As shown in Figures 2–4, the abscissa represents the number of test sets, whereas the ordinate represents the mean relative error. With the increase in the number of test sets, the number of training sets gradually decreases, and the mean relative error and variance increase. Therefore, the margin of error variation caused by the increase in the total number of samples and test sets is considered when the ratio of the training set over the test set is 7 : 3. The mean relative error and the variance of the model reach the acceptable range, and the validity of model can be tested well when three test

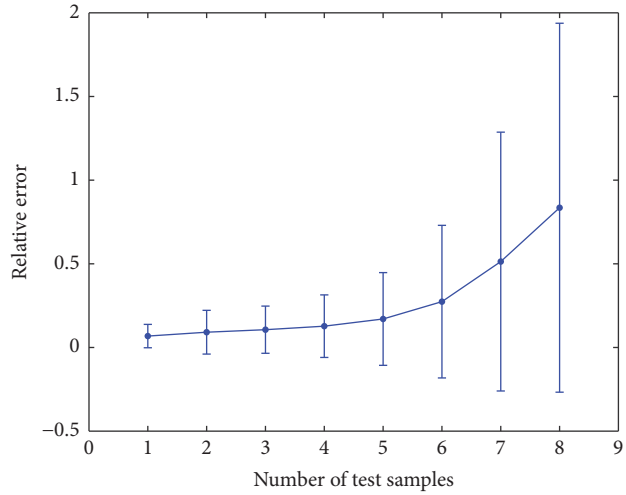


Figure 2: $h - E^h$ diagram of primary indicator.

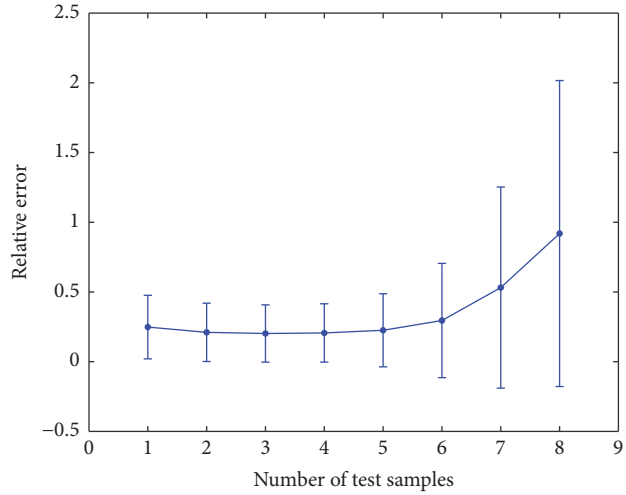


Figure 3: $h - E^h$ diagram of secondary indicator.

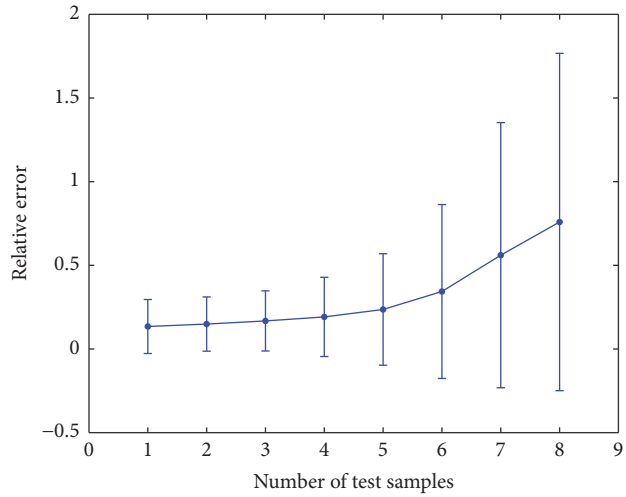


Figure 4: $h - E^h$ diagram of tertiary indicator.

sets are kept. In addition, the selected optimal proportion h^* can prove that the literature [30] sets the proportion of the training set over the test set as 3 : 1; evidence and practical experimental methods are provided.

4.3. Determination of Modeling Parameters. According to the proportion of "training set–test set" determined in Section 4.2, 7 out of the 10 samples are randomly selected as training sets. The remaining three samples are set as test sets, thereby forming C_{10}^3 "training set–test set" combinations. This study determines the optimal penalty coefficient γ and the width value γ using the MATLAB LS-SVM toolbox (version 1.8) program and fivefold cross-validation. However, allocating the LS-SVM toolbox (version 1.8) causes uncertain γ and δ interval. Therefore, this study runs the C_{10}^3 "training set–test set" combinations for 50 times ($m = 50$) to ensure the generality of the selected training set. A set of training sets that achieves the minimum relative error between the predictive value of the test set $y_{\text{smartphone}}^{v}$ and the actual value y_i is selected from the operating results and is set as the training set of the model. The γ, δ, $\vec{a}$, and b of the training set are calculated.

4.4. Model Construction. This study tests the primary indicators according to the aforementioned algorithm and calculates the model parameters; namely, $\gamma^1 = 6688.5436$, $\delta^1 = 9.3824$, $a^1 = (-1.410, 3.713, -0.1716, 0.7713, -3.6011, 1.1173, -0.4185)^\tau$, and $b^1 = 0.0904$, when they are substituted into Formula (6). Model (10) can be obtained as follows:

$$y_{\text{smartphone}}^1\left(x^1\right) = \sum_{k=1}^{n} a_k^1 K\left(x^1, x_k^1\right) + b^1. \tag{10}$$

Similarly, secondary indicators are tested, and model parameter values, namely, $\gamma^2 = 42.2070$, $\delta^2 = 94.4815$, $a^2 = (3.1259, 9.3600, -3.4464, -6.3903, -4.0192, 4.8775, -3.5075)^\tau$, and $b^2 = 0.3462$, are obtained when they are substituted into Formula (6). Model (11) can be obtained as follows:

$$y_{\text{smartphone}}^2\left(x^2\right) = \sum_{k=1}^{n} a_k^2 K\left(x^2, x_k^2\right) + b^2. \tag{11}$$

Similarly, tertiary indicators are tested, and the model parameter values, namely, $\gamma^3 = 395.1913$, $\delta^3 = 315.5547$, $a^3 = (-2.3749, 21.1087, -11.0490, -3.6557, -7.9159, 7.2021, -3.3873)^\tau$, and $b^3 = 0.2903$, are obtained when they are substituted into Formula (6). Model (12) can be obtained as follows:

$$y_{\text{smartphone}}^3\left(x^3\right) = \sum_{k=1}^{n} a_k^3 K\left(x^3, x_k^3\right) + b^3. \tag{12}$$

4.5. Result Analysis. According to Models (10)–(12), the comparison between the output values (predictive value) and the actual values of the indicators at different levels is shown in Figures 5–7.

In Figures 5–7, the ordinate represents the market share, whereas the abscissa represents the sample. In the test set

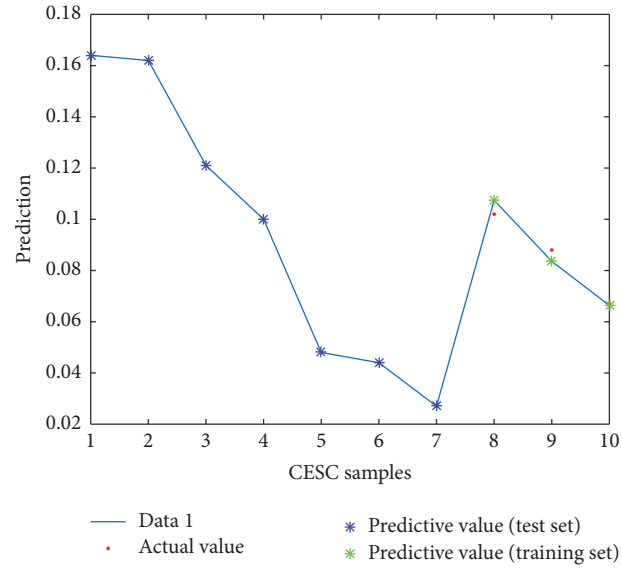


Figure 5: Output of data-driven model and actual value based on primary indicators.

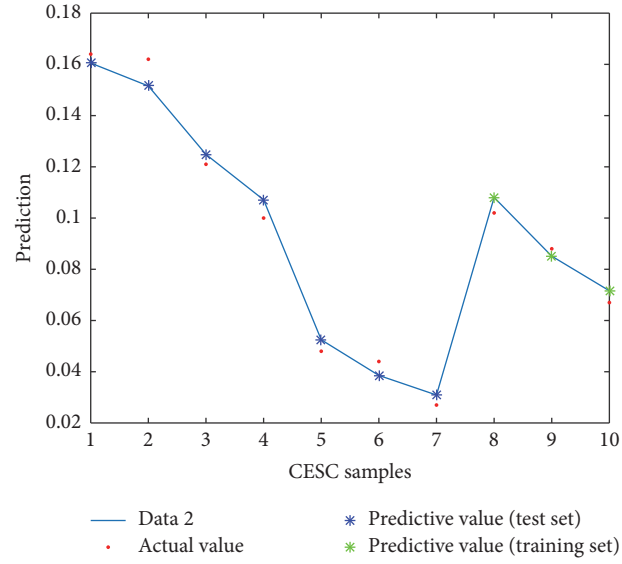


Figure 6: Output of data-driven model and actual value based on secondary indicators.

of the smartphone data-driven model based on primary, secondary, and tertiary indicators, the mean relative errors between the predictive values and the actual values are 3.76%, 5.36%, and 2.42%, respectively. From this perspective, the data-driven models (10)–(12) of the smartphone interactive value-adding system show a high prediction accuracy. With regard to the optimal combination determined in Section 4.3, the test set consists of OPPO, Samsung, and Meizu, whereas the remaining seven brands constitute the training set.

Given that the model is effective, the value-adding influence factor of the consumer electronics supply chain and the index system extracted and constructed by this study, respectively, are scientific and effective. They also

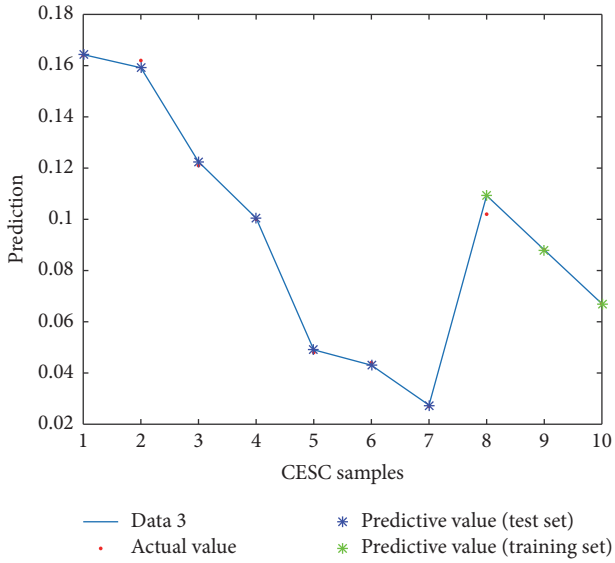


Figure 7: Output of data-driven model and actual value based on tertiary indicators.

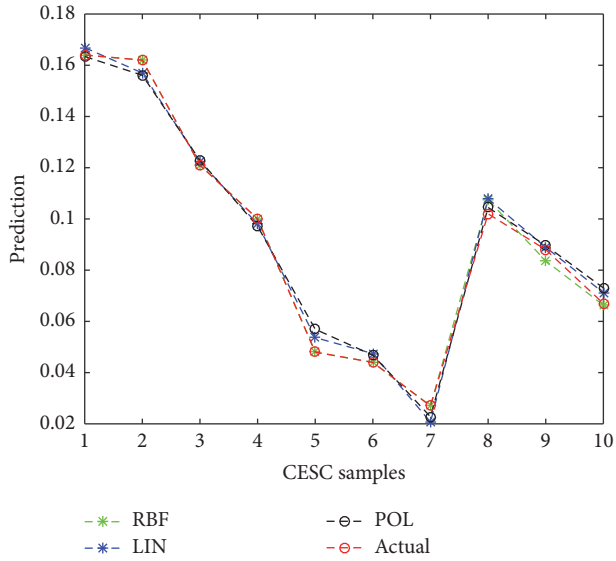


Figure 8: Effect of kernel function selection on the accuracy of the LS-SVM model (primary indicators).

appropriately reflect the value creation system of a specific industrial supply chain. In terms of prediction accuracy, the model based on tertiary indicators should be adopted when analyzing the market share of consumer electronics enterprises because this model achieves the highest precision. In terms of convenience, the model based on primary indicators is the best because this model contains few indicators and it can obtain data easily.

The scientific index system segments the value creation system of consumer electronics enterprises and demonstrates the value-adding mechanism. This system can help consumer electronics companies to determine the value-adding points accurately. Consumer electronics companies can also learn from model enterprises, compare the inputs of their interactive value-adding systems, and perform a scientific analysis on their differences to select investment strategies, improve business outputs, and reduce investment.

5. Comparative Analysis

5.1. Comparison of the Accuracies of Different Models. This study combines the training set and test set determined in Section 4.3 to compare the advantages and disadvantages of LS-SVM and other prediction methods. The prediction models based on SVM, extreme learning machines (ELM), and back propagation (BP) are also established. The predictive values of the test set in different models and their relative errors are shown in Table 3.

Table 3 shows the mean errors and standard deviations of the predictive values of the test sets of the different LS-SVM models based on indicator data at different levels and different kernel functions. They are ranked in ascending order as follows: LS-SVM < SVM < BP < ELM. The data-driven model based on the LS-SVM method shows the lowest mean error and standard deviation and is thus superior to the other models.

5.2. Effect of Kernel Function Selection on LS-SVM Model Accuracy. This section debugs the common RBF kernel, LINEAR, and POLY, which are all constructs of the LS-SVM model, to investigate the effect of kernel function selection on the data-driven model (Figures 8–10 and Table 4). Figures 8–10 show that the predictive values of the LS-SVM model based on the RBF kernel show the best fitting effect with the actual value, the minimum mean error, and the highest performance.

As shown in Table 4, when the models are based on primary indicator data, the average relative errors of the predictive values of the LS-SVM models based on the three kernel functions are ranked in ascending order as follows: RBF < POLY < LINEAR (all sets); RBF < POLY = LINEAR (test set). When the models are based on secondary indicator data, the average relative errors of the predictive values are ranked in ascending order as RBF < LINEAR < POLY (all sets) and RBF < POLY < LINEAR (test set). When the models are based on secondary tertiary indicator data, the average relative errors of the predictive values in all sets and in the test set are ranked in ascending order as follows: RBF < LINEAR < POLY. Therefore, the training set of the LS-SVM model based on the RBF kernel shows a better fitting degree than the other kernel functions. The RBF kernel performs better than the other kernel functions regardless of the prediction accuracy of all the sets or test sets.

6. Conclusions

This study establishes a value creation system of consumer electronics supply chains on the basis of the theory of value cocreation and the interactive value-adding mechanism

Table 3: Comparison of accuracies of four different models.

Actual values of test set		Predictive values (A/B)											
		Primary indicator				Secondary indicator				Tertiary indicator			
		SVM	BP	ELM	LS-SVM	SVM	BP	ELM	LS-SVM	SVM	BP	ELM	LS-SVM
Y_4	0.102	0.101/0.008	0.105/0.029	0.089/0.123	0.107/0.053	0.110/0.078	0.116/0.141	0.119/0.166	0.108/0.059	0.107/0.050	0.102/0.000	0.1038/0.018	0.109/0.071
Y_6	0.088	0.086/0.026	0.108/0.225	0.098/0.111	0.084/0.050	0.089/0.013	0.088/0.003	0.133/0.511	0.085/0.033	0.096/0.085	0.109/0.240	0.133/0.511	0.088/0.001
Y_7	0.067	0.079/0.179	0.081/0.203	0.099/0.480	0.066/0.010	0.075/0.115	0.054/0.198	0.056/0.166	0.072/0.069	0.074/0.109	0.059/0.121	0.058/0.134	0.067/0.001
Mean error		0.071	0.152	0.238	0.038	0.069	0.114	0.281	0.054	0.081	0.12	0.221	0.024
Standard deviation		0.094	0.107	0.210	0.024	0.052	0.100	0.200	0.019	0.030	0.120	0.258	0.041

A/B: A is the predictive value; B is the relative error.

Table 4: Comparison of the accuracies of the LS-SVM models based on different kernel functions.

Samples	Primary indicator			Secondary indicator			Tertiary indicator		
	RBF	LINEAR	POLY	RBF	LINEAR	POLY	RBF	LINEAR	POLY
Training set	0.000	0.018	0.003	0.021	0.021	0.040	0.002	0.003	0.003
	0.000	0.031	0.036	0.044	0.054	0.100	0.017	0.014	0.024
	0.000	0.012	0.015	0.032	0.022	0.045	0.012	0.008	0.015
	0.000	0.020	0.029	0.071	0.096	0.093	0.005	0.005	0.009
	0.000	0.118	0.186	0.093	0.094	0.188	0.022	0.022	0.041
	0.000	0.078	0.064	0.124	0.135	0.102	0.022	0.012	0.022
	0.000	0.234	0.157	0.145	0.050	0.127	0.017	0.006	0.010
Test set	0.053	0.057	0.024	0.059	0.085	0.082	0.071	0.072	0.076
	0.050	0.011	0.018	0.033	0.022	0.016	0.001	0.044	0.044
	0.010	0.063	0.091	0.069	0.190	0.134	0.001	0.062	0.075
Mean error (test set)	0.038	0.044	0.044	0.054	0.099	0.077	0.024	0.059	0.065
Mean error (all sets)	0.011	0.064	0.062	0.071	0.077	0.093	0.017	0.025	0.032

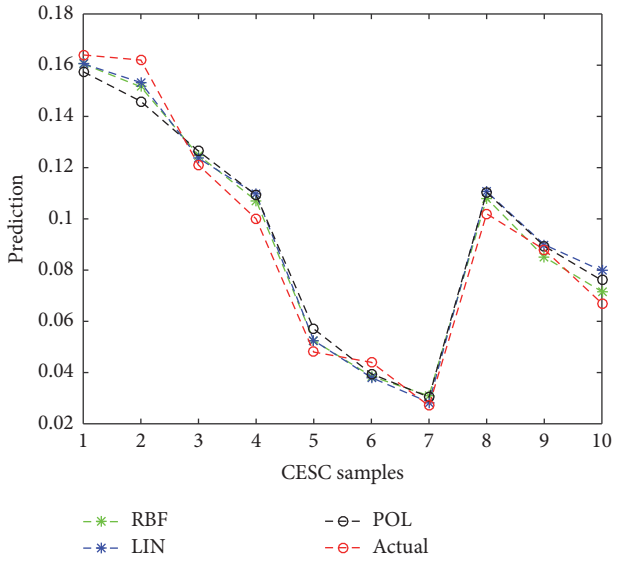


Figure 9: Effect of kernel function selection on the accuracy of the LS-SVM model (secondary indicators).

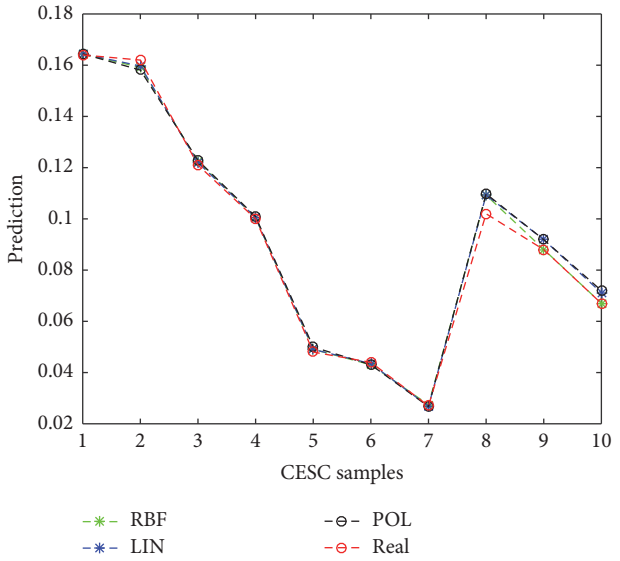


Figure 10: Effect of kernel function selection on the accuracy of the LS-SVM model (tertiary indicators).

between value creators in supply chains. Therefore, this study adopts the LS-SVM method to construct a data-driven model of consumer electronics supply chains and proposes the LS-SVM algorithm, which is suitable for limited samples for consumer electronics supply chains. On the basis of this algorithm and the marginal effect, this study presents a good proportion of the training set over the test set under the condition involving a small number of data samples. Finally, this study performs an empirical analysis of the top 10 smartphone supply chains in the Chinese market. The results indicate that the model achieves a good prediction accuracy. This research can help consumer electronics enterprises to accurately determine value-adding points and conduct market analyses. This study also provides a great practical value to improve the competitiveness of the consumer electronics supply chains of these enterprises.

The subsequent research will perform an in-depth analysis of the value creation mechanism of consumer electronics supply chains. The features will be selected according to the established index system and data-driven model, and the optimal investment strategy for consumer electronics supply chain investment will be determined.

Competing Interests

The authors declare that there are no competing interests regarding the publication of this paper.

Acknowledgments

This study is supported by the National Nature Science Foundation of China (NSFC) (Grant no. 71572096).

References

[1] S. B. Wath, A. N. Vaidya, P. S. Dutt, and T. Chakrabarti, "A roadmap for development of sustainable E-waste management system in India," *Science of the Total Environment*, vol. 409, no. 1, pp. 19–32, 2010.

[2] China Consumer Electronics Report 2015, *China Consumer Electronics Report*, Business Source Complete, EBSCO Information Services, Ipswich, Mass, USA, 2016.

[3] X. P. Shen, "Value system and competitive advantage of supply chain," *Logistics Engineering and Management*, no. 8, pp. 137–139, 2010.

[4] J. C. Anderson and J. A. Narus, *Business Market Management: Understanding, Creating, and Delivering Value*, Pearson Prentice Hall, Upper Saddle River, NJ, USA, 2nd edition, 2004.

[5] N. Mizik and R. Jacobson, "Trading off between value creation and value appropriation: the financial implications of shifts in strategic emphasis," *Journal of Marketing*, vol. 67, no. 1, pp. 63–76, 2003.

[6] S. L. Vargo and R. F. Lusch, "Evolving to a new dominant logic for marketing," *Journal of Marketing*, vol. 68, no. 1, pp. 1–17, 2004.

[7] C. K. Prahalad and V. Ramaswamy, *The Future of Competition: Co-Creating Unique Value with Customers*, Harvard Business School Press, Boston, Mass, USA, 2004.

[8] A. R. Sakamoto, C. B. Villar, and M. E. Martins, *Open Innovation and Collaborative Network in Supply Chain: The Case of Open IPTV Forum*, IGI Global, Hershey, Pa, USA, 2012.

[9] D. Aksen, N. Aras, and A. G. Karaarslan, "Design and analysis of government subsidized collection systems for incentive-dependent returns," *International Journal of Production Economics*, vol. 119, no. 2, pp. 308–327, 2009.

[10] H. H. Liu, M. Lei, H. H. Deng, G. K. Keong Leong, and T. Huang, "A dual channel, quality-based price competition model for the WEEE recycling market with government subsidy," *Omega*, vol. 59, pp. 290–302, 2016.

[11] P.-H. Qiao, X.-F. Ju, and H.-G. Fung, "Industry association networks, innovations, and firm performance in Chinese small and medium-sized enterprises," *China Economic Review*, vol. 29, no. 6, pp. 213–228, 2014.

[12] J. H. Xiao, K. Xie, Y. Wu, and X. H. Liao, "The supply chain transformation from being partner from being partner-oriented to being customer oriented: a double-case study on the supply chains in ecommerce enterprises," *Management World*, no. 4, pp. 137–154, 2015.

[13] Q. C. Meng, J. H. Dong, and C. C. Li, "Study on supply chain value maximization based on new prosumer," *Chinese Journal of Management Science*, vol. 20, no. 6, pp. 102–109, 2012.

[14] Q. C. Meng and H. H. Li, "A study on supply chain value maximization considering the competition between supply chains based on new prosumer," *Chinese Journal of Management Science*, vol. 23, no. 3, pp. 168–176, 2015.

[15] S. Huang, S. Zeng, Y. Fan, and G. Q. Huang, "Optimal service selection and composition for service-oriented manufacturing network," *International Journal of Computer Integrated Manufacturing*, vol. 24, no. 5, pp. 416–430, 2011.

[16] S. Cavalieri and G. Pezzotta, "Product-service systems engineering: state of the art and research challenges," *Computers in Industry*, vol. 63, no. 4, pp. 278–288, 2012.

[17] L. Zhen, "An analytical study on service-oriented manufacturing strategies," *International Journal of Production Economics*, vol. 139, no. 1, pp. 220–228, 2012.

[18] Y. Y. Niu, "Research on the influence of cluster supply chain network under the paradigm of open innovation," *Value Engineering*, no. 15, pp. 64–67, 2015.

[19] HignBeam Business, *IBM Opens Supply Chain Innovation Center in China*, HignBeam Business, Industry Week/IW, Chicago, Ill, USA, 2008.

[20] R. F. Lusch, S. L. Vargo, and M. O'Brien, "Competing through service: insights from service-dominant logic," *Journal of Retailing*, vol. 83, no. 1, pp. 5–18, 2007.

[21] M. S. Yadav and P. R. Varadarajan, "Understanding product migration to the electronic marketplace: a conceptual framework," *Journal of Retailing*, vol. 81, no. 2, pp. 125–140, 2005.

[22] C. Grönroos, "Service logic revisited: who creates value? And who co-creates?" *European Business Review*, vol. 20, no. 4, pp. 298–314, 2008.

[23] D. Ballantyne and R. J. Varey, "Creating value-in-use through marketing interaction: the exchange logic of relating, communicating and knowing," *Marketing Theory*, vol. 6, no. 3, pp. 335–348, 2006.

[24] G. Ramani and V. Kumar, "Interaction orientation and firm performance," *Journal of Marketing*, vol. 72, no. 1, pp. 27–45, 2008.

[25] R. T. Rust, C. Moorman, and G. Bhalla, "Rethinking marketing," *Harvard Business Review*, vol. 88, no. 4, pp. 94–101, 2010.

[26] J. C. Xu, Q. W. Ren, and Z. Z. Shen, "Prediction of the strength of concrete radiation shielding based on LS-SVM," *Annals of Nuclear Energy*, vol. 85, pp. 296–300, 2015.

[27] H. M. Khalil and M. El-Bardini, "Implementation of speed controller for rotary hydraulic motor based on LS-SVM," *Expert Systems with Applications*, vol. 38, no. 11, pp. 14249–14256, 2011.

[28] T.-C. Lin, "A multi-class dempster classifier with support vector machine for image enhancement," *International Journal of Innovative Computing, Information and Control*, vol. 11, no. 5, pp. 1639–1654, 2015.

[29] R. Langone, C. Alzate, B. D. Ketelaere, J. Vlasselaer, W. Meert, and J. A. K. Suykens, "LS-SVM based spectral clustering and regression for predicting maintenance of industrial machines," *Engineering Applications of Artificial Intelligence*, vol. 37, pp. 268–278, 2015.

[30] N. Zhang and D. Shetty, "An effective LS-SVM-based approach for surface roughness prediction in machined surfaces," *Neurocomputing*, vol. 198, pp. 35–39, 2016.

[31] X. H. Yuan, C. Chen, Y. B. Yuan, Y. H. Huang, and Q. X. Tan, "Short-term wind power prediction based on LSSVM-GSA model," *Energy Conversion and Management*, vol. 101, pp. 393–401, 2015.

[32] H. H. Yu, Y. Y. Chen, S. G. Hassan, and D. L. Li, "Prediction of the temperature in a Chinese solar greenhouse based on LSSVM optimized by improved PSO," *Computers and Electronics in Agriculture*, vol. 122, pp. 94–102, 2016.

[33] J. Wu, X. Cui, T. Nishi, and J. Li, "Game theoretic approach for coordination of telecommunication supply chain under network externality," *International Journal of Innovative Computing*, vol. 12, no. 3, pp. 915–928, 2016.

[34] H. C. Jung, J. S. Kim, and H. Heo, "Prediction of building energy consumption using an improved real coded genetic algorithm based least squares support vector machine approach," *Energy and Buildings*, vol. 90, pp. 76–84, 2015.

[35] D. A. Silva, J. P. Silva, and A. R. Rocha Neto, "Novel approaches using evolutionary computation for sparse least square support vector machines," *Neurocomputing*, vol. 168, pp. 908–916, 2015.

[36] C. Durugbo and K. Pawar, "A unified model of the co-creation process," *Expert Systems with Applications*, vol. 41, no. 9, pp. 4373–4387, 2014.

[37] A. F. Payne, K. Storbacka, and P. Frow, "Managing the co-creation of value," *Journal of the Academy of Marketing Science*, vol. 36, no. 1, pp. 83–96, 2008.

[38] H. Saarijärvi, P. K. Kannan, and H. Kuusela, "Value co-creation: theoretical approaches and practical implications," *European Business Review*, vol. 25, no. 1, pp. 6–19, 2013.

[39] N. Pinho, G. Beirão, L. Patrício, and R. P. Fisk, "Understanding value co-creation in complex services with many actors," *Journal of Service Management*, vol. 25, no. 4, pp. 470–493, 2014.

[40] X. L. Wan and Q. C. Meng, "Construction and capacity determination of supply chain interaction value-added index system from the perspective of co-creation," Working Paper, 2016.

[41] C. Cortes and V. Vapnik, "Support-vector networks," *Machine Learning*, vol. 20, no. 3, pp. 273–297, 1995.

[42] X. Juncai, R. Qingwen, and S. Zhenzhong, "Prediction of the strength of concrete radiation shielding based on LS-SVM," *Annals of Nuclear Energy*, vol. 85, pp. 296–300, 2015.

[43] HIS Technology, *China's Smartphone Market Share Report (2015Q3)*, HIS Technology, 2015.

Hindawi Publishing Corporation
Scientific Programming
Volume 2016, Article ID 8264879, 9 pages
http://dx.doi.org/10.1155/2016/8264879

Research Article

A Recovery Model for Production Scheduling: Combination of Disruption Management and Internet of Things

Yang Jiang,[1] Qiulei Ding,[2] and Xuhui Wang[2]

[1]*School of Light Industry & Chemical Engineering, Dalian Polytechnic University, Dalian 116034, China*
[2]*School of Business Administration, Dongbei University of Finance and Economics, Dalian 116025, China*

Correspondence should be addressed to Qiulei Ding; dingqiulei2016@163.com

Received 13 June 2016; Accepted 25 August 2016

Academic Editor: Chengyan Yue

It is difficult to generate the new schedule effectively for minimizing the negative impact when an unanticipated disruption occurs after a subset of tasks has been finished in production scheduling. In such cases, continuing with the original schedule may not be optimal or feasible. Based on disruption management and Internet of things (IoT), this study designs a real-time status analyzer to identify the disruption and propose a recovery model to deal with the disruption. The computational result proves that our algorithm is competitive with the existing heuristics. Furthermore, due to the tradeoff between all participators (mainly including customers, managers of production enterprise, and workers) involved in production scheduling, our model is more effective than the total rescheduling and right-shift rescheduling.

1. Introduction

Providing customers with satisfactory service is the central concern for production enterprises. However, given the complexity of processing environments, random or unanticipated events, including anomaly status of processing environment and machine breakdown, constantly occur in production scheduling. As a result, the original schedule is likely to be suboptimal and may not even be feasible. In this study, we use the term "disruption" to describe the reasons that trigger a rescheduling process.

Research on rescheduling is extensive [1, 2]. We differentiate our model from the existing findings in which the new schedule will be evaluated. The objective of most rescheduling research is to find an optimal schedule in the new environment with respect to the original objective function. In our model, we consider both the original objective function and the deviation from the original schedule. The reason for considering the deviation is that once the original schedule is generated, many preparations have already been made, such as allocating resources, distributing raw materials, and fixing customer delivery dates. Hence, if any change to the original schedule has a negative impact on these preparations, then the new schedule has to minimize the deviation [3].

In this paper, we combine disruption management and IoT to deal with the disruption. IoT is to realize the information sharing of global things from any place, which is able to monitor the process of production scheduling and identify the disruption quickly. Furthermore, disruption management aims to dynamically revise the original schedule and obtain a new schedule that reflects the constraints and objectives of the evolved environment [4].

The rest of the paper is organized as follows. Section 2 provides a brief review of the related work. Section 3 designs a real-time status analyzer based on IoT to identify the disruption. Section 4 constructs a recovery model combining quantitative model and prospect theory to measure the deviation, and Section 5 proposes an improved ant colony optimization (IACO) to obtain the new schedule quickly. Section 6 shows the validity of the demonstrated model and IACO. Finally, Section 7 provides conclusions and directions for future research.

2. Related Work

Determining how to handle disruption is a significant issue in the decision-making process. Disruption management is an example of rescheduling in real time, which is necessary in

some cases because disruption is unanticipated and cannot be embedded in making the original schedule in advance. Through disruption management, a new schedule that minimizes the negative impact can be obtained.

Disruption management was first applied to the airline industry, in which flight disruptions often entail huge cost loss [5, 6]. The successful implementation of the idea led to a growing interest in applying disruption management to other fields. In the study of production scheduling, Lee and Yu [7] studied the parallel-machine scheduling problem under the disruptive environment by minimizing the sum of weighted completion times and provided a pseudopolynomial time algorithm to solve the problem. Tang and Zhang [8] investigated the disruption caused by machine breakdown and proposed the Lagrangian relaxation approach to deal with the disruption. Wang et al. [9] discussed parallel-machine scheduling problems with a deteriorating maintenance activity and provided a polynomial time algorithm to solve the total completion time minimization problem. Liu et al. [10] studied the disrupted single-machine scheduling problem. The deviation was measured by the completion times of jobs and quantum-inspired hybrid heuristics were presented. Khedlekar et al. [11] studied the flexible managerial decision policy for disruption in production system. They solved the disruption problem analytically to determine the production period before and after disruptions. Paul et al. [12] developed the recovery model and dynamic solution approach to deal with disruptions for production-inventory system. Sarker et al. [13] focused on production scheduling under uncertain disruption and adopted an improved memetic algorithm to solve the problem.

Another stream of literature relevant to our research is the recently emerging subject of behavioral operation management [14]. The operation contexts primarily investigated were supply chain management, newsvendor problem, forecasting, and decision-making, which is concerned with the theories of individual decision-making biases, social preferences, and cultural elements [15]. Su [16] proposed a decision framework of bounded rationality applied to the classic newsvendor model. Gino and Pisano [17] explored the theoretical and practical implications of incorporating behavioral and cognitive factors into models of operations. Chen et al. [18] presented a model of bounded rationality based on the quantal response equilibrium, in which players were not perfect optimizers and faced uncertainty in the actions of their opponents. Katsikopoulos and Gigerenzer [19] briefly discussed some research on the heuristics in behavioral operation management and showed how the study of quantitative models of heuristics can complement it. Chen and Zhao [20] developed a behavioral model with different mental weights on the underage and overage costs to characterize the perception bias of a retailer on a critical fractile.

In summary, identifying the disruption was likely to be neglected in previous literature because the researchers focused on obtaining the revising schedule quickly. In addition, the existing researches assumed that humans in production scheduling were perfectly rational and emphasized the optimization of material and financial resources. They disregarded these facts that people facing a new situation often have different perceptions and the obtained solution may not be optimal or feasible. Accordingly, how to identify the disruption and measure the deviation still needs further study.

3. A Real-Time Status Analyzer Based on IoT

In production scheduling, the ambient condition subtly changes all the time. However, considering and adjusting it as a whole are worthy only when the current schedule does not work properly. Thus, a real-time status analyzer based on IoT is designed and developed to identify the disruption before adopting the recovery model.

3.1. System Architecture. The status data are sent to computer terminals (CTs) of the Internet via wireless sensor networks (WSNs) and satellite. The data processing center analyzes the data, and it is alerted when a disruption occurs. The system can achieve the following three functions.

(1) Data Acquisition. Jobs in the shop are frequently sensitive to temperature, humidity, and pressure. Thus, the status of processing environment needs to be monitored in real time. All data of the basic properties and environmental status of jobs are also sent to the radio frequency identification (RFID) reader via the WSNs. Finally, the aforementioned information is summarized and sent to the corresponding CTs.

(2) Exception Analysis. Based on the initial schedule and according to different characteristics of jobs, the corresponding thresholds for equipment status, shop temperature, humidity, and pressure are preset. Thus, the massive time-varying data can be monitored and analyzed in real time to determine whether any exception occurs.

(3) Real-Time Alert. When an exception occurs as revealed by the monitored data, the disruption occurs and the system gives out alerts immediately, thereby revising the original schedule and obtaining a new schedule that minimizes the negative impact.

3.2. Redundant Data Processing. RFID is prone to be absorbed and disrupted by the ambient condition, thereby resulting in any omission or deficiency in the reading process of the reader. This study aims to avoid such problems by proposing increasing the number of readers or the read-write frequency, thereby ensuring data integrity. However, this proposal can inevitably produce a significant amount of redundant data, increasing the burden to the RFID data transmission and storage. Thus, the system efficiency is greatly reduced. Therefore, an effective way to solve problems of redundant data processing must be determined to screen all data and to reduce the burden on the system.

Basic sorted-neighborhood method (SNM) is an efficient algorithm to cope with redundant data. However, it is difficult to control the size of smoothing window. This study presents a modified SNM algorithm with specific procedures as follows.

Step 1. The data are sorted according to the timestamp.

Step 2. In the data comparison process, a smoothing window featured with adaptive adjustment is used, and the window is resized as desired.

Step 3. The first data entry in the current window is compared with the next, and whether it is repeated is determined. If it is repeated, Step 4 follows; otherwise, Step 5 is next.

Step 4. When the repeated entry is found, the entry with the oldest timestamp is deleted and that with the latest timestamp is retained.

Step 5. Whether the first entry in the window is the last is determined. If it is the last, Step 6 follows; otherwise, Step 3 is next.

Step 6. Whether any unprocessed data exists is determined. If so, Step 1 follows; otherwise, the entire workflow ends.

4. Recovery Model for Production Scheduling

Considering its wide application in production and manufacturing industries, the Job-Shop Scheduling Problem (JSP) is chosen as an example to construct the recovery model.

4.1. Construction of the Model for Original Schedule

4.1.1. Problem Definition. The problem is confined to the following conditions. Given n jobs, each job has m operations and must be processed on m machines. The objective is to find a schedule of minimal time to complete all jobs, where

(1) the machine can only process one job at a time,

(2) every job is available for processing at time 0,

(3) once processing is initiated, the operation must be completed on the machine without interruption.

4.1.2. Notations

n: the number of jobs.

m: the number of machines.

V: the set of jobs, $V = \{v_1, v_2, \ldots, v_n\}$.

U: the set of machines, $U = \{u_1, u_2, \ldots, u_m\}$.

c_{ik}: the completion time of job i in machine k.

p_{ik}: the processing time of job i in machine k.

d_i: the due date of v_i.

M: a large positive number.

$$a_{ihk} = \begin{cases} 1, & v_i \text{ is processed in } u_h \text{ before } u_k \\ 0, & \text{otherwise,} \end{cases}$$

$$x_{ijk} = \begin{cases} 1, & v_i \text{ is processed before } v_j \text{ in } u_k \\ 0, & \text{otherwise.} \end{cases} \quad (1)$$

4.1.3. Mathematical Model. The JSP model is constructed as follows:

$$\min \quad \max_{1\le k\le m} \left\{ \max_{1\le i\le n} c_{ik} \right\} \quad (2)$$

$$c_{jk} - p_{jk} + M\left(1 - x_{ijk}\right) \ge c_{ik} \qquad i, j = 1, 2, \ldots, n;\ k = 1, 2, \ldots, m \quad (3)$$

$$c_{ik} - p_{ik} + M\left(1 - a_{ihk}\right) \ge c_{ih} \qquad i = 1, 2, \ldots, n;\ h, k = 1, 2, \ldots, m \quad (4)$$

$$d_i \le \max_{1\le k\le m} c_{ik} \quad i = 1, 2, \ldots, n. \quad (5)$$

In this model, the objective function (2) aims to minimize the makespan, which is the minimal time to complete all jobs. Formulas (3) and (4) are the technological and processing constraints. Formula (5) ensures that all jobs are completed before their due date.

4.2. Construction of the Recovery Model. When the optimal schedule is executed, the disruption occurs and is identified by the real-time status analyzer. Consequently, we need to construct the recovery model and obtain a new schedule.

Machine breakdown is the typical disruption in production scheduling. In this paper, the recovery model is illustrated by the example of machine breakdown.

4.2.1. Problem Hypothesis

(1) The original schedule is known.

(2) The time when disruption ends is regarded as 0.

(3) The processing job when disruption occurs must be reprocessed after the disruption ends. In other words, jobs are uninterruptible.

4.2.2. Notations

w: the number of unprocessed jobs.

V': a subset of unprocessed jobs, $V' = \{v_1, v_2, \ldots, v_w\}$.

A_{ij}^1: the set of operations processed before operation j in the original schedule for machine i.

A_{ij}^2: the set of operations processed after operation j in the recovery schedule for machine i.

A_{ij}: the intersection of A_{ij}^1 and A_{ij}^2, $A_{ij} = A_{ij}^1 \cap A_{ij}^2$.

$\overline{\overline{A_{ij}}}$: the cardinality of subset A_{ij}.

c_i^0: completion time of v_i in the original schedule.

c_i: completion time of v_i in the recovery schedule.

f^0: the makespan in the original schedule.

f: the makespan in the recovery schedule.

g: the sequence deviation between the original schedule and the recovery schedule.

μ_1^i: unsatisfaction degree of customer i for completion time deviation.

μ_2: unsatisfaction degree of managers for operational cost deviation.

μ_3: unsatisfaction degree of workers for sequence deviation.

α_1, β_1: parameters related to gains and losses for customers.

α_2, β_2: parameters related to gains and losses for managers.

α_3, β_3: parameters related to gains and losses for workers.

λ_1, λ_2, λ_3: loss aversion coefficient for customers, managers, and workers, respectively.

The other notations are the same as the ones in previous sections.

4.2.3. Impact of Disruption. The general process of one schedule is as follows: (i) managers of production enterprises generate an optimal schedule according to the requirements of customers, (ii) workers execute the schedule to process jobs, and then (iii) customers receive their products within the required time. The impact of disruption on the above three participators (i.e., managers, workers, and customers) is analyzed below.

(1) Managers. After disruption occurs, the processing sequence of unprocessed jobs will change, leading to the increase of makespan. The makespan is relevant to the operational cost, which is the central concern of managers. Accordingly, minimizing the makespan of unprocessed jobs is the key objective of managers when generating the recovery schedule.

(2) Workers. The processing sequence of unprocessed jobs will be changed after disruption occurs. Considering that many preparations have been made in advance based on the original schedule, workers have to reallocate raw materials and tools, which will lead to the increase of extra workload. Therefore, the recovery schedule should minimize the sequence deviation to reduce the negative impact on workers.

(3) Customers. Once the environment is interrupted by disruption, a chain reaction will be generated to affect the completion time of unprocessed jobs, and a number of customers will not be served within the required time. Thus, the recovery schedule should reduce the deviation of completion time to improve the satisfaction and loyalty of customers.

4.2.4. Function of Measuring Deviation. Prospect theory [21] is one of the most important decision-making theories. It can perfectly describe the decision-making based on bounded rationality under the uncertainty condition. Hence, on the basis of prospect theory, this paper measures the deviation of participators [22].

(1) Measuring the Deviation of Customers. The value function of customer i can be described as follows:

$$V_1^i(x) = \begin{cases} x^{\alpha_1}, & x \geq 0 \\ -\lambda_1 (-x)^{\beta_1}, & x < 0 \end{cases} \quad i = 1, 2, \ldots, w, \tag{6}$$

where the reference point is c_i^0 ($c_i^0 = \max_{1 \leq k \leq m} c_{ik}$). $c_i > c_i^0$ means losses ($x < 0$) for customer i; otherwise, it means gains ($x \geq 0$) for customer i.

The unsatisfied membership function of customer i for completion time deviation can be given as follows:

$$\mu_1^i(c_i) = \begin{cases} 1, & c_i \geq R_{1i} \\ \lambda_1 \left(c_i - c_i^0\right)^{\beta_1}, & c_i^0 \leq c_i < R_{1i} \\ 0, & 0 \leq c_i < c_i^0 \end{cases} \quad i = 1, 2, \ldots, w, \tag{7}$$

where $R_{1i} = [c_i^0 + (1/\lambda_1)^{1/\beta_1}]$ ($i = 1, 2, \ldots, w$).

(2) Measuring the Deviation of Managers. The value function of managers can be constructed as follows:

$$V_2(x) = \begin{cases} x^{\alpha_2}, & x \geq 0 \\ -\lambda_2 (-x)^{\beta_2}, & x < 0, \end{cases} \tag{8}$$

where the reference point is f^0. $f > f^0$ means losses ($x < 0$) for managers; otherwise, it means gains ($x \geq 0$) for managers.

The unsatisfied membership function of managers for operational cost deviation can be given as follows:

$$\mu_2(f) = \begin{cases} 1, & f \geq R_2 \\ \lambda_2 \left(f - f^0\right)^{\beta_2}, & f^0 \leq f < R_2 \\ 0, & 0 \leq f < f^0, \end{cases} \tag{9}$$

where $R_2 = f^0 + (1/\lambda_2)^{1/\beta_2}$.

(3) Measuring the Deviation of Workers. The value function of workers can be constructed as follows:

$$V_3(x) = -\lambda_3 (-x)^{\beta_3}, \quad x < 0, \tag{10}$$

where the reference point is 0 because the sequence deviation does not exist in the original schedule. In recovery schedule, g is equal to $\sum_{i=1}^{m} \sum_{j=1}^{w} \overline{\overline{A_{ij}}}$. $g > 0$ means losses ($x < 0$) for workers.

The unsatisfied membership function of workers for sequence deviation can be given as follows:

$$\mu_3(g) = \begin{cases} 1, & g \geq R_3 \\ \lambda_3 g^{\beta_3}, & 0 \leq g < R_3, \end{cases} \tag{11}$$

where $R_3 = (1/\lambda_3)^{1/\beta_3}$.

4.2.5. Recovery Model. On the basis of the above functions of measuring deviation, the objective function of the recovery model is constructed through lexicographic multiple goal programming as follows:

$$\begin{aligned} \min \quad & \text{Lex} = P_1 : \sum_{i=1}^{w} \mu_1^i (c_i) \\ & P_2 : \mu_2 (f) \\ & P_3 : \mu_3 (g). \end{aligned} \tag{12}$$

Formula (12) aims to minimize the sum of the unsatisfaction degrees of customers, managers, and workers. P_1, P_2, and P_3 represent the different priorities, which can be flexibly adjusted in the practical application.

5. IACO for the Recovery Model

The proposed model is NP-hard, which has an optimal solution that is difficult to obtain. Ant colony optimization (ACO) is an evolutionary computation technique developed by Colorni et al. [23]. Compared with other heuristics, ACO has the characteristics of positive feedback, distributed computation, and constructive greedy heuristic. In particular, the ACO has been shown to be an efficient algorithm for solving NP-hard problems [24–26]. However, the algorithm still has the weaknesses of local optimization and low search speed. In this study, we illustrate IACO to solve the recovery model.

5.1. Introduction of Adjusting the Pheromone Trail

(1) In ACO, pheromone trails left by ants do not always show the evolutionary direction, and the pheromone deviating from the optimal solution has the probability of increasing, which leads to premature convergence. This study proposes an approach to enhance the global search capability of ACO by adjusting the pheromone trail adaptively with the evolutionary process.

(2) The updated pheromone trail may reach the maximum value $\tau_{\max}$ or minimum value $\tau_{\min}$ after every search process is completed. $\tau_{\max}$ leads to premature convergence, and $\tau_{\min}$ increases the computational time. Thus, IACO limits the pheromone trail τ_{ij} in the interval $(\tau_{\min}, \tau_{\max})$. After the pheromone trails are updated, τ_{ij} is replaced by $\tau_{\max}$ when $\tau_{ij} > \tau_{\max}$ or by $(\tau_{\min} + \tau_{\max})/2$ when $\tau_{ij} < \tau_{\min}$.

(3) ACO has difficulty obtaining the optimal solution because the trail persistence ρ is fixed. The smaller the value of ρ is, the better the global search capability will be. However, the computational time will increase gradually. On the contrary, increasing ρ can improve the computational time, but the algorithm can easily be trapped in the local optimal solution. Therefore, IACO adopts a dynamic ρ rather than a fixed value.

5.2. Approach of the Crossover and Mutation. The crossover and mutation operations can increase the variance of the population and search the solution space completely in genetic algorithm (GA). Thus, these two operations are adopted in IACO to avoid premature convergence.

(1) Crossover Operation. In the paper, a schedule s is represented by the following encoding structure:

$$s = \left\{ \underbrace{\text{solu}_1, \ldots, \text{solu}_m}_{s_1}, \underbrace{\text{solu}_{m+1}, \ldots, \text{solu}_{2m}}_{s_2}, \ldots, \underbrace{\text{solu}_{(w-1)m+1}, \ldots, \text{solu}_{wm}}_{s_w} \right\}, \tag{13}$$

where s_i represents the sequence of operations of job $i = 1, 2, \ldots, w$. Each solu in s_i is an unrepeated integer number in the interval $[1, w]$.

When the search becomes trapped in the local optimal solution, crossover operation is conducted as follows:

(i) Assume that encoding of schedule a is $\{a_1, a_2, \ldots, a_w\}$ and encoding of schedule b is $\{b_1, b_2, \ldots, b_w\}$.

(ii) Randomly generate an integer number j in the interval $[1, w]$. Then, exchange a_j and b_j, and generate two new encodings, that is, $\{a_1, a_2, \ldots, b_j, \ldots, a_w\}$ and $\{b_1, b_2, \ldots, a_j, \ldots, b_w\}$.

(iii) Update the encoding of optimal solution.

(2) Mutation Operation. Mutation plays a significant part in improving the diversity of the population. Similar to GA, IACO is designed to avoid premature convergence and obtain a better solution by significantly decreasing pheromone trails in any path of the local optimization routing. These paths will be selected by small random probabilities because the distribution of the pheromone trails will be destroyed by too many mutation operations.

5.3. Combination with Other Heuristics. ACO has the characteristic of strong coupling, which makes it easy to combine with other heuristics. Exchange mechanism is originally applied in the vehicle routing problem with time windows. It can obtain a new solution from the current solution by exchanging the nodes. We demonstrate a neighborhood exchange mechanism to significantly improve the convergence speed of ACO.

(1) Operators of the Neighborhood Exchange Mechanism. Two neighborhood operators (among those listed below) are included in this study.

(i) *Random Swaps of Subsequences.* This operator randomly selects i and j with $i \neq j$ and swaps s_i and s_j in the solution sequence. Figure 1 shows an example of this operator.

(ii) *Random Insertions of Subsequences.* This operator randomly selects i and j with $i \neq j$ and puts s_j in front of s_i. An example is shown in Figure 2.

Table 1: Computational results of solving different instances.

Algorithm	TA05	TA11	TA20	TA37	TA43	DMU12	DMU26	DMU45	DMU61	DMU77
i-TSAB	1224	1361	1351	1778	1859	3519	4679	3321	5294	6930
GES	1224	1357	1348	1779	1870	3518	4667	3273	5293	7006
AlgFix	1224	1358	1348	1784	1869	3522	4688	3273	5310	6949
IACO	1224	1357	1348	1779	1858	3520	4665	3275	5277	6908

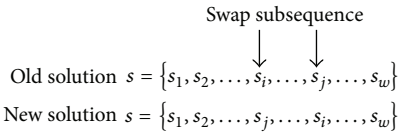


Figure 1: Random swaps of subsequences.

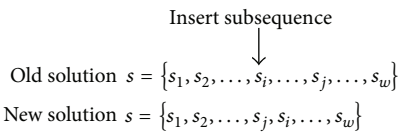


Figure 2: Random insertions of subsequences.

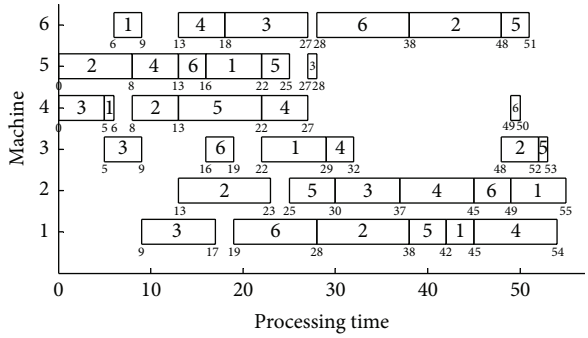


Figure 3: Original schedule.

(2) Rules of the Neighborhood Exchange Mechanism

(i) To avoid generating too many unavailable solutions, the operations in the same machine can be exchanged.

(ii) To avoid destroying the distribution of the previous pheromone trails, one neighborhood operator is chosen randomly from the above operators and applied once to the current solution.

6. Numerical Examples

In this section, we illustrate computational experiments to validate the effectiveness of IACO and recovery model presented in the previous sections.

6.1. Computational Experiment for IACO

(1) Computational Results. The IACO is tested using the classical sets of JSP, which are TA [27] and DMU [28]. For each instance, the proposed algorithm is independently executed 10 times to compute the average value. We then conduct a performance comparison between IACO and other heuristics, including i-TSAB [29], GES [30], and AlgFix [31]. The results are illustrated in Table 1.

(2) Comparison among Different Heuristics. The comparison obtained from the above results can be summarized as follows:

(i) Compared with i-TSAB, IACO obtains better or close solutions in 8 out of 10 problems (80%).

(ii) Compared with GES, IACO obtains better or close solutions in 8 out of 10 problems (80%).

(iii) Compared with AlgFix, IACO obtains better or close solutions in 9 out of 10 problems (90%).

The comparison indicates that the IACO is competitive with the existing heuristics. Moreover, it can improve the best solutions known for a number of examples, especially for large size problems, such as DMU61 and DMU77. It is noted that those parameters affect the overall performance of IACO, so the solution obtained by our algorithm would be improved further if every parameter is set more reasonably.

6.2. Computational Experiment for the Recovery Model

(1) Case Description. In the given job-shop, $n = 6$ and $m = 6$. The parameters of each job are shown in Table 2. Figure 3 illustrates the original schedule where the makespan is 55 h.

(2) Computational Results. After 10 h, the disruption suddenly occurs and the duration time is 3 h. The remaining jobs, including the processing job when disruption occurs, must be reprocessed after the disruption ends.

Following Tversky and Kahneman [32], we set $\beta = 0.88$ and $\lambda = 2.25$. By adopting our model, total rescheduling, and right-shift rescheduling [33], we obtain the results shown in Figures 4, 5, and 6, respectively. Table 3 shows the results from different approaches.

(3) Comparison among Different Approaches. The results obtained from Table 3 can be summarized as follows:

(i) From the aspect of the deviation of customers, the result of our model is much better than the results of the other two approaches. In other words, our model plays an obvious role in reducing the unsatisfaction degree of customers.

Table 2: Parameters of each job.

	Job 1	Job 2	Job 3	Job 4	Job 5	Job 6
Operation 1	M4:T1	M5:T8	M4:T5	M5:T5	M4:T9	M5:T3
Operation 2	M6:T3	M4:T5	M3:T4	M6:T5	M5:T3	M3:T3
Operation 3	M5:T6	M2:T10	M1:T8	M4:T5	M2:T5	M1:T9
Operation 4	M3:T7	M1:T10	M6:T9	M3:T3	M1:T4	M6:T10
Operation 5	M1:T3	M6:T10	M5:T1	M2:T8	M6:T3	M2:T4
Operation 6	M2:T6	M3:T4	M2:T7	M1:T9	M3:T1	M4:T1
Due date	60 h	56 h	60 h	65 h	55 h	50 h

Note: M*i*:T*j* means that the process time of the job's operation is *j* hours in machine *i*.

Table 3: Results from different approaches.

	Customers' deviation	Managers' deviation	Workers' deviation
Our model	2	1	5
Total rescheduling	3	1	4
Right-shift rescheduling	6	1	0

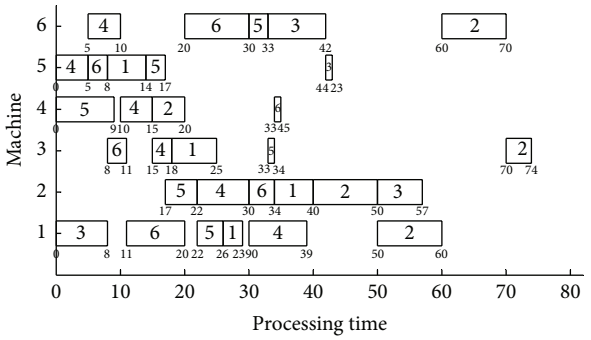


Figure 4: Recovery schedule of our model.

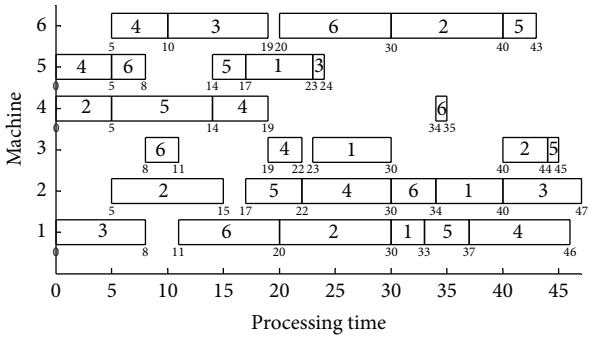


Figure 5: Recovery schedule of total rescheduling.

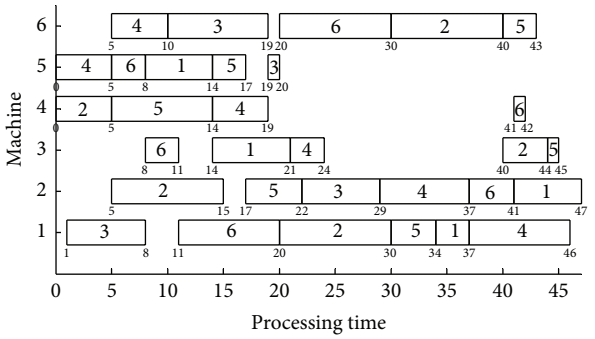


Figure 6: Recovery schedule of right-shift rescheduling.

(ii) From the aspect of the deviation of managers, our model, total rescheduling, and right-shift rescheduling obtain the same result. Therefore, our model is not worse than the other two approaches in reducing operational cost.

(iii) From the aspect of the deviation of workers, right-shift rescheduling obtains the best result based on the original schedule. The results of our model and total rescheduling are relatively poor.

In summary, compared with total rescheduling and right-shift rescheduling, our model increases the deviation of workers. By contrast, our model decreases the deviation of customers and thus significantly improves customer satisfaction. Furthermore, the production enterprise is subject to improving the loyalty of customers to expand the influence and attract more new customers. Such improvement helps enhance the potential profit and promote the development of enterprise in the long run. Therefore, our model is more reasonable and valid than the other two approaches.

7. Conclusions

This study designs a real-time status analyzer to identify the disruption and provides a recovery model to measure the deviation in production scheduling by combining disruption management with IoT. It aims to effectively identify the disruption and obtain a new schedule to minimize the negative impact on the participators (customers, managers, and workers). Meanwhile, IACO is presented to solve the above model.

To validate the effectiveness of our model in handling the disruption, we use an example to test different approaches. The computational result proves that our model is competitive with existing rescheduling and can obtain a relatively satisfactory schedule.

A limitation of this research is that we conducted the computational experiment for the recovery model by adopting the values of β and λ provided by Tversky and Kahneman [32]. Further research will concentrate on deriving the actual values of those parameters in production scheduling.

Competing Interests

The authors declare that they have no competing interests.

Acknowledgments

This work has been partially supported by the grants from the National Natural Science Foundation of China (nos. 71301020, 71272050), Program for Liaoning Excellent Talents in University (no. WJQ2014030), and Program for Liaoning Scientific Research in University (no. L2014216).

References

[1] H. Aytug, M. A. Lawley, K. McKay, S. Mohan, and R. Uzsoy, "Executing production schedules in the face of uncertainties: a review and some future directions," *European Journal of Operational Research*, vol. 161, no. 1, pp. 86–110, 2005.

[2] M. Z. Liu, H. Shan, Z. Q. Jiang, M. Ge, J. Hu, and M. Zhang, "Dynamic rescheduling optimization of job-shop under uncertain conditions," *Journal of Mechanical Engineering*, vol. 45, no. 10, pp. 137–142, 2009.

[3] X. Qi, J. F. Bard, and G. Yu, "Disruption management for machine scheduling: the case of SPT schedules," *International Journal of Production Economics*, vol. 103, no. 1, pp. 166–184, 2006.

[4] G. Yu and X. Qi, *Disruption Management: Framework, Models and Applications*, World Scientific Publishing Company Incorporated, 2004.

[5] M. Løve, K. R. Sørensen, J. Larsen, and J. Clausen, "Disruption management for an airline—rescheduling of aircraft," in *Applications of Evolutionary Computing*, S. Cagnoni, J. Gottlieb, E. Hart, M. Middendorf, and G. R. Raidl, Eds., vol. 2279 of *Lecture Notes in Computer Science*, pp. 315–324, Springer, Berlin, Germany, 2002.

[6] G. Yu, M. Argüello, G. Song, S. M. McCowan, and A. White, "A new era for crew recovery at continental airlines," *Interfaces*, vol. 33, no. 1, pp. 5–22, 2003.

[7] C.-Y. Lee and G. Yu, "Parallel-machine scheduling under potential disruption," *Optimization Letters*, vol. 2, no. 1, pp. 27–37, 2008.

[8] L. Tang and Y. Zhang, "Parallel machine scheduling under the disruption of machine breakdown," *Industrial & Engineering Chemistry Research*, vol. 48, no. 14, pp. 6660–6667, 2009.

[9] J.-J. Wang, J.-B. Wang, and F. Liu, "Parallel machines scheduling with a deteriorating maintenance activity," *Journal of the Operational Research Society*, vol. 62, no. 10, pp. 1898–1902, 2011.

[10] F. Liu, J.-J. Wang, and D.-L. Yang, "Solving single machine scheduling under disruption with discounted costs by quantum-inspired hybrid heuristics," *Journal of Manufacturing Systems*, vol. 32, no. 4, pp. 715–723, 2013.

[11] U. K. Khedlekar, D. Shukla, and R. P. S. Chandel, "Computational study for disrupted production system with time dependent demand," *Journal of Scientific & Industrial Research*, vol. 73, no. 5, pp. 294–301, 2014.

[12] S. K. Paul, R. Sarker, and D. Essam, "Managing disruption in an imperfect production–inventory system," *Computers & Industrial Engineering*, vol. 84, pp. 101–112, 2015.

[13] R. Sarker, D. Essam, S. M. K. Hasan, and A. N. M. Karim, "Managing risk in production scheduling under uncertain disruption," *Artificial Intelligence for Engineering Design, Analysis and Manufacturing*, vol. 30, no. 3, pp. 289–299, 2016.

[14] E. Bendoly, K. Donohue, and K. L. Schultz, "Behavior in operations management: assessing recent findings and revisiting old assumptions," *Journal of Operations Management*, vol. 24, no. 6, pp. 737–752, 2006.

[15] C. H. Loch and Y. Wu, *Behavioral Operations Management*, Now Publishers Inc, 2007.

[16] X. Su, "Bounded rationality in newsvendor models," *Manufacturing & Service Operations Management*, vol. 10, no. 4, pp. 566–589, 2008.

[17] F. Gino and G. Pisano, "Toward a theory of behavioral operations," *Manufacturing & Service Operations Management*, vol. 10, no. 4, pp. 676–691, 2008.

[18] Y. Chen, X. Su, and X. Zhao, "Modeling bounded rationality in capacity allocation games with the quantal response equilibrium," *Management Science*, vol. 58, no. 10, pp. 1952–1962, 2012.

[19] K. V. Katsikopoulos and G. Gigerenzer, "Behavioral operations management: a blind spot and a research program," *Journal of Supply Chain Management*, vol. 49, no. 1, pp. 3–7, 2013.

[20] Y. Chen and X. Zhao, "Decision bias in capacity allocation games with uncertain demand," *Production and Operations Management*, vol. 24, no. 4, pp. 634–646, 2015.

[21] D. Kahneman and A. Tversky, "Prospect theory: an analysis of decision under risk," *Econometrica*, vol. 47, no. 2, pp. 263–292, 1979.

[22] Q. L. Ding, X. P. Hu, and Y. Jiang, "A model of disruption management based on prospect theory in logistic distribution," *Journal of Management Sciences in China*, vol. 17, no. 11, pp. 1–9, 2014.

[23] A. Colorni, M. Dorigo, and V. Maniezzo, "Distributed optimization by ant colonies," in *Proceedings of the European Conference on Artificial Life*, pp. 134–142, Paris, France, 1991.

[24] S. R. Balseiro, I. Loiseau, and J. Ramonet, "An Ant Colony algorithm hybridized with insertion heuristics for the time dependent vehicle routing problem with time windows," *Computers & Operations Research*, vol. 38, no. 6, pp. 954–966, 2011.

[25] R.-H. Huang, C.-L. Yang, and W.-C. Cheng, "Flexible job shop scheduling with due window—a two-pheromone ant colony approach," *International Journal of Production Economics*, vol. 141, no. 2, pp. 685–697, 2013.

[26] M. Faisal, H. Mathkour, and M. Alsulaiman, "AntStar: enhancing optimization problems by integrating an ant system and A* algorithm," *Scientific Programming*, vol. 2016, Article ID 5136327, 12 pages, 2016.

[27] É. D. Taillard, "Parallel taboo search techniques for the job shop scheduling problem," *INFORMS Journal on Computing*, vol. 6, no. 2, pp. 108–117, 1994.

[28] E. Demirkol, S. Mehta, and R. Uzsoy, "A computational study of shifting bottleneck procedures for shop scheduling problems," *Journal of Heuristics*, vol. 3, no. 2, pp. 111–137, 1997.

[29] E. Nowicki and C. Smutnicki, "An advanced tabu search algorithm for the job shop problem," *Journal of Scheduling*, vol. 8, no. 2, pp. 145–159, 2005.

[30] P. M. Pardalos and O. V. Shylo, "An algorithm for the job shop scheduling problem based on global equilibrium search

techniques," *Computational Management Science*, vol. 3, no. 4, pp. 331–348, 2006.

[31] P. M. Pardalos, O. V. Shylo, and A. Vazacopoulos, "Solving job shop scheduling problems utilizing the properties of backbone and 'big valley,'" *Computational Optimization and Applications*, vol. 47, no. 1, pp. 61–76, 2010.

[32] A. Tversky and D. Kahneman, "Advances in prospect theory: cumulative representation of uncertainty," *Journal of Risk and Uncertainty*, vol. 5, no. 4, pp. 297–323, 1992.

[33] R. J. Abumaizar and J. A. Svestka, "Rescheduling job shops under random disruptions," *International Journal of Production Research*, vol. 35, no. 7, pp. 2065–2082, 1997.

Hindawi Publishing Corporation
Scientific Programming
Volume 2016, Article ID 1812094, 10 pages
http://dx.doi.org/10.1155/2016/1812094

Research Article

A Novel Evaluation Approach for Tourist Choice of Destination Based on Grey Relation Analysis

Xudong Guo[1,2] and Zilai Sun[2]

[1] *Tourism and Historical Culture College, Zhaoqing University, Yingbin Avenue, Zhaoqing 526061, China*
[2] *Institute of Systems Engineering, Dalian University of Technology, No. 2, Linggong Road, Ganjingzi District, Dalian 116024, China*

Correspondence should be addressed to Zilai Sun; sunzilai@mail.dlut.edu.cn

Received 16 June 2016; Revised 16 August 2016; Accepted 21 August 2016

Academic Editor: Guo Chen

The decision-making process of choosing an ideal tourism destination is influenced by a number of psychological and nonpsychological variables. Tourists need a method to quickly and easily select a suitable destination. Driven by this practical decision issue, a novel approach of tourist destination evaluation, grey relation analysis (GRA), is developed and applied to the ranking evaluation of Taiwan tourism destinations in China. In the evaluating process, we apply entropy to calculate the weight of each index, which is a more objective method of calculating weights. The results of the study indicate that although the same size is small and the distribution of data is unknown, GRA can still be successfully used in evaluating tourist destinations. In addition, we compare the GRA results with the Technique for Order Preference by Similarity to Ideal Solution (TOPSIS) and show that more accurate ranking results can be obtained.

1. Introduction

With the development of social economy and the improvement of people's living standard, more public holidays, faster transportation (e.g., convenient air networks and high-speed railways), and greater openness toward the world, China's domestic and outbound demand for tourism have grown unprecedentedly since the 2000s [1]. According to the National Bureau of Statistics of China, the number of domestic tourists reached 3,611 million in 2014. In addition, the number of overseas visitor arrivals was 128.4983 million and the number of Chinese outbound visitors was 107.2755 million in 2014. Along with the significant growth in tourism, tourism expenditure and foreign exchange earnings in China have also increased significantly. In 2014, China's earnings from domestic tourism reached RMB 3,031.18 billion, and the foreign exchange earnings from international tourism reached USD 56,913 million. In tandem with the tremendous development of China's tourism economy, China's government has committed to making tourism development a major policy, while China already has so much to offer to international and domestic travelers: its natural beauty; a rich culture; great cuisine; amazing technological advancement; and friendly citizens. According to the China National Tourism Administration, by 2014, China has 184 5A-class tourist scenic spots, 862 5-star hotels, and 600 travel agencies [2], as shown in Figure 1. Tourism has become one of the strongest and largest industries and turned into the period of buyer's market in China's process of development.

One of the biggest issues facing the tourist today is how to evaluate and rank the tourism destination, as well as the need to understand what factors influence their choices [3]. However, the decision-making process of choosing an appropriate travel destination is a very complex process, and choosing a viable method for effective evaluation of tourism destination is not an easy task. There is a substantial body of literature discussing different research methods applied to evaluation and choice of tourism destination. These methods include (1) Multinomial Logit Model (MNL) [4–7]; (2) Binary Regression Model [8–11]; (3) Multinomial Probit Model [12]; (4) Scobit Based Discrete-Continuous Choice Model [13]; (5) Structural Equation Model [14–19]; (6) Meta-analysis

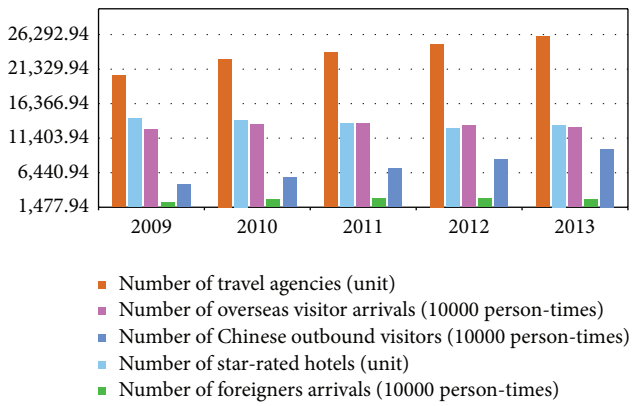


Figure 1: China's tourism development situation for nearly five years.

[20]; (7) Technique for Order Preference by Similarity to Ideal Solution (TOPSIS) [21–24]; (8) Principal Component Analysis [25–30]; and (9) Analytic Hierarchy Process (AHP) [31–37]. Some of the methods may have already been known to the public. Other methods were simply borrowed from the domain of industrial study and applied to tourism. Some are still in the embryonic stage. Each of the above nine methods can be independently applied to evaluate tourism destination. However, none of them is perfect. Researchers can only choose a method to evaluate tourism destination that has the least amount of drawbacks for that study's particular situation. Some evaluation methods require the high quality data; that is, the data should obey a certain distribution or the data volume should be large. In addition, the other evaluation methods have strong subjectivity.

Therefore, a viable method for effective evaluation of tourism destination is aimed at providing solutions for issues with multiple variables and targets. However, in studying the evaluation of tourism destination, survey data are often incomplete or unclear, and this paper therefore is bound by realistic limits, confining itself to a situation where the amount of data is small and its significance indefinite. Given these considerations, grey relation analysis (GRA) may be the best method for this study and will be used to explore practical procedure of tourist destination evaluation and choice to identify the various features that need to reflect preferences of tourists for destinations in the context of 8 tourist spots by utilizing the index system and data in literature [21]. On the basis of such preliminary analysis, this paper intends to achieve twofold aims. The first applies entropy to find the relative weights of each index of the 8 tourist spots. The second is to rank the tourist spots by grey relation analysis (GRA) and compare the ranking results with literature [21].

The rest of this paper is organized as follows. After a brief review of relevant literature in Section 2, we present hierarchy of tourism destination selection and evaluation index system in Section 3. The grey relation analysis evaluation model for tourism destinations selection is presented in Section 4. Section 5 illustrates the results obtained from the grey relation analysis evaluation model and then, we compare and analyze these results with those obtained from the Technique for Order Preference by Similarity to Ideal Solution (TOPSIS) evaluation method used by literature [21]. Conclusions are drawn in Section 6.

2. Literature Review

Evaluation and choice of tourism destination is one of the important topics frequently investigated by scholars [38]. These evaluation and choice studies have been linked to the issues of decision rules, decision-making processes, and choice factors [5, 39–45]. Hsu et al. [21] proposed a 4-level AHP model, consisting of 22 attributes on the 4th level and tested it using data collected from tourists visiting Taiwan to establish the relative importance of preselected factors (criteria). In this study, visiting friends/relatives and personal safety appear to be the 2 most important factors for inbound tourists to Taiwan, price is the least important, and Taipei 101 is the first priority for travelers. Huybers [46] applied discrete choice modeling technique to study the short-break destination choices of prospective Sydney residents. The study results show that a destination's attractiveness to prospective short-break visitors from Sydney is enhanced, to different degrees, by lower prices, being moderately busy, having a moderate level of night life, being visited during spring/summer, being accessible within two hours' travel time, and offering a mix of natural and cultural/heritage attractions. Prayag and Ryan [14] used confirmatory factor analysis to explore the relationships among four constructs, namely, destination image, place attachment, personal involvement, and visitors' satisfaction as antecedents of loyalty. The research findings indicate that destination image, personal involvement, and place attachment are antecedents of visitors' loyalty but this relationship is mediated by satisfaction levels.

Dellaert et al. [47] pointed out that tourists' choices are complex multifaceted decisions where the choices for different elements are interrelated and evolved in a decision process over time. The decision-making process is influenced by a number of psychological (internal) and nonpsychological (external) variables. Sirakaya and Woodside [48] provided a comprehensive qualitative review of the tourist decision-making literature and integrated the main conceptual and empirical work that has been reported in the tourism literature. Lam and Hsu [49] argued that the complex decision-making process leading to the choice of a travel destination had not been well researched. Past studies related to destination choice mainly focus on identifying important attributes affecting destination choice; professional judgment and factor analysis are the main methods [50–53]. Richards and Wilson [54] in their work focused on independent youth and student travel and used factor analysis to help identify four main motivating factors, including experience seeking, relaxation seeking, and sociability and contributing to the destination. Tomiš et al. [55] studied the factors considered

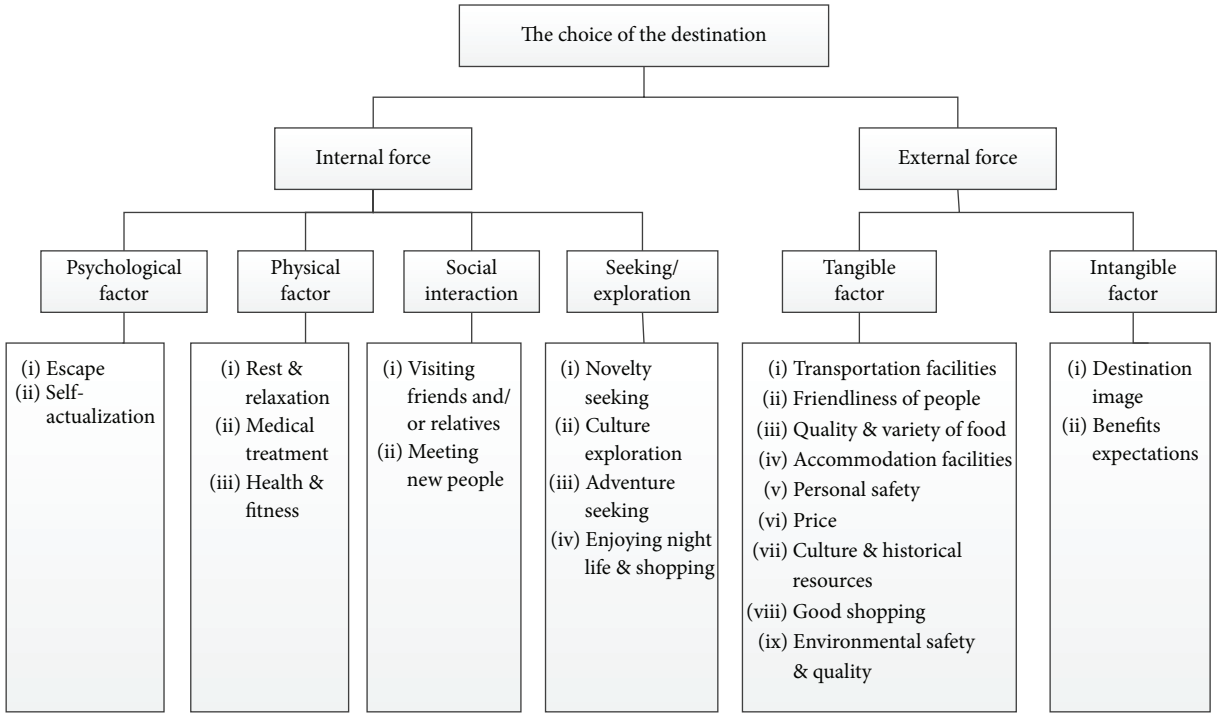


Figure 2: The hierarchy of destination selection (source: Hsu et al. [21]).

by young people when choosing a city destination in Europe and identified seven major factors: partying and having fun, accessibility to destination info, easy and cheap travel organization, outdoor activities, socializing with the local people, good shopping places, and exploring the unknown. Further analysis showed that there are significant differences among several motivation factors when it comes to Danish and international students. Pike and Page [30] provided a comprehensive narrative analysis of the destination marketing literature.

The above studies mainly applied qualitative and quantitative methods to analyze the factors influencing the choice of tourism destinations. They pointed out that the decision-making process leading to the final choice of a travel destination is very complex. However, these studies failed to consider the preferences of tourists in the decision-making process and utilize much more complex decision-making approaches. Motivated by this, we develop a new evaluation model to evaluate and rank the tourism destinations considering personal preferences. Thus, we attempt to evaluate the preferences of tourists for destinations and recognize the key aspects influencing the tourists' choice of destination. Finally, Guo et al. [56] proposed a novel similarity-based algorithm for classifying tourism scenic spots, and this evaluation method not only has some similarities to our own but also has some key differences. They studied the problem of classifying different tourism scenic spots into different levels through calculating the relative similarity of each scenic spot to the best and worst scenic spots, but, in our model, we do not consider the worst scenic spots, which make the evaluation process simple and easy to conduct.

3. The Hierarchy of Tourism Destinations Selection

According to Hsu et al. [21], the main factors affecting tourism destination choice are grouped into two categories, namely, internal and external factors. The push motivation relates to internal forces which consist of 4 criteria (psychological, physical, social interaction, and seeking/exploration) and are further divided into 11 subcriteria. The pull motivation relates to external forces which consist of 2 criteria (tangible and intangible) with tangible criteria including 9 subcriteria and intangible criteria including 2 subcriteria (see Figure 2).

4. The Grey Relation Analysis Model of Tourism Destination Choice

Grey relation analysis (GRA) is a method that can be used in decision-making in situations where there are many criteria by ordering them as to relational grade. It is especially preferred in ordering the alternatives in situations in which the sample is small and sample distribution is not known. GRA is part of the grey system theory [57]. Gray system theory was first introduced by Ju-Long [58]. The fundamental definition of "greyness" is information being incomplete or unknown;

thus an element from an incomplete message is considered to be of "grey" element. A "grey relation" refers to the measurement of changing relations between two systems or elements that occur in a system over time. The analysis method, which measures the relations between elements based on the degree of similarity or difference of development trends among these elements, is called "grey relation analysis." More precisely, during the process of system development, should the trend of change between two elements be consistent, it then enjoys a higher grade of synchronized change and can be considered as having a greater grade of relation; otherwise, the grade of relation is smaller [59]. It has been widely applied in a variety of fields such as financial performance evaluation [60–64], motor vehicular energy consumption [65, 66], and green supplier selection [67]. As far as we know, GRA has not been applied in evaluation and choice of tourism destination. We believe we are the first to utilize the GRA to evaluate and rank the tourism destinations.

Personal preferences, like motivations, may be both intrinsic, reflecting individual likes and dislikes, and extrinsic, or socially conditioned. Pearce [68] stated that preferences are more specific than motivations and are revealed by where travelers go and what travelers do. However, tourists' preferences are psychological behaviors, which are not easy to accurately measure, having the feature of grey and fuzzy. Choosing a tourism destination is a grey system evaluation process; therefore, the grey relation analysis is a more suitable method to evaluate it. The evaluation procedure of this study consists of several steps. The detailed descriptions of each step are given in the following subsections.

4.1. Problem Definition. Assume that there are n tourism destinations $\{T_1, T_2, \ldots, T_n\}$, which are to be evaluated in terms of m evaluation criteria $\{C_1, C_2, \ldots, C_m\}$, and relative weights are denoted by ω_{ij} ($\sum_{j=1}^{m} \omega_{ij} = 1$). The objective of the decision-making is to rank these tourism destinations. According to the assumption, we can get the original index data matrix

$$R = \left(r_{ij}\right)_{n\times m} = \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1m} \\ r_{21} & r_{22} & \cdots & r_{2m} \\ \vdots & \vdots & & \vdots \\ r_{n1} & r_{n2} & \cdots & r_{nm} \end{bmatrix}, \tag{1}$$

where r_{ij} is the observed value of object T_i in terms of criteria C_j ($i = 1, 2, \ldots, n$; $j = 1, 2, \ldots, m$).

Definition 1. The best tourism destination is the referential sequence made up of the best tourism destination indexes among all the alternative tourism destinations; that is,

$$R_0 = \left(r_{01}, r_{02}, \ldots, r_{0j}\right), \tag{2}$$

where $r_{0j} = \text{Optimum}(r_{ij})$ ($i = 1, 2, \ldots, n$; $j = 1, 2, \ldots, m$).

So, we can construct initial matrix R_1:

$$R_1 = \left(r_{ij}\right)_{n\times m} = \begin{bmatrix} r_{01} & r_{02} & \cdots & r_{0m} \\ r_{11} & r_{12} & \cdots & r_{1m} \\ r_{21} & r_{22} & \cdots & r_{2m} \\ \vdots & \vdots & \cdots & \vdots \\ r_{n1} & r_{n2} & \cdots & r_{nm} \end{bmatrix}. \tag{3}$$

4.2. Normalization of Index Values. In order to make comparison between the various indicators, we normalized each of the index values in accordance with the following formulas. For larger-is-better transformation, r_{ij} can be transformed to x_{ij}. The formula is defined as

$$X_{ij} = \frac{r_{ij} - \min_j r_{ij}}{\max_j r_{ij} - \min_j r_{ij}}. \tag{4}$$

For smaller-is-better transformation, the formula to transform r_{ij} to x_{ij} is

$$X_{ij} = \frac{\max_j r_{ij} - r_{ij}}{\max_j r_{ij} - \min_j r_{ij}}. \tag{5}$$

The normalized values form a new matrix as shown in

$$X = \left(x_{ij}\right)_{n\times m} = \begin{bmatrix} x_{01} & x_{02} & \cdots & x_{0m} \\ x_{11} & x_{12} & \cdots & x_{1m} \\ x_{21} & x_{22} & \cdots & x_{2m} \\ \vdots & \vdots & \cdots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nm} \end{bmatrix}. \tag{6}$$

4.3. Calculating Grey Relation Coefficient. According to Ju-Long [69], let $X_0 = (x_{01}, x_{02}, \ldots, x_{0m})$ as referential sequence and $X_i = (x_{i1}, x_{i2}, \ldots, x_{im})$ as comparative sequence. Then x_{0j} and x_{ij} would be the values of x_0 and x_i at optimum index j. If $\xi(x_{0j}, x_{ij})$ are real numbers, then they can be defined as

$$\xi\left(x_0, x_i\right) = \frac{1}{m}\sum_{j=1}^{m} \xi\left(x_{0j}, x_{ij}\right). \tag{7}$$

The mean of $\xi(x_{0j}, x_{ij})$ needs to meet the four axioms of grey relation.

Axiom 1 (norm interval). Consider

$0 < \xi(x_{0j}, x_{ij}) \leq 1$.

$\xi(x_{0j}, x_{ij}) = 1 \leftrightarrow x_{0j} = x_{ij}$ (this is called complete relation).

$\xi(x_{0j}, x_{ij}) = 0 \leftrightarrow x_0 \in \varphi$, $x_i \in \varphi$ (this is called complete nonrelation), where φ is an empty set.

Axiom 2 (duality symmetric). Consider

$x_0, x_i \in X$.

$\xi(x_{0j}, x_{ij}) = \xi(x_{ij}, x_{0j}) \leftrightarrow X = (x_0, x_i)$.

Axiom 3 (wholeness). Consider

$x_j, x_i \in X$.

$\xi(x_{0j}, x_{ij}) \neq \xi(x_{ij}, x_{0j})$; this case almost always happened.

Axiom 4 (approachability). Consider

$\xi(x_{0j}, x_{ij})$ get larger along with $|x_{0j} - x_{ij}|$ getting smaller.

If the abovementioned four axioms could all be satisfied, $\xi(x_0, x_i)$ is designed as the grade of grey relation in x_i correspondence to x_0, and $\xi(x_{0j}, x_{ij})$ is said to be the grey relational coefficient of x_{ij} to x_{0j} at optimum index j. Professor Ju-Long [58] proposes a mathematical equation that will satisfy these four axioms of grey relation, which is as follows:

$$\xi\left(X_{0j}, X_{ij}\right) = \frac{\min_i \min_j \left|X_{0j} - X_{ij}\right| + \rho \max_i \max_j \left|X_{0j} - X_{ij}\right|}{\left|X_{0j} - X_{ij}\right| + \rho \max_i \max_j \left|X_{0j} - X_{ij}\right|}, \quad (8)$$

where $\rho \in [0, 1]$ is distinguished coefficient, the function of which is to reduce its numerical value by $\max_i \max_j |X_{0j} - X_{ij}|$ increasing, so as to affect its loss-authenticity and to heighten the significance of difference among relational coefficients, and $\rho = 0.5$ is generally used. Relational coefficients $\xi(X_{0j}, X_{ij})$ matrix can be obtained through formula (8). As a matter of convenience, we use ξ_{ij} instead of $\xi(X_{0j}, X_{ij})$ in the following pages:

$$E = \left(\xi_{ij}\right)_{nm} = \begin{bmatrix} \xi_{11} & \xi_{12} & \cdots & \xi_{1m} \\ \xi_{21} & \xi_{22} & \cdots & \xi_{2m} \\ \vdots & \vdots & & \vdots \\ \xi_{n1} & \xi_{n2} & \cdots & \xi_{nm} \end{bmatrix} \quad (i = 1, 2, \ldots, n;\ j = 1, 2, \ldots, m). \quad (9)$$

4.4. Determining the Index Weights Based on Entropy. Making normalization processing for formula (9), ξ_{ij} can be transformed to μ_{ij}, and a new matrix (9) can be obtained as follows:

$$E_0 = \left(\mu_{ij}\right)_{nm} = \begin{bmatrix} \mu_{11} & \mu_{12} & \cdots & \mu_{1m} \\ \mu_{21} & \mu_{22} & \cdots & \mu_{2m} \\ \vdots & \vdots & & \vdots \\ \mu_{n1} & \mu_{n2} & \cdots & \mu_{nm} \end{bmatrix} \quad (i = 1, 2, \ldots, n;\ j = 1, 2, \ldots, m), \quad (10)$$

where $0 \leq \mu_{ij} \leq 1$ and $\sum_{i=1}^{n} \mu_{ij} = 1$; then, the information entropy is

$$H(j) = -\sum_{i=1}^{n} \mu_{ij} \cdot \log \mu_{ij}. \quad (11)$$

Let

$$\nu_{ij} = 1 + \frac{\sum_{i=1}^{n} \mu_{ij} \cdot \log \mu_{ij}}{\log n}. \quad (12)$$

Then, we obtain the weights of indexes in terms of formula (12):

$$\omega_{ij} = \frac{\nu_{ij}}{\sum_{j=1}^{m} \nu_{ij}} \left(0 \leq \omega_{ij} \leq 1,\ \sum_{j=1}^{m} \omega_{ij} = 1\right). \quad (13)$$

4.5. Calculating Grey Relational Grade. After obtaining the grey relational coefficient, we normally take the weighted value of the grey relational coefficient multiplied by weighting value as the grey relational grade. The grey relational grade is defined as follows [70]:

$$B = \omega_i \cdot \left(\xi_{ij}\right)_{nm}^{T} = \omega_i \cdot \begin{bmatrix} \xi_{11} & \xi_{21} & \cdots & \xi_{n1} \\ \xi_{12} & \xi_{22} & \cdots & \xi_{n2} \\ \vdots & \vdots & & \vdots \\ \xi_{1m} & \xi_{n2} & \cdots & \xi_{nm} \end{bmatrix} = \left(b_1, b_2, \ldots, b_n\right), \quad (14)$$

where b_n is grey relational grade of the tourism destination T_n which indicates the degree of influence that the comparative sequence could exert over the referential sequence. Therefore, if a particular comparative sequence is better than the other comparative sequences to the referential sequence, then the grey relational grade for that comparative sequence and referential sequence will be higher than other grey relational grades. Finally, all tourism destinations are evaluated synthetically according to the value of b_n, where when b_n is larger, the tourism destination T_n is better.

5. A Numerical Example

In this section, we use the proposed model to evaluate 8 tourism destinations in Taiwan Province of China (including Taipei 101, National Museum, Sun Moon Lake, Alishan, Yushan National Park, Taroko National Park, Love River, and Kenting National Park). Twenty-two (22) indexes identified from literature [21] as shown in Table 1 are used in the study ("a" is the best preference out of the 8 destinations).

5.1. The Relational Grade to Ideal Destination. According to the data from Table 1, we identify the optimum column as (7.16, 7.16, 7.59, 6.40, 7.15, 8.54, 7.45, 8.04, 7.87, 6.73, 8.56, 8.60, 6.95, 8.85, 7.81, 7.20, 7.36, 6.05, 8.47, 8.04, 7.94, 7.67), and in combination with Table 1, an initial matrix is obtained as follows:

Table 1: Overall preference measures of destinations.

Criteria	Taipei 101	National Museum	Sun Moon Lake	Alishan	Yushan Park	Taroko Park	Love River	Kenting Park
Escape	6.23	5.42	7.16[a]	6.80	5.94	7.08	5.99	6.92
Self-actualization	7.16[a]	6.89	6.30	6.09	5.22	6.41	5.66	6.45
Rest/relaxation	5.38	5.05	7.59[a]	7.06	6.13	7.05	6.36	7.30
Medical treatment	6.40[a]	4.41	4.30	4.34	3.02	4.09	5.38	4.73
Health and fitness	5.05	4.10	6.57	6.98	6.34	7.15[a]	5.17	6.57
Visiting friend/relative	7.96	5.54	8.54[a]	7.01	5.88	6.84	6.86	7.42
Meeting new people	7.45[a]	5.55	5.98	5.55	4.37	5.61	5.76	6.80
Novelty seeking	8.04[a]	6.08	5.35	5.06	4.16	5.22	5.61	6.29
Culture exploration	6.27	7.87[a]	6.54	6.41	5.42	6.28	5.66	6.45
Adventure seeking	3.97	3.43	5.96	6.56	6.30	6.67	4.94	6.73[a]
Enjoying night life	8.56[a]	5.53	4.49	3.95	3.43	4.41	6.45	5.80
Transportation facilities	8.60[a]	8.03	6.20	5.28	4.04	5.45	6.78	6.70
Friendliness of people	6.48	6.52	6.68	6.84	5.95	6.93	6.33	6.95[a]
Quality/variety of food	8.85[a]	6.41	5.70	5.58	4.28	5.80	7.16	6.88
Accommodation	7.81[a]	7.26	6.67	5.72	4.31	6.63	7.36	7.52
Environment safety	6.10	6.38	6.98	6.81	6.07	7.20[a]	6.24	6.73
Personal safety	7.36[a]	7.29	5.80	5.02	4.55	5.38	5.80	6.59
Price	4.39	5.12	5.55	5.64	5.88	6.05[a]	5.30	5.30
Culture and historical resources	6.11	8.47[a]	6.78	6.72	6.22	6.63	5.60	6.49
Good shopping	8.04[a]	5.99	4.89	4.79	3.63	5.40	6.13	6.11
Destination image	7.78	7.27	7.94[a]	7.06	6.53	6.92	6.38	7.14
Benefits expectations	7.67[a]	6.80	5.41	5.41	4.60	5.32	6.48	6.67

$$R_1 = \begin{bmatrix} 7.16 & 7.16 & 7.59 & 6.40 & 7.15 & 8.54 & 7.45 & 8.04 & 7.87 & 6.73 & 8.56 & 8.60 & 6.95 & 8.85 & 7.81 & 7.20 & 7.36 & 6.05 & 8.47 & 8.04 & 7.94 & 7.67 \\ 6.23 & 7.16 & 5.38 & 6.40 & 5.05 & 7.96 & 7.45 & 8.04 & 6.27 & 3.97 & 8.56 & 8.60 & 6.48 & 8.85 & 7.81 & 6.10 & 7.36 & 4.39 & 6.11 & 8.04 & 7.78 & 7.67 \\ 5.42 & 6.89 & 5.05 & 4.41 & 4.10 & 5.54 & 5.55 & 6.08 & 7.87 & 3.43 & 5.53 & 8.03 & 6.52 & 6.41 & 7.26 & 6.38 & 7.29 & 5.12 & 8.47 & 5.99 & 7.27 & 6.80 \\ 7.16 & 6.30 & 7.59 & 4.30 & 6.57 & 8.54 & 5.98 & 5.35 & 6.54 & 5.96 & 4.49 & 6.20 & 6.68 & 5.70 & 6.67 & 6.98 & 5.80 & 5.55 & 6.78 & 4.89 & 7.94 & 5.41 \\ 6.80 & 6.09 & 7.06 & 4.34 & 6.98 & 7.01 & 5.55 & 5.06 & 6.41 & 6.56 & 3.95 & 5.28 & 6.84 & 5.58 & 5.72 & 6.81 & 5.02 & 5.64 & 6.72 & 4.79 & 7.06 & 5.41 \\ 5.94 & 5.22 & 6.13 & 3.02 & 6.34 & 5.88 & 4.37 & 4.16 & 5.42 & 6.30 & 3.43 & 4.04 & 5.95 & 4.28 & 4.31 & 6.07 & 4.55 & 5.88 & 6.22 & 3.63 & 6.53 & 4.60 \\ 7.08 & 6.41 & 7.05 & 4.09 & 7.15 & 6.84 & 5.61 & 5.22 & 6.28 & 6.67 & 4.41 & 5.45 & 6.93 & 5.80 & 6.63 & 7.20 & 5.38 & 6.05 & 6.63 & 5.40 & 6.92 & 5.32 \\ 5.99 & 5.66 & 6.36 & 5.38 & 5.17 & 6.86 & 5.76 & 5.61 & 5.66 & 4.94 & 6.45 & 6.78 & 6.33 & 7.16 & 7.36 & 6.24 & 5.80 & 5.30 & 5.60 & 6.13 & 6.38 & 6.48 \\ 6.92 & 6.45 & 7.30 & 4.73 & 6.57 & 7.42 & 6.80 & 6.29 & 6.45 & 6.73 & 5.80 & 6.70 & 6.95 & 6.88 & 7.52 & 6.73 & 6.59 & 5.30 & 6.49 & 6.11 & 7.14 & 6.67 \end{bmatrix}. \quad (15)$$

Each of the index values is normalized according to formulas (4) and (5), and then a new matrix is obtained as follows:

$$X = \begin{bmatrix} 1.00 & 1.00 \\ 0.47 & 1.00 & 0.13 & 1.00 & 0.31 & 0.81 & 1.00 & 1.00 & 0.35 & 0.16 & 1.00 & 1.00 & 0.53 & 1.00 & 1.00 & 0.03 & 1.00 & 0.00 & 0.18 & 1.00 & 0.90 & 1.00 \\ 0.00 & 0.86 & 0.00 & 0.41 & 0.00 & 0.00 & 0.38 & 0.49 & 1.00 & 0.00 & 0.41 & 0.88 & 0.57 & 0.47 & 0.84 & 0.27 & 0.98 & 0.44 & 1.00 & 0.54 & 0.57 & 0.72 \\ 1.00 & 0.56 & 1.00 & 0.38 & 0.81 & 1.00 & 0.52 & 0.31 & 0.46 & 0.77 & 0.21 & 0.47 & 0.73 & 0.31 & 0.67 & 0.81 & 0.44 & 0.70 & 0.41 & 0.29 & 1.00 & 0.26 \\ 0.79 & 0.45 & 0.79 & 0.39 & 0.94 & 0.49 & 0.38 & 0.23 & 0.40 & 0.95 & 0.10 & 0.27 & 0.89 & 0.28 & 0.40 & 0.65 & 0.17 & 0.75 & 0.39 & 0.26 & 0.44 & 0.26 \\ 0.30 & 0.00 & 0.43 & 0.00 & 0.73 & 0.11 & 0.00 & 0.00 & 0.00 & 0.87 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.90 & 0.22 & 0.00 & 0.10 & 0.00 \\ 0.95 & 0.61 & 0.79 & 0.32 & 1.00 & 0.43 & 0.40 & 0.27 & 0.35 & 0.98 & 0.19 & 0.31 & 0.98 & 0.33 & 0.66 & 1.00 & 0.30 & 1.00 & 0.36 & 0.40 & 0.35 & 0.23 \\ 0.33 & 0.23 & 0.52 & 0.70 & 0.35 & 0.44 & 0.45 & 0.37 & 0.10 & 0.46 & 0.59 & 0.60 & 0.38 & 0.63 & 0.87 & 0.15 & 0.44 & 0.55 & 0.00 & 0.57 & 0.00 & 0.61 \\ 0.86 & 0.63 & 0.89 & 0.51 & 0.81 & 0.63 & 0.79 & 0.55 & 0.42 & 1.00 & 0.46 & 0.58 & 1.00 & 0.57 & 0.92 & 0.58 & 0.73 & 0.55 & 0.31 & 0.56 & 0.49 & 0.67 \end{bmatrix}. \quad (16)$$

According to formula (8), that is, $\xi_{ij} = (0 + 0.5 \times 1)/(|X_{0j} - X_{ij}| + 0.5 \times 1)$, we obtain the matrix of grey relational coefficients as follows:

$$\xi_{ij} = \begin{bmatrix} 0.48 & 1.00 & 0.36 & 1.00 & 0.42 & 0.72 & 1.00 & 1.00 & 0.43 & 0.37 & 1.00 & 1.00 & 0.52 & 1.00 & 1.00 & 0.34 & 1.00 & 0.33 & 0.38 & 1.00 & 0.83 & 1.00 \\ 0.33 & 0.78 & 0.33 & 0.46 & 0.33 & 0.33 & 0.45 & 0.50 & 1.00 & 0.33 & 0.46 & 0.80 & 0.54 & 0.48 & 0.76 & 0.41 & 0.95 & 0.47 & 1.00 & 0.52 & 0.54 & 0.64 \\ 1.00 & 0.53 & 1.00 & 0.45 & 0.72 & 1.00 & 0.51 & 0.42 & 0.48 & 0.68 & 0.39 & 0.49 & 0.65 & 0.42 & 0.61 & 0.72 & 0.47 & 0.62 & 0.46 & 0.41 & 1.00 & 0.40 \\ 0.71 & 0.48 & 0.71 & 0.45 & 0.90 & 0.50 & 0.45 & 0.39 & 0.46 & 0.91 & 0.36 & 0.41 & 0.82 & 0.41 & 0.46 & 0.59 & 0.38 & 0.67 & 0.45 & 0.40 & 0.47 & 0.40 \\ 0.42 & 0.33 & 0.47 & 0.33 & 0.65 & 0.36 & 0.33 & 0.33 & 0.33 & 0.79 & 0.33 & 0.33 & 0.33 & 0.33 & 0.33 & 0.33 & 0.33 & 0.83 & 0.39 & 0.33 & 0.36 & 0.33 \\ 0.92 & 0.56 & 0.70 & 0.42 & 1.00 & 0.47 & 0.46 & 0.41 & 0.44 & 0.96 & 0.38 & 0.42 & 0.96 & 0.43 & 0.60 & 1.00 & 0.42 & 1.00 & 0.44 & 0.46 & 0.43 & 0.40 \\ 0.43 & 0.39 & 0.51 & 0.62 & 0.44 & 0.47 & 0.48 & 0.44 & 0.36 & 0.48 & 0.55 & 0.56 & 0.45 & 0.57 & 0.80 & 0.37 & 0.47 & 0.53 & 0.33 & 0.54 & 0.33 & 0.56 \\ 0.78 & 0.58 & 0.81 & 0.50 & 0.72 & 0.57 & 0.70 & 0.53 & 0.46 & 1.00 & 0.48 & 0.55 & 1.00 & 0.54 & 0.86 & 0.55 & 0.65 & 0.53 & 0.42 & 0.53 & 0.49 & 0.61 \end{bmatrix}. \quad (17)$$

According to formulas (10)~(13), the index weights vector is calculated as follows:

$$\omega = (0.05, 0.04, 0.05, 0.04, 0.04, 0.04, 0.04, 0.05, 0.05, 0.05, 0.05, 0.04, 0.04, 0.04, 0.03, 0.05, 0.06, 0.04, 0.05, 0.04, 0.05, 0.05). \quad (18)$$

Using formula (14), we obtain the grey relational grades as follows:

$$B = (0.73, 0.57, 0.61, 0.53, 0.40, 0.60, 0.48, 0.63). \quad (19)$$

5.2. Comparison and Analysis. In this paper, we focus on developing a novel evaluation approach tourist choice of destination based on grey relation analysis, in order to demonstrate the advantages of the developed evaluation methods, and comparing the evaluation results with the evaluation method of "TOPSIS" used by literature [21]. So, after obtaining the grey relational grade to the ideal destination by using our proposed method, we compare and analyze these results with the evaluation results obtained from the Technique for Order Preference by Similarity to Ideal Solution (TOPSIS) used by literature [21]. The results are shown in Table 2.

As can be seen from Table 2, we can make the following comparative analysis.

(i) The ranking results obtained from literature [21] and the results in this paper are not completely the same, as shown in column 2 and column 4 of Table 2, which reflect the consistency and difference of two methods. In other words, the ranking results of the top three and the last one are the same but the others are not the same. Interestingly, almost all relational grades to ideal solution of our proposed approach are bigger than those of the similarity to ideal solution by the Technique for Order Preference by Similarity to Ideal Solution (TOPSIS) evaluation method, which shows that our proposed approach has the advantage over the TOPSIS approach.

(ii) Comparing the overall preference measures of destinations as shown in Table 1 of literature [21], it is observed that the Taroko National Park has three best preferences out of the 24 factors, and the National Palace Museum has two best preferences. Therefore, we intuitively think that the destination with more best preferences should be ranked higher than the destination with less best preferences, and the actual evaluation results are consistent with our intuition, which show that our proposed evaluation approach is more credible and more in line with the actual situation. The fundamental reason why our proposed evaluation approach is more in line with the actual situation lies in its own characteristics. In other words, the tourists' preferences to tourism destinations are grey and fuzzy, which is in line with the cognition rules of human being to the objective world.

(iii) No matter what kind of evaluation approach, Taipei 101 is ranked the number one favorite local destination. That is, Taipei 101 is worthy of visiting as it is the highest building in the world, located in the business center of Taipei city and a Taipei landmark. As shown in Table 1, Taipei 101 is considered the most desirable destination with respect to "self-actualization," "meeting new friends," "medical treatment," "novelty seeking," "enjoying night life," "transportation facilities," "quality and variety of food," "accommodation facilities," "good shopping," "personal safety," and "benefit expectations," which reflects the consistency of two methods.

6. Conclusions

From a methodological point of view, we proposed a grey relation analysis evaluation model for choosing tourism destinations and the findings demonstrate that the grey relation analysis evaluation approach is a useful tool to help support a decision in destination choice. It integrates the opinion

Table 2: Final ranking of destinations.

Destination	Results of literature [21]		Results in this paper	
	Rank	Similarity to ideal solution	Rank	Relational grade to ideal solution
Taipei 101	1	0.75	1	0.73
Kenting National Park	2	0.63	2	0.63
Sun Moon Lake	3	0.55	3	0.61
National Palace Museum	4	0.48	5	0.57
Love River	5	0.47	7	0.48
Taroko National Park	6	0.43	4	0.60
Alishan	7	0.37	6	0.53
Yushan National Park	8	0.16	8	0.40

and evaluation of experts and makes the complex decision-making system into a simple grey relation analysis system. In other words, our proposed evaluation approach makes the evaluation process simple and easy to conduct. What is more, it is especially preferred in ordering the alternatives in situations in which the sample is small and sample distribution is not known. In our future studies, we will apply the approach to evaluating other tourism destinations in China and make a ranking for Chinese main tourism destinations.

Competing Interests

The authors declare that there are no competing interests regarding the publication of this paper.

Acknowledgments

This study is supported by the National Natural Science Foundation of China under Grant no. 41440045.

References

[1] V. S. Lin, R. Mao, and H. Song, "Tourism expenditure patterns in China," *Annals of Tourism Research*, vol. 54, pp. 100–117, 2015.

[2] Domestic Tourism, National Bureau of Statistics of China, 2014, http://data.stats.gov.cn/english/easyquery.htm?cn=C01.

[3] S. C. S. Jang and S. Ham, "A double-hurdle analysis of travel expenditure: baby boomer seniors versus older seniors," *Tourism Management*, vol. 30, no. 3, pp. 372–380, 2009.

[4] T. Huybers and J. Bennett, "Impact of the environment on holiday destination choices of prospective UK tourists: implications for Tropical North Queensland," *Tourism Economics*, vol. 6, no. 1, pp. 21–46, 2000.

[5] H. R. Seddighi and A. L. Theocharous, "A model of tourism destination choice: a theoretical and empirical analysis," *Tourism Management*, vol. 23, no. 5, pp. 475–487, 2002.

[6] G. Lee, J. T. O'Leary, L. Sunhee, and A. Morrison, "Comparison and contrast of push and pull motivational effects on trip behavior: an application of a multinomial logistic regression model," *Tourism Analysis*, vol. 7, no. 2, pp. 89–104, 2001.

[7] H. Oppewal, T. Huybers, and G. I. Crouch, "Tourist destination and experience choice: a choice experimental analysis of decision sequence effects," *Tourism Management*, vol. 48, pp. 467–476, 2015.

[8] A. Molina and Á. Esteban, "Tourism brochures: usefulness and image," *Annals of Tourism Research*, vol. 33, no. 4, pp. 1036–1056, 2006.

[9] H. Kim, C.-K. Cheng, and J. T. O'Leary, "Understanding participation patterns and trends in tourism cultural attractions," *Tourism Management*, vol. 28, no. 5, pp. 1366–1371, 2007.

[10] L. Jin, Y. He, and H. Song, "Service customization: to upgrade or to downgrade? An investigation of how option framing affects tourists' choice of package-tour services," *Tourism Management*, vol. 33, no. 2, pp. 266–275, 2012.

[11] J. Ashwell, "Going bush? Factors which influence international tourists' decisions to travel to remote Australian destinations," *Tourism Management*, vol. 46, pp. 80–83, 2015.

[12] V. V. Can, "Estimation of travel mode choice for domestic tourists to Nha Trang using the multinomial probit model," *Transportation Research Part A: Policy and Practice*, vol. 49, pp. 149–159, 2013.

[13] L. Wu, J. Zhang, and A. Fujiwara, "Tourism participation and expenditure behaviour: analysis using a scobit based discrete-continuous choice model," *Annals of Tourism Research*, vol. 40, no. 1, pp. 1–17, 2013.

[14] G. Prayag and C. Ryan, "Antecedents of tourists' loyalty to mauritius: the role and influence of destination image, place attachment, personal involvement, and satisfaction," *Journal of Travel Research*, vol. 51, no. 3, pp. 342–356, 2012.

[15] N. K. Prebensen, E. Woo, J. S. Chen, and M. Uysal, "Motivation and involvement as antecedents of the perceived value of the destination experience," *Journal of Travel Research*, vol. 52, no. 2, pp. 253–264, 2013.

[16] H. Song, R. van der Veen, G. Li, and J. L. Chen, "The Hong Kong tourist satisfaction index," *Annals of Tourism Research*, vol. 39, no. 1, pp. 459–479, 2012.

[17] J.-Y. Wong and C. Yeh, "Tourist hesitation in destination decision making," *Annals of Tourism Research*, vol. 36, no. 1, pp. 6–23, 2009.

[18] B. M. Josiam, D. L. Spears, S. Pookulangara, K. Dutta, T. R. Kinley, and J. L. Duncan, "Using structural equation modeling to understand the impact of Bollywood movies on destination image, tourist activity, and purchasing behavior of Indians," *Journal of Vacation Marketing*, vol. 21, no. 3, pp. 251–261, 2015.

[19] R. Pinky Pawaskar and M. Goel, "Improving the efficacy of destination marketing strategies: a structural equation model for leisure travel," *Indian Journal of Science and Technology*, vol. 9, no. 15, pp. 1–11, 2016.

[20] H. Zhang, X. Fu, L. A. Cai, and L. Lu, "Destination image and tourist loyalty: a meta-analysis," *Tourism Management*, vol. 40, pp. 213–223, 2014.

[21] T.-K. Hsu, Y.-F. Tsai, and H.-H. Wu, "The preference analysis for tourist choice of destination: a case study of Taiwan," *Tourism Management*, vol. 30, no. 2, pp. 288–297, 2009.

[22] J.-H. Huang and K.-H. Peng, "Fuzzy Rasch model in TOPSIS: a new approach for generating fuzzy numbers to assess the competitiveness of the tourism industries in Asian countries," *Tourism Management*, vol. 33, no. 2, pp. 456–465, 2012.

[23] M. H. N. Salleh, R. Othman, and T. Sarmidi, "Tourist satisfaction of the environmental service quality for Tioman Island Marine Park," *Indian Journal of Geo-Marine Sciences*, vol. 41, no. 2, pp. 173–179, 2012.

[24] D. Mohamad and R. M. Jamil, "A Preference analysis model for selecting tourist destinations based on motivational factors: a Case Study in Kedah, Malaysia," *Procedia—Social and Behavioral Sciences*, vol. 65, pp. 20–25, 2012.

[25] M. Cucculelli and G. Goffi, "Does sustainability enhance tourism destination competitiveness? Evidence from Italian destinations of excellence," *Journal of Cleaner Production*, vol. 111, pp. 370–382, 2016.

[26] J. K. S. Jacobsen and A. M. Munar, "Tourist information search and destination choice in a digital age," *Tourism Management Perspectives*, vol. 1, no. 1, pp. 39–47, 2012.

[27] A. H. N. Mak, M. Lumbers, A. Eves, and R. C. Y. Chang, "An application of the repertory grid method and generalised Procrustes analysis to investigate the motivational factors of tourist food consumption," *International Journal of Hospitality Management*, vol. 35, pp. 327–338, 2013.

[28] G. Fuchs, N. Uriely, A. Reichel, and D. Maoz, "Vacationing in a terror-stricken destination: tourists' risk perceptions and rationalizations," *Journal of Travel Research*, vol. 52, no. 2, pp. 182–191, 2013.

[29] W. Elias and Y. Shiftan, "The influence of individual's risk perception and attitudes on travel behavior," *Transportation Research Part A: Policy and Practice*, vol. 46, no. 8, pp. 1241–1251, 2012.

[30] S. Pike and S. J. Page, "Destination marketing organizations and destination marketing: anarrative analysis of the literature," *Tourism Management*, vol. 41, pp. 202–227, 2014.

[31] G. I. Crouch and J. R. B. Ritchie, "Application of the analytic hierarchy process to tourism choice and decision making: a review and illustration applied to destination competitiveness," *Tourism Analysis*, vol. 10, no. 1, pp. 17–25, 2005.

[32] D. Das and K. Mukherjee, "Development of an AHP-QFD framework for designing a tourism product," *International Journal of Services and Operations Management*, vol. 4, no. 3, pp. 321–344, 2008.

[33] H. N. Ali, A. A. I. Sabri, and M. M. N. Noor, "Rating and ranking criteria for selected Islands using fuzzy analytic hierarchy process (FAHP)," *International Journal of Applied Mathematics and Informatics*, vol. 1, no. 6, pp. 57–65, 2012.

[34] G. Fan, E. D. Goodman, and Z. Liu, "AHP (analytic hierarchy process) and computer analysis software used in tourism safety," *Journal of Software*, vol. 8, no. 12, pp. 3114–3119, 2013.

[35] K. J. Pool, M. Amani, and A. Haroni, "Assessment of alternatives for tourism destinations advertising using AIDA model and AHP approach," *Asian Journal of Research in Marketing*, vol. 3, no. 3, pp. 88–97, 2014.

[36] Y. Zhou, K. Maumbe, J. Deng, and S. W. Selin, "Resource-based destination competitiveness evaluation using a hybrid analytic hierarchy process (AHP): the case study of West Virginia," *Tourism Management Perspectives*, vol. 15, pp. 72–80, 2015.

[37] X. Wang, X. R. Li, F. Zhen, and J. Zhang, "How smart is your tourist attraction?: measuring tourist preferences of smart tourism attractions via a FCEM-AHP and IPA approach," *Tourism Management*, vol. 54, pp. 309–320, 2016.

[38] J. S. Chen and D. Gursoy, "An investigation of tourists' destination loyalty and preferences," *International Journal of Contemporary Hospitality Management*, vol. 13, no. 2, pp. 79–85, 2001.

[39] M. Kozak, "Comparative analysis of tourist motivations by nationality and destinations," *Tourism Management*, vol. 23, no. 3, pp. 221–232, 2002.

[40] J. De Vos, P. L. Mokhtarian, T. Schwanen, V. Van Acker, and F. Witlox, "Travel mode choice and travel satisfaction: bridging the gap between decision utility and experienced utility," *Transportation*, vol. 43, no. 5, pp. 771–796, 2016.

[41] U. Pröbstl-Haider, W. Haider, V. Wirth, and B. Beardmore, "Will climate change increase the attractiveness of summer destinations in the European Alps? A survey of German tourists," *Journal of Outdoor Recreation and Tourism*, vol. 11, pp. 44–57, 2015.

[42] A. Garg, "Travel risks vs tourist decision making: a tourist perspective," *International Journal of Hospitality & Tourism Systems*, vol. 8, no. 1, pp. 1–9, 2015.

[43] S. Amaro and P. Duarte, "An integrative model of consumers' intentions to purchase travel online," *Tourism Management*, vol. 46, pp. 64–79, 2015.

[44] Z. Tang, "An integrated approach to evaluating the coupling coordination between tourism and the environment," *Tourism Management*, vol. 46, pp. 11–19, 2015.

[45] N. Paker and C. A. Vural, "Customer segmentation for marinas: evaluating marinas as destinations," *Tourism Management*, vol. 56, pp. 156–171, 2016.

[46] T. Huybers, "Modelling short-break holiday destination choices," *Tourism Economics*, vol. 9, no. 4, pp. 389–405, 2003.

[47] B. G. C. Dellaert, D. F. Ettema, and C. Lindh, "Multi-faceted tourist travel decisions: a constraint-based conceptual framework to describe tourists' sequential choices of travel components," *Tourism Management*, vol. 19, no. 4, pp. 313–320, 1998.

[48] E. Sirakaya and A. G. Woodside, "Building and testing theories of decision making by travellers," *Tourism Management*, vol. 26, no. 6, pp. 815–832, 2005.

[49] T. Lam and C. H. C. Hsu, "Predicting behavioral intention of choosing a travel destination," *Tourism Management*, vol. 27, no. 4, pp. 589–599, 2006.

[50] C. Goossens, "Tourism information and pleasure motivation," *Annals of Tourism Research*, vol. 27, no. 2, pp. 301–321, 2000.

[51] V. C. S. Heung, H. Qu, and R. Chu, "The relationship between vacation factors and socio-demographic and travelling characteristics: the case of Japanese leisure travellers," *Tourism Management*, vol. 22, no. 3, pp. 259–269, 2001.

[52] S. S. Kim and B. Prideaux, "Marketing implications arising from a comparative study of international pleasure tourist motivations and other travel-related characteristics of visitors to Korea," *Tourism Management*, vol. 26, no. 3, pp. 347–357, 2005.

[53] M. Kozak, "Comparative analysis of tourist motivations by nationality and destinations," *Tourism Management*, vol. 23, no. 3, pp. 221–232, 2002.

[54] G. Richards and J. Wilson, "Today's youth travellers: Tomorrow's global nomads. New horizons in independent youth and student travel," International Student Travel Confederation (ISTC), 2003.

[55] N. Tomiš, B. Kovaţeviš, N. Berber, and N. Miliš, "Factors influencing the motivation of young people when choosing a city destination in Europe—a case study from Esbjerg (Denmark)," *European Researcher*, vol. 69, no. 2, pp. 414–428, 2014.

[56] X. D. Guo, H. L. Zhang, and J. H. Ruan, "A novel similarity-based algorithm for classifying tourism scenic spots," *ICIC Express Letters, Part B: Applications*, vol. 6, no. 2, pp. 503–509, 2015.

[57] J. Morán, E. Granada, J. L. Míguez, and J. Porteiro, "Use of grey relational analysis to assess and optimize small biomass boilers," *Fuel Processing Technology*, vol. 87, no. 2, pp. 123–127, 2006.

[58] D. Ju-Long, "Control problems of grey systems," *Systems & Control Letters*, vol. 1, no. 5, pp. 288–294, 1982.

[59] R.-T. Wang, C.-T. B. Ho, and K. Oh, "Measuring production and marketing efficiency using grey relation analysis and data envelopment analysis," *International Journal of Production Research*, vol. 48, no. 1, pp. 183–199, 2010.

[60] C.-T. Ho, "Measuring bank operations performance: an approach based on Grey Relation Analysis," *Journal of the Operational Research Society*, vol. 57, no. 4, pp. 337–349, 2006.

[61] C.-Y. Kung and K.-L. Wen, "Applying Grey Relational Analysis and Grey Decision-Making to evaluate the relationship between company attributes and its financial performance—a case study of venture capital enterprises in Taiwan," *Decision Support Systems*, vol. 43, no. 3, pp. 842–852, 2007.

[62] Y.-J. Wang, "Combining grey relation analysis with FMCGDM to evaluate financial performance of Taiwan container lines," *Expert Systems with Applications*, vol. 36, no. 2, pp. 2424–2432, 2009.

[63] P. T.-W. Lee, C.-W. Lin, and S.-H. Shin, "A comparative study on financial positions of shipping companies in Taiwan and Korea using entropy and grey relation analysis," *Expert Systems with Applications*, vol. 39, no. 5, pp. 5649–5657, 2012.

[64] F. Ecer and A. Boyukaslan, "Measuring performances of football clubs using financial ratios: the gray relational analysis approach," *American Journal of Economics*, vol. 4, no. 1, pp. 62–71, 2014.

[65] S. J. Lin, I. J. Lu, and C. Lewis, "Grey relation performance correlations among economics, energy use and carbon dioxide emission in Taiwan," *Energy Policy*, vol. 35, no. 3, pp. 1948–1955, 2007.

[66] I. J. Lu, S. J. Lin, and C. Lewis, "Grey relation analysis of motor vehicular energy consumption in Taiwan," *Energy Policy*, vol. 36, no. 7, pp. 2556–2561, 2008.

[67] S. H. Hashemi, A. Karimi, and M. Tavana, "An integrated green supplier selection approach with analytic network process and improved Grey relational analysis," *International Journal of Production Economics*, vol. 159, pp. 178–191, 2015.

[68] D. G. Pearce, "Tourist time-budget," *Annals of Tourism Research*, vol. 15, no. 1, pp. 106–121, 1988.

[69] D. Ju-Long, "Introduction to grey system theory," *The Journal of Grey System*, vol. 1, no. 1, pp. 1–24, 1989.

[70] U. Çaydaş and A. Hasçalık, "Use of the grey relational analysis to determine optimum laser cutting parameters with multi-performance characteristics," *Optics & Laser Technology*, vol. 40, no. 7, pp. 987–994, 2008.

Hindawi Publishing Corporation
Scientific Programming
Volume 2016, Article ID 9613293, 6 pages
http://dx.doi.org/10.1155/2016/9613293

Research Article

Collaborative Product Design for Tasks Sorting Based on Shortest Delivery

Xuedong Liang,[1] Wenting Zhou,[1] Aijun Liu,[2] and Hongmei Xue[3]

[1]*Sichuan University, Chengdu 610065, China*
[2]*School of Economics and Management, Xidian University, Xi'an 710071, China*
[3]*School of Information and Electrical Engineering, Hebei University of Engineering, Hebei, China*

Correspondence should be addressed to Hongmei Xue; xuehongmei2004@126.com

Received 17 June 2016; Accepted 14 August 2016

Academic Editor: Junhu Ruan

The "people's innovation" can promote development in enterprises in urgent need of innovative product design. Collaborative product design can be a powerful tool for enterprises wishing to improve their market competitiveness and customer satisfaction. To reduce decision costs, improve efficiency, and solve other issues, promoting "people's innovation" can play a vital role. With this focus, this paper examines products produced through "people's innovation." A collaborative design task scheduling problem is presented. The design tasks are sorted based on a minimum delivery cost principle, which is determined using weighted shortest processing time (WSPT) rules and the shortest delivery time. The results show that distributed collaborative innovation can result in a reasonable arrangement for collaborative design tasks.

1. Introduction

With rapid societal, scientific, and technological development, modern manufacturing enterprises are facing new challenges daily. To strengthen the manufacturing sector, the Chinese government has proposed the "Made in China 2025" strategic framework, in which enterprises are seen not only as players in the main markets but also as drivers of innovation. Manufacturing enterprises are being encouraged to implement innovation-driven strategies to promote a "people's innovation" to enhance their core competitiveness and brand building capability. To win market share and improve customer satisfaction based on the "people's innovation," it is necessary to improve the efficiency of the product design process, to reduce costs, and to improve quality. However, as the product design process involves multiple concurrent tasks, to carry out these multiple tasks, to reduce the cost of decision delivery, and to efficiently complete tasks have become a significant problem. If collaborative product design is organized efficiently, costs would be reduced, and enterprise competitiveness and collaborative product design efficiency improved.

In theoretical research on the study of collaborative intelligent product design methods and technology, domestic and foreign scholars have conducted a number of design-related studies, focused mainly on three main areas. Research in the first area has examined the complexity and dynamics of task sorting in collaborative product design and online ordering methods to minimize completion time. Song et al. [1] studied the optimization of three types of decision sorting at the ordering online point as well as task scheduling problem constraints, from which a collaborative design task rule was developed. Liang et al. [2] introduced coordination theory dependent concepts, a constraints graph analysis, and graph theory to describe collaborative design task logic based on a disjunctive constraint mathematical model for a collaborative design task scheduling problem. Lu et al. [3] simplified the simulation scheduling system using simplified discrete event simulation (SDESA) and particle swarm optimization (PSO) to automatically configure minimum completion time.

The second research area has examined complex product design process goals based on empirical analyses to study collaborative product design. Cui et al. [4] proposed Cooperating Correlative Map Based on Activity (CCM_A), which

was a process model based on a predigestion of a determined display expression conflict. The research was based on a large coupled task set degree of coordination using the tear planning method that fully accounted for the complexity of the information collaboration between the design tasks. Li et al. [5] used Lite-QFD software and combined it with an example of a car to achieve quality function deployment (QFD) conversion and a House of Quality view in collaborative product design. Arsenyan and Büyüközkan [6] proposed a fuzzy quality function deployment, a fuzzy integrated IT planning axiom design method, and a fuzzy rule system for collaborative product design verification to improve IT collaboration methods for enterprises. Joglekar and Ford [7] designed tasks according to priorities to build synergies aimed at improving product design efficiency and developed an optimal resource allocation model.

The third main research area has focused on production processes to complete tasks based on task sorting decomposition. Zhou et al. [8] assigned design tasks based on comprehensive abilities, interest, and time constraints and proposed a new collaborative design task time scheduling allocation policy based on the degree of parallelism and coupling in the collaborative design process. Meng et al. [9] adopted a function and structure approach by dividing the combination of machine tool products according to the breakdown of the collaborative design task. Through the implementation, it was found that a recursive structural design matrix was able to operate independently and complete a set when coupled with normalized operating plan design tasks. As can be seen from the previous research, collaborative product design task sequencing has tended to be limited to theoretical research and there have been few in-depth studies on manufacturing and management. Further, collaborative product design delivery costs and other factors have not been taken into account. Therefore, while a variety of production sorting tasks have been examined, there has been little focus on enterprise collaborative product design in China.

To fill this gap, this article analyses the collaborative product design sorting task process in the context of the "people's innovation." In this analysis, we take the collaborative product design cost factors related to delivery into account. Using WSPT sorting rules to determine total minimum cost and minimum weighted completion time for the ordering scheme and the corresponding delivery, we determine the weight of the task based on the delivery. This operation provides production operating parameters, reduces design costs, improves design efficiency, and enhances enterprise competitiveness, all of which result in relatively good economic and social benefits as well as providing a reference for the next "people's innovation" business background collaborative product design process.

2. Products under the "People's Innovation" Collaborative Design Background

With the proposed "people's innovation" concept, to enhance innovation capacity, enterprises need collaborative product innovation process management, which is based on collaborative innovation processes and collaborative process management, as shown in Figure 1.

Collaborative product innovation process management is divided into a process level and management. Process level customer demand is the starting point followed by problem analysis, process identification, program generation, evaluation, testing, and improvement, until finally an innovative product is developed. The process management layer supports both the management and technology within the rules and constraints [10].

Collaborative information design, as well as understanding the needs of customers, clients, and designers, follows certain teamwork and consultation rules related to temporary team formation and a specific collaborative product design process which encompasses customer needs analysis, target breakdown, problem definition, case knowledge acquisition, problem representation, problem analysis, concept generation, program production, program evaluation, program output, and design completion. The coordination and management innovation platform supports the design process through the provision of product variety, negotiation rules, design constraints, and rules covering the complete collaborative design process, as shown in Figure 2 [10].

In collaborative product design, the designer assigns design tasks and, using queuing theory, task designers break these down to form a set of subtasks, each of which accomplishes a part of the task. Queuing theory is used to assign the task resources, which have been sorted in collaborative product design task order, to the machines. People rather than machines design the tasks, with the weight between the tasks being determined by delivery time. The design tasks are sorted according to the product case, the design constraints, and the innovative platform design rules. Task sorting based on delivery is affected by the length of delivery, which is influenced by many factors, such as personnel characteristics, task characteristics, human factor characteristics, environmental characteristics, and technical characteristics. Therefore, to accurately determine the collaborative product design task sort order for the shortest delivery, there is a need for additional theoretical analysis and research [10].

3. The Shortest Delivery Time for Products Based on the Sorting of the Collaborative Design Tasks

Based on the above collaborative product design process and according to the project management work breakdown structure (WBS), complex collaborative product design activity designers logically divide the activity into n subtasks, with a set denoted by $T = \{T_1, T_2, \ldots, T_n\}$, where ① $T = \bigcup_{i=1}^{n} T_i$ indicates that all subtasks should be combined for collaborative task T, with the complete collaborative task T after being broken down being equal to the original set; ② $\bigcap_{i=1}^{n} T_i \neq \varnothing$ shows that the decomposed subtasks are independent [2]. This article discusses two types of single discussion group task scheduling problems, both of which are

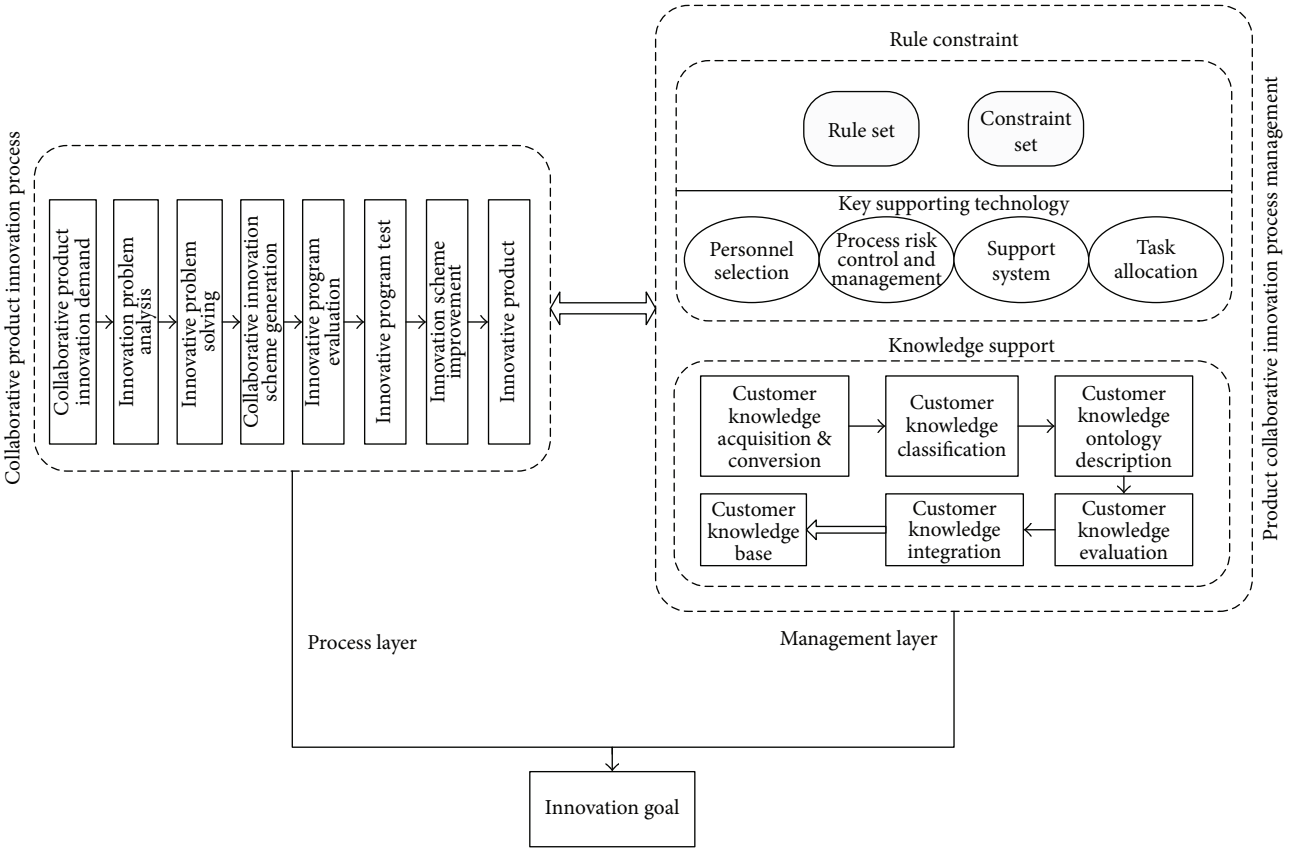


Figure 1: Customer collaborative product innovation management.

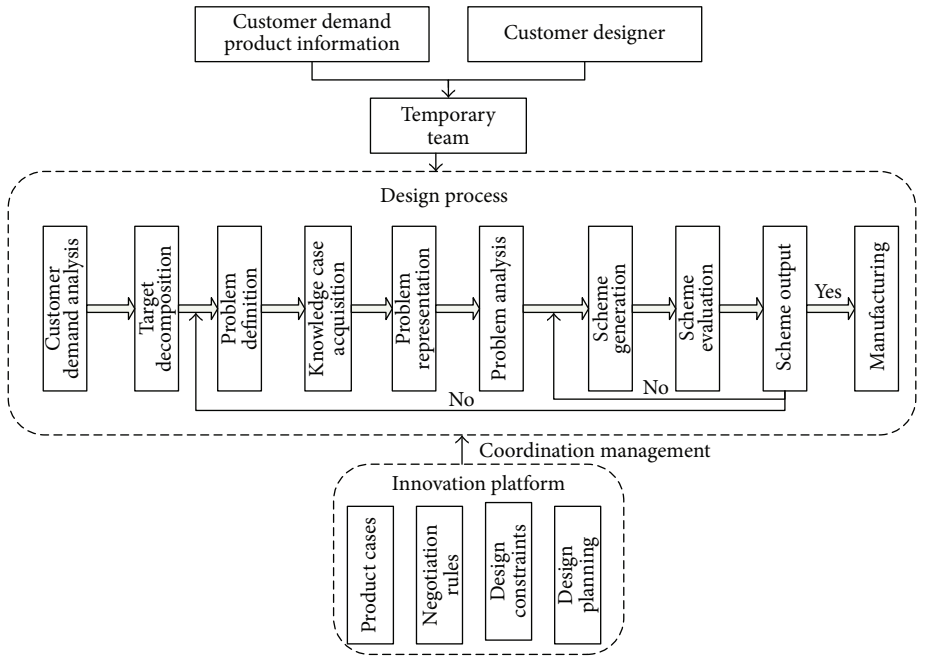


Figure 2: Collaborative product design process.

based on a minimum cost of delivery and shortest scheduling time.

The first task scheduling problem is based on the cost of delivery for the smallest single main task.

Let collaborative design task delivery be k units and the set of tasks $T = \{T_1, T_2, \ldots, T_n\}$, with $P_1, P_2, \ldots, P_n$ representing the processing time for n collaborative product design tasks. There are three kinds of costs: the cost determined by the delivery of the collaborative design tasks, the cost for early task completion, and the penalty to be paid for late task completion.

A collaborative product design task of consists of stages, and each stage has decision variables, state variables, an index function, available state transition equations, and a systems recurrence equation.

To determine the processing time required for the first stage of a task ($x_i = P_j,\ i, j = 1, 2, \ldots, n$), D_i represents the first task completion point for the state variables.

The state transition equation is

$$\begin{aligned} D_i &= D_{i-1} + x_i \quad i = 1, 2, \ldots, n, \\ D_1 &= x_i, \end{aligned} \tag{1}$$

where $A_{(k)}$ is delivery costs of the decision; $B_i(\max\{0, k - D_i\})$ is the personnel costs wasted if early task completion occurs, a nondecreasing function; $E_i(g(\max\{0, D_i - k\}))$ is the delay fine, a nondrop function:

$$g\left(\max\left\{0, D_i - k\right\}\right) = \begin{cases} 0 & \max\left\{0, D_i - k\right\} = 0 \\ 1 & \max\left\{0, D_i - k\right\} > 0. \end{cases} \tag{2}$$

Remember

$$C_i = B_i\left(\max\left\{0, k - D_i\right\}\right) + E_i\left(g\left(\max\left\{0, D_i - k\right\}\right)\right). \tag{3}$$

Then $f_i(D_i)$ represents the minimum cost from state 1 to state i in state D_i; the recurrence equation is

$$\begin{aligned} f_i\left(D_i\right) &= \min\left\{C_i + f_{i-1}\left(D_{i-1}\right)\right\} \quad i = 2, 3, \ldots, n, \\ f_1 &= \min_{x_1 = D_1} C_1. \end{aligned} \tag{4}$$

The total cost f is

$$f = A_{(k)} + f_n\left(D_n\right). \tag{5}$$

In this paper, design task delivery is uncertain. Based on the collaborative design task sort order, the calculation steps for the shortest delivery time are as follows.

Step 1. Determine the possible values:

① Determine different combinations.

② For each combination calculate the elements and assign a number $2^n - 1$, according to the sort appreciation, that, is $a_1, a_2, \ldots, a_{2^n-1}$, if it involves the same number, only one.

Step 2. Let $k_j = a_i$; solve k_j dynamic programming problems (4) and (5) using the formula to determine the corresponding optimal ordering $\{x_1^*, x_2^*, \ldots, x_n^*\}$, with no minimum total cost $f_j = A_{(k_j)} + f_n(D_n),\ j = 1, 2, \ldots, 2^n - 1$.

Step 3. $2^n - 1 f_j$: remember $f_a = \min\{f_j\}$, so A_a is the optimal delivery, $\{x_1^*, x_2^*, \ldots, x_n^*\}_a$ is the optimal sort, and f_a is the minimum cost [11].

The second task scheduling problem is the total flow for the shortest single main task.

Schedule the collaborative product design tasks using the quad $\alpha \mid \beta \mid \gamma \mid \sigma$, where α indicates that there is only single item, β is the implementation task constraints, for which there may be none or many, γ is the minimization objective, and σ indicates that it is a heavy task. Let the collaborative single main task be TASK = $(p_j, d_j, \gamma_j, \sigma_j, S_{jk}, n)$, in which p_j is the time required to complete task j, d_j is the promised task completion time for j, γ_j is main start time for task j, σ_j is the weight of the task, and S_{jk} is the sequence determined in the preparation time between task j and the task execution order; that is, if task is the first task, S_{0k} is the preparation time for the task and if task j is the last task, then S_{j0} is adjusted after task j and n is the number of tasks.

The model is then $(1|r_j, \sigma_j, S_{jk}|\sum_{j=1}^{n} C_j)$. To achieve the optimization goal, the total flow for the shortest single group task scheduling problem needs to be considered which includes all tasks associated with the task sort. The optimization target is $\sum_{j=1}^{n} C_j$ [12].

4. Applications

F enterprise manufactures digital electronic products. To retain and increase market share, the company needs to design new products every year. Therefore, suppose F enterprise needs to design a new R model computer. However, F enterprises have limited technology, resources, and staff. To improve design efficiency, they need to have a "people's innovation" response, so there is a need to involve several companies in the collaborative design to fully understand customer needs and collaboratively design the R model basic computer products. Based on the WBS principle, the R model computer main chassis design tasks can be decomposed into five separate design subtasks: "BIOS Basic Input Output System," "CPU," "Memory," "bus expansion slot," and "Chipset" design; the processing time for the five design tasks is, respectively, 20, 50, 10, 30, and 40 days; however, the collaborative design task delivery is unknown. The scheduling problem of five design tasks about R model computer's chassis is based on the following two aspects.

The first calculation is based on the cost of delivery of the decision task's smallest single main task schedule.

Set A, B_i, and E_i as the linear variable functions,

$$\begin{aligned} A_{(k)} &= \lambda_1 \cdot k, \\ B_i\left(\max\left\{0, k - D_i\right\}\right) &= \lambda_2 \max\left\{0, k - D_i\right\}, \\ E_i\left(g\left(\max\left\{0, D_i - k\right\}\right)\right) &= \lambda_3 g\left(\max\left\{0, D_i - k\right\}\right). \end{aligned} \tag{6}$$

Table 1: Calculation table.

Delivery date k	Sorting task $\{x_1^*, x_2^*, \ldots, x_n^*\}_a$	Minimum cost f
10	$\{10, 20, 30, 40, 50\} \cdots \{10, 20, 30, 50, 40\}$	410
20	$\{20, 10, 30, 40, 50\} \cdots \{20, 10, 30, 50, 40\}$	420
30	$\{20, 10, 30, 40, 50\} \cdots \{20, 10, 30, 50, 40\}$	340
40	$\{30, 10, 20, 40, 50\} \cdots \{30, 10, 20, 50, 40\}$	350
50	$\{40, 10, 20, 40, 50\} \cdots \{40, 10, 20, 50, 40\}$	360
60	{30, 20, 10, 40, 50}, {30, 20, 10, 50, 40}	330
70	{40, 20, 10, 30, 50}, {40, 20, 10, 50, 30}	350
80	{40, 30, 10, 20, 50}, {40, 30, 10, 50, 20}	370
90	{40, 30, 20, 10, 50}, {40, 30, 20, 50, 10}	400
100	{50, 30, 20, 10, 40}, {50, 30, 20, 50, 40}	420
110	{50, 30, 20, 10, 40}	440
120	{50, 40, 20, 10, 30}	470
130	{50, 40, 30, 10, 20}	500
140	{50, 40, 30, 20, 10}	540
150	{50, 40, 30, 20, 10}	610

λ_1, λ_2, and λ_3, indicate the unit costs of delivering on schedule, the unit cost for early task completion, and the unit penalty cost for late task completion. C_i of i stage is

$$C_i = \lambda_2 \max\{0, k - D_i\} + \lambda_3 g\left(\max\{0, D_i - k\}\right) \quad (7)$$

with total cost

$$\begin{aligned} f &= \lambda_1 \times K + f_n\left(D_n\right), \\ \lambda_1 &= 1, \\ \lambda_2 &= 2, \\ \lambda_3 &= 100, \\ P_1 &= 20, \\ P_2 &= 50, \\ P_3 &= 30, \\ P_4 &= 10, \\ P_5 &= 40. \end{aligned} \quad (8)$$

① Possible point values for the common species are $2^n - 1 = 25 - 1 = 31$.

② There are 31 combinations for {20, 50, 30, 10, 40}, which are {20}, {50}, {30}, {10}, {40}, {20, 50}, {20, 30}, {20, 40}, {20, 10}, {50, 30}, {50, 10}, {50, 40}, {30, 10}, {30, 40}, {10, 40}, {20, 50, 30}, {20, 50, 10}, {20, 50, 40}, {20, 30, 10}, {20, 30, 40}, {20, 10, 40}, {50, 30, 10}, {50, 30, 40}, {50, 10, 40}, {30, 10, 40}, {20, 50, 30, 10}, {20, 50, 30, 40}, {20, 50, 10, 40}, {20, 30, 10, 40}, {50, 30, 10, 40}, and {20, 50, 30, 10, 40}.

③ The elements of each combination are 20, 50, 30, 10, 40, 70, 50, 60, 30, 80, 60, 90, 40, 70, 50, 100, 80, 110, 60, 90, 70, 90, 120, 100, 80, 110, 140, 120, 100, 130, and 150; remove same 50, 30, 60, 40, 70, 80, 90, 100, 110, and 120; get results 20, 50, 30, 10, 40, 70, 60, 80, 90, 100, 110, 120, 140, 120, 130, and 150.

④ Appreciating the sort, then get 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 110, 120, 130, 140, and 150.

⑤ For 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 110, 120, 130, 140, and 150, step two is calculated according to the model. The results are shown in Table 1.

Comparison in Table 1 shows that the minimum cost of 330 when $k = 60$ is the optimal delivery point. At this time, the optimal scheduling task processing time is {30, 20, 10, 40, 50} and {30, 20, 10, 50, 40} and the best task sort is T_3-T_1-T_4-T_5-T_2 or T_3-T_1-T_4-T_2-T_5.

The second application is based on a total flow dependency shortest single main task scheduling problem. Reference [12] demonstrated that, with WSPT rules, single issues can be optimally scheduled; the rule for WSPT is to work according to the descending order $\overline{w}/p_i$, where $\overline{w}$ represents the weight of heavy task j and p_i represents the task processing time.

The design process is divided into five independent tasks. Each task has a job number, a job submission time, a task processing time, a task completion time, and a task weight. The task sequence is TASK = (ID, ST, PT, FT, DL) in which ID is the task number, ST is the task submission time, PT is the task processing time, FT is task completion time, and DL is the weight.

$1|r_j, S_{jk}|\sum_{j=1}^{n} C_j$: the weighted task matrix corresponding to the expression is

$$\text{TASK} = \begin{bmatrix} 1 & 0 & 20 & 30 & 1 \\ 2 & 0 & 50 & 80 & 3 \\ 3 & 0 & 30 & 60 & 2 \\ 4 & 0 & 10 & 50 & 1 \\ 5 & 0 & 40 & 70 & 2 \end{bmatrix}. \quad (9)$$

Prepare a zero matrix.

In the above 31 possible delivery values, frequencies of 20 to 1,50 to 3,30 to 2,10 to 1 and 40 to 2 are given which are related to delivery frequency. Weights are assigned to the tasks to align processing time with delivery. Therefore, the weights for the five design tasks are {1, 3, 2, 1, 2}.

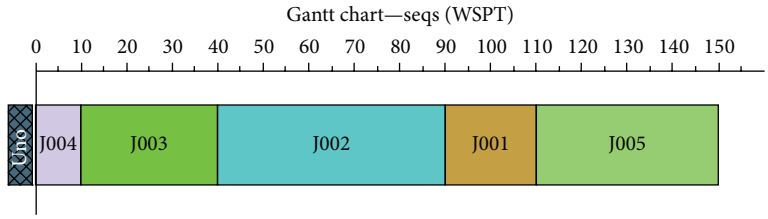


Figure 3: Lekin Sort.

From the data for the five manual design tasks corresponding to the Lekin Scheduler input system and the WSPT sorting principle, the order for the five tasks is shown in a Gantt chart (Figure 3). The program is T_4-T_3-T_2-T_1-T_5 for the weighted completion time, with a minimum target value of 770. This scenario corresponds to the shortest delivery time $k = 10$ and a minimum total cost of 410.

For the shortest delivery time $k = 10$, the R computer model ordering scheme for the five main chassis design task (T_4-T_3-T_2-T_1-T_5) costs is determined by the minimum delivery weighted completion time and minimal cost. The order for the R model main computer chassis design for the five subtasks is "bus expansion slots", "Memory", "CPU", "BIOS Basic Input Output System", and "chipset."

5. Conclusion

In this article, based on a WBS collaborative task analysis, the complex business product design process was divided into n subtasks. First, according to the minimum cost of delivery principle decision to sort the tasks for different delivery solutions, WSPT rules were used to determine a reasonable sort order to achieve a total minimum cost and a minimum weighted completion time, all of which reduced costs, improved efficiency, and enhanced core competitiveness.

In this article, the shortest delivery time was determined by sorting the collaborative product design tasks based on delivery to determine the task weight. Using the WSPT sort rules, the ordering scheme was derived to achieve total minimum cost and minimum weighted completion time. This method allows for the identification of the actual product design process parameters, which is in line with the "Made in China 2025" and "people's innovation" aims for an effective distributed collaborative design task ordering scheme. This innovation could be useful for China future industrial policies.

Competing Interests

The authors declare that they have no competing interests.

Acknowledgments

This research was funded by the National Nature Science Foundation of China (71401019), Postdoctoral Science Special Foundation of Sichuan Province, Postdoctoral Science Special Foundation of Sichuan University, Foundation of Academic Leader Training in Sichuan Province ((2015)100-6), and China Postdoctoral Science Foundation Funded Project (2016M590929).

References

[1] L. Song, Y. Yang, J. Yang, and W. Wang, "Study on task scheduling in collaborative product development," *China Mechanical Engineering*, vol. 19, no. 7, pp. 798–803, 2008.

[2] X. Liang, Y. Yang, Q. Zeng et al., "Task sequencing for distributed collaborative design based on disjunctive constraints," *Systems Engineering*, vol. 27, no. 9, pp. 65–69, 2009.

[3] M. Lu, H.-C. Lam, and F. Dai, "Resource-constrained critical path analysis based on discrete event simulation and particle swarm optimization," *Automation in Construction*, vol. 17, no. 6, pp. 670–681, 2008.

[4] W. Cui, H. Wang, and B. Yang, "The research of conflict pre-resolution in product cooperative design process," *Modular Machine Tool & Automatic Manufacturing Technique*, vol. 8, pp. 40–44, 2009.

[5] Z. Li, S.-S. Guo, and X.-X. Li, "Research on QFD method in product collaborative design," *Manufacturing Automation*, vol. 36, no. 4, pp. 125–127, 2014.

[6] J. Arsenyan and G. Büyüközkan, "An integrated fuzzy approach for information technology planning in collaborative product development," *International Journal of Production Research*, vol. 54, no. 11, pp. 3149–3169, 2016.

[7] N. R. Joglekar and D. N. Ford, "Product development resource allocation with foresight," *European Journal of Operational Research*, vol. 160, no. 1, pp. 72–87, 2005.

[8] X. Zhou, X. Li, and X. Ruan, "Study on task plan algorithm for injection product and mold collaborative design," *Journal of Mechanical Engineering*, vol. 39, no. 2, pp. 113–118, 2003.

[9] X. Meng, X. Han, and J. Cao, "Research on collaborative task schedule and management of machine tool," *Computer Engineering*, vol. 32, no. 20, pp. 238–241, 2006.

[10] X. Liang, *Customer Collaborative Product Innovation Process Management and Its Key Technology*, Sichuan University Press, Chengdu, China, 2013.

[11] B. Liu and X. Lu, "The method of due date determination and job sequence optimization for single machine scheduling," *Journal of Nanchang University (Natural Science)*, vol. 20, no. 2, pp. 142–147, 1996.

[12] M. Pinedo, *Scheduling: Theory, Algorithms and Systems*, Tsinghua University Press, Beijing, China, 2007.

Hindawi Publishing Corporation
Scientific Programming
Volume 2016, Article ID 5642856, 9 pages
http://dx.doi.org/10.1155/2016/5642856

Research Article

Research on Healthy Anomaly Detection Model Based on Deep Learning from Multiple Time-Series Physiological Signals

Kai Wang,[1] Youjin Zhao,[1] Qingyu Xiong,[2,3] Min Fan,[1] Guotan Sun,[1] Longkun Ma,[1] and Tong Liu[1]

[1]*School of Automation, Chongqing University, Chongqing, China*
[2]*Key Laboratory of Dependable Service Computing in Cyber Physical Society, Chongqing University, Ministry of Education, Chongqing, China*
[3]*School of Software Engineering, Chongqing University, Chongqing, China*

Correspondence should be addressed to Kai Wang; akyle@163.com

Received 16 June 2016; Accepted 17 August 2016

Academic Editor: X. Wang

Health is vital to every human being. To further improve its already respectable medical technology, the medical community is transitioning towards a proactive approach which anticipates and mitigates risks before getting ill. This approach requires measuring the physiological signals of human and analyzes these data at regular intervals. In this paper, we present a novel approach to apply deep learning in physiological signals analysis that allows doctor to identify latent risks. However, extracting high level information from physiological time-series data is a hard problem faced by the machine learning communities. Therefore, in this approach, we apply model based on convolutional neural network that can automatically learn features from raw physiological signals in an unsupervised manner and then based on the learned features use multivariate Gauss distribution anomaly detection method to detect anomaly data. Our experiment is shown to have a significant performance in physiological signals anomaly detection. So it is a promising tool for doctor to identify early signs of illness even if the criteria are unknown a priori.

1. Introduction

Over the years, all kinds of intelligent devices and modern life are more and more inseparable [1]. People can record various kinds of physiological time-series data through those devices at any time and any places [2, 3]. Analyzing those physiological time-series data [4] gets a lot of information about our body. Although countries invested heavily in the development of biomedicine, the incidence of various types of chronic noncommunicable diseases is increasing. So the medical community is transitioning towards a proactive direction. Different from the previous one, the approach aims at analyzing physiological time-series data and identifying potential risk of illness and mitigation measures are taken before getting ill.

Many methods can be used to help us get a better understanding of our physical condition. Machine learning [5] is a fundamental and significant research in many fields. It is widely used in industry [6], power system [7, 8], weather forecast [9], transit systems [10], computer-aided detection and diagnosis systems [11], and so on. Some companies have also launched deep learning related project via collecting and analyzing massive amounts of data and applied to anomaly detection or others applications. It is also an important assistant means for medical and has important application value in the field of medical care [12]. In this paper, we propose a lightweight approach for detecting the anomaly data by analyzing the physiological signals. Empirically, physiological signals can be obtained from biosensors in various ways.

Though anomaly detection is widely used in other fields, the problem of physiological signals anomaly detecting in the context of human-computer interaction still remains complex and largely unexplored. The anomaly detection of physiological signals is, primarily, by using machine learning techniques for learning features [13] from physiological signals and then constructing computational models [14]

of anomaly detection. The main components of the model consists of two parts: processing of input signals (learned features) and detecting anomaly data.

Feature extraction and feature selections are the key in understanding and training an anomaly detection algorithm [15, 16]. Physiological signals are usually correlated to time and space [17]; they belong to high-dimensional time-series data. Time-series data is high-dimensional and complex with unique properties that make them challenge to analyze and model. One of the major challenges in healthy anomaly detection is to extract features in multivariate physiological signals, which can be used to detect the anomaly data correctly [18]. Traditional approaches for modeling sequential data include the estimation of parameters from an assumed time-series model such as autoregressive models and Linear Dynamical Systems (LDS) and the popular Hidden Markov Model (HMM). The estimated parameters can then be used as features in a classifier to perform classification. The restricted Boltzmann machine (RBM) is a generative probabilistic model between input units (visible), x, and latent units (hidden). Several RBMs can be stacked and trained in a greedy manner to form so-called Deep Belief Networks (DBN), which are probabilistic generative neural network composed of multiple layers of restricted Boltzmann machine. DBNs are graphical models which learn to extract a deep hierarchical representation of the training data. Another model that had been used for modeling sequential data is the Recurrent Neural Network (RNN). Generally, an RNN is obtained from the feed forward network by connecting the neurons' output to their inputs. Hand-designed feature extractors require a human expert to find the suitable data manipulations that will lead to good evaluation performance. To determine important features and pick the effective ones to handle a new application may be labor-intensive and time-consuming. It inherit a number of critical limitations that make their use cumbersome in highly complex multimodal input spaces. In this paper, we present a new approach to the automatic physiological signals anomaly detection. The focus is to develop unsupervised feature learning method to learn meaningful feature representations from unlabeled physiological signals. Our hypothesis is that use of nonlinear unsupervised and multivariate Gauss distribution model methods [19–21] relying on the principle of deep learning can eliminate the limitations of the current feature extraction and feature selection in physiological signals anomaly detection. Unsupervised feature learning techniques [4, 22] are a way of learning feature representations [23] that a human expert might not be aware of and could reflect the essence the healthy states feature representations. A secondary contribution of the proposed method is using the features extracting from convolutional neural network (CNN) feed to anomaly detection model and giving the anomaly data to doctor to evaluate mitigating risks before illness. CNN's convolution and pooling can help us deal with high-dimensional data more quickly. By analyzing physiological signals, if anomaly data occurs, the special prevention should be done in advance to reduce the risk of disease.

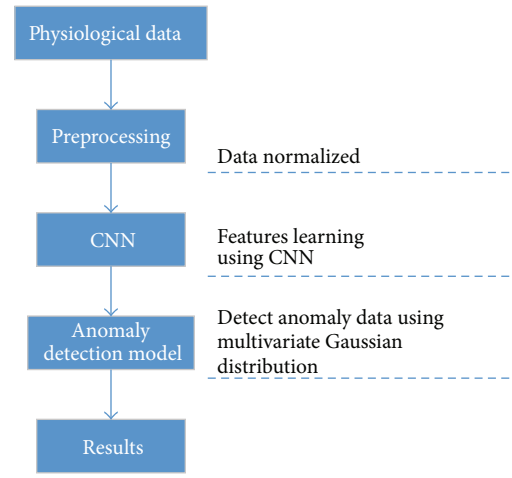


Figure 1: The block diagram of the algorithm.

2. Materials and Methods

In this section, we describe a common framework used for feature learning. For concreteness, we will focus on the application of these algorithms to learn features from physiological signals, though our approach is applicable to other forms of data as well. Furthermore, the studies cover the two key research pillars of this paper: (1) defining feature set to extract relevant bits of information from objective data signals; (2) creating models that map a feature set into multivariate Gaussian anomaly detection model to predict the anomaly physiological signals.

At a high level, our algorithm performs the following steps (see Figure 1) to learn feature representation:

(a) Dividing physiological signals into a number of segments from unlabeled training data.

(b) Applying a preprocessing stage to the segments and normalizing the raw data.

(c) Extracting high level information using an unsupervised learning algorithm.

(d) Using Gauss model to detect the anomaly physiological signals.

Now we describe the components of this pipeline and its parameters in more detail.

2.1. Feature Learning. In the context of healthy state anomaly detection, feature learning refers to the process of transforming the raw signals captured by the hardware into a set of inputs suitable for a computational evaluation of anomaly data. Usually, learning feature from one-dimensional continuous signals is simple statistical features [24, 25], such as average and standard deviation values, calculated on the time or frequency domains of the raw or the normalized signals. Physiological signals anomaly detection based on signals with more than one dimension typically boils down to physiological anomaly data detection from all kinds of physiological signals. The focus of this paper is on convolutional

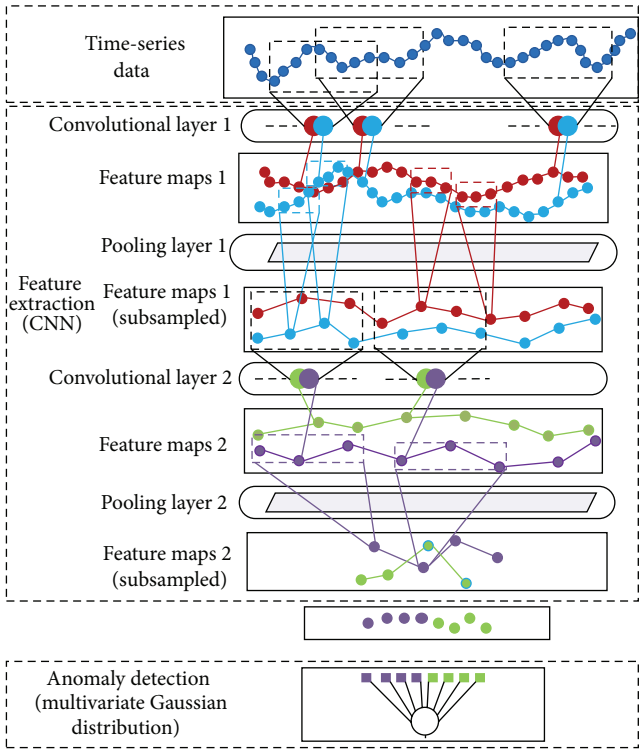


Figure 2: The structure of deep ANN architecture.

neural networks [5, 13, 26] methods that can automatically extract new features or unknown features in an unsupervised manner from those data.

Convolutional neural networks, as a popular technique, could be used in many fields such as image and video classification natural language processing, pedestrian detection, generic visual recognition, face recognition, and image recognition [10, 27]. They are very similar to ordinary neural networks. It is composed of a number of neurons that have learnable weights and biases. Each neuron receives some inputs and performs a dot product with a nonlinearity function Sigmoid. A convolutional neural network is comprised of one or more convolutional layers (often with a subsampling step) and then followed by one or more fully connected layers as in a standard multilayer neural network. The architecture of an ANN is designed to include both the feature extractor and the anomaly detection, as shown in Figure 2.

Usually, in convolutional neural networks, each layer is composed of two operations: convolution and max-pooling. At a convolution layer, the previous layer feature maps are convolved with learnable kernels and then put through the activation function to form the output feature map. Each output map may combine convolutions with multiple input maps. Supposing X denote the input signal is 2-dimensional data of size $m \times n$, where m and n are positive integers. A feature map is obtained by convolution of the input signal with a linear filter, adding a bias term and then applying a nonlinear function f. If we denote the kth feature map at a given layer as h^k, whose filters are determined by the weights w^k and bias b^k, M_j means the selection of input maps in layer $k-1$. Then the feature map h^k is obtained as follows:

$$h_j^k = f\left(\sum_{i \in M_j} X_i^{k-1} * W_{ij}^k + b^k\right). \tag{1}$$

Sensitivity computation process:

$$\delta_j^k = \beta_j^{k+1}\left(f' \circ \mathrm{up}\left(\delta_j^{k+1}\right)\right). \tag{2}$$

up$(\cdot)$ function is exactly an inverse process of the downsampling.

Compute the gradient,

$$\frac{\partial J}{\partial W_{ij}^k} = \sum \delta_j^k * p_i^{k-1},$$
$$\frac{\partial J}{\partial b_j^k} = \sum \delta_j^k. \tag{3}$$

$J(\cdot)$ is cost function. The new variable p_i^{k-1} here, which means the patch in X_i^{k-1} that was multiplied element wise by W_{ij}^k.

Gradients in subsampling layers:

$$h_j^k = f\left(\beta_j^k \,\mathrm{down}\left(h_j^{k-1}\right) + b_j^k\right). \tag{4}$$

Here, down$(\cdot)$ means the subsampling function, such as average-pooling or max-pooling.

$$\delta_j^k = f' \circ q_j^k. \tag{5}$$

q_j^k is a matrix which has the same size of δ_j^k, and its element is the convolution between all sensitivities of nodes in layer $k+1$ that have connection with the node in layer k and the weights are defined as W_j^k, and we could compute the gradients for β and b in subsampling layers as the following equations:

$$\frac{\partial J}{\partial \beta_j^k} = \sum \delta_j^k \circ \mathrm{down}\left(h_j^k\right),$$
$$\frac{\partial J}{\partial b_j} = \sum \delta_j^k. \tag{6}$$

The input of each neuron determines the size of patch. Each neuron contains a number of trainable weights equal to the number of its inputs and additional bias parameter; the output is calculated by applying an activation function to the weighted sum of the input and bias. (Activation function: tanh, ReLU, Sigmoid, and Softplus. tanh: $y = (\exp(x) - \exp(-x))/(\exp(x) + \exp(-x))$. The real worth to compress input ranges in -1 to 1, so that it is substantially 0 mean; ReLU: $y = \max(0, x)$. The network can be introduced into the sparsity, and the performance of ReLU is better than other activation functions in the case of no pretraining; Sigmoid: $y = (1 + \exp(-x))^{-1}$. The activation function in the neural network learning can push towards key features to the central and the nonkey features to both sides of the

zone; Softplus: $y = \log(1 + \exp(x))$. The activation function makes the relationship between output and input keep the nonlinear monotonic rise and fall and fault tolerance of neural network is good.) Each neuron scans sequentially the input, assessing at each patch location the similarity to the pattern encoded on weights. The consecutive outputs generated at every location of the input assemble a feature map. The output of convolution layers is the set of feature maps repeated application of a nonfunction across subregions of entire input. In the following, we provide details on CNN architectures in Figure 2. The neurons in the convolutional layer take as input a patch 8×1 on the input time-series data. The input patch is applied with a sliding-window stride of 8 along time-axis. Each of the neurons calculates h^k and then obtains features.

As soon as feature maps have been generated, a pooling layer aggregates consecutive values of the feature maps resulting from the previous convolution layer, reducing their resolution with a pooling function. The maximum or average values are the two most commonly used pooling functions providing max-pooling and average-pooling layers, respectively. It is common to periodically insert a pooling layer in between successive convolutional layer in a ConvNet architecture. Its function is to progressively reduce the spatial size of the representation to reduce the amount of parameters and computation in the network and hence to also control overfitting. In this paper, the pooling layer operates independently on every depth slice of the input and resizes it spatially, using the MAX operation. The most common form is a pooling layer with filters of size 2×2 applied with a stride of 2 downsamples in every depth slice in the input by 2 along both width and height, discarding 75% of the activation. Because physiological signals are time-series data, we designed the filters of size 2×1 with a stride of 2 downsamples along time-axis.

As we described above, a simple CNN is a sequence of layers, and each layer of a CNN transforms one volume of activation to another through a differentiable function. We use three main types of layers to build CNN architectures: convolutional layer, pooling layer, and fully connected layer (exactly as seen in regular neural networks). In our work, it is a hierarchical model that alternate convolution and pooling layers in order to process large input spaces in which a spatial or temporal relation among the input exists such as time-series data, speech, or physiological signals. Therefore, hierarchical analysis and learning architectures are the key to success in anomaly detection.

2.2. Autoencoders. We must train our network weights with a kind of unsupervised method because our training set is unlabeled. Usually physiological signals are unlabeled, so we need to take other methods to train ConvNet weights. An autoencoder neural network (see Figure 3) is an unsupervised learning algorithm that applies backpropagation, setting the target values to be equal to the inputs [28]. In this paper, we use autoencoders to train all convolution layers of our CNN. Now we have only a set of unlabeled training examples $\{x^{(1)}, x^{(2)}, x^{(3)}, \ldots\}$, where $x^{(i)} \in R^n$. An autoencoder takes an input $x \in [1, 0]^d$ and first maps it (with an encoder) to

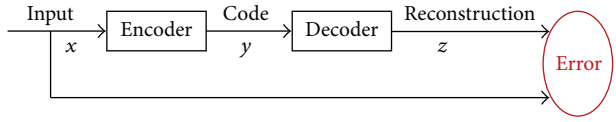


Figure 3: The block diagram of autoencoders.

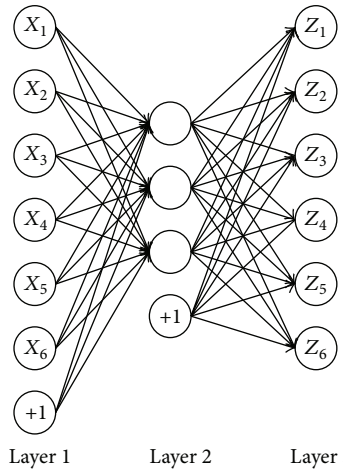


Figure 4: The structure of autoencoder.

a hidden representation $y \in [1, 0]^{d'}$ through a deterministic mapping:

$$y = s(Wx + b), \tag{7}$$

where s is a nonlinearity function Sigmoid. The latent representation y or code is then mapped back (with a decoder) into construction z of the same shape as x. The mapping happens through a similar transformation:

$$z = s\left(W' y + b'\right). \tag{8}$$

z should be seen as a prediction of x, given the code y.

An autoencoder is a model that transforms an input space into a new distributed representation by applying a deterministic parameterized function called the encoder (see Figure 4). The autoencoder learns how to map the output of the encoder into the input space, with a parameterized decoder, to have small reconstruction error on the training examples; that is, the original and corresponding decoded inputs are similar. The encoder weights (used to obtain the output representation) are also used to reconstruct the input. By defining the reconstruction error as the sum of squared differences between the inputs and the reconstructed input, then use gradient descent method such as backpropagation to train the weights of the ConvNet. The reconstruction error can be measured in many ways, depending on the appropriate distributional assumptions on the input given the code. The traditional squared error cost function is given by

$$J(xz) = \|x - z\|^2. \tag{9}$$

Here x is raw data and z is reconstruction data.

In this paper, autoencoders are used to train unsupervised CNN to transpose subsets of the raw input signals into learned features. In turn, the outputs of learned features extracted from the input layer may feed any function approximation or classifier that attempts to find a mapping between the input signal and a target output. In this paper, we use multivariate Gaussian anomaly detection model to detect the anomaly physiological signals for a user based on the learned features of his physiological signals.

2.3. Multivariate Gaussian Distribution. The multivariate Gaussian distribution is a generalization of the univariate normal to two or more variables. It is a distribution for random vectors of correlated variables, each element of which has a univariate normal distribution. A vector-valued random variable $X = [X_1, X_2, \ldots, X_n]^T$ is said to have a multivariate Gaussian distribution with mean vector $\mu \in R^n$ and covariance matrix Σ. Its probability density function is given by

$$p(x, \mu, \Sigma) = \frac{1}{(2\pi)^{n/2} |\Sigma|^{-1}} \exp\left(-\frac{1}{2}(x-\mu)^T \Sigma^{-1} (x-\mu)\right), \quad (10)$$
$$\Sigma = E\left[(X-\mu)(X-\mu)^T\right].$$

We write this as $X \sim N(\mu, \Sigma)$.

Anomaly detection is an unsupervised learning method, using density estimation to evaluate data is normal or not. The expression is as follows:

$$\text{if } p(x) \begin{cases} \leq \varepsilon & \text{anomaly} \\ > \varepsilon & \text{normal.} \end{cases} \quad (11)$$

When $p(x)$ is greater than threshold ε, the data is normal and $p(x)$ is less than threshold ε, and the data is anomaly.

3. Results and Discussion

In the experiments, we focused on evaluating the efficacy of using CNN to construct a model of physiological signals anomaly detection and we test our algorithm on eight physiological signals on DEAP dataset, a dataset for emotion analysis using EEG, physiological, and video signals. We expect that information relevant to anomaly detection can be extracted more effectively using CNN methods directly on the raw physiological signals automatic selection of features than on a set of designer-selected extracted features. The hardware and software environment in experiment are as follows: hardware environment: Intel(R) Core(TM) i3-2330 CPU @ 2.2 GHz RAM 2.00 GB; software environment: Windows 7, Python 2.7, and Matlab R2014a.

3.1. Training Models of Physiological Signals Anomaly Detection. In the approach presented here, we investigate an effective method of learning models that map signals of user physiological to detect anomaly data. In the feature extraction stage, we use a deep model composed from a multilayer convolutional neural network that transforms the raw signals into reduced set features. In the anomaly detection stage, we feed those features to an anomaly detection model which uses the multivariate Gaussian distribution to detect anomaly physiological signals (see Figure 2). Before new unlabeled time-series physiological signals enter the model, first, make the time-series physiological signals normal. Then extract features in the original network parameters. Then last, use multivariate Gaussian distribution to detect anomaly data in new unlabeled time-series physiological signals. The deep ANN architecture contains two convolutional layers, two pooling layers, and a multivariate Gaussian anomaly detection model. The first convolutional layer (patch length of 12 raw physiological signals) processes physiological signals, which is then propagated forward to a maximum-pooling layer (window length of 2 features). The second convolutional layer (patch length of 5 subsampled features) processes the subsampled feature maps and the resulting feature maps of the second pooling layer (window length of 2 features). The final subsampled feature maps form the outputs of the CNN which provides a number of learning features feeding the input of Gaussian anomaly detection model. Our hypothesis is that the automation of feature extraction via deep learning will yield anomaly physiological signals of higher predictive power, which, in turn, deliver evaluation models of higher accuracy.

Before feeding the raw data to CNN, in order to cause the reconstruction error convergence, we normalized the raw data use:

$$X\_\text{norm} = \frac{X - \mu}{\text{std}(X)}. \quad (12)$$

μ and std, respectively, represent the average value of raw data and the standard deviation of raw data. Then feed the normalized data X_norm to CNN learning features of raw data. By analyzing the reconstruction error Jw between the inputs data X and the reconstruction inputs Z and the number of iterations, we can obtain preferable learned features. In theory, the greater the iterations number, the smaller the reconstruction error Jw.

With the increase of the number of iterations, the cost function Jw is tending towards zeros and then keeping stability, we can draw a conclusion that the reconstruction data are nearly the same. Therefore, the learned features from raw data are effective and we can view those features as high level expression of the raw data. As Figure 5 shows, we take 5000 iterations, the cost function Jw is tending to zeros, and we can get the weights of the ConvNet and high level features of the eight physiological signals.

As soon as features have been learned, feed them to Gaussian mixtures models to detect the anomaly features and get the coordinate. We consider the data and the learned features to be subject to Gauss distribution.

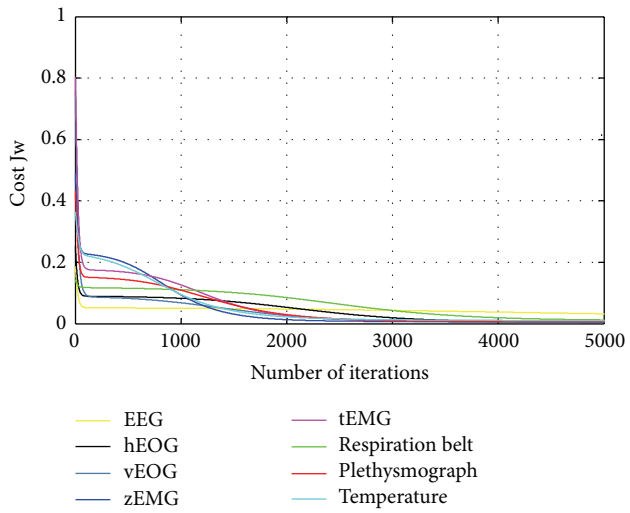


Figure 5: Learned features of the best-performing convolutional neural networks. Lines plotted are the values of cost function for each physiological signal. The x-axis displays the number of iterations.

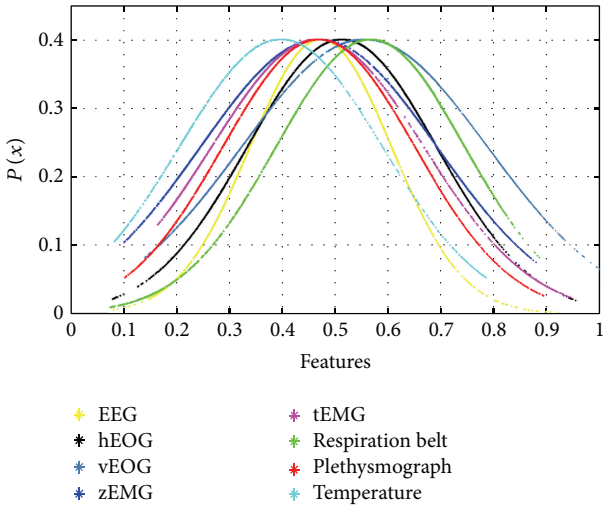


Figure 6: Gauss distribution of features.

As Figure 6 shows, the features of the eight physiological signals are subject to Gauss distribution signals.

3.2. Results. Then we choose a fine threshold ε to get the features which is $p(x_i) \leq \varepsilon$ and then get the coordinate and the corresponding raw data. All anomaly detection algorithms can be set at different threshold ε, which may make results correspond to reality. If the ratio of anomaly data is 1% and threshold $\varepsilon = 0.2$ we can get the anomaly features and the coordinate of raw physiological signals. In addition, if the ratio of anomaly data is 5% and threshold $\varepsilon = 0.3$ we can get other anomaly features and the coordinate of raw physiological.

In this paper, we use a series of detection thresholds ε 0.2, 0.23, 0.25, and 0.26, and the ratios of anomaly data are 1%, 2%, 3%, and 5%, respectively. The percentage controls the number of physiological signals to be considered anomaly ranked by severity. We get four sets of anomaly data as Figure 7 shows, (a) and (b), (c) and (d), (e) and (f), and (g) and (h). Finally, it is easy to obtain the anomaly raw physiological signals according to anomaly features point coordinate and doctor can quickly analyze those anomaly physiological signals to help users understand the healthy states at present.

3.3. Discussion. The testing on dataset showed that the method can detect anomaly physiological signals and some may exhibit early signs of illness. Therefore, the method could be a tool to help doctor identify the underlying disease.

There is no "medical instance" for performance evaluation or a benchmark dataset in which every physiological signal is labeled "normal or anomaly" in definitive terms. In this paper, the detection threshold ε is artificially chosen to evaluate physiological signals normally or anomaly. Therefore, one limitation of this paper is that we do not provide an evaluation of the detection accuracy in a traditional sense, such as false negatives and false positives. A new method is needed to evaluate the performance of an anomaly detection method that does not rely on preexisting criteria but is capable of detecting unknown issues.

4. Conclusions

This paper introduces the application of deep learning to the construction of an anomaly detection model built on physiological signals manifestations of anomaly data. To detect the anomaly data, key point is learning effective features in the raw feature. The algorithm proposed employs a number of convolution layers that learn to extract relevant features from the input signal and then feed those features to multivariate Gaussian distribution to detect anomaly features. The algorithm was tested on eight physiological signals. Result, in general, suggests that algorithm is highly efficient to learn high level features from raw physiological signals and multivariate Gaussian distribution anomaly detection.

Feature learning directions in our algorithm can be outlined as follows. First, a wide range of datasets each of which has different characteristics from different part of body should be employed in order to demonstrate the effectiveness of the method. Second, since the database is a key in our method, the collecting of data is still going on. We cannot really evaluate the results because of the database being unlabeled. Therefore, collecting of some labeled physiological signals is a way to enhance the performance of the algorithm. Further research is needed to comprehensively evaluate the performance of the algorithm in detecting unknown issues. Based on this paper, future research can be from the following aspects: first, collecting of some labeled physiological signals to comprehensively evaluate and improve the performance of the algorithm; second, because of lack of comparative tests, it is necessary to do comparative tests with other algorithms to verify the performance of the algorithm.

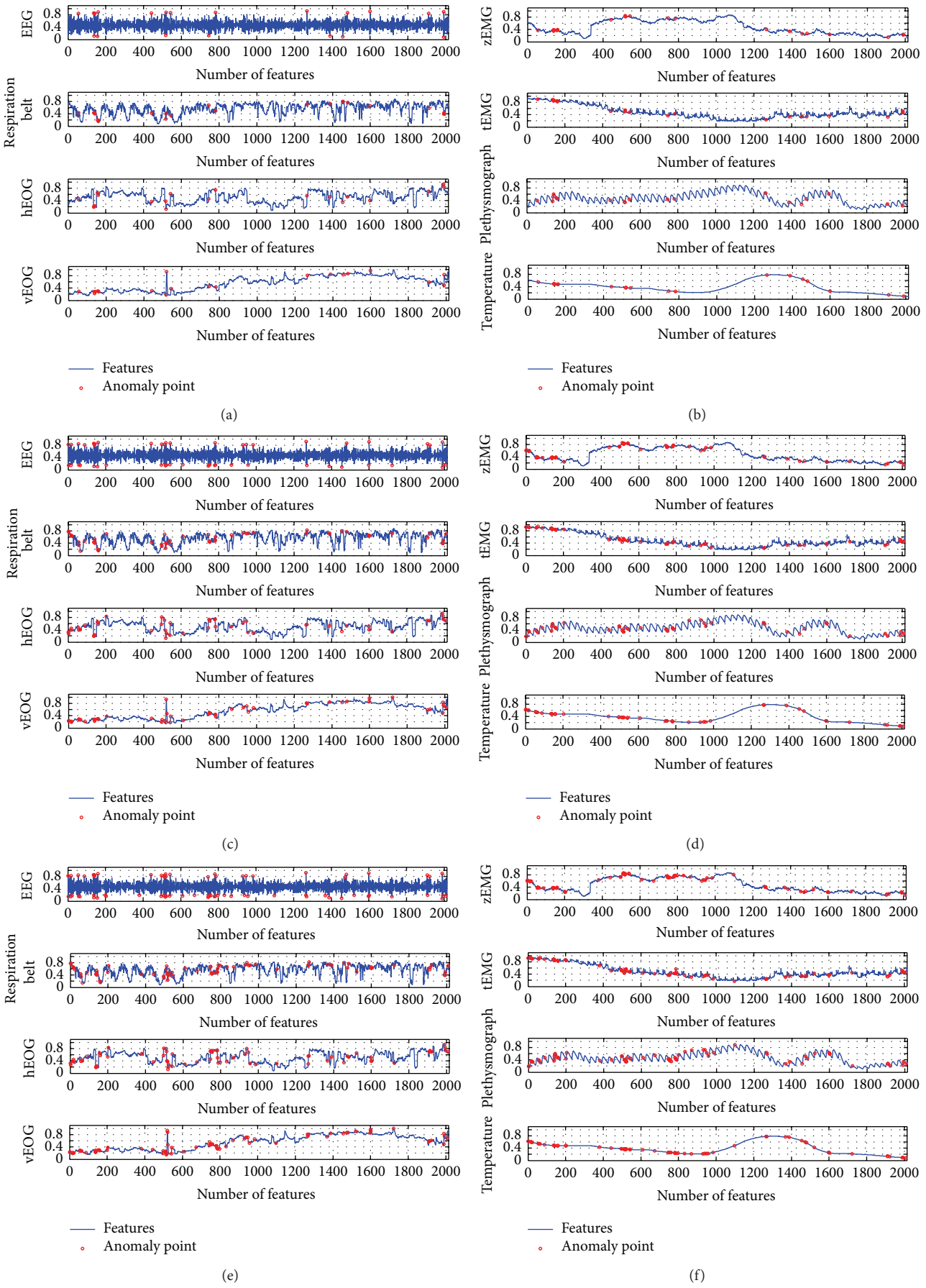


Figure 7: Continued.

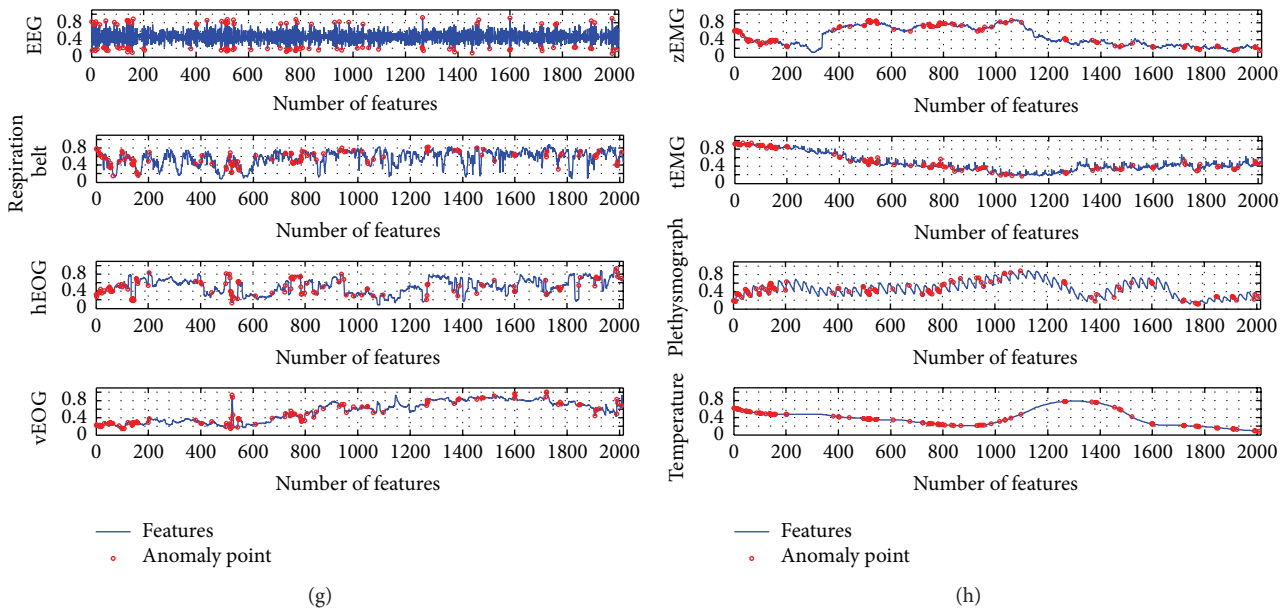


Figure 7: Anomaly features detection with different thresholds.

Competing Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

This work is supported by Major State Basic Research Development Program of China (973 Program) under Project no. 2013CB328903 and by State Scholarship Fund of China (CSC) under File no. 201306055018. The physiological signals are provided by DEAP dataset, and the URL is http://www.eecs.qmul.ac.uk/mmv/datasets/deap/.

References

[1] A. Szczepański and K. Saeed, "A mobile device system for early warning of ECG anomalies," *Sensors*, vol. 14, no. 6, pp. 11031–11044, 2014.

[2] H. Zhao, R. Xu, M. Shu, and J. Hu, "Physiological-signal-based key negotiation protocols for body sensor networks: a survey," *Simulation Modelling Practice and Theory*, vol. 65, pp. 32–44, 2016.

[3] T. Li, Z. Wang, D. Zhao, and L. Huang, "Development of a portable multi-channel system for plant physiological signal recording," *Information Processing in Agriculture*, vol. 3, no. 2, pp. 124–132, 2016.

[4] M. Längkvist, L. Karlsson, and A. Loutfi, "A review of unsupervised feature learning and deep learning for time-series modeling," *Pattern Recognition Letters*, vol. 42, no. 1, pp. 11–24, 2014.

[5] J. Schmidhuber, "Deep learning in neural networks: an overview," *Neural Networks*, vol. 61, pp. 85–117, 2015.

[6] M. Zhang, X. Liu, and Z. Zhang, "A soft sensor for industrial melt index prediction based on evolutionary extreme learning machine," *Chinese Journal of Chemical Engineering*, vol. 24, pp. 1013–1019, 2016.

[7] J. Heinermann and O. Kramer, "Machine learning ensembles for wind power prediction," *Renewable Energy*, vol. 89, pp. 671–679, 2016.

[8] P. Tang, D. Chen, and Y. Hou, "Entropy method combined with extreme learning machine method for the short-term photovoltaic power generation forecasting," *Chaos, Solitons & Fractals*, vol. 89, pp. 243–248, 2016.

[9] M. Gocic, D. Petković, S. Shamshirband, and A. Kamsin, "Comparative analysis of reference evapotranspiration equations modelling by extreme learning machine," *Computers and Electronics in Agriculture*, vol. 127, pp. 56–63, 2016.

[10] D. Cireşan, U. Meier, J. Masci, and J. Schmidhuber, "Multi-column deep neural network for traffic sign classification," *Neural Networks*, vol. 32, pp. 333–338, 2012.

[11] J. Y. Choi, D. H. Kim, K. N. Plataniotis, and Y. M. Ro, "Classifier ensemble generation and selection with multiple feature representations for classification applications in computer-aided detection and diagnosis on mammography," *Expert Systems with Applications*, vol. 46, pp. 106–121, 2016.

[12] W. Huang, S. Zeng, M. Wan, and G. Chen, "Medical media analytics via ranking and big learning: a multi-modality image-based disease severity prediction study," *Neurocomputing*, vol. 204, pp. 125–134, 2016.

[13] I. Mrazova and M. Kukacka, "Can deep neural networks discover meaningful pattern features?" *Procedia Computer Science*, vol. 12, pp. 194–199, 2012.

[14] S. Aras and T. D. Kocakoç, "A new model selection strategy in time series forecasting with artificial neural networks: IHTS," *Neurocomputing*, vol. 174, pp. 974–987, 2016.

[15] I. J. Goodfellow, D. Erhan, P. L. Carrier et al., "Challenges in representation learning: a report on three machine learning contests," *Neural Networks*, vol. 64, pp. 59–63, 2015.

[16] K. Kim, H. Lin, J. Y. Choi, and K. Choi, "A design framework for hierarchical ensemble of multiple feature extractors and multiple classifiers," *Pattern Recognition*, vol. 52, pp. 1–16, 2016.

[17] A. L. V. Coelho and C. A. M. Lima, "Assessing fractal dimension methods as feature extractors for EMG signal classification," *Engineering Applications of Artificial Intelligence*, vol. 36, pp. 81–98, 2014.

[18] S. M. Erfani, S. Rajasegarar, S. Karunasekera, and C. Leckie, "High-dimensional and large-scale anomaly detection using a linear one-class SVM with deep learning," *Pattern Recognition*, vol. 58, pp. 121–134, 2016.

[19] A. Aiyer, K. Pyun, Y.-Z. Huang, D. B. O'Brien, and R. M. Gray, "Lloyd clustering of Gauss mixture models for image compression and classification," *Signal Processing: Image Communication*, vol. 20, no. 5, pp. 459–485, 2005.

[20] B. Konomi, G. Karagiannis, and G. Lin, "On the Bayesian treed multivariate Gaussian process with linear model of coregionalization," *Journal of Statistical Planning and Inference*, vol. 157/158, pp. 1–15, 2015.

[21] A. O'Hagan, T. B. Murphy, I. C. Gormley, P. D. McNicholas, and D. Karlis, "Clustering with the multivariate normal inverse Gaussian distribution," *Computational Statistics & Data Analysis*, vol. 93, pp. 18–30, 2016.

[22] M. Siddula, F. Dai, Y. Ye, and J. Fan, "Unsupervised feature learning for objects of interest detection in cluttered construction roof site images," *Procedia Engineering*, vol. 145, pp. 428–435, 2016.

[23] B. Zoefel and R. VanRullen, "EEG oscillations entrain their phase to high-level features of speech sound," *NeuroImage*, vol. 124, pp. 16–23, 2016.

[24] R. Jegadeeshwaran and V. Sugumaran, "Fault diagnosis of automobile hydraulic brake system using statistical features and support vector machines," *Mechanical Systems and Signal Processing*, vol. 52-53, no. 1, pp. 436–446, 2015.

[25] N. Sengupta, M. Sahidullah, and G. Saha, "Lung sound classification using cepstral-based statistical features," *Computers in Biology and Medicine*, vol. 75, pp. 118–129, 2016.

[26] K. Audhkhasi, O. Osoba, and B. Kosko, "Noise-enhanced convolutional neural networks," *Neural Networks*, vol. 78, pp. 15–23, 2016.

[27] X.-X. Niu and C. Y. Suen, "A novel hybrid CNN-SVM classifier for recognizing handwritten digits," *Pattern Recognition*, vol. 45, no. 4, pp. 1318–1325, 2012.

[28] Y. Wang, Z. Xie, K. Xu, Y. Dou, and Y. Lei, "An efficient and effective convolutional auto-encoder extreme learning machine network for 3d feature learning," *Neurocomputing*, vol. 174, pp. 988–998, 2016.

Hindawi Publishing Corporation
Scientific Programming
Volume 2016, Article ID 4180817, 10 pages
http://dx.doi.org/10.1155/2016/4180817

Research Article

Optimal Control for Bufferbloat Queue Management Using Indirect Method with Parametric Optimization

Amr Radwan,[1] Hoang-Linh To,[1] and Won-Joo Hwang[2]

[1]*Department of Information and Communication System, Inje University, Gimhae, Gyeongnam 50834, Republic of Korea*
[2]*Department of Information and Communications Engineering, HSV-TRC, Inje University, Gimhae, Gyeongnam 50834, Republic of Korea*

Correspondence should be addressed to Won-Joo Hwang; ichwang@inje.ac.kr

Received 20 May 2016; Accepted 31 July 2016

Academic Editor: Junhu Ruan

Because memory buffers become larger and cheaper, they have been put into network devices to reduce the number of loss packets and improve network performance. However, the consequences of large buffers are long queues at network bottlenecks and throughput saturation, which has been recently noticed in research community as bufferbloat phenomenon. To address such issues, in this article, we design a forward-backward optimal control queue algorithm based on an indirect approach with parametric optimization. The cost function which we want to minimize represents a trade-off between queue length and packet loss rate performance. Through the integration of an indirect approach with parametric optimization, our proposal has advantages of scalability and accuracy compared to direct approaches, while still maintaining good throughput and shorter queue length than several existing queue management algorithms. All numerical analysis, simulation in ns-2, and experiment results are provided to solidify the efficiency of our proposal. In detailed comparisons to other conventional algorithms, the proposed procedure can run much faster than direct collocation methods while maintaining a desired short queue (≈40 packets in simulation and 80 (ms) in experiment test).

1. Introduction

Nowadays, modern computer networks are incredibly complex and we rely on them to transport huge quantities of data across the globe in seconds. Although this works well, there are some foreseen issues that need to be tackled. As bandwidth-heavy applications such as peer-to-peer networks and websites relying on user-generated content have become more prevalent, especially the relatively slow residential broadband links have been used at full capacity, and interruptions in connectivity have become more common [1]. Recent research has shown that the culprit is the buffers built into network equipment [2]. Accordingly, *bufferbloat* term has been used to describe related issues whenever these buffers misbehave to produce unnecessary latency [3].

In order to efficiently manage queues which are generated due to the *bufferbloat* phenomenon, active queue management (AQM) algorithms have been recommended to use in network equipment [4]. Most of the existing approaches to AQM design exploit feedback control theory with the linearized TCP core model that was proposed by Hollot et al. [5]. The well-known AQMs that monitor the average queue size and drops (or marks) packets based on statistical probabilities are Random Early Detection (RED) [6] or Random Early Marking (REM) [7]. If the buffer is empty, all incoming packets are accepted. When the buffer is full, the probability has reached 1 and all incoming packets are dropped. When queue is growing, the probability grows according to a piecewise linear function and RED (or REM) drops (or marks) packets using the updated probability. The main drawback of RED or REM is that it is sensitive to network parameter changes and requires careful tuning of its parameters in order to provide optimal performance in any scenarios. Recently, in [8], a Proportional Integral Enhanced (PIE) controller as a lightweight AQM is proposed, without the need of per-packet extra processing. Such PI-type controllers are known to provide queue control with zero offset (the mean queue length converges to the target value)

but are consequently less stable and slower reacting. There have been several other control-theoretic solutions based on feedback fluid-flow model proposed in [9–12] to improve stable and robust control mechanisms, but none of them tackle the issue of system optimality in terms of minimizing both queue length and packet dropping rate performance and searching for an optimal control trajectory.

To address such issues, a promising design direction is to reformulate the network queuing problem as an *optimal* control queue (OCQ) problem [13], where the main state variable is queue length and the control variable is the actual input rate to form the queue [14] or packet dropping rate. Then one of the approaches to solve OCQ is that we can priorly discretize the governing ordinary differential equations (ODEs) and the integral terms in the cost function or constraint functions and thereby replace the infinite dimensional optimal control problem with a large nonlinear optimization problem (NOP). This is known as the *direct method* in the literature for solving OCQ. This approach is typically easier to use, especially for OCQ with a state equality or inequality constraints. The main difference among direct approaches is how to handle the constraints corresponding to the system dynamics. The three most common direct approaches are direct single shooting, direct multiple shooting, and direct collocation. Direct methods have been used with references in [15–18].

Alternatively, one can first form the optimality conditions, using the calculus of variations and Pontryagin minimum principle, and then solve the resulting boundary value problem. This is known as the *indirect method* for solving OCQ. The references present just a small sample of the work that discusses or applies indirect methods for the solution of optimal control problems [15, 16, 18, 19]. In rare cases, the solution can be obtained in closed form from the optimality conditions, but, in general, approximation methods are used to solve the problem numerically. The optimality conditions of these problems generally take the form of differential algebraic equations (DAEs) with boundary conditions (BCs). The approximate solution to the OCQ can be obtained by using a boundary value problem (BVP) solver. Perhaps the most popular methods are multiple shooting and collocation. More recently, a combination of direct and indirect methods was proposed leading to hybrid methods [16].

In this paper, we firstly derive methods to solve OCQ using both direct and indirect approaches. We show how to apply the direct collocation approach to solve OCQ problem which can be solved in popular optimization solvers such as JModelica [20] and GAlib [21]. Indirect method with forward-backward for optimal control queue algorithm (FB-OCQ) is designed as an alternative to tackle it as well. Our key difference from existing works is that we provide a novel method to update the control step δu by solving a parametric optimization subproblem. This method is scalable which means that we can expand for larger problems with more variables as well. The numerical results for both direct and indirect approaches are discussed in Section 5.1 to emphasize our choice of integrating an indirect method with parametric optimization for active queue management design, which is demonstrated more efficiently (i.e., faster reaction and more stable) than direct methods. Finally, we evaluate the proposed algorithm using network simulator ns-2 and compare it to other AQMs including RED, REM, and PI. The dropping feature of the proposed algorithm makes the average queue length shortened and stabilized compared with others, while throughput is not reduced so much.

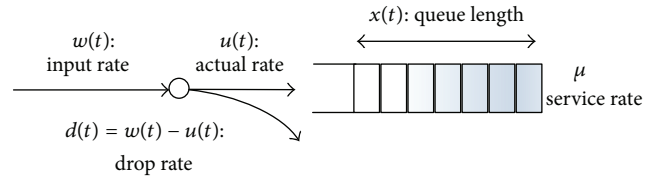


Figure 1: OCQ model.

Our Contributions

(i) We present both indirect and direct methods for an optimal control problem which is applied in queue management field (Section 3). The step update δu of control variable is calculated by solving a subproblem of parametric optimization with the advantage of scalability. Numerical results show that all of them bring nearly similar results, but indirect method (FB-OCQ) is much faster than other solvers and gives the best cost function value (Section 5.1).

(ii) We evaluate FB-OCQ in simulation and show that a desired small queue length value at 40 packets can be obtained. Nevertheless, we cannot avoid the trade-off between queue length and throughput. FB-OCQ's throughput is slightly smaller than RED's one (0.075 versus 0.082 Mbps); however, it can be an acceptable value when compared to REM's and PI's (Section 5.2).

(iii) We implement FB-OCQ in Linux kernel (Ubuntu 16.04) and test it in the worst case using Realtime Response Under Load (RRUL) test suite. The experiment result shows that our algorithm brings low latency ping value compared with the other existing algorithms in Linux kernel (DropTail and RED).

2. Problem Formulation

We consider an optimal control model of queue management problem, named (OCQ), in [14] (Figure 1). The main idea is to minimize the cost function which implies a trade-off between queue length and dropping rate, subject to a dynamic constraint of queue length along time t_0 to t_f as follows:

$$\begin{aligned} \underset{x}{\text{minimize}} \quad & J = \int_{t=0}^{t_f} L\left(x\left(t\right), u\left(t\right), t\right) \\ \text{subject to} \quad & \dot{x} = f\left(x\left(t\right), u\left(t\right), t\right), \\ & 0 \le u\left(t\right) \le w\left(t\right), \end{aligned} \tag{OCQ}$$

where $L(x,u,t) = x(t) + R(w - u(t))$ is cost function; $f(x(t),u(t),t) = u(t) - \mu x/(a + x)$; R is weight on dropping rate; t_f is final time; μ is service rate (bandwidth capacity); and a is parameter for different types of queuing model; for example, when $a = 1$, we obtain an M/M/1 queue.

3. Numerical Methods for OCQ

3.1. Direct Method: Collocation Method. In this section, we present how to apply the direct collocation [18] for OCQ. In this method, we discretize the time interval $t_0 = t_1 < t_2 < t_3 < \cdots < t_k = t_f$ into N elements. The state and the total number of packets and control variables at each node are $x_j = x(t_j)$ and $w_j = w(t_j)$ and $u_j = u(t_j)$, such that the state, control, and packets variables at the nodes are defined as nonlinear programming (NLP) variables:

$$Y = [x(t_1), \ldots, x(t_k), u(t_1), \ldots, u(t_k), w(t_1), \ldots, w(t_k)]. \quad (1)$$

The controls are chosen as piecewise linear interpolating functions between $u(t_j)$ and $u(t_{j+1})$ for $t_j \le t \le t_{j+1}$ as follows:

$$u_{\text{app}}(t) = u(t_j) + \frac{t - t_j}{h_j}\left[u(t_{j+1}) - u(t_j)\right], \quad h_j = t_{j+1} - t_j. \quad (2)$$

The value of the control variables at the center $t_{j+1/2}$ is given by

$$u(t_{j+1/2}) = \frac{u(t_j) + u(t_{j+1})}{2}, \quad j = 1, \ldots, k-1. \quad (3)$$

The piecewise linear interpolation is used to prepare for the possibility of discontinuous solutions in control. Similarly, we can derive the approximate of the total number of the packets $w_{\text{app}}(t)$ and $w(t_{j+1/2})$ as above. The state variable $x(t)$ is approximated by a continuously differentiable and piecewise Hermite-Simpson cubic polynomial between $x(t_j)$ and $x(t_{j+1})$ on the interval $t_j \le t \le t_{j+1}$ of length h_j:

$$\begin{aligned} x_{\text{app}}(t) &= \sum_{r=0}^{3} c_r^j \frac{t - t_j}{h_j}, \\ c_0^j &= x(t_j), \\ c_1^j &= h_j f_j, \\ c_2^j &= -3x(t_j) - 2h_j f_j + 3x(t_{j+1}) - (h_j) f_{j+1}, \\ c_3^j &= 2x(t_j) + h_j f_j - 2x(t_{j+1}) + (h_j) f_{j+1}, \end{aligned} \quad (4)$$

where

$$f_j = f(x(t_j), u(t_j), w(t_j), t_j) \quad t_j \le t \le t_{j+1},\ j = 1, \ldots, k-1. \quad (5)$$

The value of the state variables at the center point $t_{j+1/2}$ of the cubic approximation is

$$x(t_{j+1/2}) = \frac{x(t_j) + x(t_{j+1})}{2} + \frac{t_{j+1} + t_j}{8}(f_j + f_{j+1}), \quad j = 1, \ldots, k-1, \quad (6)$$

and the derivative is

$$\dot{x}(t_{j+1/2}) = \frac{3(x_j + x_{j+1})}{2(t_{j+1} + t_j)} - \frac{1}{4}(f_j + f_{j+1}), \quad j = 1, \ldots, k-1. \quad (7)$$

In addition, the chosen interpolating polynomial for the state and control variables must satisfy the midpoint conditions for the differential equations as follows:

$$f(x_{\text{app}}(t_{j+1/2}), u_{\text{app}}(t_{j+1/2}), w_{\text{app}}(t_{j+1/2}), t_{j+1/2}) - \dot{x}_{\text{app}}(t_{j+1/2}) = 0. \quad (8)$$

Equations (OCQ) can now be defined as a discretized problem as follows:

$$\min \quad f(Y) \quad (9)$$

$$\begin{aligned} \text{subject to} \quad & f(x_{\text{app}}(t), u_{\text{app}}(t), w_{\text{app}}(t), t) - \dot{x}_{\text{app}} = 0, \\ & x_{\text{app}}(t_1) - x_1 = 0, \\ & 0 \le u_{\text{app}}(t) \le w_{\text{app}}(t), \end{aligned} \quad (10)$$

where x_{app}, u_{app}, and w_{app} are the approximations of the state, the control, and the total number of packets, constituting Y in (9). This above discretization problem (9)-(10) can be solved using the following:

(i) JModelica, which is a package for simulation and optimization of Modelica models (for more details see [20]);

(ii) GAlib, which is C++ library of genetic algorithm (for more details see [21]).

3.2. Indirect Method: Forward-Backward Sweeping. We solve OCQ by using *indirect method* approach:

(i) forming optimality conditions;

(ii) solving BVP by First-Order Sweeping algorithm.

Let $\overline{x}(t)$ be the adjoint variable. At time t, let $u^*(t)$ denote the optimum controls and let $x^*(t)$ and $\overline{x}^*(t)$ denote the state and adjoint evaluated at the optimum. Using Pontryagin minimum principle we get the following equations.

Hamiltonian Function

$$H(x(t), u(t), \overline{x}(t)) = x(t) + R(w(t) - u(t)) + \overline{x}(t)\left(u(t) - \frac{\mu x(t)}{a + x(t)}\right). \quad (11)$$

Adjoint Equations

$$\dot{\overline{x}}(t) = -1 + u(t)\frac{a\mu}{(a + x(t))^2}. \quad (12)$$

Transversality Condition

$$\overline{x}\left(t_f\right) = 0. \tag{13}$$

Hamiltonian Minimization Condition. Derivative of the Hamiltonian is evaluated to zero at interior points; hence

$$\begin{aligned} &H_u\left(x\left(t\right), u\left(t\right), \overline{x}\left(t\right), \lambda\left(t\right), t\right) \\ &\quad = -R + \overline{x}\left(t\right) - \lambda_1\left(t\right) + \lambda_2\left(t\right) = 0. \end{aligned} \tag{14}$$

If $u^*(t)$ is optimal in (OCQ), then it must satisfy the minimum condition:

$$H\left(x^*, u^*, \overline{x}^*, t\right) = \min_{v \in U} \quad H\left(x, v, \overline{x}, t\right) \quad \forall t \in N. \tag{15}$$

Furthermore, we assume the following:

(H1) the Hamiltonian is strictly convex with respect to control variable u;

(H2) $\widetilde{u}(t) = \arg\min_{u \in U} H(x, u, \overline{x})$, $t \in [t_0, t_f]$ is continuous on $[t_0, t_f]$;

(H3) $u \in C^r[t_0, t_f]$.

Note that $u(t)$ is determined in a unique way for each t since U is convex and H is convex. Now we consider the problem of minimizing the Hamiltonian with respect to p:

$$\min_{u \in U} \quad H\left(x, u, \overline{x}\right) \quad \text{for each } t \in \left[t_0, t_f\right]. \tag{16}$$

Usually in the literature, u is found explicitly as a function $u = u(\overline{x}, x, t)$ and, after substituting it into the system, the problem reduces to the boundary value problem.

Theorem 1. *Assume that conditions (H1)–(H3) hold and problem has an optimal solution (u^*, x^*). Then for a given ϵ there exists a finite discretization*

$$t_0 = \tau_0 < \tau_1 < \cdots < \tau_i < \cdots < \tau_N = t_f \tag{17}$$

and an approximate solution $\widetilde{u}(t), t \in [t_0, t_f]$, such that

$$\left\| u^*\left(t_i\right) - \widetilde{u}\left(t_i\right) \right\| < \epsilon, \quad i = 1, 2, \ldots, N. \tag{18}$$

Proof. Let u^* be an optimal solution of problem. Then (p^*, x^*) satisfies the conditions:

$$\begin{aligned} \dot{x}_i^* &= \frac{\partial H\left(x^*, u^*, \overline{x}^*\right)}{\partial u_i}, \\ \dot{\overline{x}}_i^* &= -\frac{\partial H\left(\left(x^*, u^*, \overline{x}^*\right)\right)}{\partial x_i}, \\ &\qquad i = 1, 2, \ldots, N, \end{aligned} \tag{19}$$

where

$$\begin{aligned} H\left(x, u, \overline{x}\right) &= \sum_{i=1}^{n} \overline{x}_i\left(t\right) f_i\left(x, u\right) - f_0\left(x, u\right), \\ &\qquad t \in \left[t_0, t_f\right] \end{aligned} \tag{20}$$

and (u^*, x^*) satisfies the minimum principle:

$$H\left(x^*, u^*, \overline{x}^*\right) = \min_{u \in U} \quad H\left(x^*, u, \overline{x}^*\right), \quad t \in \left[t_0, t_f\right]. \tag{21}$$

□

(a) First-Order Sweeping Method. We linearize the Hamiltonian around a reference solution $u(t)$ and obtain the following equation for variation $\delta u(t)$:

$$0 = H_u{}^{\top} + H_{uu}\delta u\left(t\right). \tag{22}$$

With the strong Legendre-Clebsch condition, one can approximate the above equation as

$$\delta u\left(t\right) = \arg \min_{\delta u \in \{U - u\}} \quad H_{\text{aug}}\left(x, u + \delta u, \overline{x}, t\right), \tag{23}$$

where

$$H_{\text{aug}}\left(x, u + \delta u, \overline{x}, t\right) = H\left(x, u + \delta u, \overline{x}, t\right) + \beta \left\|\delta u\right\|^2 \tag{24}$$

is the augmented Hamiltonian function and $\beta\|\delta u\|^2$ is a penalty term. This is really parametric optimization, with $\beta = 0$ initially; if U is a convex set and H is a convex function with respect to u, then δu is a descent direction. If $u(t) + \delta u(t)$ does not yield a reduction, then we set $\beta = 1$ and then double it repeatedly until the objective is really reduced.

Remark 2. An alternative way to ensure descent is to apply Backtracking Line-Search Procedure.

(b) Parametric Optimization to OCQ. Now we consider problem (23) as one parametric minimization problem. Since $H_{\text{aug}}(\overline{x}^*, x^*, u^*, \delta u, t)$ is twice differentiable in δu and assumptions (H1)–(H3) hold, we can apply Theorem 1 to the problem. Then as a result, it generates a discretization, $t_0 = \tau_0 < \tau_1 < \cdots < \tau_i < \cdots < \tau_N = t_f$, and corresponding points $\widetilde{u}_i = \widetilde{u}(t_i)$ such that

$$\left\| \delta u^*\left(t_i\right) - \delta\widetilde{u}\left(t_i\right) \right\| < \epsilon, \quad i = 1, 2, \ldots, N, \tag{25}$$

which proves the assertion.

Remark 3. Parametric optimization also can be applied in finding nominal optimal control given in [22]. It is easy to see that, at each iteration k, the Hamiltonian function is a scalar function of $u \in U \subset R^r$ and $t \in T = [t_0, t_f]$; that is,

$$G_k\left(\delta u, t\right) = H_{\text{aug}}\left(\overline{x}^k\left(t\right), x^k\left(t\right), \delta u, t\right). \tag{26}$$

The latter states that $\delta\widehat{u}^k(t)$ must be a minimizer of the following problem:

$$\min_{\delta u \in U} \quad G_k\left(\delta u, t\right), \quad t \in T \tag{27}$$

which is a problem of parametric optimization as formulated in various papers from [23], where the independent variable t is now considered as unknown parameter $t \in T = [t_0, t_f]$. We can also consider a case when the set of admissible controls is time-varying; that is, $U = U(t)$, $t \in T = [t_0, t_f]$. In this case, a general theory of parametric optimization is also applicable for finding the nominal optimal controls.

Let $(x^*, \delta u^*)$ be an optimal process in problem. Introduce the function $f : U \times R \to R$:

$$f(\delta u, t) = -H(x^*, \delta u, \overline{x}^*, t), \quad t \in [t_0, t_f], \tag{28}$$

where a parametric optimization problem is defined as

$$\begin{aligned} &\min_{\delta u \in U} \quad f(\delta u, t), \quad t \in [t_0, t_f], \\ &U = \{\delta u \in R^r : g_i(\delta u) \le 0,\ i \in J\}, \\ &J = \{1, 2, \ldots, s\}. \end{aligned} \tag{29}$$

The KKT conditions for the problem state that

$$\begin{aligned} &D_{\delta u} f(\delta u, t) + \sum_{j \in J} \mu_i D_\delta u g_i(\delta u, t) = 0, \\ &g_i(\delta u, t)\, \mu_i \le 0, \quad \mu_i \ge 0,\ i \in J, \\ &\mu_i g_i(\delta u, t) = 0, \\ &i \in J,\ t \in [t_0, t_f], \end{aligned} \tag{30}$$

where

$$D_{\delta u} f(\delta u, t) = \left(\frac{\partial f(\delta u, t)}{\partial \delta u_1}, \ldots, \frac{\partial f(\delta u, t)}{\partial \delta u_r} \right). \tag{31}$$

Consider the auxiliary parametric optimization problem:

$$\begin{aligned} &\min \quad f(\delta u, t), \quad t \in [t_0, t_f] \\ &\text{subject to} \quad g_i(\delta u) = 0, \quad i \in \widetilde{J} \subset J. \end{aligned} \tag{32}$$

Let $v^0 = (\delta u^0(t), \mu^0(t))$ satisfy the KKT conditions for problem with $\widetilde{J} = J_0$. This system can be written in the following compact notation:

$$F(\delta u, t) = 0, \quad t \in [t_0, t_f], \tag{33}$$

where $v = (\delta u^0(t), \mu^0(t))$. In order to apply Newton's method to system, we have to solve a linear system with $D_{\delta u} F(v(t), t)$ as matrix. The same matrix is used to compute $\dot{v}(t)$:

$$D_{\delta u} F(\delta u(t), t)\, \dot{v}(t) = D_{\delta u} f(\delta u(t), t). \tag{34}$$

Therefore, using the Newton method as corrector, we have

$$\begin{aligned} &D_{\delta u} F\left(\delta u_i^{k_i - 1}, t_i\right)\left(\delta u_i^{k_i} - \delta u_i^{k_i - 1}\right) = -F\left(\delta u_i^{k_i - 1}, t_i\right), \\ &D_{\delta u} F\left(\delta u_i^{k_i - 1}, t_i\right)\left(\dot{\delta} u_i^{k_i - 1}\right) = -\Delta_t F\left(\delta u_i^{k_i - 1}, t_i\right). \end{aligned} \tag{35}$$

4. Proposed Algorithm: FB-OCQ

The *first-order indirect* approach motivates us to design FB-OCQ algorithm. Shortly, this algorithm uses gradient and does forward-backward searching for the optimal control solution. Let us choose an initial control trajectory:

$$(u_0, \ldots, u_N), (\delta u_0, \ldots, \delta u_N);$$

$$(u_{+0} = u_0, \ldots, u_{+N} = U_N);$$

$$J_- = \infty, \gamma, \text{Iter} = 1, \Delta_0, \text{TOL}, \text{MaxIter}, \rho \in [0, 1].$$

From Algorithm 1 (FB-OCQ), it is obvious that the algorithm terminates if the norm of the gradient of the Hamiltonian with respect to u, $\gamma = \|H_u\|_2$, during the run time of the program, is smaller than the tolerance and the parameter β must change at the inner iterations if we do not have a descent direction, so we must divide β by parameter $\rho \in (0, 1)$ until we get a reduction in the cost function. Figure 2 explains our algorithm steps: forward sweep (control variable and cost function value) and backward sweep (adjoint variable) in detail.

(1) **Initialization:**
  $u(t), \delta u(t) = 0, u_+(t) = u(t), \forall t \in [t_0, t_f]$;
  $\mathcal{J}_- = \infty, \gamma = 0, \beta, \rho \in (0, 1)$;
  Iter = 1, MaxIter, TOL.
(2) **while** Iter ≤ MaxIter **do**
(3) **procedure** Forward Sweep
(4) $x_+(t_0) = x_0, \mathcal{J}_+(t_0) = 0.$
(5) **Integrate Forward** $t : t_0 \to t_f$
(6) **if** Iter > 1 **then**
(7) $\delta u(t) = \arg\min_{\delta u} H_{\text{aug}}(x(t), u(t) + \delta u, \overline{x}(t), t);$
(8) $u_+(t) = u(t) + \delta u(t).$
(9) **end if**
(10) $\dot{x}_+(t) = f(x_+(t), u_+(t), t).$
(11) $\dot{\mathcal{J}}_+(t) = l(x_+(t), u_+(t), t).$
(12) **end procedure**
(13) $\mathcal{J}_+ = \mathcal{J}_+(t_f).$
(14) **if** $\mathcal{J}_+ < \mathcal{J}_-$ **then**
(15) $\mathcal{J}_- = \mathcal{J}_+, \beta = \beta\rho.$
(16) **else**
(17) $\beta = \beta/\rho$; go to **Forward Sweep**
(18) **end if**
(19) **procedure** Adjoint Sweep
(20) $\overline{x}(t_f) = \phi_x^\top(x_+(t_f)), \gamma(t_0) = 0.$
(21) **Integrate Backward** $t : t_f \to t_0.$
(22) $x(t) = x_+(t), u(t) = u_+(t).$
(23) $\dot{\overline{x}}(t)^\top = -(\partial H(x(t), u(t), \overline{x}(t), t)/\partial x)$
(24) $\overline{u}(t) = \partial H(x(t), u(t), \overline{x}(t), t)/\partial u.$
(25) $\dot{\gamma}(t) = \|\overline{u}(t)\|^2.$
(26) $\gamma = \gamma(t_f).$
(27) **end procedure**
(28) **if** $\|\gamma\|$ < TOL **then**
(29) stop;
(30) **end if**
(31) Iter ← Iter + 1.
(32) **end while**

Algorithm 1: FB-OCQ.

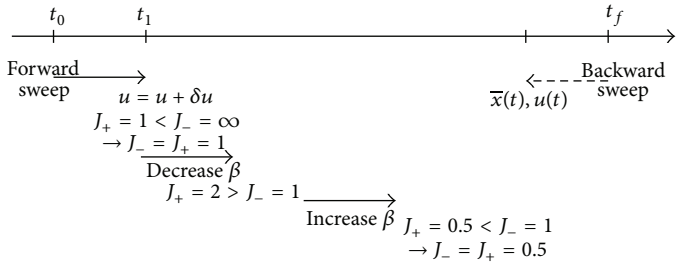


Figure 2: FB-OCQ algorithm explanation.

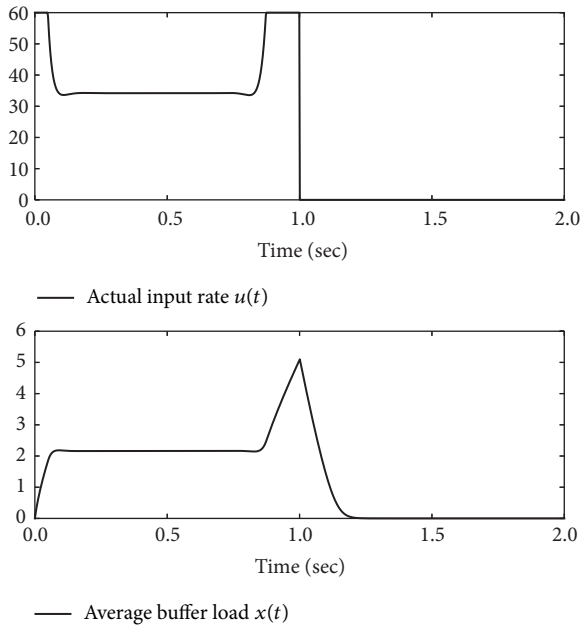


Figure 3: Indirect method: FB-OCQ.

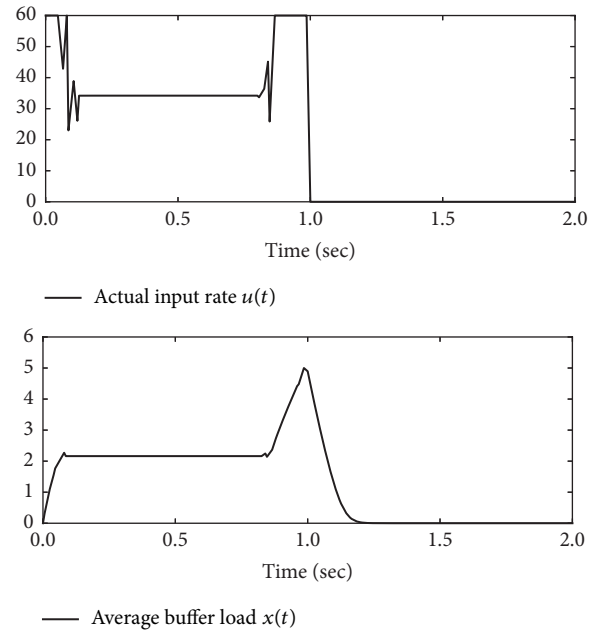


Figure 4: Direct method: nonlinear optimization with JModelica solver.

5. Performance Evaluation

5.1. Numerical Results. In this section, we provide some iteration results for both direct and indirect approaches as follows.

Indirect Method: FB-OCQ. The ODE was solved on equidistant discretization with 1000 discretization points. The optimal control and optimal state are depicted in Figure 3. We realize, by looking at the final control, that there exist some points where the set of active constraints changed due to the singularity; that is, the optimal control is of the bang-bang type with the possibility of a singular arc. To the meaning of such singular arc for queue management, it presents the sudden changes of input rate from 60 (packets/sec) to 35 (packets/sec) which accordingly result in the changes of buffer load (or queue length). In the β history, in Figure 3, the parameter β changed at the inner iterations because we do not have a descent direction, so we must divide β by $\rho \in (0, 1)$ until we get a reduction in the cost function $\mathcal{J} = 6.81280316$. The processing time of central processing unit (CPU) for this algorithm is 0.04 s and the norm of the gradient of the Hamiltonian function $\|H_u\|_2$ with respect to the control u goes to approximately $10e - 7$.

Direct Collocation Method: JModelica and GAlib Solvers. With $N = 1000$ and the number of interpolation points being 3, we tested the discretized problem (9)-(10) by JModelica and GAlib solvers (see [20, 21]) as follows.

(i) JModelica Solver. The cost function is reduced to $\mathcal{J} = 6.8678362$ and the CPU processing time is 0.52 s. The optimal control and optimal state trajectories are obtained during JModelica running and illustrated in Figure 4. An advantage of this approach is that we do not need to derive the adjoint equations.

(ii) GAlib Solver. For this problem, we use the number of generations $n_{\text{gen}} = 2500$, and the population size is 200. The optimal control and optimal state are depicted in Figure 5 obtained during the run of the GAlib. The cost function is reduced to $\mathcal{J} = 7.06744$ during the run time of the program. The CPU processing time is 1.290 s. Although the genetic

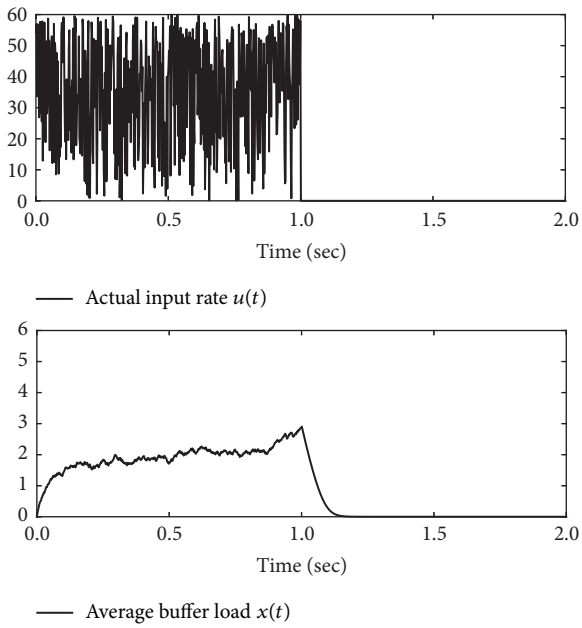


Figure 5: Direct method: genetic algorithm with GAlib solver.

Table 1: Numerical result summary.

Criteria/methods	Direct GAlib	Direct JModelica	Indirect FB-OCQ
Cost function value	$\mathcal{J} = 7.067$	$\mathcal{J} = 6.867$	$\mathcal{J} = 6.812$
CPU process time	1.29 (sec)	0.52 (sec)	0.04 (sec)

algorithm has an advantage that it needs no derivatives or Hessian's information, the control functions produced by it are useless and the convergence is very slow in comparison to Algorithm 1 (FB-OCQ) and the one using JModelica solver.

Table 1 summarizes and compares the three methods that we develop numerically. We conclude that Algorithm 1 is much faster than the other solvers and gives the best cost function. By using genetic library GAlib, the cost function does not reach the local solution as it claims in finding the global solution. The limitation of the indirect methods (FB-OCQ) is that one should derive the adjoint equations which are not easy to derive in some applications.

5.2. Simulation Results. In this section, the performance of the obtained algorithm (FB-OCQ) is evaluated by comparing it with some popular AQMs including RED, REM, and PI. The credibility of results is confirmed using ns-2 simulator [24]. We investigate a network topology with 40 sources, an intermediate router, and one destination (Figure 6). All sources simultaneously send packets to the destination through the router. Hence, a large queue is built up at the router or bottleneck point. Maximum buffer size of the router is 100 packets. All of the compared RED, REM, and PI algorithms are configured to obtain the desired queue length at 40 packets. Simulation lasts for 60 seconds and is repeated using a built-in random generator in ns-2 to obtain more credible results.

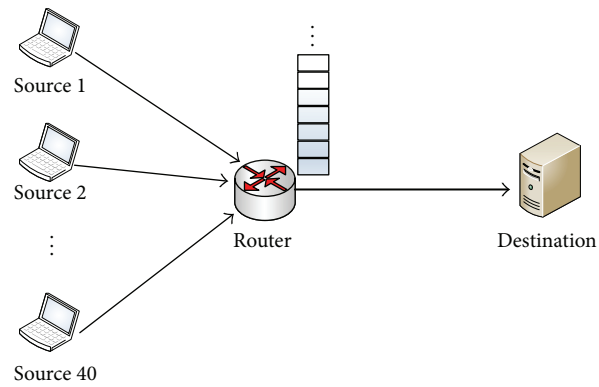


Figure 6: Simulation topology.

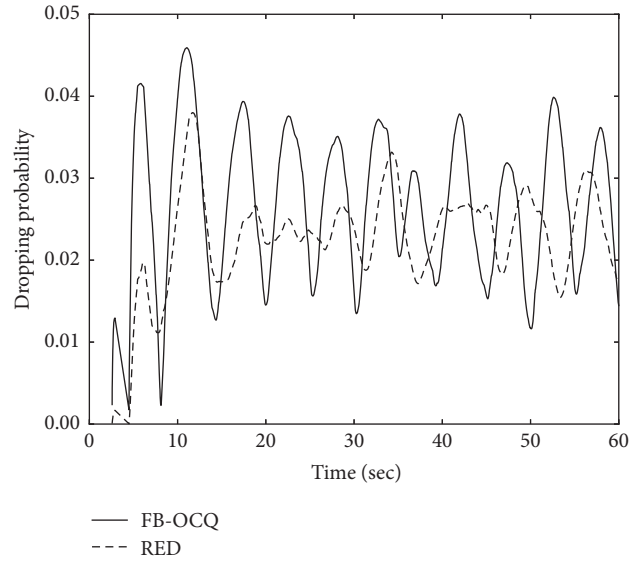


Figure 7: Dropping probability.

Figure 7 shows dropping probability of FB-OCQ and RED algorithms. With the proposed square-root drop function, FB-OCQ drops queuing packets more aggressively than RED although dropping frequency is nearly the same. Maximum dropping ratio is 0.045 for FB-OCQ and 0.039 for RED. In fact, when the algorithm drops more packets, the queue stability will be increased but we will have to sacrifice the system throughput performance. We can see this trade-off in next results.

Figure 8 presents average queue length values measured at the bottleneck link from router to destination for different algorithms. Due to aggressive dropping, FB-OCQ maintains the shortest queue at 40 packets which is also the desired value. Only REM can obtain the same value but in a longer time, ≈50 seconds. PI, in fact, can achieve the same performance only if its parameters are well configured and be dynamically changed to different network scenarios. In Figure 8, we can see that PI performs not so well, even we set the desired queue length variable at 40 packets and exploit default parameters in ns-2 for PI controller.

Finally, we investigate throughput performance of proposed FB-OCQ algorithm. It is confirmed from Figure 9 that there exists a trade-off in the relationship between

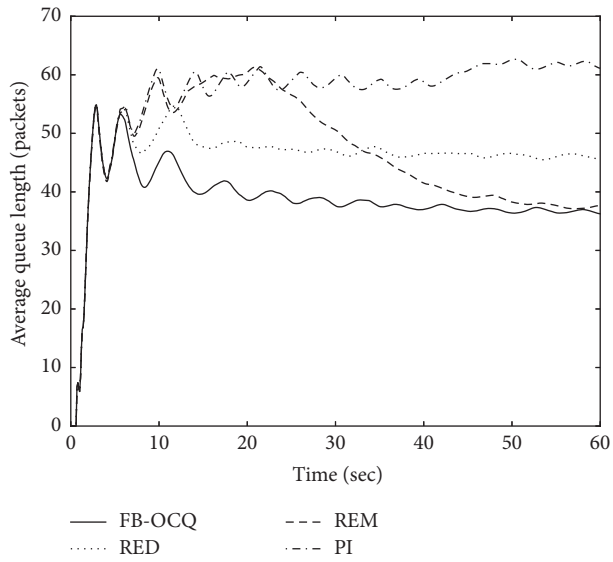


Figure 8: Average queue length versus time.

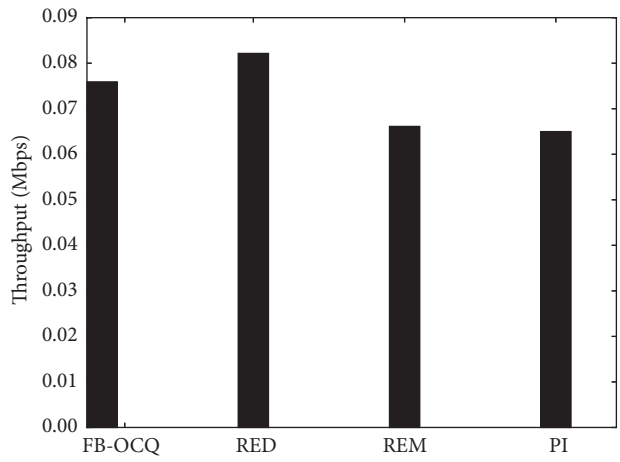


Figure 9: Throughput comparison.

throughput and queue length/dropping probability. PI and REM achieve nearly the same throughput at 0.06 (Mbps) while FB-OCQ and RED obtain the better throughput performance at 0.08 (Mbps) value. In Figure 8, although REM can achieve the desired queue length 40 packets in this scenario, REM still maintains a large queue during simulation time from 0 (sec) up to 25 (sec). That reason leads to throughput performance of REM being lower than RED and FB-OCQ. Our proposed FB-OCQ drops aggressively the packets so that its throughput value (0.075904 Mbps) is slightly less than RED (0.082146 Mbps).

5.3. Experiment Results. We examine the effectiveness of our state-of-the-art optimal control queue method in the Linux kernel (e.g., Ubuntu 16.04). We conduct experiment from a desktop computer through the Internet gateway to an outside server (Figure 10). To manage working queuing disciplines (qdiscs), we use the scheduler qdisc in the Linux kernel. The chosen server is a dedicated bufferbloat server,

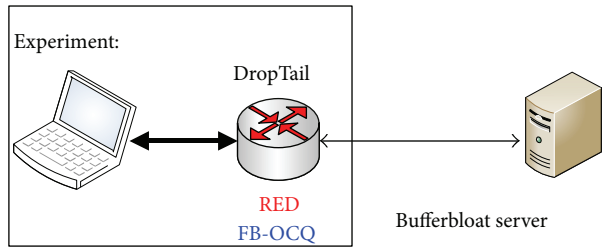


Figure 10: Experiment test setup, from the test client to the test server.

which is able to stand very high congestion due to many data flows at the same time. We exploit the Flent: FLExible Network Tester [25] and Realtime Response Under Load (RRUL) test [26] to evaluate our proposal. RRUL test puts a network under worst case conditions, reliably saturates the network link, and thus recreates bufferbloat phenomenon for queue algorithm testing. Simulation time duration is 60 seconds.

We compare latency under load test with queue being handled by different AQM schemes (DropTail, RED, and FB-OCQ) in turn. DropTail queuing method is by far the simplest approach to network router queue management. The router accepts and forwards all the packets that arrive as long as its buffer space is available for the next incoming packets. If a packet arrives and the queue is currently full, the incoming packet will be dropped. The sender then detects the packet lost event and shrinks its sending window. While it is the most widely used due to simplicity and relatively high efficiency, DropTail has some weakness such as the bad fairness sharing among TCP connections, and throughput and bottleneck link efficiency suffer severe degradation if congestion is getting worse.

RED [5, 6] was presented with the objective of minimizing packet loss and queuing delay. Moreover, it can compensate the weakness of DropTail by avoiding global synchronization of TCP sources so that it improves fairness. To achieve these goals, RED utilizes two thresholds, $\min_{th}$ and $\max_{th}$, and an exponentially weighted moving average (EWMA) formula to estimate average queue length [27]. When the average queue length exceeds a predefined threshold, the link is implied to be in congested state and drop action is taken. A temporary increase in the queue length notifies the transient congestion, while an increase in the average queue length reflects long-lived congestion. Based on such information, RED router sends randomized feedback signals to the senders to make decision of decreasing their congestion windows. RED has good fairness among connections because of the feedback randomized mechanism [28].

Figure 11 presents latency ping results under RRUL test suite. Ping is a networking utility and operates by sending Internet Control Message Protocol (ICMP) echo request packets to the target server and waits for an ICMP echo reply. The program measures the round-trip time from transmission to reception, reporting errors and packet loss. Our proposed algorithm FB-OCQ achieves the lowest packet latency compared with the other two algorithms inside Linux

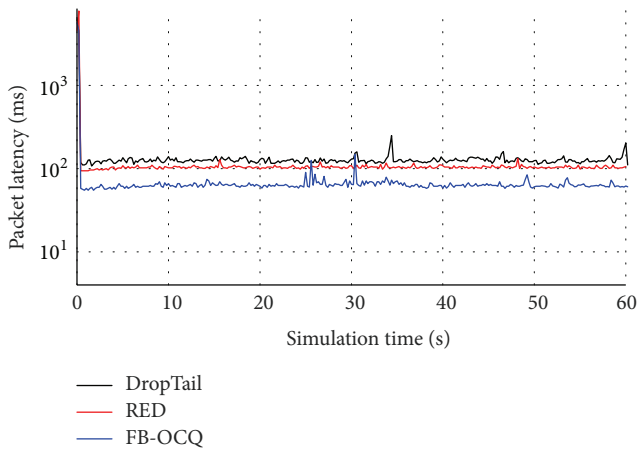


Figure 11: ICMP ping result using Realtime Response Under Load test.

kernel. Specifically, the packet latency when using FB-OCQ is about 80 (ms), while about 120 (ms) if using DropTail (pfifo_fast in Ubuntu) and 100 ms if using RED algorithm.

6. Conclusions

We proposed a queue management algorithm named forward-backward optimal control queue (FB-OCQ) to solve the OCQ problem. Derived from indirect approaches in dynamic optimization, this algorithm demonstrates faster reaction while still achieving the same performance in numerical analysis compared to direct methods. Employing under network simulation ns-2, we see that the proposed algorithm drops packets more aggressively than the traditional RED algorithm, in a higher frequency and magnitude. As a result, average queue length can be reduced much more, while an acceptable value of throughput still can be maintained. In future works, we try to investigate the memory efficiency of FB-OCQ under wireless sensor networks.

Competing Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korean government (MEST) (no. NRF-2016R1A2B1013733).

References

[1] T. Hoiland-Jorgensen, "Battling bufferbloat: an experimental comparison of four apapproach to queue management in linux," Tech. Rep., Roskilde University, Roskilde, Denmark, 2012.

[2] J. Gettys, "Bufferbloat: dark buffers in the Internet," *IEEE Internet Computing*, vol. 15, no. 3, pp. 95–96, 2011.

[3] J. Gettys, "Bufferbloat: views of the elephant," in *Proceedings of the International Summit for Community Wireless Networks*, October 2012.

[4] B. Braden, D. Clark, J. Crowcroft et al., "Recommendations on queue management and congestion avoidance in the internet," RFC 2309, IETF, San Francisco, Calif, USA, 1998.

[5] C. Hollot, V. Misra, D. Towsley, and W.-B. Gong, "A control theoretic analysis of red," in *Proceedings of the 20th Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM '01)*, vol. 3, pp. 1510–1519, IEEE, Anchorage, Alaska, USA, 2001.

[6] S. Floyd and V. Jacobson, "Random early detection gateways for congestion avoidance," *IEEE/ACM Transactions on Networking*, vol. 1, no. 4, pp. 397–413, 1993.

[7] S. Athuraliya, S. Low, and D. Lapsley, "Random early marking," in *Quality of Future Internet Services*, J. Crowcroft, J. Roberts, and M. I. Smirnov, Eds., vol. 1922 of *Lecture Notes in Computer Science*, pp. 43–54, Springer, Berlin, Germany, 2000.

[8] R. Pan, P. Natarajan, C. Piglione et al., "PIE: a lightweight control scheme to address the bufferbloat problem," in *Proceedings of the IEEE 14th International Conference on High Performance Switching and Routing (HPSR '13)*, pp. 148–155, Taipei, Taiwan, July 2013.

[9] F. Ren, C. Lin, and B. Wei, "Design a robust controller for active queue management in large delay networks," in *Proceedings of the 9th International Symposium on Computers and Communications (ISCC '04)*, vol. 2, pp. 748–754, June-July 2004.

[10] R. Zhu, H. Teng, and J. Fu, "A predictive PID controller for AQM router supporting TCP with ECN," in *Proceedings of the 5th International Symposium on Multi-Dimensional Mobile Communications. The 2004 Joint Conference of the 10th Asia-Pacific Conference on Communications*, vol. 1, pp. 356–360, August-September 2004.

[11] J. Sun, M. Zukerman, and M. Palaniswami, "Stabilizing RED using a fuzzy controller," in *Proceedings of the IEEE International Conference on Communications (ICC '07)*, pp. 266–271, Glasgow, Scotland, June 2007.

[12] H.-L. To, T. M. Thi, and W.-J. Hwang, "Cascade probability control to mitigate bufferbloat under multiple real-world TCP stacks," *Mathematical Problems in Engineering*, vol. 2015, Article ID 628583, 13 pages, 2015.

[13] M. Hassan and H. Sirisena, "Optimal control of queues in computer networks," in *Proceedings of the International Conference on Communications (ICC '01)*, vol. 2, pp. 637–641, June 2001.

[14] V. Guffens and G. Bastin, "Optimal adaptive feedback control of a network buffer," in *Proceedings of the American Control Conference (ACC '05)*, vol. 3, pp. 1835–1840, Portland, Ore, USA, June 2005.

[15] J. T. Betts, *Practical Methods for Optimal Control Using Nonlinear Programming*, Society for Industrial and Applied Mathematics, Philadelphia, Pa, USA, 2001.

[16] R. Bulirsch, E. Nerz, H. Pesch, and O. von Stryk, "Combining direct and indirect methods in optimal control: Range maximization of a hang glider," in *Optimal Control*, R. Bulirsch, A. Miele, J. Stoer, and K. Well, Eds., vol. 111 of *ISNM International Series of Numerical Mathematics*, pp. 273–288, Birkhuser, Basel, Switzerland, 1993.

[17] A. Radwan, *Utilization of parametric programming and evolutionary computing in optimal control [Ph.D. dissertation]*, Humboldt-Universität zu Berlin, 2012.

[18] O. von Stryk and R. Bulirsch, "Direct and indirect methods for trajectory optimization," *Annals of Operations Research*, vol. 37, no. 1, pp. 357–373, 1992.

[19] A. Bryson and Y. Ho, *Applied Optimal Control: Optimization, Estimation and Control*, Hemisphere: CRC Press, Washington, DC, USA, 1975.

[20] M. Ab, Jmodelica.org user guide 1.6.0, 2011, http://www.jmodelica.org/page/236.

[21] M. Wall, *GAlib: A C++ Library of Genetic Algorithm Components*, Version 2.4 Documentation Revision B, Massachusetts Institute of Technology, 1996.

[22] A. Radwan, O. Vasilieva, R. Enkhbat, A. Griewank, and J. Guddat, "Parametric approach to optimal control," *Optimization Letters*, vol. 6, no. 7, pp. 1303–1316, 2012.

[23] J. Guddat, F. Guerra Vazquez, and H. T. Jongen, *Parametric Optimization: Singularities, Pathfollowing and Jumps*, John Wiley & Sons, 1990.

[24] T. Issariyakul and E. Hossain, *Introduction to Network Simulator NS2*, Springer, Berlin, Germany, 1st edition, 2008.

[25] T. Hiland-Jrgensen, Flent: The flexible network tester, 2012–2015, https://flent.org.

[26] D. Tht, "Realtime response under load (rrul) test," 2013, https://github.com/dtaht/deBloat/blob/master/spec/rrule.doc.

[27] J. Koo, S. Ahn, and J. Chung, "A comparative study of queue, delay, and loss characteristics of AQM schemes in QoS-enabled networks," *Computing and Informatics*, vol. 23, no. 4, pp. 317–335, 2004.

[28] T. B. Reddy and A. Ahammed, "Performance comparison of active queue management techniques," *Journal of Computer Science*, vol. 4, no. 12, pp. 1020–1023, 2008.

Hindawi Publishing Corporation
Scientific Programming
Volume 2016, Article ID 9751243, 9 pages
http://dx.doi.org/10.1155/2016/9751243

Research Article

An Optimal Decision Assessment Model Based on the Acceptable Maximum LGD of Commercial Banks and Its Application

Baofeng Shi,[1] Bin Meng,[2] and Jing Wang[1]

[1]*College of Economics & Management, Northwest A&F University, Yangling, Shaanxi 712100, China*
[2]*Collaborative Innovation Center for Transport Studies, Dalian Maritime University, Dalian 116026, China*

Correspondence should be addressed to Bin Meng; mengbinfly@163.com

Received 18 June 2016; Revised 8 August 2016; Accepted 15 August 2016

Academic Editor: Guo Chen

This paper introduces a novel decision assessment method which is suitable for customers' credit risk evaluation and credit decision. First of all, the paper creates an optimal credit rating model, and it consisted of an objective function and two constraint conditions. The first constraint condition of the strictly increasing LGDs eliminates the unreasonable phenomenon that the higher the credit rating is, the higher the LGD (loss given default) is. Secondly, on the basis of the credit rating results, a credit decision-making assessment model based on measuring the acceptable maximum LGD of commercial banks is established. Thirdly, empirical results using the data on 2817 farmers' microfinance of a Chinese commercial bank suggest that the proposed approach can accurately find out the good customers from all the loan applications. Moreover, our approach contributes to providing a reference for decision assessment of customers in other commercial banks in the world.

1. Introduction

There are a lot of decision assessment issues in reality, such as production scheduling assessment, transportation routing assessment, economy development evaluation, and risk assessment. In order to cope with these problems, lots of decision-making optimization and evaluation approaches have been created in the existing studies [1–4]. However, only a small part of them have been widely used in the real world. Much more attention should be paid to the application of the decision-making optimization and evaluation models. For example, in credit risk evaluation of commercial banks, an accurate optimal decision-making evaluation model is in urgent need of selecting the good customers from all the loan applications, which can help the bankers, the society, and the bond investors to reduce loss [5].

The main proposed references to the evaluation of credit risk of commercial banks can be divided into two categories. The first category concentrates on the establishment of credit scoring models, and the second category focuses on the division of credit ratings or the ascertaining method of credit threshold. Early studies included the establishment of a five-variable Z-Score credit risk evaluation model, and the Zeta rating model can optimize the Z-Score model by using statistics discrimination techniques [6, 7]. Min and Lee proposed a credit scoring model based on data envelopment analysis (DEA) methods [8]. Using the data information of the New York Stock Exchange, Hwang et al. established an ordered semi-parametric Probit credit rating model by substituting ordered semi-parametric function for the linear regression function [9]. Gupton et al. utilized transition matrices to describe debtor's probability of credit rating change, thereby creating an optimal CreditMetrics credit rating model [10]. Recent methods have advanced Bayesian Analysis Techniques to help deal with the issue of loss from bad loans [11]. Expert information systems and artificial intelligence techniques are other important approaches for credit risk assessment of commercial banks. For example, predicted accuracy techniques, like multiple classifiers and classifier ensembles, were a major testing methodology with expert information systems and artificial intelligence techniques [12–14].

Research on how to issue loans requires the division of credit ratings and the ascertaining method of credit threshold. For the past few years, more and more scholars and bankers began to monitor and divide the customers' credit ratings, which deepens the associated theoretical research

and practical processes of credit risk management. Gorzałczany and Rudziński classified customers' credit levels into 10 different grades by using a multiobjective genetic optimization approach [15]. By creating the objective function maximizing the deviation matrix **A** between default customers and nondefault customers and minimizing the deviation matrix **B** among indices in the same group (default customer group or nondefault customer group), the optimal weights had been obtained. Then, Shi et al. proposed a credit risk assessment model, which was suitable for the credit rating classification of imbalanced data [5]. According to the distribution of the number of customers, some researchers divided the loan customers into 9 credit ratings, namely, AAA, AA, A, BBB, BB, B, CCC, CC, and C [16, 17]. Florez-Lopez estimated customers' PD (probability of default) through MDA (multiple discriminant analysis), Logit, and decision tree methods and then classified customers into 5 credit grades by setting a variable for each grade (AA = $1, \ldots, \mathrm{B} = 5$) [18]. After getting the numerical solution for loan pricing through mathematical analysis, Pascucci et al. (2013) calculated the credit threshold and then created a customer loan pricing model [19]. Chen and Cheng graded the credit level by combining a rough set classifier with the minimum entropy principle approach. Then, they determined the credit ratings for issuing loans from their subjective experience [20]. Guo et al. established an evaluation model of customers' credit risk by linking the personal loan on P2P lending platform with portfolio optimization [21]. Jones et al. predicted the variation tendency of customer's credit level through binary classifier and then ascertained the credit threshold [22].

The above studies have made great progress in handling decision evaluation issues of credit risk of commercial banks. However, further studies on the following problems are still needed. First of all, the existing credit risk assessment results have the unreasonable phenomenon that the higher credit rating customers also have the higher loss given default (i.e., LGD). It will generate a result that a bank can issue loans to the customers who are under the break-even point of the bank, which can increase the bank's loan loss. Secondly, the existing credit rating classification results do not establish the connection between the credit rating classification results and the break-even point of commercial banks, so the classification methods do not have the functions of credit decisions. Bankers or credit clerks cannot locate the qualified customers who can realize the bank's target profit.

To fill in the above gaps, this study advances in three aspects. First, this paper establishes a nonlinear credit risk assessment model which consisted of an objective function and two constraint conditions. The first constraint condition of the strictly increasing LGDs avoids the unreasonable phenomenon, like higher ratings with higher LGDs. Second, by measuring the break-even point threshold of commercial banks, this paper works out the maximum annual LGD which the bank can bear and then introduces the break-even point into the credit rating results. Thus, the relation between the target profit of bank and the credit threshold can be known, which makes up for the deficiency of previous researches. Third, an empirical study on the decision-making data of 2817 farmers' microfinance of a large-scale Chinese commercial bank in 28 provinces is carried out in order to verify the accuracy of the proposed model.

The rest of the paper is organized as follows. Section 2 introduces the methodology of the paper. Section 3 presents the data and empirical analysis of our decision assessment model for 2817 farmers' microfinance in China. Section 4 is the conclusions.

2. Design and Methodology of the Study

In this section, we introduce an optimal credit decision assessment model based on the acceptable maximum LGD of commercial banks. According to the relation between the annual owed loan capital and interest and annual receivable loan capital and interest of customers' loan, together with the credit scores of customers, the customers are classified into 9 credit ratings in this paper based on the principle that the customer with a higher credit rating is lower in annual LGD. A step-by-step instruction is provided.

Step 1. Sequence the customers in a descending order according to their credit scores, and calculate the customer's LGD.

Step 2. Divide the customers' credit ratings in terms of the principle that the higher the credit rating is, the lower the LGD is.

Step 3. Calculate the break-even point threshold.

Step 4. Make credit decisions.

2.1. Calculation of the Customer's LGD. According to the Basel Capital Accord III [23], it will be deemed to be a defaulter if the client fails to repay the loan in full or on time. Then, the loan customers who fail to pay off the principal and interest within 90 days after the maturity date T of loan should be regarded as delinquent customers in this paper. The annual LGD of customers in a certain credit grade can be obtained from the ratio of the total annual owed loan capital and interest in this grade to its total annual receivable loan capital and interest. The total of annual owed loan capital and interest of customers refers to the total amount of annual loss (outstanding principal and interest) customers that have caused the bank 90 days after the maturity date T of loan. The total of annual receivable loan capital and interest of customers refers to the total amount of loan (principal and interest) that customers should repay to the bank each year.

Let LGD_j denote the annual loss given default of the jth credit rating ($j = 1, 2, \ldots, 9$), let L_j denote the total of the annual owed loan capital and interest of microfinance of all the customers in the jth credit rating, let R_j denote the total of the annual receivable loan capital and interest of microfinance of all the customers in the jth credit rating, let n denote the total number of customers, let R_{ij} denote the annual receivable loan capital and interest of the microfinance

of the jth credit rating on the ith customer, and let L_{ij} denote the annual owed loan capital and interest of the microfinance of the jth credit rating on the ith customer. Then, the LGD_j can be expressed as

$$\mathrm{LGD}_j = \frac{L_j}{R_j}, \tag{1}$$

where R_j and L_j can be computed by using (2) and (3), respectively:

$$R_j = \sum_{i=1}^{n} R_{ij}, \tag{2}$$

$$L_j = \sum_{i=1}^{n} L_{ij}. \tag{3}$$

2.2. Establishment of an Optimal Credit Rating Model. In this section, we will create a nonlinear programming model to classify the customers' credit ratings. According to the credit rating requirements of the commercial bank and the China Banking Regulatory Commission (CBRC) [24, 25], we divide the loan customers into nine ratings.

(1) Establish an Objective Function. Let LGD_j denote the loss given default of the jth credit rating, and $j = 1, 2, \ldots, 9$, respectively, represents nine credit ratings, that is, AAA, AA, ..., C. Establish an objective function, that is, the minimum distance of all adjacent rating LGDs,

$$\begin{aligned} \text{Obj: } \min f = {} & \left(\mathrm{LGD}_9 - \mathrm{LGD}_8\right)^2 + \left(\mathrm{LGD}_8 - \mathrm{LGD}_7\right)^2 \\ & + \cdots + \left(\mathrm{LGD}_2 - \mathrm{LGD}_1\right)^2. \end{aligned} \tag{4}$$

The specialty of (4) is that by creating the objective function minimizing the distance of LGDs of all adjacent credit ratings, it ensures that the LGD distribution of the selected credit rating result is the closest to an isosceles triangle.

(2) Constraint Condition 1. The LGDs increase strictly. Namely,

$$0 < \mathrm{LGD}_1 < \mathrm{LGD}_2 < \cdots < \mathrm{LGD}_9 \le 1. \tag{5}$$

As we all know, only national debt's LGD equals 0, and LGD = 1 shows that all of the credit rating customers are default customers; so the value of LGD which is more than 0 and less than or equal to 1 is logical in (5).

(3) Constraint Condition 2. It is the equality constraint to calculate LGD_j of the jth credit rating. The LGD_j can be expressed as (1).

In a word, the optimal credit rating model can be written as (6). It should be pointed out that if researchers want to grade the loan customers into five ratings [25], they can substitute the number nine for the number five in (6). Then, the calculation results can be obtained through the C++ software,

$$\begin{aligned} \min \quad & f \\ & = \left(\mathrm{LGD}_9 - \mathrm{LGD}_8\right)^2 + \left(\mathrm{LGD}_8 - \mathrm{LGD}_7\right)^2 + \cdots \\ & \quad + \left(\mathrm{LGD}_2 - \mathrm{LGD}_1\right)^2 \\ \text{s.t.} \quad & 0 < \mathrm{LGD}_1 < \mathrm{LGD}_2 < \cdots < \mathrm{LGD}_9 \le 1 \\ & \mathrm{LGD}_j = \frac{\sum_{i=1}^{n} R_{ij}}{\sum_{i=1}^{n} L_{ij}}, \quad j = 1, 2, \ldots, 9. \end{aligned} \tag{6}$$

2.3. Calculation of the Break-Even Point Threshold. In order to compute the break-even point threshold of customers' loan, we have to calculate the ratio of financial costs of customers' microfinance r_C, the minimum target rate of return of customers' microfinance r_P, the ratio of the annual owed loan capital and interest to the annual receivable loan capital, and interest r_D.

(1) The Ascertaining of the Ratio of Financial Costs r_C. Let r_C denote the ratio of financial costs, F_1 be the expenses of interest on deposit, F_2 be the charge against revenue, and P be the total loan, we have Financial Costs = (Interest Expenses on Deposit + Charge against Revenue)/Total Loan; that is,

$$r_C = \frac{\left(F_1 + F_2\right)}{P}. \tag{7}$$

(a) Calculation of the Expenses of Interest on Deposit F_1. Not all the deposits in the bank can be used for loan. To ensure the deposit withdrawal of customer and the operation of capital, the bank has to deposit some money in the People's Bank of China as required reserves. Therefore, the deposit used for loan by the bank does not include the part deposited in the People's Bank of China. The People's Bank of China should pay the interest of required reserves to the bank so as to make up the loss of bank for not operating the capital. Therefore, expenses of interest on deposit F_1 is calculated by adding the deposit interests of loan principal and the deposit interests of the corresponding required reserves of loan principal required by the People's Bank of China and deducting the interest income of the corresponding required reserves of loan principal deposited in the People's Bank of China.

Let F_1 denote the expenses of interest on deposit, P denote the total loan, r_1 denote the one-year deposit rate, L_{re} denote the corresponding required reserves of total loan required by the People's Bank of China, and r_{re} denote the deposit reserve rate. Then,

$$F_1 = P \times r_1 + L_{\mathrm{re}} \times r_1 - L_{\mathrm{re}} \times r_{\mathrm{re}}. \tag{8}$$

Given P = total loan, L_D = the corresponding deposit of total loan, and ε = rate of legal deposit reserve, we have

$$P = L_D \times (1 - \varepsilon). \tag{9}$$

Then, the corresponding deposit of total loan L_D can be written as

$$L_D = \frac{P}{(1 - \varepsilon)}. \tag{10}$$

The corresponding deposit of total loan L_D times the rate of legal deposit reserve ε is the corresponding required reserves of total loan required by the People's Bank of China L_{re}:

$$L_{\text{re}} = L_D \times \varepsilon. \tag{11}$$

Substituting (10) into (11), we have

$$L_{\text{re}} = \frac{P \times \varepsilon}{1 - \varepsilon}. \tag{12}$$

Equation (12) is the calculation formula of the corresponding required reserves of total loan required by the People's Bank of China L_{re}.

Substitute (11) into (8), and then

$$\begin{aligned} F_1 &= P \times r_1 + L_{\text{re}} \times r_1 - L_{\text{re}} \times r_{\text{re}} \\ &= P \times r_1 + \frac{P \times \varepsilon}{1 - \varepsilon} \times r_1 - \frac{P \times \varepsilon}{1 - \varepsilon} \times r_{\text{re}}. \end{aligned} \tag{13}$$

(b) Calculation of the Charge against Revenue F_2. Charge against revenue refers to all the expenses the bank has to pay in business operation and management except the expenses of interest on deposit. It is made up of operating expenses, depreciation charges of fixed assets, and other business expenses.

Let F_2 be the charge against revenue, E_1 be the operating expenses, E_2 be the depreciation charges of fixed assets, and E_3 be other business expenses, and then

$$F_2 = E_1 + E_2 + E_3. \tag{14}$$

(2) The Ascertaining of the Minimum Target Rate of Return r_P. Let C denote the core capital of bank, let C_d denote the deduction of core capital, let α_i denote the risk weight of risk asset i, let A_i denote the line of risk asset i, let M_c denote the market risk capital, and let $\beta_{\min}$ denote the minimum core capital adequacy ratio stipulated by CBRC [26]. We have

$$\frac{C - C_d}{\sum \alpha_i A_i + 12.5 M_c} \geq \beta_{\min}, \tag{15}$$

where $\sum \alpha_i A_i$ is the sum of different risk-weighted assets and $(\sum \alpha_i A_i + 12.5 M_c)$ is the sum of risk-weighted assets and 12.5 times market risk capital.

Let ROE be the rate of return on common stockholders' equity, GP′ be the total of the target net profit of bank, and C' be the net asset of bank, and then

$$C' = \frac{\text{GP}'}{\text{ROE}}. \tag{16}$$

The net asset C' includes capital stock, capital reserve, surplus reserve, statutory public welfare fund, and undistributed profit. The core capital C includes capital stock, capital reserve, surplus reserve, and undistributed profit. Let us suppose that the net asset of bank C' equals to the core capital of bank C. After substituting (16) into (15), we have

$$\begin{aligned} \text{GP}' &\geq \text{ROE} \times \left[\left(\alpha_i A_i + 12.5 M_c \right) \times \beta_{\min} + C_d \right] = \text{ROE} \\ &\times \left[\left(\alpha_i A_i + 12.5 \times \sum A_i \times \frac{M_c}{\sum A_i} \right) \times \beta_{\min} + \sum A_i \right. \\ &\left. \times \frac{C_d}{\sum A_i} \right], \end{aligned} \tag{17}$$

where $M_c / \sum A_i$ refers to the market risk capital of the average asset of bank and $C_d / \sum A_i$ refers to the deduction of core capital of the average asset of bank. When the bank issues a loan, the sum of different risk-weighted assets $\sum \alpha_i A_i$ can be replaced by the risk asset αP, and the total assets of bank $\sum A_i$ can be replaced by the total loan P. Then, there is

$$\begin{aligned} \text{GP}' &\geq \text{ROE} \\ &\times \left[\left(\alpha P + 12.5 \times P \times \frac{M_c}{\sum A_i} \right) \times \beta_{\min} + P \times \frac{C_d}{\sum A_i} \right]. \end{aligned} \tag{18}$$

Divide the maximum net profit GP′ of a loan by the total loan P, and we get the target rate of return r_P,

$$\begin{aligned} r_P &= \frac{\text{ROE} \times \left[\left(\alpha P + 12.5 \times P \times \left(M_c / \sum A_i \right) \right) \times \beta_{\min} + P \times \left(C_d / \sum A_i \right) \right]}{P} = \text{ROE} \\ &\times \left[\left(\alpha + 12.5 \times \frac{M_c}{\sum A_i} \right) \times \beta_{\min} + \frac{C_d}{\sum A_i} \right], \end{aligned} \tag{19}$$

from which the minimum target rate of return r_P can be obtained. Equation (19) is to get the minimum legal target rate of return r_P of bank according to the stipulations of Basel New Capital Accord III and CBRC on minimum core capital adequacy ratio.

(3) The Ascertaining of the Ratio r_D. Generally, the loan rate r based on cost plus is obtained by adding the financial costs, target rate of return, and loss given default irrespective of business tax [27]. Let r denote the lending rate, let r_D denote the loss given default, and then

$$r = r_C + r_P + r_D. \tag{20}$$

We have

$$r_D = r - r_C - r_P. \tag{21}$$

Table 1: Original data of farmers for credit decision assessment.

(1) Index	(2) Farmer number	(3) Credit scores	(4) Receivable loan capital and interest R_{ij} (Yuan)	(5) Owed loan capital and interest L_{ij} (Yuan)
1	Farmer 2154	100	3259642	0
2	Farmer 2155	100	2173095	0
3	Farmer 2825	99.24	20074635	0
4	Farmer 2943	99.13	10605947	0
⋮	⋮	⋮	⋮	⋮
416	Farmer 218	77.16	4470876	535791
417	Farmer 96	76.82	3551334	0
⋮	⋮	⋮	⋮	⋮
2814	Farmer 1109	1.69	8538718	8071240
2815	Farmer 1102	0.00	236826	236826
2816	Farmer 1103	0.00	255343.3	255343.3
2817	Farmer 1104	0.00	78904.72	78904.72

The loss given default r_D refers to the ratio of the annual owed loan capital and interest to the annual receivable loan capital and interest, which shows the risk compensation of customer's default.

2.4. Making Credit Decisions Based on the Acceptable Maximum LGD of Commercial Banks. Through the financial costs of a bank and the minimum target rate of return deduced from the minimum core capital adequacy ratio stipulated by CBRC, this paper works out the maximum annual LGD which the bank can bear. Then, the threshold of customers' default can be obtained by comparing the calculated annual LGD of the customers in different credit ratings with the acceptable maximum annual LGD of bank. Loans will not be issued to those customers whose credit rating is lower than the break-even point threshold. Customers with a credit rating lower than the threshold are unable to get a loan from the bank.

3. Empirical Study of the Proposed Model Based on Chinese Farmer Sample

3.1. Sample and Data Source. In consideration of research purpose of decision assessment, we implement empirical study based on the 2817 farmers' microfinance data collected and screened from a Chinese state-owned commercial bank. Our project started with a talk between a chair manager of a Chinese state-owned commercial bank and one coauthor's supervisor about credit decision evaluation of farmers' microfinance loan when we attended the referee conference of major research project organized by National Science Foundation of China in September, 2009. After that, several meetings and communications of various forms have been made, and they lead to the formal start of this project. As a major participant, one coauthor is in charge of evaluation of farmer credit in the entire project. In order to guarantee the correctness and genuineness, all the sample data have been recertified in the bank's credit information system. The 2817 farmer samples include 28 provincial administrative provinces except Beijing Municipality, Yunnan Province, Tibet Autonomous Region, Taiwan Province, Hong Kong Autonomous Region, and Macao Autonomous Region in China, which guarantee data diversity, authenticity, and applicability [24]. It should be pointed out that the sample adopted in this paper is based on the evaluation index system and credit scoring model of farmer's microfinance. The indicators are selected by utilizing multicollinearity analysis. The credit scoring model is built by using logistic regression. Besides, the credit scores have been normalized. As the focus of this paper lies in the rating of credit and the lending decision, the detailed processes of index selecting and credit score calculation are not given (see details in [24]). The original data of farmers for credit decision assessment is shown in Table 1 [24].

3.2. Application of the Decision Assessment Model

(1) Divide the Farmers' Credit Ratings. Substitute the owed loan capital and interest R_{ij} in the fourth column of Table 1 and the receivable loan capital and interest L_{ij} in the fifth column of Table 1 into (1) to (6), and then the credit rating result can be obtained, as shown in Table 2. The second column of Table 2 is the sample size of each credit rating. The third column is the ratio of the sample size of each credit rating to the total sample size. The fourth column is the score intervals of different credit ratings. The fifth column is the interval length of credit scores. The sixth column is the annual LGD of different credit ratings.

Figure 1 shows the distribution of the sample size of different credit ratings. Figure 2 is the pie chart of the sample size of different credit ratings. Figure 3 is the distribution of the annual loss given default of each credit rating. It is shown in Figures 1 and 2 that the majority of farmers are in Grade

Table 2: Results of farmer's credit rating.

(1) Credit rating	(2) Sample size	(3) The ratio of the sample size to the total sample size	(4) Intervals of credit scores	(5) Interval length of credit scores	(6) Annual LGD_j
AAA	416	14.77%	$77.16 \le S < 100$	22.84	0.114%
AA	1821	64.64%	$43.95 \le S < 77.16$	33.21	0.959%
A	145	5.15%	$43.39 \le S < 43.95$	0.56	1.598%
BBB	201	7.14%	$39.21 \le S < 43.39$	4.18	2.875%
BB	90	3.19%	$33.03 \le S < 39.21$	6.18	4.276%
B	34	1.21%	$31.46 \le S < 33.03$	1.57	4.919%
CCC	83	2.95%	$24.26 \le S < 31.46$	7.2	6.641%
CC	14	0.50%	$9.36 \le S < 21.46$	14.9	10.981%
C	13	0.46%	$0 \le S < 9.36$	9.36	14.797%

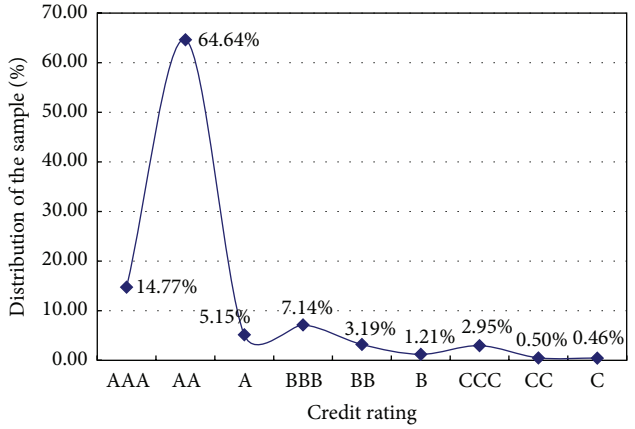


Figure 1: The distribution of the sample size of different credit ratings.

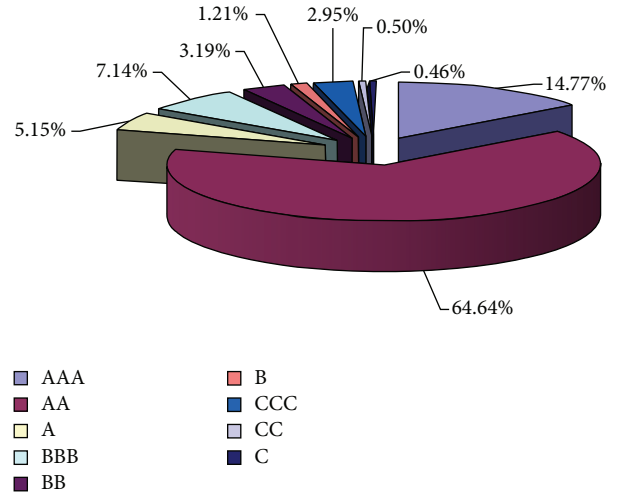


Figure 2: Pie chart of the sample size of different credit ratings.

AA, which accounts for 64.64%, and only 0.50% and 0.46% farmers are in CC and C, respectively. The grades of farmers that the bank has issued a loan for are high in credit ratings. Therefore, the risk of the bank will be low. As is shown in Figure 3, the slopes of the distribution of the annual LGD of different credit ratings are basically the same. That is to say, the change in the slope between the annual LGD of each credit rating is small, which accords with the practical situation. So, the results of credit rating are reasonable.

(2) Calculate the Break-Even Point of Farmers' Microfinance. The related data for calculating the break-even point of farmers' microfinance comes from the Chinese state-owned commercial bank headquarters [24], as shown in Table 3. The descriptions of data are as follows. The loan rate r in Row 1 is 10.5285%, which is the highest lending rate of the Chinese commercial bank in history. The one-year deposit rate r_1 in Row 2 is from the interest table of 2014 benchmark deposit rate of RMB of financing institutions. The rate of legal deposit reserve ε in Row 3 and the deposit reserve rate r_{re} in Row 4 are from the People's Bank of China. The parameters in Row 5 to Row 9 are from the 2014 annual report of the Chinese

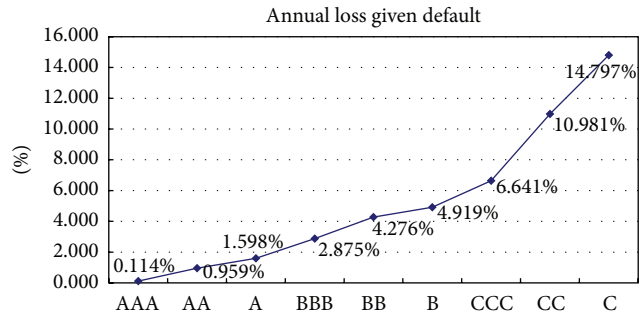


Figure 3: The distribution of the annual LGD of each credit rating.

commercial bank. The market risk capital M_c equals 0 in Row 10 (according to [26], the commercial bank, whose total cash in the trading account is 10% higher or 8.5 billion more than its total assets, can calculate their market risk capital; the market risk capital M_c equals 0). The core capital adequacy ratio β equals 6% in Row 11 [26]. The parameters in Rows

Table 3: The ascertaining of the acceptable maximum annual LGD.

(1) Index	(2) Indicator	Indicators for calculating credit threshold (3) Intermediate variable 1	(4) Intermediate variable 2	(5) Indicator data
1	Loan rate	—	—	10.5285%
2	Financial costs $r_C = 5.57\%$	Expenses of interest on deposit F_1 (3318498 thousand Yuan)	One-year deposit rate r_1	3.00%
3			Rate of legal deposit reserve ε	16.5%
4			Deposit reserve rate r_{re}	1.62%
5		Charge against revenue F_2 (2334148 thousand Yuan)	Operating expenses E_1	2161448 (thousand Yuan)
6			Depreciation charges of fixed assets E_2	166427 (thousand Yuan)
7			Other business expenses E_3	6273 (thousand Yuan)
8		Total loan P		101399562 (thousand Yuan)
9	Target rate of return $r_P = 0.71\%$	Weighted average ROE		17.65%
10		Market risk capital M_c		0 (thousand Yuan)
11		Minimum core capital adequacy ratio β		6%
12		Deductions of core capital C_d		85762 (thousand Yuan)
13		Assets $\sum A_i$		256800388 (thousand Yuan)
14		Weight of the risk assets for loan α		1

12 and 13 are from the 2014 annual report of the Chinese commercial bank. The risk weight α equals 0 in Row 14 [26].

(a) Calculate the Ratio of Financial Costs r_C. Substitute the data of Rows 2, 3, and 4 in the fifth column of Table 3 into (13), and we get the expenses of interest on deposit F_1:

$$\begin{aligned} F_1 &= Pr_1 + \frac{P\varepsilon}{1-\varepsilon} \times r_1 - \frac{P\varepsilon}{1-\varepsilon} \times r_{re} \\ &= 101399562 \times 3.00\% + \frac{101399562 \times 16.5\%}{1-16.5\%} \\ &\quad \times 3.00\% - \frac{101399562 \times 16.5\%}{1-16.5\%} \times 1.62\% \\ &= 3318498 \end{aligned} \tag{22}$$

thousand Yuan, which is listed in the third column of Table 3.

Substitute the data in Rows 5, 6, and 7 in the fifth column of Table 3 into (14), and we get the charge against revenue $F_2 = E_1 + E_2 + E_3 = 2161448 + 166427 + 6273 = 2334148$ thousand Yuan, which is listed in the third column of Table 3.

Substitute the expenses of interest on deposit $F_1 = 3318498$ thousand Yuan, the charge against revenue $F_2 = 2334148$ thousand Yuan, and the total loan $P = 101399562$ thousand Yuan into (7), and we get the financial costs $r_C = (F_1 + F_2)/P = 5.57\%$, which is listed in the second column of Table 3.

(b) Calculate the Minimum Target Rate of Return r_P. Substitute the data in Rows 8–14 in the fifth column of Table 3 into (19), and we have

$$\begin{aligned} r_P &= \mathrm{ROE} \times \left[\left(\alpha + 12.5 \frac{M_c}{\sum A_i} \right) \beta_{\min} + \frac{C_d}{\sum A_i} \right] \\ &= 0.1765 \times \left[\left(1 + 12.5 \times \frac{0}{256800388} \right) \times 0.04 \right. \\ &\quad \left. + \frac{85762}{256800388} \right] = 0.71\%, \end{aligned} \tag{23}$$

which is listed in the second column of Table 3.

(c) Calculate the Acceptable Maximum Annual LGD (i.e., r_D). Substitute $r = 10.5285\%$, $r_C = 5.57\%$, and $r_P = 0.71\%$ into (21), and then $r_{D1} = r - r_C - r_P = 10.5285\% - 5.57\% - 0.71\% = 4.25\%$ can be obtained. That is to say, on the premise of guaranteed target profit, the acceptable maximum annual LGD is 4.25%.

Substitute $r = 10.5285\%$, $r_C = 5.57\%$, $r_P = 0\%$ in break-even conditions into (21), and we get the acceptable maximum annual LGD r_D of loan in break-even conditions: $r_{D2} = r - r_C - r_P = 10.5285\% - 5.57\% - 0\% = 4.96\%$. It means

Table 4: The decision assessment results for farmers' microfinance.

(1) Credit rating	(2) Annual LGD_j	(3) Contrast 1	(4) Loan or not for contrast 1	(5) Contrast 2	(6) Loan or not for contrast 2
AAA	0.114%	0.114% < 4.25%	Agree to lend	0.114% < 4.96%	Agree to lend
AA	0.959%	0.959% < 4.25%		0.959% < 4.96%	
A	1.598%	1.598% < 4.25%		1.598% < 4.96%	
BBB	2.875%	2.875% < 4.25%		2.875% < 4.96%	
BB	4.276%	4.276% > 4.25%	Reject lending	4.276% < 4.96%	Agree to lend
B	4.919%	4.919% > 4.25%		4.919% < 4.96%	
CCC	6.641%	6.641% > 4.25%	Reject lending	6.641% > 4.96%	Reject lending
CC	10.981%	10.981% > 4.25%		10.981% > 4.96%	
C	14.797%	14.797% > 4.25%		14.797% > 4.96%	

that, on the premise of break-even, the acceptable maximum annual LGD is 4.96%.

(3) Make the Credit Decision of Farmers' Microfinance. (a) The annual LGD of farmers in Ratings AAA, AA, A, and BBB are 0.114%, 0.959%, 1.598%, and 2.875%, respectively, which are lower than the acceptable maximum annual LGD 4.25% on the premise of guaranteed target profit, as shown in the third column of Table 4. Therefore, the target profit of the bank can be ensured when the bank issues loans to the farmers in those four credit ratings.

(b) The annual LGD of farmers in Ratings BB and B are 4.276% and 4.919%, respectively, which are higher than the acceptable maximum annual LGD 4.25% on the premise of guaranteed target profit whereas they are lower than the acceptable maximum annual loss given default 4.96% on the premise of break-even, as shown in the fifth column of Table 4. Therefore, the break-even of the bank, rather than the target profit, can be ensured when the bank issues loans to the farmers in Credit Ratings BB and B.

(c) The annual LGD of farmers in Ratings CCC, CC, and C are 6.641%, 10.981%, and 14.797%, respectively, which are higher than the acceptable maximum annual loss given default 4.96% on the premise of break-even. Therefore, the bank is likely to suffer a loss when it issues loans to the farmers in Credit Ratings CCC, CC, and C.

4. Conclusions

Decision assessment theory is one of the hot topics in the literature. Lots of decision-making optimization and evaluation methods have been proposed in the existing studies. Nonetheless, only a small part of them have been widely used in reality, especially in credit risk evaluation of commercial banks. In order to help the bankers or credit clerks to locate the good customers which can realize the bank's target profit, we conduct the credit risk decision evaluation of commercial banks. First of all, this paper creates an optimal credit rating model which consisted of an objective function and two constraint conditions. In the model, the first constraint condition of the strictly increasing LGDs eliminates the unreasonable phenomenon that the higher the credit rating is, the higher the LGD is. Secondly, the paper proposes a credit decision-making assessment method on the basis of measuring the acceptable maximum LGD of commercial banks. Finally, the proposed model is verified by the data from 2817 farmers' microfinance of a large-scale Chinese commercial bank. Empirical analysis results are provided below. (1) The proposed approach can accurately find out the good customers from all of the customers who applied for loans. (2) The target profit of bank can be ensured when the bank issues loans to the farmers in Ratings AAA, AA, A, and BBB. (3) The bank is likely to suffer a loss when it issues loans to the farmers in Ratings CCC, CC, and C.

The contributions of the paper lie in three aspects. First, the decision assessment of the credit risk of commercial banks from the loss given default (LGD) of customers, the break-even point of customers' loan, and the target profit of commercial bank perspective, instead of a more traditional credit risk perspective, seems to offer a new insight into the credit rating of customers. Second, this paper proposes a novel decision assessment method which is suitable for customers' credit risk evaluation and credit decision. Third, our research not only has practical significance for decision assessment of the 2817 farmers' microfinance, but also provides a reference for credit risk evaluation and lending decision of customers in other commercial banks in the world.

This paper has some interesting extensions. Farmers' loan pricing can be included in the future extended model based on the proposed decision assessment approach. Moreover, the empirical results are derived from the 2817 farmers' microfinance data, which may not be generalized enough for all loan farmers in China. When investigating sustainability issues, there are also concerns about more relevant data. In addition, generalizing this decision assessment approach to other types of loan customers is not too difficult and researchers can easily conduct credit risk evaluation and credit decision through cases and empirical studies.

Competing Interests

The authors declare that there is no conflict of interests regarding the publication of the paper.

Acknowledgments

The research was supported by the National Natural Science Foundation of China (nos. 71503199, 71471027, 71373207,

71301017, and 71273037), the Project Funded by China Postdoctoral Science Foundation (nos. 2015M572608 and 2016T90957), the Shaanxi Province Postdoctoral Science Foundation Funded Project, the Basic Business Project of Humanities Social Sciences for Central University (no. 2015RWYB09), the Natural Science Basic Research Project in Shaanxi Province (no. 2016JQ7005), and the China Ministry of Education Social Sciences and Humanities Research Youth Fund Project (no. 16YJC630102). The authors thank the organizations mentioned above.

References

[1] T. R. Hawkins, B. Singh, G. Majeau-Bettez, and A. H. Strømman, "Comparative environmental life cycle assessment of conventional and electric vehicles," *Journal of Industrial Ecology*, vol. 17, no. 1, pp. 53–64, 2013.

[2] A. Chen, H. Yang, H. K. Lo, and W. H. Tang, "Capacity reliability of a road network: an assessment methodology and numerical results," *Transportation Research Part B: Methodological*, vol. 36, no. 3, pp. 225–252, 2002.

[3] B. Shi, H. Yang, J. Wang, and J. Zhao, "City green economy evaluation: empirical evidence from 15 sub-provincial cities in China," *Sustainability*, vol. 8, no. 6, article 551, 2016.

[4] T. Aven, "Risk assessment and risk management: review of recent advances on their foundation," *European Journal of Operational Research*, vol. 253, no. 1, pp. 1–13, 2016.

[5] B. F. Shi, J. Wang, J. Y. Qi, and Y. Q. Cheng, "A novel imbalanced data classification approach based on logistic regression and Fisher discriminant," *Mathematical Problems in Engineering*, vol. 2015, Article ID 945359, 12 pages, 2015.

[6] E. I. Altman, "Financial ratios, discriminant analysis and the prediction of corporate bankruptcy," *The Journal of Finance*, vol. 23, no. 4, pp. 589–609, 1968.

[7] E. I. Altman, R. G. Haldeman, and P. Narayanan, "ZETATM analysis: a new model to identify bankruptcy risk of corporations," *Journal of Banking and Finance*, vol. 1, no. 1, pp. 29–54, 1977.

[8] J. H. Min and Y.-C. Lee, "A practical approach to credit scoring," *Expert Systems with Applications*, vol. 35, no. 4, pp. 1762–1770, 2008.

[9] R.-C. Hwang, H. Chung, and C. K. Chu, "Predicting issuer credit ratings using a semiparametric method," *Journal of Empirical Finance*, vol. 17, no. 1, pp. 120–137, 2010.

[10] G. M. Gupton, C. C. Finger, and M. Bhatia, *Creditmetrics: Technical Document*, J. P. Morgan & Co. Incorporated, New York, NY, USA, 1997.

[11] K. Bijak and L. C. Thomas, "Modelling LGD for unsecured retail loans using Bayesian methods," *Journal of the Operational Research Society*, vol. 66, no. 2, pp. 342–352, 2015.

[12] B. Twala, "Multiple classifier application to credit risk assessment," *Expert Systems with Applications*, vol. 37, no. 4, pp. 3326–3336, 2010.

[13] S. Finlay, "Multiple classifier architectures and their application to credit risk assessment," *European Journal of Operational Research*, vol. 210, no. 2, pp. 368–378, 2011.

[14] J. Abellán and C. J. Mantas, "Improving experimental studies about ensembles of classifiers for bankruptcy prediction and credit scoring," *Expert Systems with Applications*, vol. 41, no. 8, pp. 3825–3830, 2014.

[15] M. B. Gorzałczany and F. Rudziński, "A multi-objective genetic optimization for fast, fuzzy rule-based credit classification with balanced accuracy and interpretability," *Applied Soft Computing*, vol. 40, pp. 206–220, 2016.

[16] S. H. Han, S. X. Lu, and S. C. H. Leung, "Segmentation of telecom customers based on customer value by decision tree model," *Expert Systems with Applications*, vol. 39, no. 4, pp. 3964–3973, 2012.

[17] C. Ljungwall, X. Yi, and Y. T. Zou, "Central bank financial strength and the cost of sterilization in China," *China Economic Review*, vol. 25, no. 1, pp. 105–116, 2013.

[18] R. Florez-Lopez, "Modelling of insurers' rating determinants: an application of machine learning techniques and statistical models," *European Journal of Operational Research*, vol. 183, no. 3, pp. 1488–1512, 2007.

[19] A. Pascucci, M. Suárez-Taboada, and C. Vázquez, "Mathematical analysis and numerical methods for a PDE model of a stock loan pricing problem," *Journal of Mathematical Analysis and Applications*, vol. 403, no. 1, pp. 38–53, 2013.

[20] Y.-S. Chen and C.-H. Cheng, "Hybrid models based on rough set classifiers for setting credit rating decision rules in the global banking industry," *Knowledge-Based Systems*, vol. 39, no. 4, pp. 224–239, 2013.

[21] Y. H. Guo, W. J. Zhou, C. Y. Luo, C. R. Liu, and H. Xiong, "Instance-based credit risk assessment for investment decisions in P2P Lending," *European Journal of Operational Research*, vol. 249, no. 2, pp. 417–426, 2016.

[22] S. Jones, D. Johnstone, and R. Wilson, "An empirical evaluation of the performance of binary classifiers in the prediction of credit ratings changes," *Journal of Banking & Finance*, vol. 56, no. 4, pp. 72–85, 2015.

[23] Basel Committee on Banking Supervision, *Basel III: A Global Regulatory Framework for More Resilient Banks and Banking Systems*, Bank for International Settlements, 2010.

[24] The Postal Savings Bank of China (PSBC), *Credit Risks Evaluation and Decision System of Microcredit for Farmers for the Postal Savings Bank of China*, 2nd edition, 2011.

[25] The China Banking Regulatory Commission (CBRC), "Rules for regulating the risk rating system of joint-stock commercial banks," http://www.cbrc.gov.cn/chinese/home/docDOC_ReadView/301.html.

[26] The China Banking Regulatory Commission (CBRC), "Rules for regulating the capital adequacy requirement of commercial banks," http://www.cbrc.gov.cn/chinese/home/docView/189.html.

[27] R. M. Stein, "The relationship between default prediction and lending profits: integrating ROC analysis and loan pricing," *Journal of Banking & Finance*, vol. 29, no. 5, pp. 1213–1236, 2005.

Hindawi Publishing Corporation
Scientific Programming
Volume 2016, Article ID 3943859, 10 pages
http://dx.doi.org/10.1155/2016/3943859

Research Article

A Multiple Classifier Fusion Algorithm Using Weighted Decision Templates

Aizhong Mi, Lei Wang, and Junyan Qi

School of Computer Science and Technology, Henan Polytechnic University, Jiaozuo 454000, China

Correspondence should be addressed to Lei Wang; wang_leiqjy@163.com

Received 2 June 2016; Revised 5 August 2016; Accepted 10 August 2016

Academic Editor: Guo Chen

Fusing classifiers' decisions can improve the performance of a pattern recognition system. Many applications areas have adopted the methods of multiple classifier fusion to increase the classification accuracy in the recognition process. From fully considering the classifier performance differences and the training sample information, a multiple classifier fusion algorithm using weighted decision templates is proposed in this paper. The algorithm uses a statistical vector to measure the classifier's performance and makes a weighed transform on each classifier according to the reliability of its output. To make a decision, the information in the training samples around an input sample is used by the k-nearest-neighbor rule if the algorithm evaluates the sample as being highly likely to be misclassified. An experimental comparison was performed on 15 data sets from the KDD'99, UCI, and ELENA databases. The experimental results indicate that the algorithm can achieve better classification performance. Next, the algorithm was applied to cataract grading in the cataract ultrasonic phacoemulsification operation. The application result indicates that the proposed algorithm is effective and can meet the practical requirements of the operation.

1. Introduction

A single classifier was always used in a traditional pattern recognition system. However, in recent years, it has been found that the samples that were wrongly classified by distinct classifiers were usually not the same in many experiments. This finding means that complementary information about the object to be recognized can be potentially offered by different classifiers and effective fusion of the complementary information is expected to considerably improve the performance of a pattern recognition system. When the member classifiers are diverse or complementary, multiple classifier systems can usually obtain higher classification accuracies compared with a single classifier [1]. Through a large number of experiments and applications, it has been proven that fusing classifiers' decisions can achieve better performance than the best single classifier and improve the efficiency and robustness of a pattern recognition system. Currently, many applications areas have adopted the methods of multiple classifier fusion, such as fault diagnosis [2, 3], radar emitter recognition [4], remote sensing image recognition [5–7], medical diagnosis [8], face recognition [9], and intrusion detection [10]. Some of the fusion methods have good generality and show good classification performance in certain applications. However, similar to the classifier, there is no fusion method that can obtain the optimal classification performance for all of the applications. Therefore, the study of multiple classifier fusion is still an open problem. For a specific application, this approach requires further research on a more suitable algorithm for multiple classifier fusion.

Multiple classifier fusion assumes that all of the classifiers are equally "experienced" over the entire feature space. Thus, all of the outputs of the classifiers are fused in a certain way to achieve the final decision. According to the different outputs of the classifiers, they can be further divided into three categories. When the output is shown in a decision form, Naive Bayes [11], majority voting, and Behavior Knowledge Space (BKS) [12] are the representative fusion algorithms. When the output is shown in a ranked form, Borda Count [13, 14] is a typical fusion algorithm. When the output is shown in a measured form, the fusion algorithms mainly include Max, Min, Sum, Product, Median, Support Vector Machine (SVM)

fusion [15–17], Neural Networks fusion [18], Dempster-Shafer theory [19], and decision templates (DT) [20].

The decision templates (DT) algorithm is a simple and effective fusion algorithm that has measure outputs and that is researched and used widely [21–26]. During the training stage, the DT algorithm calculates the average of the decision profiles that correspond to the training samples that belong to each class according to the outputs of all of the classifiers. The averages are the decision templates (one per class). Then, the input sample's class is determined by evaluating the similarity between its decision profile and various decision templates. The algorithm has the following advantages [27]: there is a simple training process; it requires no strict assumptions compared to probability-based algorithms; it is less sensitive to the size of the training set than other algorithms and overtraining rarely happens; it is also intuitive, having a small number of calculations; and it is not time-consuming. However, the DT algorithm still has two problems: (1) a decision template is only the average of the decision profiles that belong to a class and does not fully reflect the differences in the classifier's performance. (2) It uses only the classification information of the decision template and does not take full advantage of the training samples' information.

Therefore, an improved DT algorithm (called VWDT) is proposed here. This algorithm measures the classifier's performance by a statistical vector and assigns different weights to each classifier according to their outputs. Reliable output is assigned to a large weight in such a way that the output represents a significant share of the decision templates. For a sample that is easily misclassified, the information in the surrounding training samples is used to make a decision in addition to using the similarities between it and the decision templates.

This algorithm was compared with the DT algorithm on 15 data sets, from the KDD'99, UCI, and ELENA databases. The experimental results show that the algorithm can achieve better classification performance. Then, the algorithm was applied to cataract grading in the cataract ultrasonic phacoemulsification operation. The application effect indicates that the proposed algorithm is effective and can meet the practical requirements of the operation. Thus, the algorithm can be applied to the computer-aided cataract recognition system of the ultrasonic emulsification apparatus to automatically recognize the hardness grade of the cataract, which thus lowers the operation difficulty level, shortens the study period, and improves the safety of the operation.

The organization of this paper is as follows: Section 2 contains the description of the multiple classifier fusion algorithm proposed in this paper. The experiment test of the algorithm is given in Section 3, and Section 4 provides the validation of the practical application, including the application background and the application description and effect. Finally, Section 5 presents the conclusions and summarizes the contents of this paper.

2. VWDT Algorithm Description

2.1. Relative Expression. The structure of the multiple classifier fusion used in this paper is shown in Figure 1.

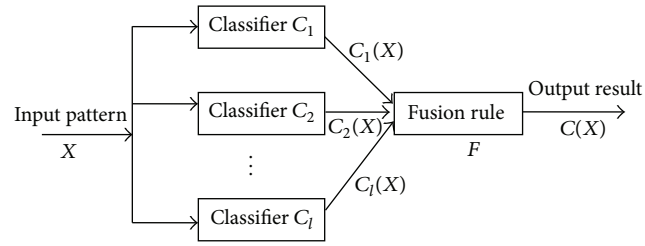


Figure 1: The structure of the multiple classifier fusion.

Let R^n be the n-dimensional feature space; let $X = [x_1, x_2, \ldots, x_n]^T$ be the n-dimensional feature vector, $X \in R^n$; let $\Omega = \{\omega_1, \omega_2, \ldots, \omega_m\}$ be the set of potential class labels; and let $C = \{C_1, C_2, \ldots, C_l\}$ be the set of trained classifiers for decision fusion. Given the input pattern X, the output of the ith classifier is denoted as in

$$C_i(X) = \left[c_{i,1}(X), c_{i,2}(X), \ldots, c_{i,m}(X)\right]^T, \quad (1)$$

where $c_{i,j}(X)$, $i = 1, 2, \ldots, l$, $j = 1, 2, \ldots, m$, represents the measure value of the possibility that classifier C_i considers that X belongs to class ω_j.

The fused output of l classifiers is constructed as in

$$C(X) = F\left(C_1(X), C_2(X), \ldots, C_l(X)\right), \quad (2)$$

where F is the fusion rule.

The output of all of the classifiers can be represented as the decision profile, $l \times m$ matrix as shown in

$$\mathrm{DP}(X) = \begin{bmatrix} c_{1,1}(X) & \cdots & c_{1,j}(X) & \cdots & c_{1,m}(X) \\ \vdots & & \vdots & & \vdots \\ c_{i,1}(X) & \cdots & c_{i,j}(X) & \cdots & c_{i,m}(X) \\ \vdots & & \vdots & & \vdots \\ c_{l,1}(X) & \cdots & c_{l,j}(X) & \cdots & c_{l,m}(X) \end{bmatrix}. \quad (3)$$

The ith row is the measure layer output of the ith classifier $C_i(X)$, which is given according to (1). The jth column is the possibility measure value that l classifiers consider that the input pattern X belongs to class ω_j. In addition, the fusion result $C(X)$ is an m-dimensional vector that is represented by the measure layer form, which is denoted as in

$$C(X) = \left[d_1(X), d_2(X), \ldots, d_m(X)\right]^T, \quad (4)$$

where $d_i(X)$, $i = 1, 2, \ldots, m$, shows the possibility measure value that X belongs to class ω_i after fusion.

After acquiring the fusion result of (4), a certain rule is used to judge which class the input pattern belongs to. Typically, the rule of using the maximum value is adopted; namely, if $d_K(X) = \max_{i=1}^{m} d_i(X)$, then it shall be deemed that $X \in \omega_K$.

The simple fusion rules (Max, Min, Sum, Product, and Median) acquire the system output by operating each column of DP(X).

(1) Max rule is

$$d_j(X) = \max_{i=1}^{l} c_{i,j}(X), \quad j = 1, 2, \ldots, m. \tag{5}$$

The Max rule takes the maximum of every DP(X) column as the fused output $C(X)$.

(2) Min rule is

$$d_j(X) = \min_{i=1}^{l} c_{i,j}(X), \quad j = 1, 2, \ldots, m. \tag{6}$$

The Min rule takes the minimum of every DP(X) column as the fused output $C(X)$.

(3) Sum rule is

$$d_j(X) = \sum_{i=1}^{l} c_{i,j}(X), \quad j = 1, 2, \ldots, m. \tag{7}$$

The Sum rule computes the sum of every DP(X) column as the fused output $C(X)$. It is also called the mean rule when it computes the mean; these are simply two forms of the same rule.

(4) Product rule is

$$d_j(X) = \prod_{i=1}^{l} c_{i,j}(X), \quad j = 1, 2, \ldots, m. \tag{8}$$

The Product rule computes the product of every DP(X) column as the fused output $C(X)$.

(5) Median rule is

$$d_j(X) = \operatorname*{median}_{i=1}^{l} c_{i,j}(X), \quad j = 1, 2, \ldots, m. \tag{9}$$

The Median rule takes the median of every DP(X) column as the fused output $C(X)$. If l is an even number, then the mean of two medians is taken as the result of a column.

The DT algorithm acquires the system output using the entire DP(X) and calculates the corresponding decision template DT_i, $i = 1, 2, \ldots, m$, on the training set Z, as in

$$\mathrm{DT}_i = \frac{1}{N_z} \sum \mathrm{DP}\left(z_j\right), \quad z_j \in Z,\ z_j \in \omega_i, \tag{10}$$

where z_j represents the samples that belong to class ω_i in the training set Z and N_z represents the number of z_j.

Then, the input sample's class is determined by evaluating the similarity between its decision profile and various decision templates. Here, the squared Euclidean distance is used for calculating the similarity, but other measures can also be applied. For a training sample $P \in Z$, the calculation equation is shown in

$$d_E\left(\mathrm{DP}(P), \mathrm{DT}_i\right) = \sum_{j=1}^{m} \sum_{k=1}^{l} \left(c_{k,j}(P) - \mathrm{dt}_i(k, j)\right)^2, \tag{11}$$

where $\mathrm{dt}_i(k, j)$ represents the element at the intersection of the kth row and jth column in DT_i.

2.2. Algorithm Idea. Due to the existence of the performance differences of the classifiers, even using the same classifier, the distinguishing capacity for different classes of data in a data set is different. When the classifier output reserves detailed information about a class, the DT algorithm's performance will be quite good. However, when the classifier is very sensitive to certain features of the input space, the result is a major change to certain information in the classes in the output space, and the performance of the algorithm is reduced to an obvious extent. For this problem, the VWDT algorithm measures the performance of various classifiers by a statistical vector and self-adaptively assigns weights for various classifiers according to the specific output status of the classifiers. The output of a classifier that has better performance occupies a larger proportion in the constructing process of the decision template, to acquire a more reliable decision template and improve the classification accuracy.

For certain samples, even being compared with the samples in the same class, they still have obvious differences. These samples are often the outliers of each class of samples, and their decision profiles are quite different from most of the samples. When they are rare, in the DT algorithm, the decision template acquired by the calculation is quite different from their decision profiles. Moreover, the DT algorithm evaluates only the similarity between the decision profile and decision template, which makes some samples from various classes misclassified due to having a small similarity or the fact of being at an overlapping region of multiple classes. With respect to this problem, the VWDT algorithm takes advantage of the information in various training samples that are around the test sample. It searches for several nearest-neighbors of the test sample and combines the training sample's information and the calculation of the decision template similarity together, thus avoiding the deviation that results from total dependence on the decision template to some extent and making the final class judgment more reliable.

2.3. Algorithm Steps. Firstly, a statistical vector is used to measure the performance of various classifiers, and the decision profile takes a weighed transform.

The classification error of the kth classifier C_k can be represented by an m-dimensional statistical vector V^k, as in

$$V^k = \left[n_1^k, n_2^k, \ldots, n_m^k\right]^T, \tag{12}$$

where the element n_i^k, $i = 1, 2, \ldots, m$, represents the number of training samples that are in the ith class that are correctly recognized by classifier C_k. Let the total number of training samples be N. The vector V^k is normalized by dividing

by N. At this time, the meaning of each element changes into its corresponding percentage. The normalized vector is represented as in

$$V^{k/N} = \left[n_1^{k/N}, n_2^{k/N}, \ldots, n_m^{k/N}\right]^T. \quad (13)$$

Here, the reliability of the output vector of classifier C_k can be weighed by $n_h^{k/N}$, and the value of h should satisfy the following condition, as shown in

$$c_{k,h} = \max_{j=1}^{m}\left(c_{k,j}\right). \quad (14)$$

Thus, the reliable output of classifier C_k can be represented as in

$$C_k'(X) = n_h^{k/N} * C_k(X). \quad (15)$$

From (1), (15) is denoted as in

$$\begin{aligned}&\left[c_{k,1}'(X), c_{k,2}'(X), \ldots, c_{k,m}'(X)\right]\\ &\quad = n_h^{k/N} * \left[c_{k,1}(X), c_{k,2}(X), \ldots, c_{k,m}(X)\right].\end{aligned} \quad (16)$$

According to the reliable output vectors of all of the classifiers calculated by (16), the new decision profile $\mathrm{DP}'(X)$ can be acquired as in

$$\mathrm{DP}'(X) = \begin{bmatrix} c_{1,1}'(X) & \cdots & c_{1,j}'(X) & \cdots & c_{1,m}'(X) \\ \vdots & & \vdots & & \vdots \\ c_{i,1}'(X) & \cdots & c_{i,j}'(X) & \cdots & c_{i,m}'(X) \\ \vdots & & \vdots & & \vdots \\ c_{l,1}'(X) & \cdots & c_{l,j}'(X) & \cdots & c_{l,m}'(X) \end{bmatrix}. \quad (17)$$

Using $\mathrm{DP}'(X)$ to replace $\mathrm{DP}(X)$ and adopting the Euclidean distance as the similarity measure, the VWDT algorithm is described as follows.

2.3.1. Training Process

(1) Use each classifier to classify the training samples, acquire each classifier's statistical vector V^k, $k = 1, 2, \ldots, l$, and construct the set D that corresponds to the decision profile $\mathrm{DP}'(X)$; from (10), the corresponding decision template DT_i', $i = 1, 2, \ldots, m$, can be acquired, as shown in

$$\mathrm{DT}_i' = \frac{1}{N_z}\sum \mathrm{DP}'\left(z_j\right), \quad z_j \in Z,\ z_j \in \omega_i. \quad (18)$$

(2) For a training sample $P \in Z$ and from (11), the square value of the Euclidean distance between the decision profile $\mathrm{DP}'(P)$ and DT_i' can be acquired, as shown in

$$d_E\left(\mathrm{DP}'(P), \mathrm{DT}_i'\right) = \sum_{j=1}^{m}\sum_{k=1}^{l}\left(c_{k,j}'(P) - \mathrm{dt}_i'(k, j)\right)^2. \quad (19)$$

(3) Calculate the possibility measure values of training sample P belonging to various classes, as shown in

$$d_i(P) = 1 - \frac{1}{l \times m} d_E\left(\mathrm{DP}'(P), \mathrm{DT}_i'\right). \quad (20)$$

If $d_K(P) = \max_{i=1}^{m} d_i(P)$, then it shall be deemed that $P \in \omega_K$.

(4) Repeat steps (2)-(3) until all of the training samples are classified. According to the classification results, acquire a weighted decision profile set D_e ($D_e \subseteq D$) that corresponds to the training samples that are misclassified.

(5) Set $D_e \neq \Phi$, calculate the average values of all of the elements in set D_w as an additional decision template, which shall be denoted as DT_{m+1}', and use the set D_e to train a k-NN classifier C_{knn}.

(6) Return all of the decision templates DT_i', $i = 1, 2, \ldots, m+1$, the statistical vector V^k, $k = 1, 2, \ldots, l$, and the classifier C_{knn}.

2.3.2. Classification Process

(1) For an input pattern X, use the statistical vectors that are acquired in the above training process to construct its decision profile $\mathrm{DP}'(X)$.

(2) Calculate the squared value of the Euclidean distance between $\mathrm{DP}'(X)$ and DT_i' by (19), and, then, calculate the possibility measure value $d_i(X)$, $i = 1, 2, \ldots, m+1$, of the various classes that the sample X belongs to by (20).

(3) Judge the class of X:

(a) Let $d_K(X) = \max_{i=1}^{m+1}(d_i(X))$; if $K \leq m$, then it shall be deemed that $X \in \omega_K$.

(b) Otherwise, use C_{knn} to classify X, and let its measure layer output be $[c_1(X), c_2(X), \ldots, c_m(X)]^T$ if $d_K(X) = \max_{i=1}^{m}(d_i(X) + c_i(X))$, $X \in \omega_K$.

The algorithm measures the classifier's performance by a statistical vector, which is acquired from the statistical data on the training sample set. Different weights are self-adaptively assigned to the classifiers' output vectors according to their judgments, which makes the decisions from different classifiers and the decisions of different classes from a classifier have different proportions in the decision templates. The weights can be acquired independently according to the prior information of the training samples and the classifiers' decisions on the current input sample. According to the statistical vector, if a classifier has better classification performance for a certain class of samples, then its output can be considered to be more reliable when it judges that the current input sample belongs to this class, and it is assigned a larger weight. If so, more reliable output has a higher proportion in the decision templates, which makes them more reliable. The training samples that are wrongly classified by the multiple

classifier system are used to construct an additional decision template. When an input sample is nearest to the center of this template, it can be reasonably considered that the sample is easily misclassified by the decision templates. At this time, the nearest-neighbor rule is adopted to make the judgment, which combines the information in the training samples near the input sample and the distance data between the input sample and various templates, which thus improves the accuracy.

3. Experimental Analyses

Currently, there are many practical classification algorithms, such as the nearest-neighbor algorithm, linear classifiers, the minimum distance algorithm, support vector machines, and artificial neural networks. The nearest-neighbor algorithm is very intuitive, wherein the basic idea is to find a sample from the training set that has the minimum distance to the object to be recognized and take the sample's class as the recognition result. The k-nearest-neighbor algorithm, as a typical representative of nonparametric algorithms, is used as the base classifier in this paper because it is simple and reliable. The experiments in this paper were realized with Matlab. During the programming process of the specific codes, the pattern recognition toolbox "PRTools4" [28] was used. In the experiments, the code of the k-nearest (k-NN) classifier is from the toolbox. To simplify the experiments, the parameter k is selected automatically and optimally by the leave-one-out method. The measure of the classifier's output is the posterior probability. To sufficiently test the performance and generality of the proposed algorithm, 15 data sets from the KDD'99, UCI, and ELENA databases were selected for the experiments. These databases are comparatively authoritative test data from related research fields, and the selected data sets show relatively good representation.

In the experiment results, if the sample data has separate training set and test set and the test result is unchanged, the statistical equation of the error rate is shown in

$$E = \frac{N_e}{N_t}, \tag{21}$$

where E represents the error rate, N_e represents the number of misclassified test samples, and N_t represents the number of test samples.

If the sample data has only one set and M-fold cross-validation is used, the test result is the average of the error rates of M validations, and the statistical equation of the average error rate is shown in

$$\overline{E} = \frac{1}{M}\sum_{i=1}^{M} E_i, \tag{22}$$

where $\overline{E}$ represents the average error rate.

3.1. Using the KDD'99 Data. The samples are from the KDD'99 intrusion detection database (http://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html). The available database includes a large number of network connection patterns that were extracted from TCP/IP packets. Each connection pattern is a 41-dimensional feature vector and belongs to one out of five classes (i.e., normal, R2L attacks, DoS attacks, U2R attacks, and probing attacks). The feature can be divided into three categories: intrinsic feature, traffic feature, and content feature.

Table 1: Description of the KDD data sets.

Data set name	Number of classes	Number of samples	Number of features
ftp5c	5	A 800	29
ftp2c	2	B 5212	
http4c	4	A 410	28
http2c	2	B 3380	

Table 2: Performance on the KDD data sets (in %).

Data set name	Error rate	
	DT	VWDT
ftp5c	5.794	5.411
ftp2c	1.055	1.036
http4c	3.047	0.237
http2c	0.237	0.148

The connection patterns related to the ftp service were selected from two data sets with corrected labels. Each pattern was processed by dimension reduction (these features whose values were always constants were abandoned) and quantification of the symbolic features and normalization (a linear transformation was applied to make each feature within the range $[0, 1]$). Training set A and test set B were obtained correspondingly. This group of data sets with five-class data was denoted by ftp5c. Combining the four attack classes as an abnormal class, a group of data sets with two-class data was obtained and denoted by ftp2c.

Through a similar process, the connection patterns that were related to the http service were also selected. Two groups of data sets were acquired, which were denoted by http4c and http2c. The former was four-class data, and the latter was two-class data. Because the number of http connection patterns was large, a restricted condition in the selection process was added: the duration was above zero. Table 1 is the description of the KDD data sets. Three k-NN classifiers were trained using a distinct category of features on each training set A and constituted the classifier set. The original test results of the DT and VWDT algorithms are shown in Table 2 and Figure 2 (ER represents the error rate).

It can be observed that on the four data sets, the performance of the VWDT algorithm is better than that of the DT algorithm. Especially on the multiple-class data sets of ftp5c and http4c, the error rate of the algorithm is reduced more obviously. The reason is that the more the classes that exist, the more obvious the difference that is in the classifier's performance to various classes, which is reflected in the decision template, and the reliability of the decision template is higher. At the same time, the difference between the misclassified samples and other samples is more

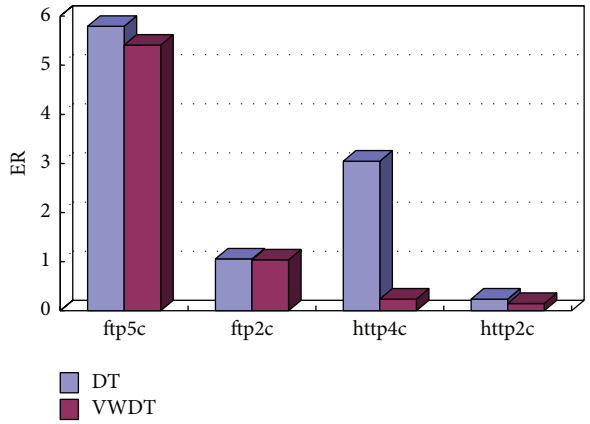


Figure 2: Performance comparison on the KDD data sets.

Table 3: Description of the UCI data sets.

Data set name	Number of samples	Number of features	Number of classes
Glass	214	9	6
Bupa	345	6	2
Pima	768	8	2
Thyroid	215	5	3
Wine	178	13	3
German	1000	24	2
Heart	270	13	2
Vehicle	846	18	4

obvious, and the complementarity is better by combining the information in the training samples around the input sample and the calculated similarity between the decision templates and the input sample's decision profile.

3.2. Using the UCI Data. A total of 8 data sets from the UCI machine learning database (http://www.ics.uci.edu/~mlearn/MLRepository.html/) were selected for experimental analysis. Due to the study demand for the algorithms, the selected data sets were all in digital form without errors or missing feature values. Normalization was performed on the selected data. The basic information about the selected data sets is in Table 3. The name of each data set is only the first word of its name in the original database.

For each data set, 6-fold cross-validation was used: the data set was divided into 6 subsets with similar size and distribution. Every time, a subset was selected as the test set, and the remainder comprised the training set. On each training subset, a k-NN classifier was trained. The corresponding 5 classifiers constituted the classifier set for the fusion. The test results are the average of 6 times (see Table 4 and Figure 3; AER represents the average error rate).

It can be observed that, of the 8 data sets, the VWDT algorithm is better than the DT algorithm on 7 of the data sets. On the data sets (thyroid, wine, and vehicle), which have more classes, the performances of the VWDT algorithm are improved more obviously. However, the data set of glass is

Table 4: Performance on the UCI data sets (in %).

Data set name	Average error rate	
	DT	VWDT
Glass	41.367	42.538
Bupa	24.751	20.526
Pima	26.042	23.828
Thyroid	14.484	3.320
Wine	6.825	2.857
German	27.564	24.154
Heart	17.407	15.926
Vehicle	33.394	26.195

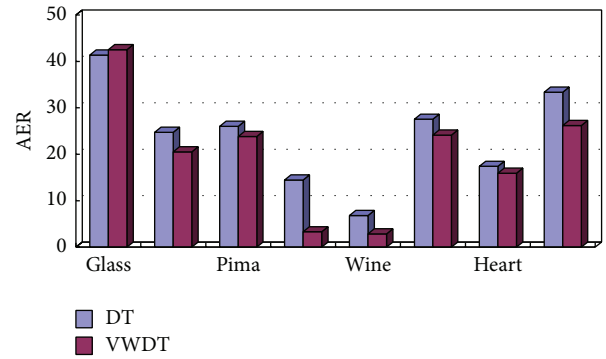


Figure 3: Performance comparison on the UCI data sets.

an exceptional instance, mainly because there are not many samples but lots of classes in the data set, and the distributions of the samples of the different classes are obviously varied. Under such a situation, taking the samples' decision profile average of a class as the representative of the samples of that class could result in a larger error, which is also proven by the DT algorithm's performance. For the VWDT algorithm, an additional decision template is divided, which further lowers the performance. However, it can be observed that, on the data set of glass, compared with the DT algorithm, the error rate increase of the VWDT algorithm is not obvious (2.83%). On the other data sets, the minimum error rate reduction of this algorithm is 8.5% (data set pima).

3.3. Using the ELENA Data. A total of 3 data sets from the ELENA database (https://www.elen.ucl.ac.be/neural-nets/Research/Projects/ELENA/elena.htm) were selected for experimental analysis. Normalization was also made for the selected data. The basic information of the selected data sets is in Table 5. For each data set, 6-fold cross-validation was also used, and the test results are in Table 6 and Figure 4.

It can also be observed that the VWDT algorithm is better than the DT algorithm. On the data set of satimage, which has more classes, the performance of the VWDT algorithm is improved more obviously.

The experiment results show that the VWDT algorithm is effective for improving the classification performance. It must be noted that k-NN classifiers are adopted in the experiments of this paper. When the classifier set is composed

Table 5: Description of the ELENA data sets.

Data set name	Number of samples	Number of features	Number of classes
Clouds	5000	2	2
Satimage	6435	36	6
Phoneme	5404	5	2

Table 6: Performance on the ELENA data sets (in %).

Data set name	Average error rate	
	DT	VWDT
Clouds	10.808	7.205
Satimage	9.925	4.659
Phoneme	8.315	5.387

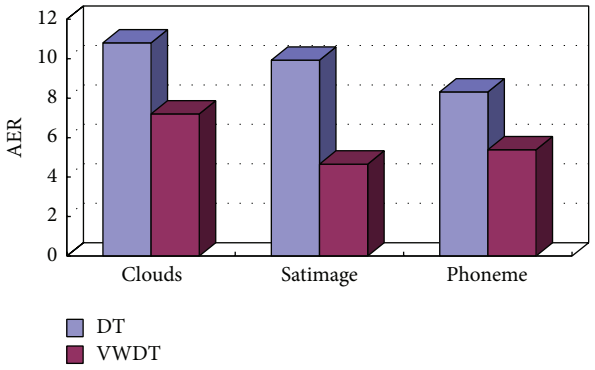


Figure 4: Performance comparison on the ELENA data sets.

of classifiers of a different type, because the physical meanings and measures of the classifiers' outputs are different, the performances of both the DT and VWDT algorithms could be obviously impacted. At this time, the outputs of all of the classifiers should be transformed into unified reliability.

4. Practical Application

4.1. Application Background. Age-related cataract is the leading cause of blindness in the world. Now, there are 25 million people with cataract-induced blindness, and, among them, almost 2.5 million are in China and account for 10 percent. Because the population is aging, there would be more than one million cataract patients yearly in China. Until now, an operation is the only effective method to cure cataract blindness. However, a cataract operation could not fully treat the newly increased patients, and, thus, the future appears to be rather gloomy for China's cataract blindness. The ultrasonic phacoemulsification operation is considered to be the main treatment for cataracts and is widely accepted because of its small cut, short operation time, including no suture, no need of being in hospital, and no limitation for any activities, and fast recovery of eyesight.

Before performing the operation, strict training shall be conducted for the doctor in charge of a case, to make the

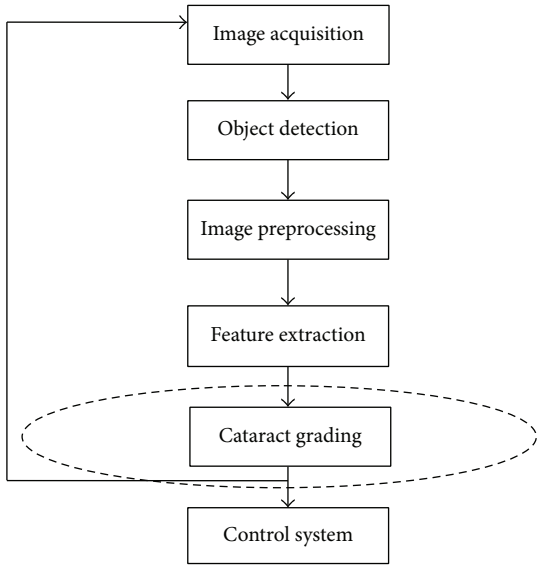


Figure 5: Module structure of the cataract recognition system.

doctor very familiar with the cataract grading. However, because the hardness grades of the cataracts are different for different individuals, the problem of "how to correctly implement a proper oscillation frequency according to a different lens nucleus" requires long-term practice for the operation, to accumulate rich operation experience toward reaching the best operation outcome. The operation is impossible to promote over a large area because of its low automatization, operation difficulty, and long learning cycle. That is the vital reason why this operation cannot be popularized.

Along with the development of information technology, computer-aided healthcare by integrating medical devices and healthcare information systems to improve healthcare quality and productivity is receiving more and more attention. In particular, a computer-aided healthcare system could provide a solution in the developing areas where medical resources are scarce [29]. Image processing and pattern recognition technologies can be applied to the computer-aided cataract recognition system of the ultrasonic emulsification apparatus, to automatically recognize the hardness grade of the cataract and, thus, to lower the operation's difficulty level, shorten the study period, and improve the safety of the operation.

Such a cataract recognition system is mainly composed of five modules, as shown in Figure 5 (the position of the research content of this paper is in the dash-dotted ellipse). The image acquisition module has a real-time operation video via camera on the operating table, achieves single-frame images by sampling at a certain frequency, and switches them into formats that can be processed by the system; the object detection module detects the emulsification pinhead from the image and locates the image area for recognition, which is the first key problem of the system; the image preprocessing module conducts the preprocessing for the image area, such as filtering and denoising, to eliminate the noise; and the feature extraction module extracts the features of the image

area. The extracted features should effectively represent the tissue and reserve the original information of the emulsification pinhead, cataract tissues, and normal tissues as much as possible; the cataract grading module adopts a certain classification algorithm according to the extracted features to recognize the tissue and submits the result to the control system. Finding a suitable recognition algorithm to ensure accurate judgment for the cataract grading is the second key problem of the system. If the recognition result is a cataract with certain hardness, the control system will emit different signals to control the ultrasonic frequency according to different hardness grades. Otherwise, no ultrasonic signal can be sent out.

4.2. Application Description. The cataract grading is mainly based on Emery and Little's grade standard, which judges the color of the lens nucleus to grade its hardness according to the examination results under the slit-lamp. Human eyes can directly distinguish normal tissue from diseased tissue and recognize different hardness of the lens nucleus based on the appearance of different colors. Considering the practical situation of lens nucleus recognition, color is selected as the image feature for the system, and the RGB color model is used in this paper.

The image of the area near the emulsification pinhead was partitioned into $m \times n$ pieces, and the color value of a piece was the average of all of the pixels in the piece. This paper selected $m = 15$ and $n = 5$ as the parameters. Thus, each image was partitioned into 75 pieces, and each piece had a similar number of pixels. The average R, G, and B value of all of the pixels in a piece were separately calculated as the feature value of that piece. In other words, every image corresponds to a 225-dimensional feature vector.

The software development environment of the recognition system was Visual C++. On the training set, three k-NN classifiers were trained separately by using R, G, and B features; that is, the input of each classifier is a feature vector of 75 dimensions. The decisions of three single feature classifiers were fused by the DT and VWDT algorithms. For a comparison, a k-NN classifier was trained by using the RGB feature; that is, the input of the classifier is a 225-dimensional feature vector. To test the effect of the parameter k on the results, the parameter is selected to be $k = 3, 5, 8$. In the VWDT algorithm, the parameter k of classifier C_{knn} is selected to be $k = 1$.

4.3. Application Effect. The cataract images were intercepted from the operation videos provided by Beijing Tongren Hospital. The size of the image had three categories: 20×20, 30×30, and 40×40. The training image set had 649 images, and the test image set had 647 images. Each set had 6 types of images, normal tissues, and cataracts with I to V hardness grades. The class of each image was confirmed by the ophthalmologists from the hospital. The numeric distribution of the different classes of images is shown in Table 7. The original application results are shown in Table 8, and the average error rates (AER) are shown in Figure 6.

It can be observed that the classifiers that use the R feature have the highest error rates. The classifiers that use the G feature have the lowest error rate among the single color feature classifiers, which are slightly lower than the classifiers that use the RGB feature. The error rate of the VWDT algorithm is lower than that of all of the single classifiers and the DT algorithm, which shows the effectiveness of the algorithm again. Among the images that were misclassified by all of the algorithms, there is no normal tissue, and more than 80% of them are divided into lower grates. This result is ideal, and it will not cause adverse effects on the patients. The doctor must only properly increase the output signal intensity according to experience. The VWDT algorithm has the highest average recognition rate, 94.75%, which reaches the actual demand of the cataract ultrasonic phacoemulsification operation.

When the parameter k takes on different values, the error rates change accordingly but not very obviously (as shown in Figure 7). When $k = 5$, the error rates of all of the algorithms (except for the VWDT algorithm, when $k = 3$) are slightly lower. Thus, in practical applications of the VWDT algorithm, the parameter k can be determined by experience (a fifteenth of the number of training samples is a suggestion).

Table 7: Number distribution of the different classes of images.

Class of the tissue	Number of training images	Number of test images
Normal	95	91
I	293	296
II	184	183
III	37	33
IV	28	30
V	12	14

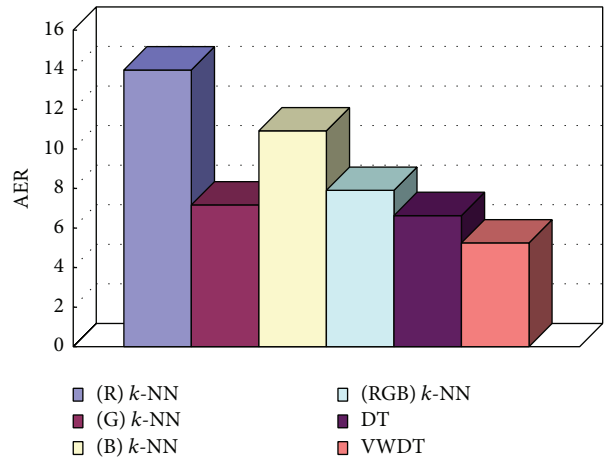


Figure 6: Comparison of the average error rates.

5. Conclusions

In this paper, a multiple classifier fusion algorithm is proposed and applied for recognition of the hardness of a cataract lens nucleus. The algorithm considers the difference in the classifiers' performances and makes full use of the training

Table 8: Application results (in %).

Feature	Method	Error rate	Method	Error rate	Method	Error rate
R	3-NN	13.75	5-NN	13.75	8-NN	14.47
G	3-NN	7.48	5-NN	6.88	8-NN	7.15
B	3-NN	10.96	5-NN	10.12	8-NN	11.65
RGB	3-NN	8.36	5-NN	6.98	8-NN	8.36
	DT	6.72	DT	6.44	DT	6.72
	VWDT	4.95	VWDT	5.26	VWDT	5.54

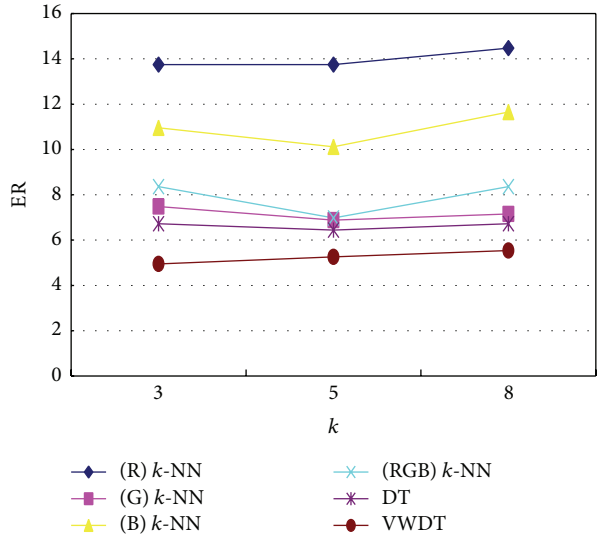


Figure 7: Comparison of the error rates.

sample information. It uses a statistical vector to measure the classifier's performance, and an appropriate weight is assigned to each classifier according to the reliability of its output. Different color components of cataract images were used to train the corresponding classifiers, and the decisions of the classifiers using a single color feature were fused to judge the hardness. The experimental results show that the algorithm can achieve a high classification performance that meets the actual demands of the cataract ultrasonic phacoemulsification operation.

Competing Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

This work was financially supported by the Key Scientific Research Project of Colleges and Universities of Henan Province (no. 15A520017), Soft Science Research Project of Henan Province (no. 152400410513), the Doctoral Fund of Henan Polytechnic University (nos. B2010-046 and B2013-040), the Basic Research Development Program of Henan Province (132300410333), and Key Science and Technology Project of Henan Province (152102210102).

References

[1] S.-Y. Liang, D.-Q. Han, and C.-Z. Han, "A novel diversity measure based on geometric relationship and its application to design of multiple classifier systems," *Acta Automatica Sinica*, vol. 40, no. 3, pp. 449–458, 2014.

[2] D.-H. Zhou and Y.-Y. Hu, "Fault diagnosis techniques for dynamic systems," *Acta Automatica Sinica*, vol. 35, no. 6, pp. 748–758, 2009.

[3] J.-T. Huang, M.-H. Wang, W.-J. Li, and B. Gu, "Multiple classifier fault diagnosis system based on dynamic weight," *Acta Electronica Sinica*, vol. 40, no. 4, pp. 734–738, 2012.

[4] L. Li and H. Ji, "Radar emitter recognition based on multiple classifier fusion," *Journal of Data Acquisition and Processing*, vol. 25, no. 3, pp. 396–400, 2010.

[5] L. Sun, C.-Z. Han, J.-J. Shen, and N. Dai, "Generalized rough set method for ensemble feature selection and multiple classifier fusion," *Acta Automatica Sinica*, vol. 34, no. 3, pp. 298–304, 2008.

[6] A. A. López-Caloca, "Data fusion approach for employing multiple classifiers to improve lake shoreline analysis," in *Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications*, E. Bayro-Corrochano and E. Hancock, Eds., vol. 8827 of *Lecture Notes in Computer Science*, pp. 1022–1029, Springer, Berlin, Germany, 2014.

[7] M. Woźniak, M. Graña, and E. Corchado, "A survey of multiple classifier systems as hybrid systems," *Information Fusion*, vol. 16, no. 1, pp. 3–17, 2014.

[8] V. G. Giannoglou and J. B. Theocharis, "Decision fusion of multiple classifiers for coronary plaque characterization from IVUS images," *International Journal on Artificial Intelligence Tools*, vol. 23, no. 3, Article ID 1460005, 2014.

[9] M. De-La-Torre, E. Granger, P. V. W. Radtke, R. Sabourin, and D. O. Gorodnichy, "Partially-supervised learning from facial trajectories for face recognition in video surveillance," *Information Fusion*, vol. 24, pp. 31–53, 2015.

[10] M. Panda, A. Abraham, and M. R. Patra, "Hybrid intelligent systems for detecting network intrusions," *Security and Communication Networks*, vol. 8, no. 16, pp. 2741–2749, 2015.

[11] H. Altinçay, "On naive Bayesian fusion of dependent classifiers," *Pattern Recognition Letters*, vol. 26, no. 15, pp. 2463–2473, 2005.

[12] A. Dainotti, A. Pescapé, C. Sansone, and A. Quintavalle, "Using a behaviour knowledge space approach for detecting unknown IP traffic flows," in *Multiple Classifier Systems*, vol. 6713 of *Lecture Notes in Computer Science*, pp. 360–369, Springer, Berlin, Germany, 2011.

[13] J. Zou, Q. Ji, and G. Nagy, "A comparative study of local matching approach for face recognition," *IEEE Transactions on Image Processing*, vol. 16, no. 10, pp. 2617–2628, 2007.

[14] M. A. Zahid and H. de Swart, "The borda majority count," *Information Sciences*, vol. 295, pp. 429–440, 2015.

[15] J. P. Hwang, S. Park, and E. Kim, "A new weighted approach to imbalanced data classification problem via support vector machine with quadratic cost function," *Expert Systems with Applications*, vol. 38, no. 7, pp. 8580–8585, 2011.

[16] L. Zhang and Y. Dong, "Research on diagnosis of AC engine wear fault based on support vector machine and information fusion," *Journal of Computers*, vol. 7, no. 9, pp. 2292–2297, 2012.

[17] Z. Wenbo, J. Hongbing, and W. Lei, "Adaptive weighted feature fusion classification method," *Systems Engineering and Electronics*, vol. 35, no. 6, pp. 1133–1137, 2013.

[18] A. C. Ben Abdallah, H. Frigui, and P. Gader, "Adaptive local fusion with neural networks," in *Artificial Neural Networks-ICANN 2010*, vol. 6353 of *Lecture Notes in Computer Science*, pp. 486–491, Springer, Berlin, Germany, 2010.

[19] Y. Yang, D. Han, and C. Han, "A novel diversity measure of multiple classifier systems based on distance of evidence," *Acta Aeronautica et Astronautica Sinica*, vol. 33, no. 6, pp. 1093–1099, 2012.

[20] L. I. Kuncheva, J. C. Bezdek, and R. P. W. Duin, "Decision templates for multiple classifier fusion: an experimental comparison," *Pattern Recognition*, vol. 34, no. 2, pp. 299–314, 2001.

[21] R. Ebrahimpour and S. Hamedi, "Hand written digit recognition by multiple classifier fusion based on decision templates approach," *World Academy of Science, Engineering and Technology*, vol. 33, pp. 560–565, 2009.

[22] W. Chen, S.-W. Zhang, Y.-M. Cheng, and Q. Pan, "Prediction of protein-protein interaction types using the decision templates based on multiple classier fusion," *Mathematical and Computer Modelling*, vol. 52, no. 11-12, pp. 2075–2084, 2010.

[23] M. Re and G. Valentini, "Integration of heterogeneous data sources for gene function prediction using decision templates and ensembles of learning machines," *Neurocomputing*, vol. 73, no. 7–9, pp. 1533–1537, 2010.

[24] M. S. Haghighi, A. Vahedian, and H. S. Yazdi, "Extended decision template presentation for combining classifiers," *Expert Systems with Applications*, vol. 38, no. 7, pp. 8414–8418, 2011.

[25] R. Ebrahimpour, R. D. Vahid, and B. M. Nezhad, "Decision templates with gradient based features for Farsi handwritten word recognition," *International Journal of International Journal of Hybrid Information Technology*, vol. 4, no. 1, pp. 1–12, 2011.

[26] B. Bigdeli, F. Samadzadegan, and P. Reinartz, "Feature grouping-based multiple fuzzy classifier system for fusion of hyperspectral and LIDAR data," *Journal of Applied Remote Sensing*, vol. 8, no. 1, Article ID 083509, 2014.

[27] L. Valet, E. Ramasso, and S. Teyssier, "Quality evaluation of insulating parts by fusion of classifiers issued from tomographic images," *Information Fusion*, vol. 9, no. 2, pp. 211–222, 2008.

[28] P. W. R. Duin, P. Juszczak, P. Paclik et al., *PRTools4, A Matlab Toolbox for Pattern Recognition*, Delft University of Technology, Delft, Netherlands, 2007.

[29] L. Guo, J.-J. Yang, L. Peng, J. Li, and Q. Liang, "A computer-aided healthcare system for cataract classification and grading based on fundus image analysis," *Computers in Industry*, vol. 69, pp. 72–80, 2015.

Hindawi Publishing Corporation
Scientific Programming
Volume 2016, Article ID 4589723, 8 pages
http://dx.doi.org/10.1155/2016/4589723

Research Article

Intelligence in Ecology: How Internet of Things Expands Insights into the Missing CO_2 Sink

Wenfeng Wang,[1] Xi Chen,[1] Hongwei Zheng,[1] Zhihan Lv,[2] Zhengjia Liu,[3] Jing Qian,[4] and Ping Hu[1]

[1] *State Key Laboratory of Desert and Oasis Ecology, Xinjiang Institute of Ecology and Geography, Chinese Academy of Sciences, Urumqi 830011, China*
[2] *High Performance Computing Center, Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen 518055, China*
[3] *State Key Laboratory of Remote Sensing Science, Institute of Remote Sensing and Digital Earth, Chinese Academy of Sciences, Beijing 100101, China*
[4] *Center for Geospatial Information, Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen 518055, China*

Correspondence should be addressed to Xi Chen; cx@ms.xjb.ac.cn

Received 13 April 2016; Accepted 5 June 2016

Academic Editor: Guo Chen

Arid region characterizes more than 30% of the Earth's total land surface area and the area is still increasing due to the trends of desertification, yet the extent to which it modulates the global C balance has been inadequately studied. As an emerging technology, IoT monitoring can combine researchers, instruments, and field sites and generate archival data for a better understanding of soil abiotic CO_2 uptake in arid region. Images' similarity analyses based on IoT monitoring can help ecologists to find sites where the abiotic uptake can temporally dominate and how the negative soil respiration fluxes were produced, while IoT monitoring with a set of intelligent video recognition algorithms enables ecologists to revisit these sites and the experiments details through the videos. Therefore, IoT monitoring of geospatial images, videos, and associated optimization and control algorithms should be a research priority towards expanding insights for soil abiotic CO_2 uptake and a better understanding of how the uptake happens in arid region. Nevertheless, there are still considerable uncertainties and difficulties in determining the overall perspective of IoT monitoring for insights into the missing CO_2 sink.

1. Introduction

Largely because of human activities after the Industrial Revolution and the produced substantial climate changes, atmospheric CO_2 levels have increased more than 30% in the past century [1]. This major environmental issue has motivated scientists to carry out a huge effort to quantify the sources and sinks of the atmospheric CO_2, and the existence of a "missing CO_2 sink" is finally concluded [1–7]. Numerous scientists ever claimed to have located the "missing sink," but each of them was finally denied [8–20]. Location of the missing CO_2 sink has become a long-sought challenge in ecology.

Recent studies of the arid and semiarid ecosystems suggest that the missing CO_2 sink can be partly attributed to unneglectable soil abiotic CO_2 uptake in arid region [21–24]. Such uptake has been long-term overlooked in estimating the net ecosystem exchange of CO_2 [NEE] around the world. The global "CO_2 flux towers" employed in current micrometeorological measurements interpret NEE as biological fluxes, exactly defined as the direct sum of photosynthetic and respiratory components [20]. Arid region characterizes more than 30% of the Earth's total land surface area and the area is still increasing due to the trends of global desertification, yet the extent to which it modulates the global C balance has been inadequately studied [25–33].

Estimates of the overall contribution of such abiotic CO_2 uptake are essentially emergent for expanding insights into the missing CO_2 sink, which further requires common huge efforts of the world scientific communities [23]. The current estimates based on very limited data collected from a few sites within several typical desert ecosystems were thought to be not convincing and even problematic [24]. Ecologists were cautioned to keep discreet minds in both data collection and the determination of the whole story of soil abiotic CO_2 uptake in arid region. Such abiotic uptake can be varying with predominant processes, site location, and climatic conditions. These are important factors affecting experimental designs because spatial-temporal heterogeneity must be taken into account. To treat these disturbances and simplify experimental designs, it is hence imperative to implement intelligent methods for ecologists to collect both convincing data and further evidences.

In previous publications for insights into the missing CO_2 sink and especially for insights into soil abiotic CO_2 uptake in arid regions, the utilized technologies are rather old. The emerging information technologies were hardly employed. These unemployed technologies include the wireless sensing networks [34–36], Internet of Things (IoT) [37–40], and cloud computing [41–46]. Particularly, IoT has been further integrated with the surveillance systems and the integrated system was termed as IoT monitoring [47–50]. Since IoT monitoring can generate images, videos, and other archival data, it is necessary to investigate whether IoT monitoring can serve for a better understanding of soil abiotic CO_2 uptake in arid region. The currently published studies are very limited and were thought to be not convincing. Geospatial images and videos from IoT monitoring help us to explain at which sites soil abiotic CO_2 uptake was observed and present more details of the whole experimental process.

Our objectives in this study were to examine the potentials of IoT monitoring as an emerging technology for insights into soil abiotic CO_2 uptake and in turn for expanding insights into the missing CO_2 sink in the unneglectable arid region. Utilizing geospatial, archival data, intelligent algorithms on videos and images were performed to theoretically expand insights into soil abiotic CO_2 uptake in unneglectable arid region, which has been overlooked for a long period. Additionally, the existing uncertainties and unresolved issues to develop such a thematic IoT monitoring system are also discussed.

2. Materials and Methods

2.1. Collection of Geospatial Data. Analyses of the potentials of IoT monitoring for insights into the missing CO_2 sink in the present study are based on the collected geospatial images and videos from the field sites at the south edge of the Gurbantunggut Desert in the north of Xinjiang Uygur Autonomous Region, China (Figure 1).

These field sites were chosen because it has been confirmed that soil abiotic CO_2 uptake can temporally dominate and cause the apparent negative soil respiration fluxes at these sites [51–53]. Collecting the geospatial images of these sites

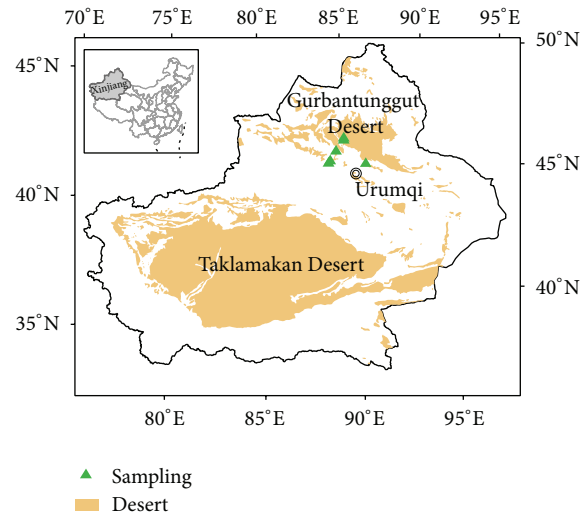


Figure 1: Distribution of the field sites where geospatial images and videos are collected in this study.

from IoT monitoring helps us to explain at which sites soil abiotic CO_2 uptake was observed and present more details of the experimental sites. Overall, geospatial images and videos were collected from 19 field sites, 18 of which are distributed within the Manas River Basin. These sites are close to each other. Another field site is located in the Sangong River Basin [51, 52].

A mobile communication tool (Redmi Note 4, with MATLAB software installed to operate the algorithms) was employed for the collection of geospatial images and videos. In total 70 geospatial images of these field sites were collected and 36 images were chosen to build the first database of geospatial images for the sites where soil abiotic CO_2 uptake can temporally dominate (Figure 2). As a first example of the utilization of geospatial videos in analyzing soil abiotic CO_2 uptake, a special experiment was designed to expand insights into soil texture at those sites where abiotic CO_2 uptake can temporally dominate in soil respiration fluxes.

The details of this experimental design are as follows. We aim to collect a video to record the process when one inserts the WET sensors of HH2 Moisture meter (Delta-T Devices Ltd., Cambridge, UK) into the soil and then utilize video tracking algorithms to analyze the movements of the sensors beneath the soil surface. This is really a challenge because we realized that the time of the collected video may be too short. However, the soil texture cannot be objectively displayed if we deliberately slowly insert the sensors. Therefore, the daughter of the first author (Wenfeng Wang), who is 6 years old and named Yanbo Wang, was invited to join the "scientific game." She saw this as an interesting game and naturally tried her best. A short video was collected when she was inserting the WET sensors into the soil.

2.2. Optimization and Control. A histogram-based image similarity algorithm [54–56] was further optimized and employed to analyze the match degree between the test image

Figure 2: The first images database of the sites where soil abiotic CO_2 uptake can temporally dominate.

and each image from the first database of geospatial images of the sites where soil abiotic CO_2 uptake can temporally dominate. This helps in finding the best match of the test image in the database. In a previous publication [54], to optimize the performance of the algorithm, the histograms of the Red-band H(R), the Green-band H(G), and the Blue-band H(B), respectively, were used. In the present study, the algorithm was further optimized by taking into account the weights of H(R), H(G), and H(B) to each image, where the weights/contributions were determined by calculating the information entropy [57–59]. R-G-B-weighted average correlation-efficient parameters that were employed to evaluate the histogram-based image similarity between the test image and each image from the database were calculated.

In order to objectively evaluate the potentials of IoT monitoring for insights into the missing CO_2 sink, a real challenge was carried out. The video object tracking algorithm was performed on the collected short video for the real-time video tracking of the WET sensors. Traditional algorithms, such as the mean-shift algorithm [60–62], are unsuitable for this video object tracking. Therefore, we previously specialized the video target for tracking by morphological segmentation [63], which helps to improve the performance of mean-shift algorithm.

3. Results and Discussions

3.1. Images' Similarity Analyses Based on IoT Monitoring. The further optimized histogram-based images similarity algorithm was applied to search field sites where soil abiotic CO_2 uptake can temporally dominate. The performance of the similarity detection algorithm worked out the best match of the test image among images in the first images database of the sites where the abiotic CO_2 uptake can temporally dominate. Results show that the match degree between test image and best match is approximated to 90%. Therefore, the histogram-based image similarity analyses based on IoT monitoring confirmed that the test image represents a site where soil abiotic CO_2 uptake can temporally dominate in soil respiration and cause negative soil respiration fluxes.

Through further reviews of the details of the test image, it is easy to find an obvious salt accumulation on the soil surface at the test site (Figure 3). The test image can be joined to the database and the extended information can be utilized. Exactly, some previous reports of negative soil respiration fluxes in arid region do not emphasize the role of salt accumulation [21].

This helps in convincing the ecologists who were not convinced by the previous reports since they may realize that the soil and groundwater are alkaline, which is advantageous to the subterranean fixation of CO_2. Taking into account the abiotic flux components, the soil CO_2 flux can be further reconciled as

$$\begin{aligned} F_s &= R - F_{DIC} - F_{SIC} = F_e - F_i, \\ F_e &= R, \\ F_i &= F_{DIC} + F_{SIC}, \end{aligned} \tag{1}$$

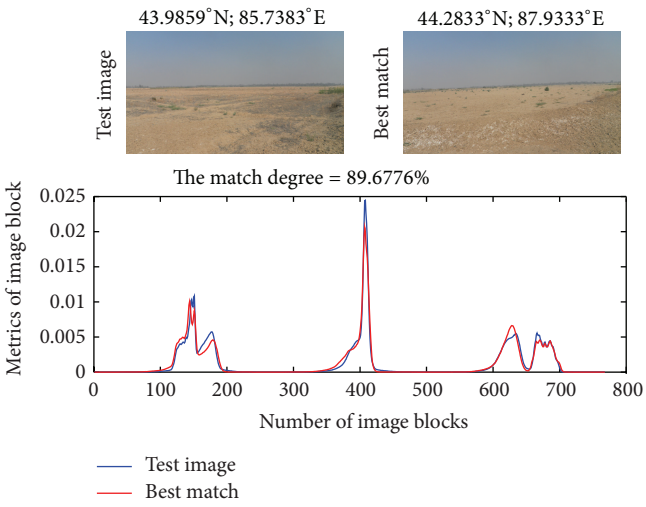


FIGURE 3: To find the best match for test image in the first images database of the sites where soil abiotic CO_2 uptake can temporally dominate, utilizing the image analysis algorithm referred to in this study.

where R is the CO_2 release from roots and soil microbial respiration and F_{DIC} and F_{SIC} are the net CO_2 fixation in the groundwater and the soil [in inorganic forms], respectively. F_e and F_i are the net soil CO_2 influx and the net soil CO_2 efflux, respectively [52].

A sketch of soil CO_2 flux formation in arid region can be hence expanded by further mathematical analyses. First, review the mechanism of how the soil CO_2 analyzer (e.g., LI-8100; see [53]) works. Assume that, after per unit time T, the CO_2 analyzer abstracts air of volume V_1 from a gas room of volume V and then supplies air of the same volume sampled from atmosphere for the CO_2 pressure balance in the gas room. Go round and begin again. To compute CO_2 flux, the following is used:

$$F_c = \frac{dC(t)}{dt}, \tag{2}$$

where $C(t)$ is the CO_2 concentration in the gas room at time t [64].

Let q be CO_2 concentration in the atmosphere. For the nth measured value, the input and output of CO_2 are $F_{input} = r_n$ and $F_{output}/r_n = p_n$, respectively, taking average within nth time interval $[nT, (n+1)T]$. The dynamic of CO_2 concentration in the gas room should be as follows.

Input-output balance equation:

$$C(nT+T) - C(nT) = \frac{(V_1 q + r_n)\cdot T - \int_{nT}^{nT+T} V_1 \cdot C(s) + r_n p_n ds}{V}, \tag{3}$$
$$C(0) = C_0,$$

where C_0 is the CO_2 concentration at starting time point.

Thus the nth measured value of soil CO_2 flux is

$$F_{c\text{-}n\text{th}} = \frac{C(nT+T) - C(nT)}{T} = \frac{V_1 q + r_n(1-p_n) - C(\xi_n)\cdot V_1}{V}, \tag{4}$$

where $C(\xi_n)$ is the CO_2 concentration from *mean value theorem of integrals*.

Negative soil respiration CO_2 fluxes are observed if

$$p_n > \frac{V_1[q - C(\xi_n)]}{r_n} + 1. \tag{5}$$

Finally, it must be cautioned that infimum of the negative values of soil CO_2 flux may exist. Let $T \to 0$; we obtain

$$\frac{dC(t)}{dt} = \frac{V_1 q + r_n - r_n p_n}{V} - \frac{V_1}{V}\cdot C(t), \quad C(0) = C_0. \tag{6}$$

Hence,

$$C(t) = \frac{V_1 q + rp}{V_1} + \left(C_0 - \frac{V_1 q + rp}{V_1}\right)\cdot e^{-V_1 t/V}. \tag{7}$$

Stable negative fluxes may happen within a small measurement interval T when

$$\frac{V_1 q + rp}{V_1} > C_0. \tag{8}$$

Let $t \to \infty$; we get

$$\lim_{t\to\infty} C(t) = \frac{V_1 p + r}{V_1}. \tag{9}$$

This is the infimum of the CO_2 concentration.

3.2. IoT Monitoring with Intelligent Video Recognition Algorithm. The trajectory analysis of soil sensors is realized in performance of IoT monitoring with intelligent video recognition algorithm. Such video object tracking algorithm not only enables ecologists to revisit these sites and the experiments details by geospatial videos, but also helps ecologists to further understand the compact soil texture so that the whole process costs 21 seconds. The footprint of the WET sensors revealed that the process is difficult for this little girl (Figure 4).

Consequently, a part of soil respiration (R) temporally gathers in soil (R_g) or is ventilated in subterranean cavity (R_v) and contributes to the abiotic release later. This also is advantageous for a chemical fixation of CO_2 in the soil-groundwater system (Figure 5).

This expands a perspective frame of IoT for insights into the missing CO_2 sink (Figure 6).

Therefore, the potentials of IoT monitoring for insights into soil abiotic CO_2 uptake and hence for the insights into the missing CO_2 sink are highlighted. A part of soil inorganic CO_2 (SIC) remained in soil layers [SR] and a part of DIC is carried away and might go out at the terminal of the

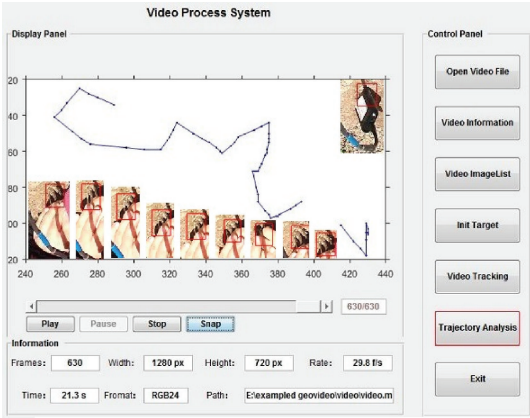


FIGURE 4: Trajectory analysis of soil sensors by the video object tracking algorithm referred to in this study.

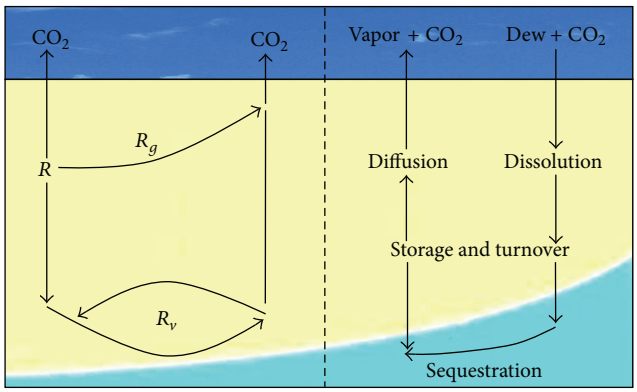


FIGURE 5: Integrated story of soil abiotic CO_2 uptake/release, where parts of soil respiration (R) temporally gather in soil (R_g) or are ventilated in subterranean cavity (R_v) and then contribute to abiotic release later.

groundwater-soil system [TO], while the other parts of SIC and DIC form the final absorption in groundwater [GA]. SR falls into three phases, solid SR [SSR forms hydrogen carbonate and changes molar number of carbon atoms], liquid SR [LSR dissolved part of SR], and gaseous SR [GSR increases the CO_2 concentration in soil pores]; GA is a single phase: liquid phase [consisting of liquor diverse carbon species]; TO falls into two phases: liquid TO [LTO recharged DIC] and gaseous TO [GTO released CO_2]. Assignment of missing carbon should be formulated as follows:

$$\begin{aligned} C_{\text{missing}} &= \text{SR} + \text{GA} + \text{TO} \\ &= \text{SSR} + \text{LSR} + \text{GA} + \text{GSR} + \text{LTO} + \text{GTO}. \end{aligned} \tag{10}$$

The carbon assignment equation can be further hypothetically expanded. We can classify soil pores as three types: dry pore [DP], small water pore [SWSP], and big water pore [BWSP] according to their size and water content. DP is distributed in shallow soil layers and can absorb CO_2 if coupled with the condensing of vapor or the infiltration of precipitation; SWSP is distributed in moist layers around the roots system, dissolving CO_2 in it; BWSP is distributed in deep layers, dissolving CO_2 and then migrating it into groundwater. Note that these three types of soil pores may convert to each other with the changes or movements of soil water.

The balance equations can be represented as

$$\begin{aligned} \Delta\text{GSR} &= \Delta\text{DP}, \\ \Delta\text{SSR} &= \Delta\text{SWSP}, \end{aligned} \tag{11}$$

where the groundwater recharge/discharge is the major regulator of the balance.

4. Conclusions and Outstanding Remarks

As an emerging technology, IoT monitoring combines researchers, instruments, and field sites and generates archival data for a better understanding of soil abiotic CO_2 uptake in arid region and in turn has great potentials for insights into the missing CO_2 sink. By histogram-based image similarity analyses of image data collected from IoT monitoring, ecologists can easily find field sites where soil abiotic uptake of CO_2 can temporally dominate and further improve their understanding of the negative soil respiration flux values. Video object tracking algorithms based on IoT monitoring not only enable ecologists to revisit these sites and the experiments details by geospatial videos, but also help the ecologists to further understand other details, such as the footprint of soil sensors, which in turn can help ecologists to understand the integrated story of soil abiotic CO_2 uptake. In subsequent studies, the employed algorithms can be more and more complex and the uncertainties of the presented algorithms must be explicitly discussed.

Nevertheless, it must be pointed out that there are still considerable uncertainties and difficulties in developing such a thematic IoT monitoring system. One major challenge is how to conceptualize ecosystem as a volume with explicitly defined top, bottom, and sides. The other major challenge is how to estimate the SIC/DIC assignment proportion of the carbon fluxes in soil layers and groundwater, which should be also analyzed in complicated cases due to great difference in soil types and the groundwater levels. The possible scheme is relating the field sites, instruments, and researchers together by a stable IoT monitoring system and conceptualizing each block of terrestrial ecosystem. In this case, net ecosystem carbon balance equals the total C input minus the total C output from the ecosystem over a specified time interval.

To reduce the increased complexity, one can analyze the situation in the different layers of the local groundwater-soil system. A research priority is the explicit characterization of the situation in different layers of the local groundwater-soil system, which deserves subsequent studies on the field collection of geospatial data for soil abiotic CO_2 uptake [65–73], the visualization of CO_2 footprints [74], and 2D-3D video treatments technology to enhance the visualization effect [75].

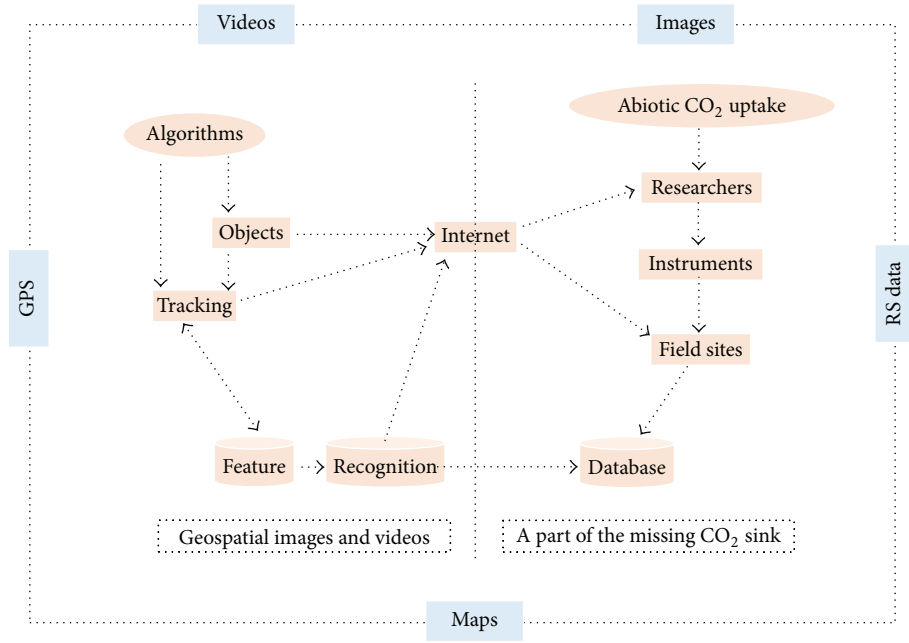


Figure 6: The first perspective frame of IoT monitoring for insights into the missing CO_2 sink.

Competing Interests

The authors declare that they have no competing interests.

Acknowledgments

This research was financially supported by the CAS "Light of West China" Program (XBBS-2014-16), the "Thousand Talents" plan (Y474161), the National Natural Science Foundation of China (41571299), and the Shenzhen Basic Research Project (JCYJ20150630114942260). Essential geodata were collected from National Science & Technology Infrastructure Center and Data Sharing Infrastructure of Earth System Science, <http://www.geodata.cn/>.

References

[1] R. F. Keeling, S. C. Piper, and M. Heimann, "Global and hemispheric CO_2 sinks deduced from changes in atmospheric O_2 concentration," *Nature*, vol. 381, no. 6579, pp. 218–221, 1996.

[2] R. P. Detwiler and C. A. S. Hall, "Tropical forests and the global carbon cycle," *Science*, vol. 239, no. 4835, pp. 42–47, 1988.

[3] P. P. Tans, I. Y. Fung, and T. Takahashi, "Observational constraints on the global atmospheric CO_2 budget," *Science*, vol. 247, no. 4949, pp. 1431–1438, 1990.

[4] D. W. Schindler, "The mysterious missing sink," *Nature*, vol. 398, no. 6723, pp. 105–107, 1999.

[5] S. C. Wofsy, "Where has all the carbon gone?" *Science*, vol. 292, no. 5525, pp. 2261–2263, 2001.

[6] U. Siegenthaler and J. L. Sarmiento, "Atmospheric carbon dioxide and the ocean," *Nature*, vol. 365, no. 6442, pp. 119–125, 1993.

[7] T. Erbrecht and W. Lucht, "Impacts of large-scale climatic disturbances on the terrestrial carbon cycle," *Carbon Balance & Management*, vol. 1, no. 1, pp. 1–7, 2006.

[8] R. B. Myneni, J. Dong, C. J. Tucker et al., "A large carbon sink in the woody biomass of northern forests," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 98, no. 26, pp. 14784–14789, 2001.

[9] D. S. Schimel, J. I. House, K. A. Hibbard et al., "Recent patterns and mechanisms of carbon exchange by terrestrial ecosystems," *Nature*, vol. 414, no. 6860, pp. 169–172, 2001.

[10] H. A. de Wit, T. Palosuo, G. Hylen, and J. Liski, "A carbon budget of forest biomass and soils in southeast Norway calculated using a widely applicable method," *Forest Ecology & Management*, vol. 225, no. 1–3, pp. 15–26, 2006.

[11] P. B. Woodbury, J. E. Smith, and L. S. Heath, "Carbon sequestration in the U.S. forest sector from 1990 to 2010," *Forest Ecology & Management*, vol. 241, no. 1–3, pp. 14–27, 2007.

[12] M. J. Fisher, I. M. Rao, M. A. Ayarza et al., "Carbon storage by introduced deep-rooted grasses in the South American savannas," *Nature*, vol. 371, no. 6494, pp. 236–238, 1994.

[13] C. L. Goodale and E. A. Davidson, "Uncertain sinks in the shrubs," *Nature*, vol. 418, no. 6898, pp. 593–594, 2002.

[14] T. J. Kessler and C. F. Harvey, "The global flux of carbon dioxide into groundwater," *Geophysical Research Letters*, vol. 28, no. 2, pp. 279–282, 2001.

[15] J. Grace and Y. Malhi, "Carbon dioxide goes with the flow," *Nature*, vol. 416, no. 6881, pp. 594–595, 2002.

[16] R. Lal, "Soil erosion and the global carbon budget," *Environment International*, vol. 29, no. 4, pp. 437–450, 2003.

[17] J. Leifeld and J. Fuhrer, "Organic farming and soil carbon sequestration: what do we really know about the benefits?" *Ambio*, vol. 39, no. 8, pp. 585–599, 2010.

[18] W. H. Schlesinger, "Carbon sequestration in soils: some cautions amidst optimism," *Agriculture, Ecosystems & Environment*, vol. 82, no. 1–3, pp. 121–127, 2000.

[19] J. J. C. Dawson and P. Smith, "Carbon losses from soil and its consequences for land-use management," *Science of the Total Environment*, vol. 382, no. 2-3, pp. 165–190, 2007.

[20] S. Inouye, A. Nakazawa, and T. Nakazawa, "Soil CO_2 efflux in a beech forest: the contribution of root respiration," *Annals of Forest Science*, vol. 56, no. 4, pp. 289–295, 1999.

[21] G. Wohlfahrt, L. F. Fenstermaker, and J. A. Arnone, "Large annual net ecosystem CO_2 uptake of a Mojave Desert ecosystem," *Global Change Biology*, vol. 14, no. 7, pp. 1475–1487, 2008.

[22] J. X. Xie, Y. Li, C. X. Zhai, C. H. Li, and Z. D. Lan, "CO_2 absorption by alkaline soils and its implication to the global carbon cycle," *Environmental Geology*, vol. 56, no. 5, pp. 953–961, 2009.

[23] R. Stone, "Have desert researchers discovered a hidden loop in the carbon cycle?" *Science*, vol. 320, no. 5882, pp. 1409–1410, 2008.

[24] W. H. Schlesinger, J. Belnap, and G. Marion, "On carbon sequestration in desert ecosystems," *Global Change Biology*, vol. 15, no. 6, pp. 1488–1490, 2009.

[25] R. L. Jasoni, S. D. Smith, and J. A. Arnone III, "Net ecosystem CO_2 exchange in Mojave Desert shrublands during the eighth year of exposure to elevated CO_2," *Global Change Biology*, vol. 11, no. 5, pp. 749–756, 2005.

[26] B. E. Law, O. J. Sun, J. Campbell, S. Van Tuyl, and P. E. Thornton, "Changes in carbon storage and fluxes in a chronosequence of ponderosa pine," *Global Change Biology*, vol. 9, no. 4, pp. 510–524, 2003.

[27] M. Pansu, Y. Martineau, and B. Saugier, "A modelling method to quantify in situ the input of carbon from roots and the resulting C turnover in soil," *Plant & Soil*, vol. 317, no. 1-2, pp. 103–120, 2009.

[28] J. Kern, H. J. Hellebrand, V. Scholz, and B. Linke, "Assessment of nitrogen fertilization for the CO_2 balance during the production of poplar and rye," *Renewable & Sustainable Energy Reviews*, vol. 14, no. 5, pp. 1453–1460, 2010.

[29] B. A. Ball and R. A. Virginia, "Controls on diel soil CO_2 flux across moisture gradients in a polar desert," *Antarctic Science*, vol. 27, no. 6, pp. 527–534, 2015.

[30] R. M. Chew and A. E. Chew, "The primary productivity of a desert-shrub (*Larrea tridentata*) community," *Ecological Monographs*, vol. 35, no. 4, pp. 355–375, 1965.

[31] C. D. Keeling, J. F. S. Chin, and T. P. Whorf, "Increased activity of northern vegetation inferred from atmospheric CO_2 measurements," *Nature*, vol. 382, no. 6587, pp. 146–149, 1996.

[32] M. L. Imhoff and L. Bounoua, "Exploring global patterns of net primary production carbon supply and demand using satellite observations and statistical data," *Journal of Geophysical Research Atmospheres*, vol. 111, no. 22, 2006.

[33] L. T. C. Bonten, G. J. Reinds, and M. Posch, "A model to calculate effects of atmospheric deposition on soil acidification, eutrophication and carbon sequestration," *Environmental Modelling & Software*, vol. 79, pp. 75–84, 2016.

[34] Y. Zhang and J. N. Harb, "Characterization of battery behavior for battery-aware design of Wireless Sensing Networks," *Journal of the Electrochemical Society*, vol. 162, no. 6, pp. A820–A826, 2015.

[35] A. Nayebi-Astaneh, N. Pariz, and M.-B. Naghibi-Sistani, "Adaptive node scheduling under accuracy constraint forwireless sensor nodes with multiple bearings-only sensing units," *IEEE Transactions on Aerospace & Electronic Systems*, vol. 51, no. 2, pp. 1547–1557, 2015.

[36] D. Zordan, T. Melodia, and M. Rossi, "On the design of temporal compression strategies for energy harvesting sensor networks," *IEEE Transactions on Wireless Communications*, vol. 15, no. 2, pp. 1336–1352, 2016.

[37] A. Sheth, "Internet of things to smart IoT through semantic, cognitive, and perceptual computing," *IEEE Intelligent Systems*, vol. 31, no. 2, pp. 108–112, 2016.

[38] S. Li, "The internet of things: a security point of view," *Internet Research*, vol. 26, no. 2, pp. 337–359, 2016.

[39] J. Gozalvez, "New 3GPP standard for IoT [mobile radio]," *IEEE Vehicular Technology Magazine*, vol. 11, no. 1, pp. 14–20, 2016.

[40] V. Cerf and M. Senges, "Taking the internet to the next physical level," *Computer*, vol. 49, no. 2, pp. 80–86, 2016.

[41] S. Dustdar, "Cloud computing," *Computer*, vol. 49, no. 2, pp. 12–13, 2016.

[42] M. García-Valls, T. Cucinotta, and C. Lu, "Challenges in real-time virtualization and predictable cloud computing," *Journal of Systems Architecture*, vol. 60, no. 9, pp. 726–740, 2014.

[43] J. M. Alcaraz Calero and J. G. Aguado, "MonPaaS: an adaptive monitoring platformas a service for cloud computing infrastructures and services," *IEEE Transactions on Services Computing*, vol. 8, no. 1, pp. 65–78, 2015.

[44] C.-S. Chen, W.-Y. Liang, and H.-Y. Hsu, "A cloud computing platform for ERP applications," *Applied Soft Computing*, vol. 27, pp. 127–136, 2015.

[45] I. Gomez-Miguelez, V. Marojevic, and A. Gelonch, "Deployment and management of SDR cloud computing resources: problem definition and fundamental limits," *Environmental Science & Technology*, vol. 49, no. 8, pp. 1–11, 2015.

[46] S. P. Crago and J. P. Walters, "Heterogeneous cloud computing: the way forward," *Computer*, vol. 48, no. 1, Article ID 7030253, pp. 59–61, 2015.

[47] Y. Tang, B. Ma, and H. Yan, "Intelligent video surveillance system for elderly people living alone based on ODVS," *Advances in Internet of Things*, vol. 3, no. 2, pp. 44–52, 2013.

[48] J. Cui, J. Chen, and B. Zheng, "Energy-aware distributed scheduling for multimedia streaming over Internet of Things," *International Journal of Ad Hoc & Ubiquitous Computing*, vol. 13, no. 3-4, pp. 176–186, 2013.

[49] H.-Y. Kang, "Design and realization of internet of things based on embedded system used in intelligent campus," *International Journal of Advancements in Computing Technology*, vol. 3, no. 11, pp. 291–298, 2011.

[50] K.-D. Chang and J.-L. Chen, "A survey of trust management in WSNs, internet of things and future internet," *KSII Transactions on Internet and Information Systems*, vol. 6, no. 1, pp. 5–23, 2012.

[51] W. F. Wang, X. Chen, G. P. Luo, and L. H. Li, "Modeling the contribution of abiotic exchange to CO_2 flux in alkaline soils of arid areas," *Journal of Arid Land*, vol. 6, no. 1, pp. 27–36, 2014.

[52] W. Wang, X. Chen, and Z. Pu, "Negative soil respiration fluxes in unneglectable arid regions," *Polish Journal of Environmental Studies*, vol. 24, no. 2, pp. 905–908, 2015.

[53] X. Chen, W.-F. Wang, G.-P. Luo, L.-H. Li, and Y. Li, "Time lag between carbon dioxide influx to and efflux from bare saline-alkali soil detected by the explicit partitioning and reconciling of soil CO_2 flux," *Stochastic Environmental Research and Risk Assessment*, vol. 27, no. 3, pp. 737–745, 2013.

[54] Y. Zou and B. Lei, "Median-based thresholding, minimum error thresholding, and their relationships with histogram-based image similarity," in *6th International Conference on Digital Image Processing (ICDIP 2014)*, vol. 9159 of *Proceedings*

of SPIE, pp. 4177–4180, The International Society for Optical Engineering, April 2014.

[55] D. M. Ballesteros L, D. Renza, and R. Rincon, "Gray-scale images within color images using similarity histogram-based selection and replacement algorithm," *Journal of Information Hiding and Multimedia Signal Processing*, vol. 6, no. 6, pp. 1156–1166, 2015.

[56] P. Zhao, S. Li, L. Zhou, L. Li, and X. Guo, "Detecting affine-distorted duplicated regions in images by color histograms," *Journal of Information Hiding and Multimedia Signal Processing*, vol. 6, no. 1, pp. 163–174, 2015.

[57] M. Parsani, M. H. Carpenter, and E. J. Nielsen, "Entropy stable wall boundary conditions for the three-dimensional compressible Navier-Stokes equations," *Journal of Computational Physics*, vol. 292, pp. 88–113, 2015.

[58] M. Kang, J. Kim, and J. M. Kim, "Reliable fault diagnosis for incipient low-speed bearings using fault feature analysis based on a binary bat algorithm," *Information Sciences*, vol. 311, pp. 205–206, 2015.

[59] P. Sevastjanov and L. Dymova, "Generalised operations on hesitant fuzzy values in the framework of Dempster-Shafer theory," *Information Sciences*, vol. 311, pp. 39–58, 2015.

[60] O. Debeir, P. V. Ham, R. Kiss, and C. Decaestecker, "Tracking of migrating cells under phase-contrast video microscopy with combined mean-shift processes," *IEEE Transactions on Medical Imaging*, vol. 24, no. 6, pp. 697–711, 2005.

[61] A. Mayer and H. Greenspan, "An adaptive mean-shift framework for MRI brain segmentation," *IEEE Transactions on Medical Imaging*, vol. 28, no. 8, pp. 1238–1250, 2009.

[62] J. Ning, L. Zhang, D. Zhang, and C. Wu, "Robust mean-shift tracking with corrected background-weighted histogram," *IET Computer Vision*, vol. 6, no. 1, pp. 62–69, 2012.

[63] F. Meyer and S. Beucher, "Morphological segmentation," *Journal of Visual Communication & Image Representation*, vol. 1, no. 1, pp. 21–46, 1990.

[64] K. E. Savage and E. A. Davidson, "A comparison of manual and automated systems for soil CO_2 flux measurements: trade-offs between spatial and temporal resolution," *Journal of Experimental Botany*, vol. 54, no. 384, pp. 891–899, 2003.

[65] Y. Q. Cui, J. Y. Ma, and W. Sun, "Application of stable isotope techniques to the study of soil salinization," *Journal of Arid Land*, vol. 3, no. 4, pp. 285–291, 2011.

[66] S. Danierhan, A. Shalamu, H. Tumaerbai, and D. H. Guan, "Effects of emitter discharge rates on soil salinity distribution and cotton (*Gossypium hirsutum* L.) yield under drip irrigation with plastic mulch in an arid region of Northwest China," *Journal of Arid Land*, vol. 5, no. 1, pp. 51–59, 2013.

[67] R. Cakir, "Estimating the effect of controlled drainage on soil salinity and irrigation efficiency in the harran plain using saltmod," *Turkish Journal of Agriculture & Forestry*, vol. 32, no. 2, pp. 101–109, 2008.

[68] A. El-Keblawy, S. Gairola, and A. Bhatt, "Maternal salinity environment affects salt tolerance during germination in *Anabasis setifera*: a facultative desert halophyte," *Journal of Arid Land*, vol. 8, no. 2, pp. 254–263, 2016.

[69] X. H. Yang, A. Mamtimin, Q. He, X. C. Liu, and W. Huo, "Observation of saltation activity at Tazhong area in Taklimakan Desert, China," *Journal of Arid Land*, vol. 5, no. 1, pp. 32–41, 2013.

[70] W. Zeng, C. Xu, J. Wu, and J. Huang, "Soil salt leaching under different irrigation regimes: HYDRUS-1D modelling and analysis," *Journal of Arid Land*, vol. 6, no. 1, pp. 44–58, 2014.

[71] L. J. Chen and Q. Feng, "Soil water and salt distribution under furrow irrigation of saline water with plastic mulch on ridge," *Journal of Arid Land*, vol. 5, no. 1, pp. 60–70, 2013.

[72] X. Zhao, H. Xu, P. Zhang, and Y. Bai, "Distribution of soil moisture and salinity in shelterbelts and its relationship with groundwater level in extreme arid area, Northwest of China," *Water & Environment Journal*, vol. 27, no. 4, pp. 453–461, 2013.

[73] J. Abuduwaili, Y. Tang, M. Abulimiti, D. W. Liu, and L. Ma, "Spatial distribution of soil moisture, salinity and organic matter in Manas River watershed, Xinjiang, China," *Journal of Arid Land*, vol. 4, no. 4, pp. 441–449, 2012.

[74] Z. Lv, A. Tek, F. Da Silva, C. Empereur-Mot, M. Chavent, and M. Baaden, "Game on, science—how video game technology may help biologists tackle visualization challenges," *PLoS ONE*, vol. 8, no. 3, Article ID e57990, 2013.

[75] Z. Lv, J. Chirivella, and P. Gagliardo, "Bigdata oriented multimedia mobile health applications," *Journal of Medical Systems*, vol. 40, no. 5, pp. 1–10, 2016.

Hindawi Publishing Corporation
Scientific Programming
Volume 2016, Article ID 4801784, 16 pages
http://dx.doi.org/10.1155/2016/4801784

Research Article

Two-Phase Algorithm for Optimal Camera Placement

Jun-Woo Ahn,[1] Tai-Woo Chang,[2] Sung-Hee Lee,[3] and Yong Won Seo[1]

[1]*College of Business & Economics, Chung-Ang University, 84 Heukseok-ro, Dongjak-gu, Seoul 06974, Republic of Korea*
[2]*Department of Industrial & Management Engineering, Kyonggi University, 154-42 Gwanggyosan-ro, Yeongtong-gu, Suwon, Gyeonggi 16227, Republic of Korea*
[3]*Department of Business Administration, Hoseo University, 12 Hoseodae-gil, Dongnam-gu, Cheonan-si, Chungcheongnam-do 31066, Republic of Korea*

Correspondence should be addressed to Yong Won Seo; seoyw@cau.ac.kr

Received 17 May 2016; Accepted 19 July 2016

Academic Editor: Minsoo Kim

As markers for visual sensor networks have become larger, interest in the optimal camera placement problem has continued to increase. The most featured solution for the optimal camera placement problem is based on binary integer programming (BIP). Due to the NP-hard characteristic of the optimal camera placement problem, however, it is difficult to find a solution for a complex, real-world problem using BIP. Many approximation algorithms have been developed to solve this problem. In this paper, a two-phase algorithm is proposed as an approximation algorithm based on BIP that can solve the optimal camera placement problem for a placement space larger than in current studies. This study solves the problem in three-dimensional space for a real-world structure.

1. Introduction

The global surveillance camera market is rapidly growing. According to the 2013 IMS Research data shown in Figure 1, the surveillance camera market is expected to grow by 1.5 times or more in the next five years. This is because surveillance cameras are used for more than simply preventing and solving crime or managing traffic. They are now needed for production assembly lines or observing natural disasters [1, 2]. Moreover, with the development in big data image-processing techniques, it is also possible not only to watch the images but also to extract the necessary data from them [3].

Along with the growth of the surveillance camera market, interest in efficient camera placement has also been increasing. If the placement of cameras is inefficient, even with many installed cameras the effect can be unsatisfactory. For efficient placement of surveillance cameras, several studies [4–15] have investigated the optimal camera placement problem. The optimal camera placement problem, sometimes called the camera network deployment problem, is defined as how to adequately place cameras to maximize the coverage under certain conditions [6, 10]. This optimal camera placement problem consists of finding the minimum number of cameras that satisfies a specific coverage or finding the maximum coverage with a given number of cameras [4].

Current studies [6, 7, 10, 11] hypothesized a continuous space that is simplified as a two-dimensional (2D) grid of points. Here, the grid points are discrete points on x- and y-axes by the minimum distance Δ, which takes into account the spatial sampling frequency ($f_x = f_y : \Delta = 1/f_x = 1/f_y$) after simplifying real space into 2D [6]. When modeling a fixed-area terrain using the above method, the solution quality of the optimal camera placement problem with a higher resolution tends to be better than that with a lower resolution, because the ratio of the real-world terrain that is reflected in the modeling area with a high resolution (large f_a; small Δ) using a larger number of grid points is higher than that with a low resolution (small f_a; large Δ) using fewer grid points. Thus Hörster and Lienhart [6] claimed that considering a large number of grid points is necessary.

Because the optimal camera placement problem is NP-hard [16], existing studies have focused on finding efficient and effective approximation algorithms rather than finding an optimal solution.

Table 1: Publications on the optimal camera placement problem.

Source	Year	Methods	x grid	y grid	Number of grid points
[10] Surveillance of a 2D plane area with 3D deployed cameras	2014	PSO	4	4	16
[11] Particle swarm optimization inspired probability algorithm for optimal camera network placement	2012	PSO	11	11	121
[6] Approximating optimal visual sensor placement	2006	IP	12	12	144
[13] Grid coverage for surveillance and target location in distributed sensor networks	2002	IP	13	13	169
[7] Optimal camera placement for total coverage	2009	IP	12	6	72
Model simulation 1 in this paper		Greedy	200	200	40,000
Model simulation 2 (Appendix) in this paper		Greedy	500	500	250,000

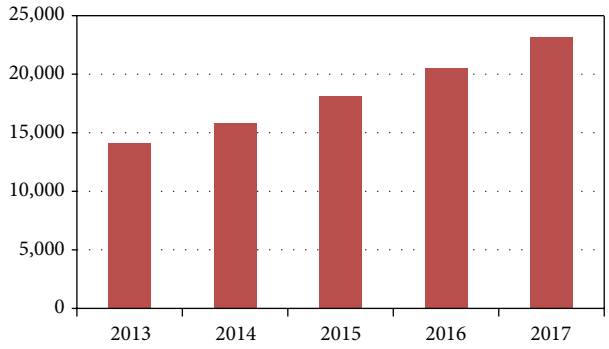


Figure 1: Surveillance Camera Market Size Prediction [20].

The approximation algorithm proposed in previous studies solves the problem directly at the high resolution of the desired level. On the other hand, our study proposes a method of finding a solution under a low resolution using BIP, then solving the problem correctly at its desired high resolution based on the found solution. The proposed method decreases the complexity of the calculation, which can lead to faster problem-solving at a high resolution than existing methods.

The reliability of setting the start point, which can cause a localized optimization in the approximate algorithm, is also improved. As a result, under the same conditions, the confidence of the proposed solution increases when compared to solving the problem at a high resolution to begin with.

Additionally, rather than using the virtual modeling area generally used in existing studies, this study uses a real-world modeling area from geographic information system (GIS) data of actual terrain. The data came from the satellite pictures. Three-dimensional (3D) camera placement was selected to provide more practicality, instead of 2D camera placement which is unrealistic to apply.

This paper is organized as follows. Section 2 analyzes the relevant studies. Section 3 explains the spatial configuration required for the camera placement and the calculation method for the surveillance camera view and also describes the algorithm that solves the actual problem. Section 4 compares the quality of the solutions obtained from binary integer programming and from the proposed method. Section 5 presents the conclusion.

2. Literature Review

The art gallery problem (AGP), studied in the field of computational geometry, is the problem of placing at least one security guard to check every area of a museum or gallery. Because AGP finds the optimal placement point within the restricted viewpoint of the security guard and the optimal camera placement problem finds the optimal placement point within the restricted viewpoint of the camera, solving the optimal camera placement problem is very similar to solving AGP [17, 18].

This optimal camera placement problem has been studied to solve both MIN problem, which finds the minimum number of cameras and placement conditions to satisfy the target coverage under the given conditions, and the FIX problem, which maximizes the coverage with a fixed number of cameras under the given conditions [4].

In the meantime, looking from the methodological viewpoint of problem-solving, previous studies on solving the optimal camera placement problem generally have been based on binary integer programming (BIP) [5–9]. BIP offers the global optimal solution; however, the studies based on BIP only answer problems with limited, simple conditions due to the NP-hard property of the problem [4].

Therefore, studies have approached the problem from various directions to solve the optimal camera placement problem within a modeling area that can reflect reality with complex conditions, and many approximation algorithms have been suggested as a result [4–15]. Previous literature in the modeling area and the camera installation area has its roots in 2D-based studies [12]. The greedy algorithm [8, 14], genetic algorithm (GA) [10, 15], particle swarm optimization (PSO) [11, 12], and so on have been used in existing studies as approximation algorithms to solve the problem. However, all the studies mentioned above have high computational complexity, for they found the solution directly at a high resolution. Table 1 lists the approximation algorithms suggested in previous studies.

Moreover, the 2D model is too simple to compute a real-world case of the optimal camera placement problem [12]; methods to solve the problem using 3D were studied in

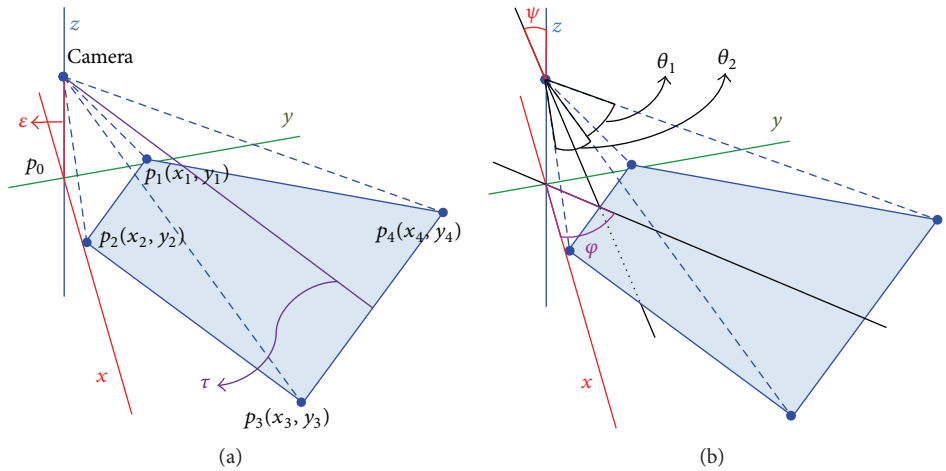


Figure 2: FOV computation parameters.

[11, 12]. However, 3D problem-solving exacerbates the issue of high computational complexity.

Previous studies have consistently reported the issue of high computational complexity as they continue to use problem-solving methods at high resolution. To remove this issue, phase 1 of the two-phase algorithm proposed in this study uses BIP to find the global optimal solution of the MIN problem within the low-resolution area (small number of grid points), and phase 2 uses an approximation algorithm, hill climbing method, to solve the FIX problem at a high resolution (large number of grid points).

With this process, the solution for a wider high-resolution area can be found based on the verified global optimal solution found in the low-resolution. Existing studies mainly used methods to avoid local optima, such as the genetic algorithm, particle swarm optimization, and simulated annealing, though they have high computational complexity [4]. Therefore, this study proposes using a hill climbing method, known to have low computational complexity. In general, greedy algorithms like a hill climbing method can find local optima if they are assigned the wrong starting point; however, this study proposes using the starting point found by BIP. The low computational complexity can reflect the modeling area of a large number of grid points with the same condition. Thus, this study proposes an approximation algorithm that is more likely to be used for real-world cases.

3. Model and Solution

This paper proposes a two-phase algorithm and assumes a 3D camera installation in a 2D modeling area. Phase 1 solves the problem using BIP, which offers an optimal solution by configuring the modeling area with a low-resolution grid for simple execution. Phase 2 finds a real-world applicable answer by setting the starting point from the low-resolution solution of phase 1 and then using the hill climbing method [19] for the modeling area configured with high-resolution grids.

3.1. Modeling Space. This paper assumes the surveillance of a plane area without obstacles. The surveillance area is divided into grid points, as shown in [13], and a grid point is captured by the camera if it is observed from the camera. As mentioned above, grid points refer to discrete points on x- and y-axes, separated by minimum distance Δ for the spatial sampling frequency [6]. Later, the plane area is divided into camera-installable and not camera-installable areas, and the surveillance area is assigned.

3.2. Modeling Surveillance Area. As in previous studies [4–12], field of view (FOV) modeling is proposed prior to explaining the placement method. Finding a solution for the optimal camera placement problem is equivalent to finding the conditions that create the FOV of each properly placed camera; the problem can be solved only if the method of computing the FOV is defined beforehand.

Like the study in [6], this study assumes a camera that is fixed in a certain direction so that it only surveils the same spot; therefore, a single camera has a fixed FOV depending on its installation condition. The FOV of the surveillance camera has a trapezoidal shape on the surveilled plane area, corresponding to the installation location (x_0, y_0), horizontal angle (φ), vertical angle (ψ), installation height (ε), horizontal and vertical angles of camera view (θ_1, θ_2), and maximum recognition distance (T). The horizontal and vertical angles of camera view mean the horizontal and vertical viewing angles of the scene captured by the camera.

Figure 2(a) shows the location of a camera which is installed at the ground coordinate $P_0(x_0, y_0)$ with the height ε and the recognition distance τ. Note that the actual recognition distance (τ) is less than or equal to the maximum recognition distance $(\tau \leq T)$. Figure 2(b) shows the horizontal view angle (θ_1) and the vertical view angle (θ_2), as well as the horizontal angle (φ) and the vertical angle (ψ). Here, the horizontal angle (φ) of the camera means the direction in which the camera watches. The vertical angle (ψ) is the watching angle of the camera, measured from a line perpendicular to the ground at the installation point.

Table 2: Camera specifications for the comparison test.

Consideration	Number of options	Details
Horizontal orientations (°)	8	$0, 45, 90, \ldots, 270, 315$
Vertical orientations (°)	15	$1, 3, \ldots, 27, 29$
Height (m)	1	7
Camera type		
Horizontal angle of view (°)	1	80
Vertical angle of view (°)	1	80
Available sight range (m)	1	60

Based on the given camera conditions, the algorithm to compute the coordinates of the trapezoid vertices, which are the FOV of the camera, is described as follows.

Step 0. $\tau = \varepsilon/\{\cos(\theta_2 + \psi)\}$. If $\theta_2 + \psi < 90$ or $\tau > T$, then stop the calculation.

Step 1. If $P_0 = (0, 0)$ and $\varphi = 0$, FOV is made of four vertices $\{p_1', p_2', p_3', p_4'\}$ (each point p_i' is made of $\{x_i', y_i'\}$):

$$\begin{bmatrix} h \times \tan\psi \\ \frac{h}{\cos\psi} \times \left\{+\tan\left(\frac{\theta_1}{2}\right)\right\} \\ h \times \tan\psi \\ \frac{h}{\cos\psi} \times \left\{-\tan\left(\frac{\theta_1}{2}\right)\right\} \\ h \times \tan(\psi + \theta_2) \\ \frac{h}{\cos(\psi + \theta_2)} \times \left\{+\tan\left(\frac{\theta_1}{2}\right)\right\} \\ h \times \tan(\psi + \theta_2) \\ \frac{h}{\cos(\psi + \theta_2)} \times \left\{-\tan\left(\frac{\theta_1}{2}\right)\right\} \end{bmatrix} = \begin{bmatrix} x_1' \\ y_1' \\ x_2' \\ y_2' \\ x_3' \\ y_3' \\ x_4' \\ y_4' \end{bmatrix}. \quad (1)$$

Step 2. Calculate vertex p_i'' by rotating p_i' by φ

$$\begin{bmatrix} \cos\varphi & -\sin\varphi \\ \sin\varphi & \cos\varphi \end{bmatrix} \begin{bmatrix} x_i' \\ y_i' \end{bmatrix} = \begin{bmatrix} x_i'' \\ y_i'' \end{bmatrix}. \quad (2)$$

Step 3. Actual camera installation information is added to each p_i''

$$\begin{bmatrix} x_i'' \\ y_i'' \end{bmatrix} + \begin{bmatrix} x_0 \\ y_0 \end{bmatrix} = \begin{bmatrix} x_i \\ y_i \end{bmatrix}. \quad (3)$$

Step 0 considers the maximum recognition distance (T), vertical angle (ψ), and vertical view angle (θ_2) to check whether the FOV can be computed. If τ exceeds the maximum recognition distance (T) set beforehand, FOV with such a condition does not exist and therefore is not computed. The FOV also does not exist if the sum of the vertical angle (ψ) and the vertical view angle (θ_2) exceeds 90 degrees, for the camera cannot see the floor.

Step 1 explains the calculation of the coordinates for the FOV trapezoid vertices, assuming that the surveillance camera is installed in a direction parallel to x-direction from the origin. Equation (1) takes the vertical angle (ψ) of the camera installation into account, as well as the vertical view angle (θ_2) and the horizontal view angle (θ_1).

Step 2 shows how to obtain the coordinates for the vertices of the FOV trapezoid by taking the horizontal angle (φ) of the installed camera into account, based on the value obtained from Step 1.

Step 3 includes the equation for calculating the coordinates for the vertices of the actual FOV trapezoid after adding the ground coordinates (x_0, y_0) of the installation point to the value from Step 2.

In conclusion, combining (1), (2), and (3) in Table 2 will consider the actual installation location for the camera and compute the coordinates of each vertex of the surveillance area (FOV trapezoid) of a single camera, using the matrix calculation of

$$\begin{bmatrix} \cos\varphi & -\sin\varphi & 0 & 0 & 0 & 0 & 0 & 0 \\ \sin\varphi & \cos\varphi & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \cos\varphi & -\sin\varphi & 0 & 0 & 0 & 0 \\ 0 & 0 & \sin\varphi & \cos\varphi & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \cos\varphi & -\sin\varphi & 0 & 0 \\ 0 & 0 & 0 & 0 & \sin\varphi & \cos\varphi & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \cos\varphi & -\sin\varphi \\ 0 & 0 & 0 & 0 & 0 & 0 & \sin\varphi & \cos\varphi \end{bmatrix} \begin{bmatrix} h \times \tan\psi \\ \frac{h}{\cos\psi} \times \left\{+\tan\left(\frac{\theta_1}{2}\right)\right\} \\ h \times \tan\psi \\ \frac{h}{\cos\psi} \times \left\{-\tan\left(\frac{\theta_1}{2}\right)\right\} \\ h \times \tan(\psi + \theta_2) \\ \frac{h}{\cos(\psi + \theta_2)} \times \left\{+\tan\left(\frac{\theta_1}{2}\right)\right\} \\ h \times \tan(\psi + \theta_2) \\ \frac{h}{\cos(\psi + \theta_2)} \times \left\{-\tan\left(\frac{\theta_1}{2}\right)\right\} \end{bmatrix} + \begin{bmatrix} x_0 \\ y_0 \\ x_0 \\ y_0 \\ x_0 \\ y_0 \\ x_0 \\ y_0 \end{bmatrix} = \begin{bmatrix} x_1 \\ y_1 \\ x_2 \\ y_2 \\ x_3 \\ y_3 \\ x_4 \\ y_4 \end{bmatrix}. \quad (4)$$

3.3. Two-Phase Algorithm. Our two-phase algorithm (a) generates the grid model of the candidate locations of camera installation and the target area for surveillance; (b) solves the small scale of phase 1 at a low-resolution; (c) sets the starting value of phase 2 based on the solution from the previous step; and (d) solves the large scale problem of phase 2 at a high-resolution.

3.3.1. Phase 1. In phase 1, the minimum number of cameras that satisfies the specific condition given with the grid points of the simulation area is obtained; it is then used to solve the MIN problem to find the location of each camera and the installation condition. We approached the problem using a further-developed method based on BIP formulas [6] for the existing 2D placement problem to solve the 3D placement problem. The detailed procedure is as follows.

(1) First, decision variables are assigned, just as when solving a general BIP.

$$x_{ijdet} = \begin{cases} 1, & \text{If there exists a camera at position } i \text{ with horizontal orientation } j \text{ vertical orientation } d, \text{height } e, \text{and angle of view (AOV) } t, \\ 0, & \text{Otherwise,} \end{cases} \tag{5}$$

$$y_k = \begin{cases} 1, & \text{If the target at position } k \text{ is covered by a camera,} \\ 0, & \text{Otherwise.} \end{cases}$$

Thus, x_{ijdet} is 1 if there exists a camera at position i with horizontal angle j, vertical angle d, height e, and AOV t and 0 if not and y_k equals 1 if the surveillance area k is watched with x_{ijdet} and 0 if not.

(2) Other parameters required for the formula are defined as follows:

$$v_{ijdetk} = \begin{cases} 1, & \text{if the target at position } k \text{ is visible from camera position } i \text{ with horizontal orientation } j, \text{vertical orientation } d, \text{height } e, \text{and AOV } t, \\ 0, & \text{Otherwise.} \end{cases} \tag{6}$$

N_C: number of camera positions.

N_{hD}: number of horizontal orientations.

N_{vD}: number of vertical orientations.

N_E: number of heights.

N_A: number of camera types.

N_T: number of target positions.

CVR: given minimal coverage rate.

(3) The objective function minimizes the number of cameras as follows:

$$\min \quad \sum_{i=1}^{N_C} \sum_{j=1}^{N_{hD}} \sum_{d=1}^{N_{vD}} \sum_{e=1}^{N_E} \sum_{t=1}^{N_A} x_{ijdet}. \tag{7}$$

(4) The following constraints are also necessary:

$$\sum_{i=1}^{N_C} \sum_{j=1}^{N_{hD}} \sum_{d=1}^{N_{vD}} \sum_{e=1}^{N_E} \sum_{t=1}^{N_A} v_{ijdetk} \cdot x_{ijdet} \geq y_k \quad \text{for } \forall k \left(1, \ldots, N_T\right), \tag{8}$$

$$\sum_{i=1}^{N_C} \sum_{j=1}^{N_{hD}} \sum_{d=1}^{N_{vD}} \sum_{e=1}^{N_E} \sum_{t=1}^{N_A} v_{ijdetk} \cdot x_{ijdet} \leq N_C \cdot y_k \quad \text{for } \forall k \left(1, \ldots, N_T\right), \tag{9}$$

$$\sum_{k=1}^{N_A} y_k \geq N_T \cdot \text{CVR}. \tag{10}$$

Equations (8) and (9) are the constraints to obtain y_k and (10) is a constraint such that the sum of y_k must be greater than or equal to the product of the minimum coverage rate and the target surveillance area.

3.3.2. Phase 2. Phase 2 solves the FIX problem, which finds the combination for maximum coverage with the constraint of the number of cameras determined from phase 1, with the result obtained from phase 1 as the initial value. This study uses the hill climbing method [19]. Since the objective function is nongradient, a direct-search method was applied; among the different direct-search methods, the alternating variable search method was used, as the problem has a multidimensional variable. This study proposes a hill climbing method, as [19] proposed such a method.

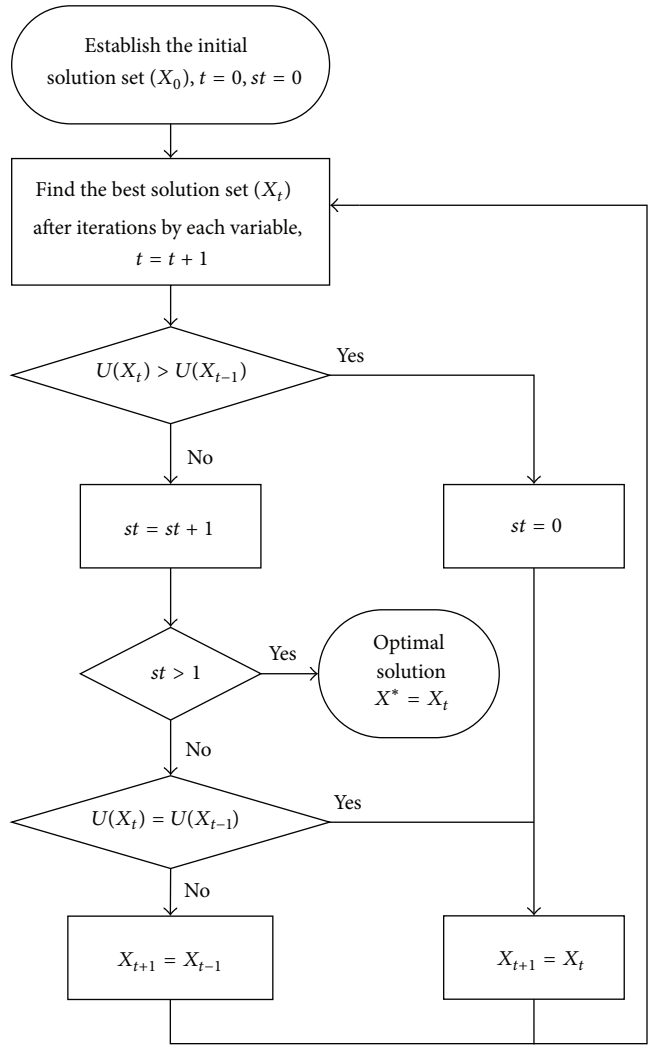


Figure 3: Flowchart of Phase 2.

Since phase 2 aims to maximize the coverage rate with the minimum number of cameras obtained in phase 1, the number of cameras does not change; only the conditions for each camera change. Each x_{ijdet} has the specification data for the x-coordinate, y-coordinate, horizontal angle (φ), vertical angle (ψ), installation height (ε), horizontal view angle (θ_1), and vertical view angle (θ_2) according to the camera type, as well as the maximum recognition distance. Each piece of information is both variable and dimensional. A flowchart of this phase is shown in Figure 3 and a hill climbing method of phase 2 is described as follows.

The notations and their meaning are as follows:

X: set of x_{ijdet}

X_t: set of optimal solutions found in the tth operation

t: the number of operations

st: the number of iterations that have an identical objective value

$U(X_t)$: coverage rate of camera set X at tth trial

n: index of a camera

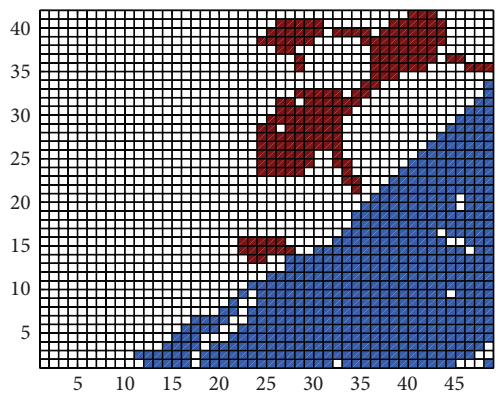
m: index of a camera setting variable (one of i, j, d, e, t). (i has both x- and y-coordinate properties, therefore treated as two different variables).

Step 0 (initialization). Establish a starting point X_0^* and set $t = 0$, $st = 0$.

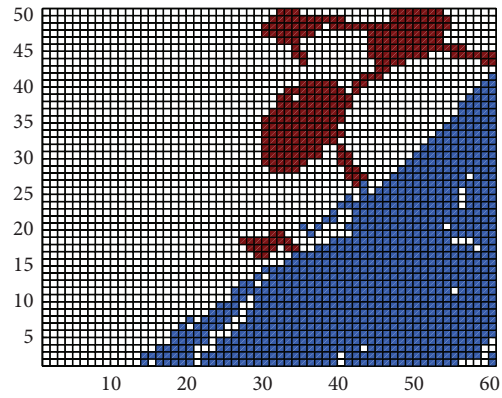
Step 1 (variant search). Set $t = t + 1$.

In X, the nth variable x_{ijdet} is selected. For each index (i, j, d, e, t) of x_{ijdet}, the variants are generated by numerically changing mth index of x_{ijdet} to the ± direction. The objective function for each variant is evaluated, and the best one among the tried variants is taken upon comparing the values of the objective function.

(a) Picture of modeling area



(b) 48 × 48 grid model



(c) 60 × 60 grid model

Figure 4: Map of Sevit Island (683 Olympic Blvd, Seocho-gu, Seoul, Korea) width 400 m; length 400 m.

Step 1.0 (initialization). Establish a starting point X_t and set $n = 0, m = 0$.

Step 1.1 (selection of a camera and calculation of the coverage rate)

For $n = 1: n(X_t)$;

Let $X_t^{\text{best}} = X_t$;
Let $x_n = n$th variable in X_t
For $m = 1:6$

Let X_{tnm+} to be a variant solution generated by numerically increase the value of the mth index of x_n, $U(X_{tnm+})$ to be the coverage rate of X_{tnm+};
Let X_{tnm-} to be a variant solution generated by numerically decrease the value of the mth index of x_n, and $U(X_{tnm-})$ to be the coverage rate of X_{tnm-};
Update X_t^{best} to be $\text{argmax}_{X \in \{X_t^{\text{best}}, X_{tnm+}, X_{tnm-}\}} (U(X))$, where in case of a tie one is randomly selected

End of For m

End of For n.

Step 1.2 (solution improvement). Set $X_t = X_t^{\text{best}}$.

Step 2 (check for improvement)

If $U(X_t) > U(X_{t-1})$, then set $X_{t+1} = X_t$, $st = 0$ and go to Step 1.

Otherwise set $st = st + 1$.

Step 3 (termination criterion)

If $st > 1$, then $X^* = X_t$; terminate the search.

Otherwise go to Step 4.

Step 4 (check for equivalent)

If $U(X_t) < U(X_{t-1})$, then set $X_{t+1} = X_{t-1}$, $st = 0$; and go to Step 1.

Otherwise set $X_{t+1} = X_t$ and go to Step 1.

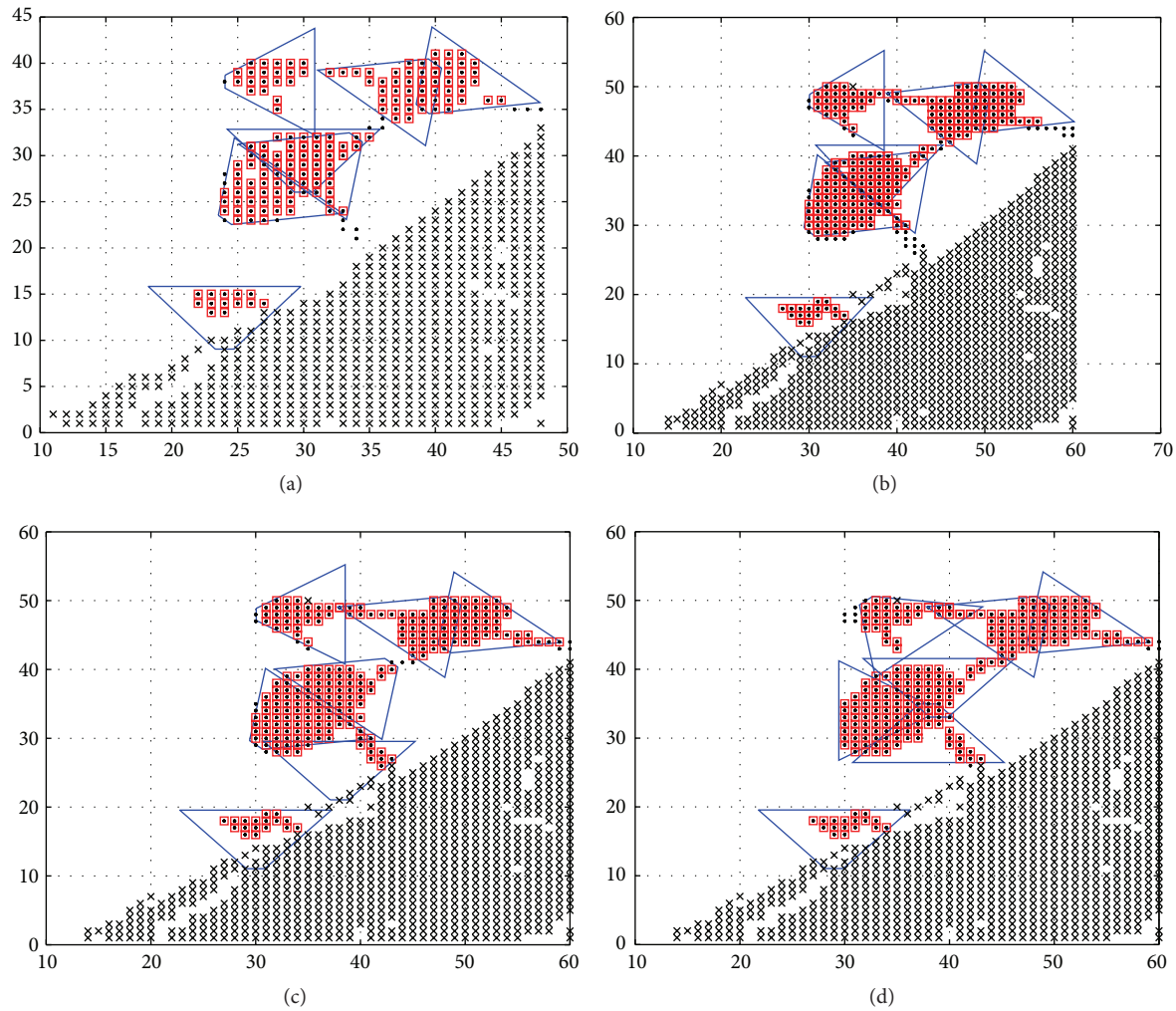


Figure 5: Solution for each step.

Step 0 sets the initial starting point. Here, t and st are set to 0.

Step 1 finds the best variant to improve the objective function, among the variants generated by increasing or decreasing the indices of each x_{ijdet} in the solution set X_t. Here, t is increased by 1. Step 1.0 initializes the current optimal solution X_t to be the starting point of this iteration. Step 1.1 finds the best solution which gives the highest coverage rate among the variants of X_t, generated by changing each of indices if x_{ijdet} is in X_t.

Step 2 checks if the coverage rate of X_t found in Step 1 actually increased compared to the coverage rate of X_{t-1}. If it is increased, X_{t+1} stays as X_t and st becomes zero. If not, st is increased by 1.

Step 3 checks whether the terminating conditions have been met. Here, if st is greater than 1, the optimal solution of phase 2 becomes $X^* = X_t$ and the solution search is completed.

Step 4 checks if the coverage rate of X_t decreased from the coverage rate of X_{t-1}. If decreased, X_{t+1} is reverted to X_{t-1} and moves to Step 1; if not, meaning that the values are the same, X_{t+1} stays at X_t and Step 1 is repeated.

The existing alternating variable search method [19] assigns the variable of the changing dimension; when it changes, the moving direction and the magnitude of change for the dimension variable where the objective function value improved the most are found. However, the method adapted for this study preassigns the magnitude of change for the dimension variable, and then finds which dimension variable in which direction improves the value of the objective function most. This method was utilized because the existing method vibrates near the optimal point, causing prolonged computation. Thus, it was more desirable to stop the computation at a proper point and use the solution than to continue the computation until it finished [19]. The method developed in this study eliminates the vibrating solution problem.

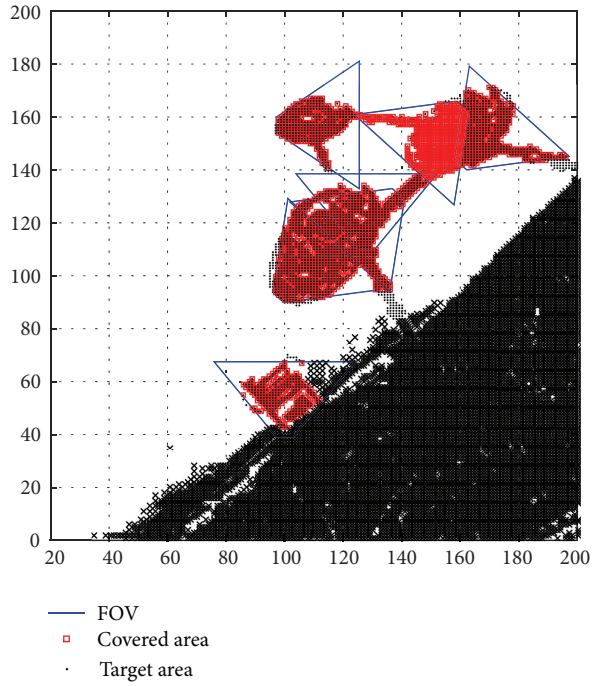


Figure 6: Solution obtained from high-resolution terrain.

4. Experimental Results

To evaluate the time efficiency and the coverage rate of the proposed two-phase algorithm, a comparison test was performed using a PC with an Intel Core i5-3337U processor, with 8 GB of DDR3 SDRAM. Matlab R2013b was used, and the BIP solution was obtained using the IBM ILOG CPLEX Optimization Studio 12.6.1. Additionally, the satellite map image of the actual landscape was transformed from an image to text (or numbers) with coordinates using Ascgen 2.0.0 from Jonathan Mathews Software.

To compare the global optimal solution obtained from BIP with a solution of the proposed two-phase algorithm, the modeling area was configured as in Figure 4. Figure 4(a) is the picture of the actual landscape of a 400 m × 400 m square for the modeling area. Figure 4(b) is the modeling area transformed into 2304 low-resolution grids (48 × 48), and Figure 4(c) is the one transformed into 3600 high-resolution grids (60 × 60). It can be seen that the modeling area in Figure 4(c) with higher resolution reflects the actual landscape of Figure 4(a) better than the modeling area in Figure 4(b) at low-resolution. Since the geometric feature of the Sevit Island shown in Figure 4 is quite complicated requiring a high-resolution modeling, it is adequate to test the effectiveness of the proposed algorithm. Experimental results for three other problems are shown in the appendix.

The specifications of cameras for the comparison test are shown in Table 2. The horizontal angle for the camera l installation has eight options, starting from 0 degrees and stepping by 45 degrees; the vertical angle has 15 options, starting from 1 degree and stepping by 2 degrees. The height of the installed camera is assumed to be 7 m, the horizontal view angle (θ_1) and vertical view angle (θ_2) are 80 degrees, and the maximum recognition distance is set to 60 m.

With the specifications shown in Table 2, a test that solves a real-world camera placement problem was carried out and the results were compared. As mentioned before, the two-phase algorithm consists of phase 1, which finds the initial solution (Figure 5(a)) using BIP with the specifications of Table 2 in the modeling area of Figure 4(b), and phase 2, which sets the starting point (Figure 5(b)) of the hill climbing method in the modeling area in Figure 4(c), based on the solution of Figure 5(b), and finds the solution (Figure 5(c)). Figure 5(d) is the solution directly obtained by BIP in the modeling area of Figure 4(c). Consequently, results from the proposed approximation algorithm and BIP can be evaluated, respectively, by comparing Figures 5(c) and 5(d).

The coverage rate of the solution and the time required for the test in case of Sevit Island area are as follows. When using the two-phase algorithm proposed in this study, the coverage rate is 94.72%, whereas it is 96.23% when using BIP. Therefore, comparing the solutions' coverage rates indicates that the proposed approximation algorithm obtains a solution with 98.43% of BIP's quality. The approximation algorithm took 15,823 ms, whereas BIP took 31,724 ms. Thus, the approximation algorithm needs about 49.88% of BIP's computing time.

As shown in these results, the coverage rate of the solution computed by the two-phase algorithm was comparable to that computed by BIP. This stems from the disadvantage of using the hill climbing method, which is simpler than the metaheuristic approximation used in previous studies, which

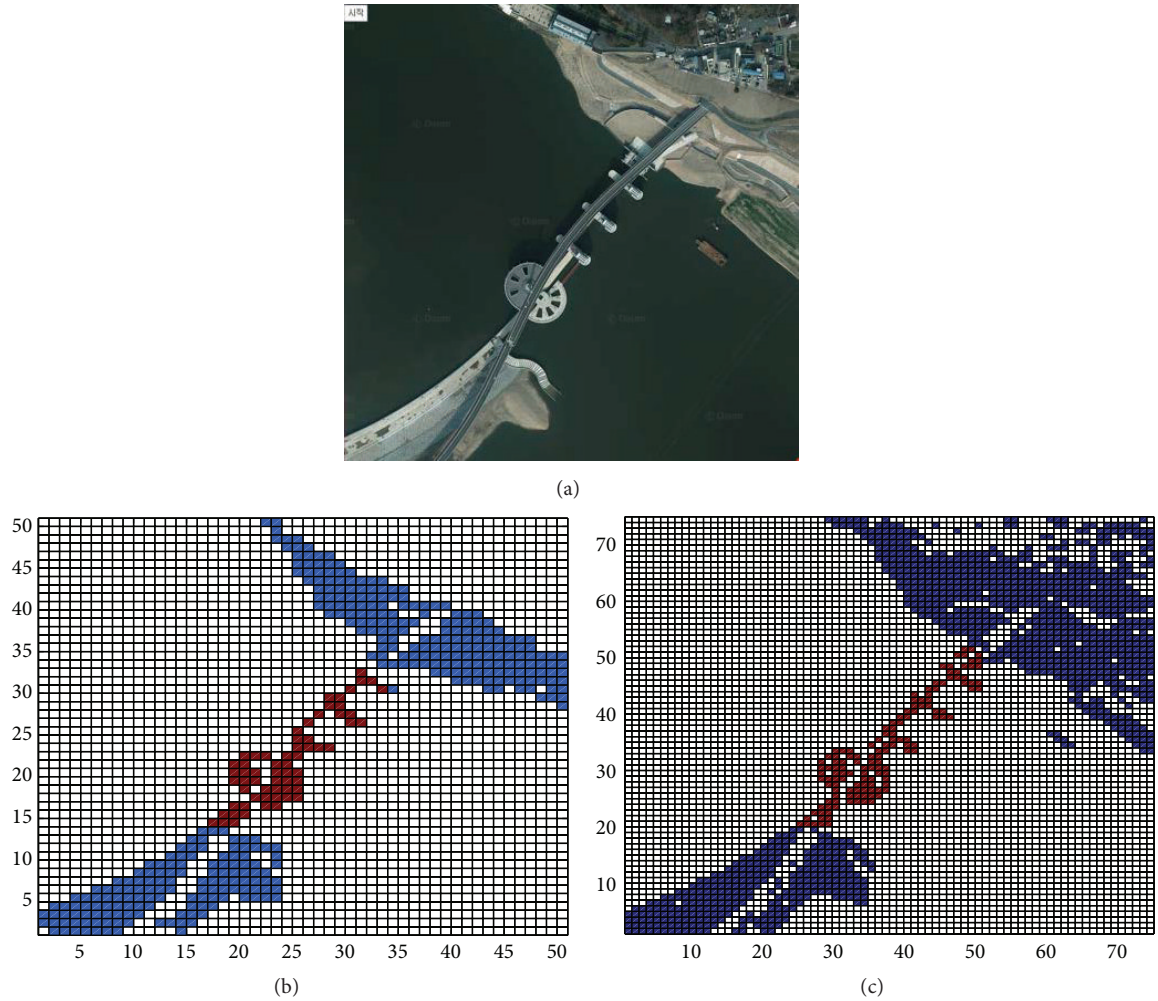


Figure 7: Map of Gangjeong Goryeong-bo (Daegu, Korea) 600 m × 600 m square.

was complemented by using the optimal solution obtained by BIP in phase 1 as the starting point.

Moreover, the two-phase algorithm proposed in this study solved the problem more quickly than BIP. This means that the computational complexity of our proposed model is lower than that of BIP and, as mentioned in Section 2, our model is more adequate to a large-area problem, as well.

For the comparison study in this passage, phase 2 was performed with the area of a 3600 grid (60 × 60), which is much smaller than the actual area that can be computed for a comparison of the solution quality. However, it would be more realistic and more accurate to approach with the higher-resolution area in phase 2, because the high-resolution terrain can reflect the actual landscape more precisely than the low-resolution terrain in the same space. Figure 6 shows the solution obtained by phase 2, performed in a higher-resolution terrain (40,000 grids (200 × 200)) based on Figure 5(b). While existing studies have a difficulty in finding a solution for such a large terrain, the approximation algorithm proposed in this study can find a solution.

We performed the test not only with Sevit Island as a modeling area, but also with other modeling areas. Table 3 shows the comparison results of the two-phase algorithm (TPA) and BIP in the other modeling areas, of which solution details are described in Appendix.

This study was able to find the solution for terrains with a large number of grid points because it used phase 1 and phase 2. Phase 1 finds the global optimal solution using BIP at a low resolution, and phase 2 elaborates on the solution offered in phase 1 at a high resolution. This study's contribution is providing an effective method to solve the optimal camera placement problem for a wide, detailed area, which can be applied in real-world situations.

5. Conclusions

This study presented a two-phase approximation algorithm to solve the optimal camera placement problem. This algorithm had lower computational complexity than existing methods and did not reduce the quality of the solution. As a result,

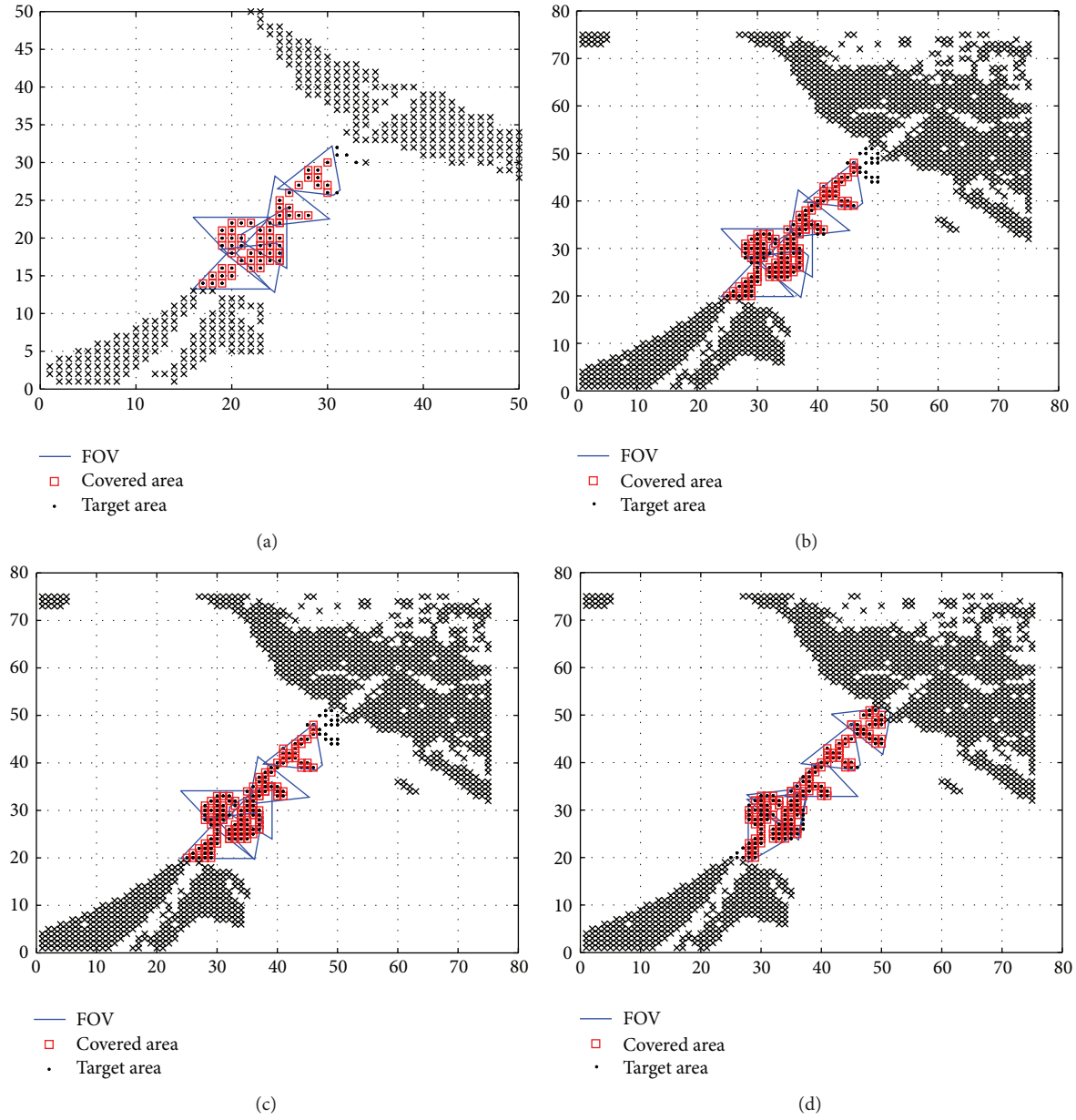


Figure 8: Solution to each step for Gangjeong Goryeong-bo (Weir) problem.

Table 3: Comparison test results.

Test area	Computing time (milliseconds)			Coverage rate (%)		
	TPA	BIP	TPA/BIP	TPA	BIP	TPA/BIP
Sevit Island	15,823	31,724	49.88%	94.72	96.23	98.43%
Gangjeong Goryeong-bo (weir)	4,378	17,030	25.71%	85.51	90.58	94.40%
Incheon port	9,896	32,665	30.30%	86.32	91.45	94.40%
Dongmyeong dock	31,403	48,785	64.37%	70.00	80.56	86.89%

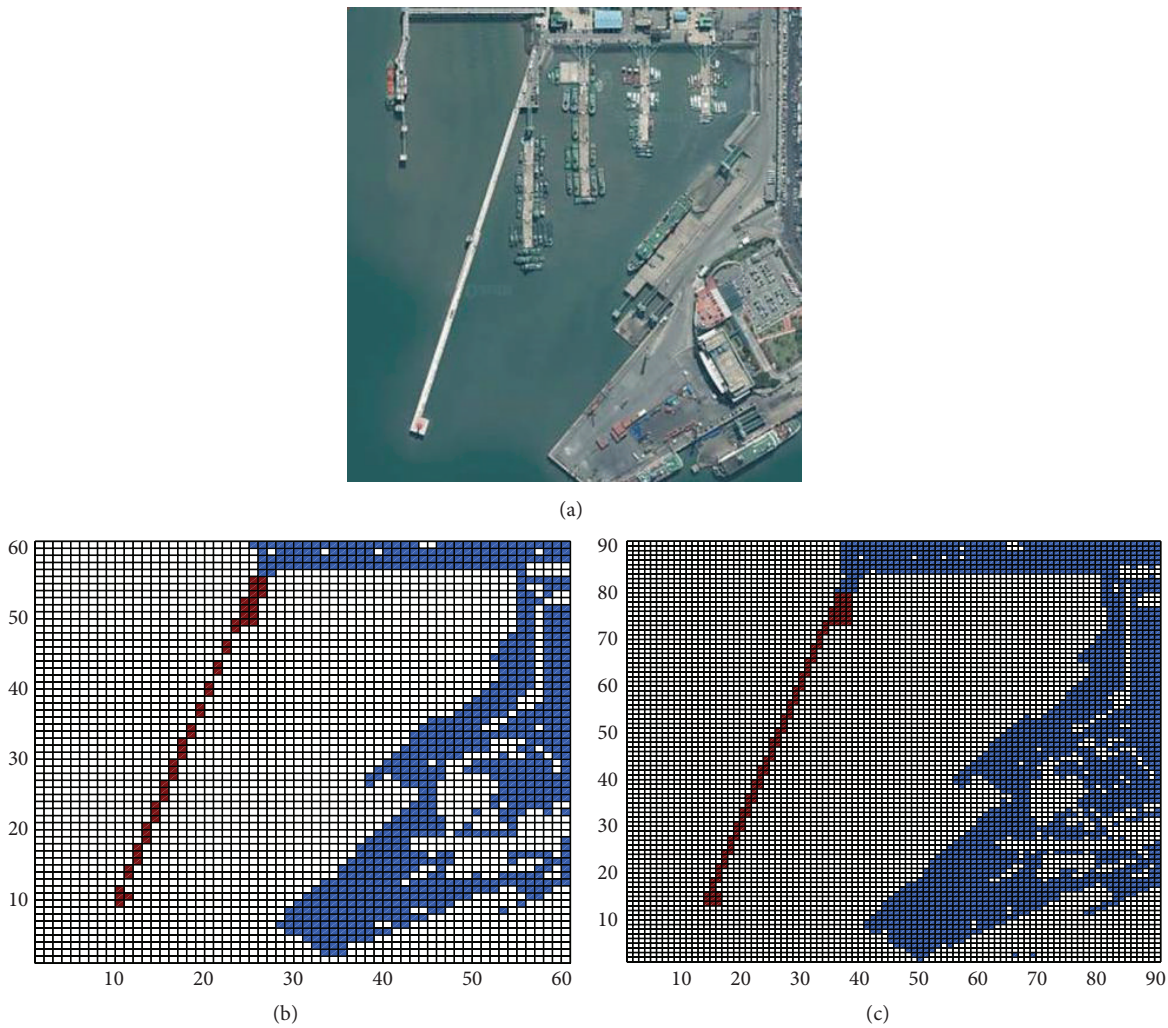


Figure 9: Map of Incheon International Airport Passenger Terminal (Incheon, Korea) 700 m × 700 m square.

the optimal camera placement problem could be solved, even with wide, real-world terrain under complex conditions that could not be solved with the existing BIP method.

The two-phase algorithm proposed in this study finds a global optimal solution in phase 1 to use as the starting point in phase 2; thus, the confidence for the starting point is large. A comparison study in Section 4 reveals that the quality of the solution did not show significant differences from BIP.

Meanwhile, the two-phase algorithm proposed in this study had too high computational complexity to solve complex problems and thus could not reflect reality. However, the low-resolution problem could be solved using BIP, which not only offered a global optimal solution but also provided the idea of applying the solution of phase 1 in higher-resolution terrains that look more like reality. Phase 2 changed the method of solving the problem so that other perspectives could be applied to problems in the future. This study used a hill climbing method in phase 2 with low computational complexity, but other methods, such as GA or PSO, could also be used in later studies.

The limitation of the study was that it used a hill climbing method in phase 2, which converges to relative local optimal solutions, instead of other approximate algorithms such as SA, GA, or PSO that have a higher chance of avoiding local optima. In this regard other approximate algorithms could also be applied in later studies. To solve the optimal camera placement problem with this modeling area with a large number of grid points under realistic restrictions, one of two approaches should be chosen: (i) to solve the problem through a simple algorithm at high resolution or (ii) to solve the problem at a relatively low resolution using a different method that needs more computational resources but has a higher change of finding a global optimal solution. This paper proposed the former to solve the problem at a high resolution in a wider terrain. Finding a proper balance point by comparing the two methods is left for future work.

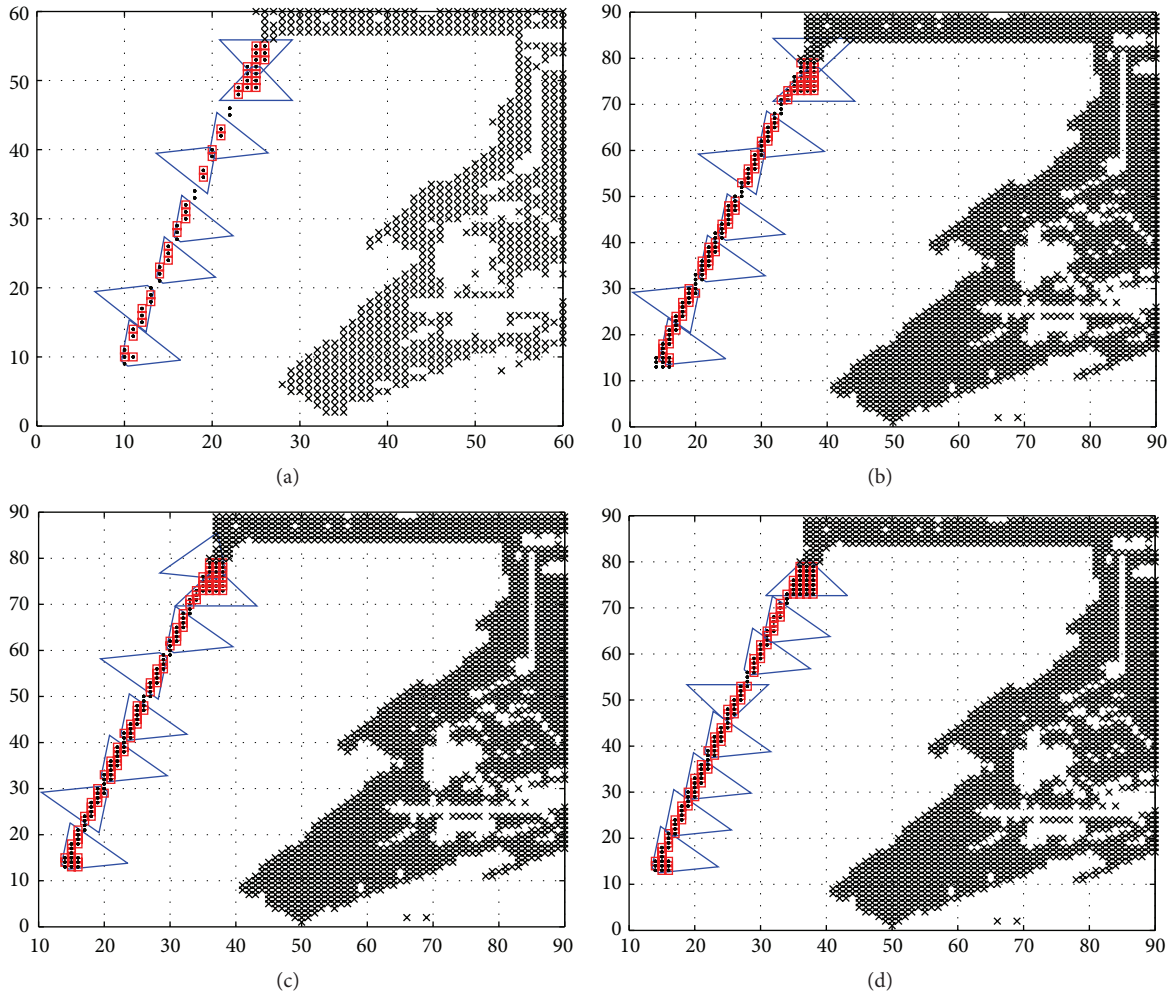


Figure 10: Solution to each step for the Incheon port problem.

Appendix

A. Additional Comparison Studies

The conditions for the additional comparison studies mentioned in Table 2 are identical to those described in Table 2.

A.1. Gangjeong Goryeong-bo (Weir) Problem. The modeling area in Figure 7 was configured to perform the model on Gangjeong Goryeong-bo (Weir) mentioned in Table 3. Figure 7(a) is the actual landscape of a 600 m × 600 m square to configure the modeling area. Figure 7(b) makes the 600 m × 600 m square terrain into a low-resolution modeling area of 2500 grids (50 × 50), and Figure 7(c) makes the same terrain into a high-resolution modeling area of 5625 grids (75 × 75).

A comparison test to solve the real-world optimal camera placement problem was carried out with the conditions mentioned above. Figure 8(a) shows the solution found in the modeling area in Figure 7(b), using BIP with the conditions of Table 2. Based on the solution of Figure 8(a), Figure 8(b) is the starting point of phase 2 for the hill climbing method in the modeling area of Figure 7(c). Figure 8(c) is the solution obtained by performing phase 2. The solution obtained by performing BIP in the modeling area of Figure 7(c) to begin with is Figure 8(d). The difference between the problems solved using the approximation algorithm proposed in this study or BIP can be studied by comparing Figures 8(c) and 8(d). The computing times and coverage rates of the result can be confirmed in Table 3.

A.2. Incheon Port Problem. The modeling area in Figure 9 was configured for the model on Incheon port mentioned in Table 3. Figure 9(a) is the actual landscape of a 700 m × 700 m square to configure the modeling area. Figure 9(b) shows the 700 m × 700 m square terrain as a low-resolution modeling area of 3600 grids (60 × 60), and Figure 9(c) shows the same terrain as a high-resolution modeling area of 6400 grids (80 × 80).

A comparison test to solve the real-world optimal camera placement problem was carried out with the conditions mentioned above. Figure 10(a) shows the solution found in the modeling area in Figure 9(b) using BIP with the conditions

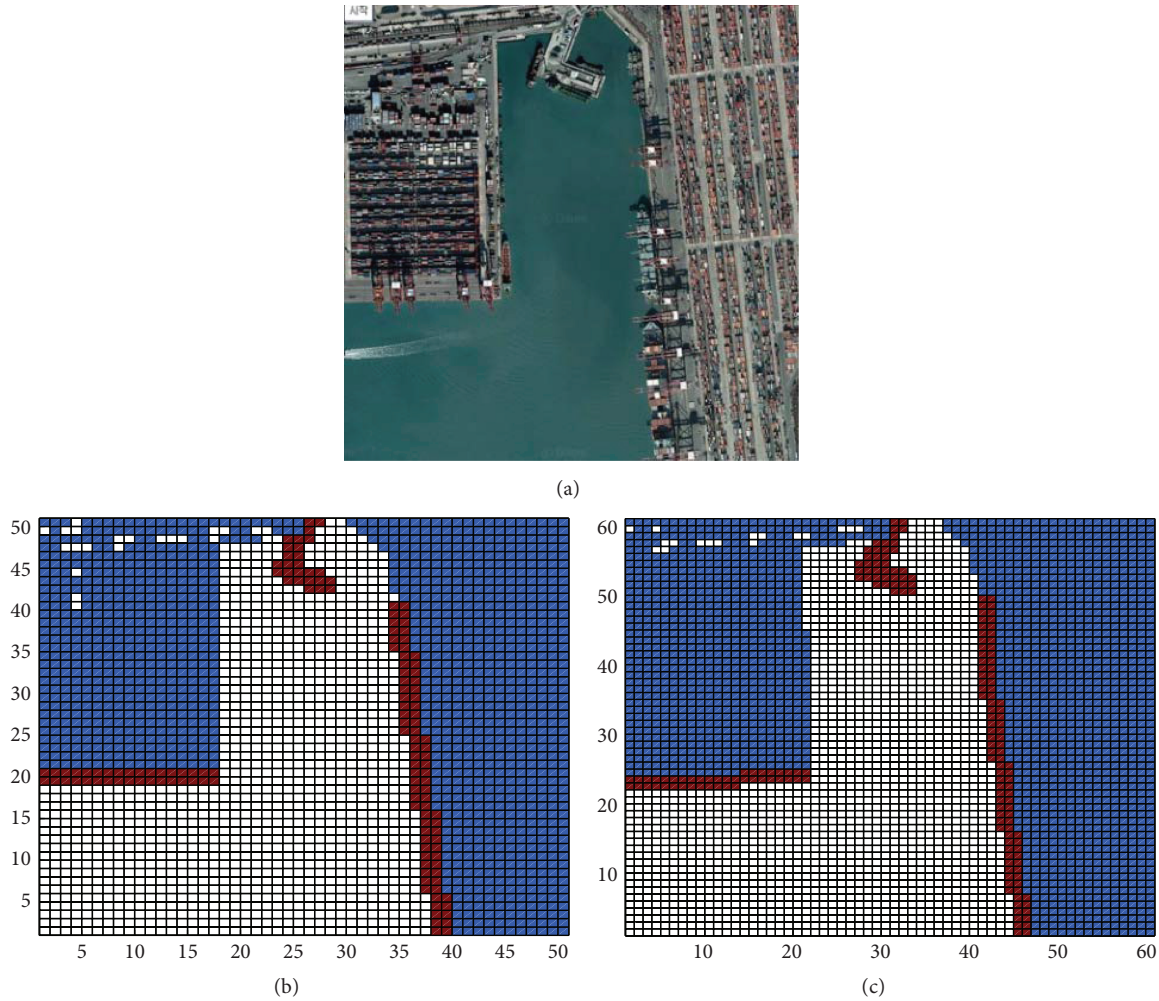


Figure 11: Map of Dongmyeong dock (Busan, Korea) 1000 m × 1000 m square.

of Table 2. Based on the solution of Figure 10(a), Figure 10(b) is the starting point of phase 2 for the hill climbing method in the modeling area of Figure 9(c). Figure 10(c) is the solution obtained by performing phase 2. The solution obtained by performing BIP in the modeling area of Figure 9(c) to begin with is Figure 10(d). The difference between the problems solved using the approximation algorithm proposed in this study or BIP can be studied by comparing Figures 10(c) and 10(d). The computing times and coverage rates of the result can be confirmed in Table 3.

A.3. Busan Dongmyeong Dock Problem. The modeling area in Figure 11 was configured to perform the model on the Dongmyeong dock mentioned in Table 3. Figure 11(a) is the actual landscape of a 1000 m × 1000 m square to configure the modeling area. Figure 11(b) shows the 1000 m × 1000 m square terrain as a low-resolution modeling area of 2500 grids (50 × 50), and Figure 11(c) shows the same terrain as a high-resolution modeling area of 3600 grids (60 × 60).

A comparison test to solve the real-world optimal camera placement problem was carried out with the conditions mentioned above. Figure 12(a) shows the solution found in the modeling area in Figure 11(b) using BIP with the conditions of Table 2. Based on the solution of Figure 12(a), Figure 12(b) is the starting point of phase 2 for the hill climbing method in the modeling area of Figure 11(c). Figure 12(c) is the solution obtained by performing phase 2. The solution obtained by performing BIP in the modeling area of Figure 11(c) to begin with is Figure 12(d). The difference between the problems solved using the approximation algorithm proposed in this study or BIP can be studied by comparing Figures 12(c) and 12(d). The computing times and coverage rates of the result can be confirmed in Table 3.

Competing Interests

The authors declare that there are no competing interests regarding the publication of this paper.

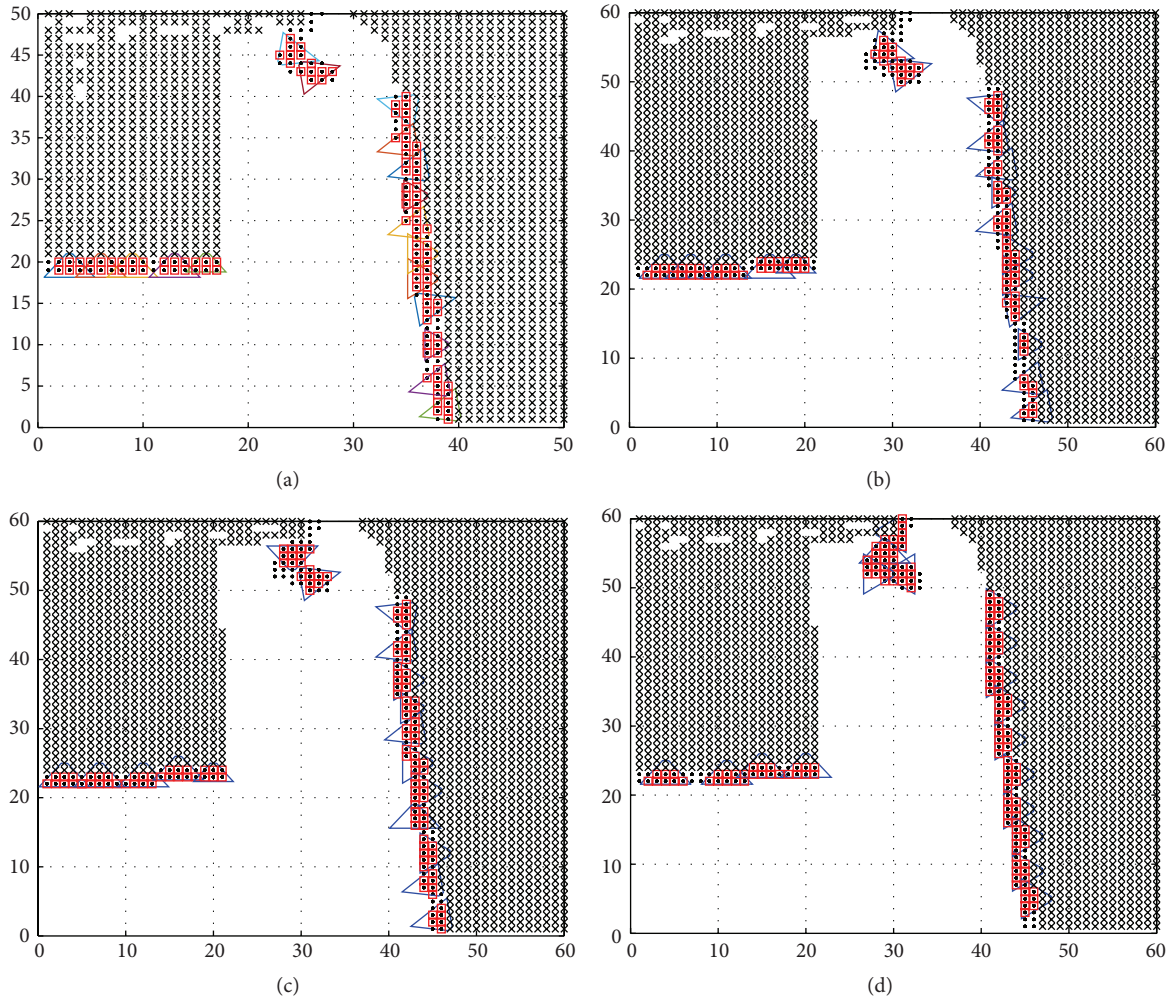


Figure 12: Solution to each step for the Busan Dongmyeong dock problem.

Acknowledgments

This research was supported by a grant (14SCIP-B065985-02) from Smart Civil Infrastructure Research Program funded by Ministry of Land, Infrastructure, and Transport (MOLIT) of Korea Government and Korea Agency for Infrastructure Technology Advancement (KAIA).

References

[1] K. Lee, K. Yim, and M. A. Mikki, "A secure framework of the surveillance video network integrating heterogeneous video formats and protocols," *Computers and Mathematics with Applications*, vol. 63, no. 2, pp. 525–535, 2012.

[2] N. Razmjooy, B. S. Mousavi, and F. Soleymani, "A real-time mathematical computer method for potato inspection using machine vision," *Computers and Mathematics with Applications*, vol. 63, no. 1, pp. 268–279, 2012.

[3] D. I. D. Staff, *DARPA's VIRAT: Video Search, with a Twist*, Defense Industry Daily, 2010.

[4] J. Zhao, R. Yoshida, S.-C. S. Cheung, and D. Haws, "Approximate techniques in solving optimal camera placement problems," *International Journal of Distributed Sensor Networks*, vol. 9, no. 11, Article ID 241913, 2013.

[5] J. Ai and A. A. Abouzeid, "Coverage by directional sensors in randomly deployed wireless sensor networks," *Journal of Combinatorial Optimization*, vol. 11, no. 1, pp. 21–41, 2006.

[6] E. Hörster and R. Lienhart, "Approximating optimal visual sensor placement," in *Proceedings of the IEEE International Conference on Multimedia and Expo (ICME '06)*, pp. 1257–1260, July 2006.

[7] J.-J. Gonzalez-Barbosa, T. García-Ramírez, J. Salas, J.-B. Hurtado-Ramos, and J.-D. Rico-Jiménez, "Optimal camera placement for total coverage," in *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA '09)*, pp. 844–848, IEEE Press, Kobe, Japan, May 2009.

[8] J. Zhao, S.-C. S. Cheung, and T. Nguyen, "Optimal visual sensor network configuration," in *Multi-Camera Networks: Principles and Applications*, H. Aghajan and A. Cavallaro, Eds., pp. 139–162, Academic Press, New York, NY, USA, 2009.

[9] Z. Fang and J. Wang, "Hybrid approximation for minimum-cost target coverage in wireless sensor networks," *Optimization Letters*, vol. 4, no. 3, pp. 371–381, 2010.

[10] A. Van Den Hengel, R. Hill, B. Ward et al., "Automatic camera placement for large scale surveillance networks," in *Proceedings*

of the Workshop on Applications of Computer Vision (WACV '09), pp. 1–6, Snowbird, Utah, USA, December 2009.

[11] Y. Morsly, N. Aouf, M. S. Djouadi, and M. Richardson, "Particle swarm optimization inspired probability algorithm for optimal camera network placement," *IEEE Sensors Journal*, vol. 12, no. 5, pp. 1402–1412, 2012.

[12] Y.-G. Fu, J. Zhou, and L. Deng, "Surveillance of a 2D plane area with 3D deployed cameras," *Sensors*, vol. 14, no. 2, pp. 1988–2011, 2014.

[13] K. Chakrabarty, S. S. Iyengar, H. Qi, and E. Cho, "Grid coverage for surveillance and target location in distributed sensor networks," *IEEE Transactions on Computers*, vol. 51, no. 12, pp. 1448–1453, 2002.

[14] M. Al Hasan, K. K. Ramachandran, and J. E. Mitchell, "Optimal placement of stereo sensors," *Optimization Letters*, vol. 2, no. 1, pp. 99–111, 2008.

[15] X. Chen and J. Davis, "Camera placement considering occlusion for robust motion capture," Tech. Rep. 2, Computer Graphics Laboratory, Stanford University, Stanford, Calif, USA, 2000.

[16] R. Cole and M. Sharir, "Visibility problems for polyhedral terrains," *Journal of Symbolic Computation*, vol. 7, no. 1, pp. 11–30, 1989.

[17] J. O'Rourke, *Art Gallery Theorems and Algorithms*, Oxford University Press, Oxford, UK, 1987.

[18] H. González-Baños, "A randomized art-gallery algorithm for sensor placement," in *Proceedings of the 17th Annual Symposium on Computational Geometry (SCG '01)*, pp. 232–240, Medford, Massachusetts, USA, June 2001.

[19] H.-P. P. Schwefel, *Evolution and Optimum Seeking: The Sixth Generation*, John Wiley & Sons, New York, NY, USA, 1993.

[20] IHS Technology, Video Surveillance & Storage—Opportunities at the Intersection of IT & Physical Security, 2014.

Hindawi Publishing Corporation
Scientific Programming
Volume 2016, Article ID 5606234, 13 pages
http://dx.doi.org/10.1155/2016/5606234

Research Article

Event-Triggered Faults Tolerant Control for Stochastic Systems with Time Delays

Ling Huang, Xuhuan Xie, and Wenbo Xie

School of Automation, Harbin University of Science and Technology, Harbin 150080, China

Correspondence should be addressed to Ling Huang; mail_huangling@163.com

Received 16 March 2016; Accepted 19 May 2016

Academic Editor: Guo Chen

This paper is concerned with the state-feedback controller design for stochastic networked control systems (NCSs) with random actuator failures and transmission delays. Firstly, an event-triggered scheme is introduced to optimize the performance of the stochastic NCSs. Secondly, stochastic NCSs under event-triggered scheme are modeled as stochastic time-delay systems. Thirdly, some less conservative delay-dependent stability criteria in terms of linear matrix inequalities for the codesign of both the controller gain and the trigger parameters are obtained by using delay-decomposition technique and convex combination approach. Finally, a numerical example is provided to show the less sampled data transmission and less conservatism of the proposed theory.

1. Introduction

Networked control systems (NCSs) are control systems wherein the control loops are closed through a certain digital communication network, which have been widely used in many fields, for example, vehicle industry, robot, and unmanned aerial vehicles, because of their benefits, for example, low installation and maintenance costs, high reliability, increased system flexibility, and decreased wiring. However, inserting a network to a control loop also brings a series of network-induced constraints, for example, time delays, packet losses, and competition of multiple nodes accessing network. The network-induced constraints will deteriorate the systems performance or even destabilize the system [1, 2]. Therefore, NCSs with network-induced constraints have gained considerable research interests and numerous results have been published [3–21].

So far, most of the analysis and synthesis for NCSs utilize time-triggered communications scheme, for example, stabilization [3, 4], filtering [5, 6], and tracking control [7]. It is well known that network-induced constraints are mainly caused by the limited network bandwidth. Moreover, it is very significant to mitigate the networked load to increase the lifespan of the battery of the nodes in wireless NCSs [8]. Therefore, to reduce the unnecessary waste of computation and transmitting sample data not only has theoretical importance but also has practical significance. In general, NCSs using event-triggered communication scheme can considerably reduce the network resource occupancy while maintaining the control performance compared to NCSs using time-triggered communication scheme [9]. So, event-triggered control for NCSs has received considerable attention in the past decade because of energy conservation and optimal performance, especially in the wireless NCSs [8], for example, event-triggered state-feedback control [10, 11], event-triggered guaranteed cost control [12], event-triggered dynamic output feedback control [13], event-triggered output tracking control [14], and event-triggered fuzzy filtering [15]. Considering the event-triggered control for NCS with stochastic perturbations [22–25], limited work has been reported in the open literatures. In [16], event-triggered stabilization for networked stochastic systems with multiplicative noise and network-induced delays has been investigated. However, the time delays in state and the effect of fault are not taken into account. On the one hand, fault tolerant control for NCSs has attracted great attention over the past decades because actuator faults caused by actuator aging or actuator in a hostile environment will have a great threat to the security and the reliability of NCSs [17–21]. On the other hand, systems with delays are ubiquitous, for example,

NCSs, biology systems, and hydraulic rolling mill systems [2, 26–28]. To the best of the authors' knowledge, there is no result reported in the open literatures on event-triggered faults tolerant control for stochastic systems with time delays. This motivates the study presented in the paper.

In the paper, we deal with the state-feedback controller design for stochastic NCSs with random actuator failures and transmission delays. The main works of this paper are as follows: (1) a triggered scheme is introduced to optimize the performance of stochastic NCSs; (2) stochastic NCSs under an event-triggered scheme are modeled as stochastic time-delays systems; (3) some less conservative stability criteria for the codesign of both the controller gain and the trigger parameters are obtained by using delay-decomposition technique and convex combination approach; (4) a practice example on event-triggered control for an unstable batch reactor is provided to show the merits of the proposed theory.

Notation. The notation used throughout the paper is fairly standard. N^T and N^{-1} denote the transpose and the inverse of matrix N, respectively. For a square matrix M, Sym$\{M\}$ is defined as $M+M^T$. The symmetric term in a matrix is denoted by $*$; for example, $\left[\begin{smallmatrix} X & Y \\ * & Z \end{smallmatrix}\right] = \left[\begin{smallmatrix} X & Y \\ Y^T & Z \end{smallmatrix}\right]$. The notation $P = P^T > 0$ ($P = P^T \geq 0$) means that the matrix P is real symmetric positive definite (positive semidefinite). $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\geq 0}, \mathbb{P})$ is a complete probability space and $E\{\cdot\}$ stands for the expectation operator with respect to the probability measure. diag$\{\cdots\}$ stands for a block diagonal matrix. Matrices, if not explicitly stated, are assumed to have appropriate dimensions.

2. Problem Formulation

Consider a linear system with state delay and Itô process, which is represented by the following stochastic differential equation:

$$\begin{aligned} dx(t) &= \left[Ax(t) + A_d x(t-\tau_d) + Bu(t)\right] dt \\ &\quad + \left[Hx(t) + H_d x(t-\tau_d)\right] d\varpi(t), \quad t > 0, \\ x(t) &= \varphi(t), \quad t \in [-\tau_d, 0], \end{aligned} \tag{1}$$

where $x(t) \in \mathbb{R}^n$ and $u(t) \in \mathbb{R}^m$ denote the state vector and the control input of the system, respectively, τ_d is a constant delay, A, A_d, B, H, and H_d are some constant matrices with appropriate dimensions, $\varpi(t)$ is a scalar standard Brownian Motion defined on a complete probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\geq 0}, \mathbb{P})$ with a natural filtration $\{\mathcal{F}_t\}_{t\geq 0}$ and satisfies $E\{d\varpi(t)\} = 0$ and $E\{d\varpi(t)^2\} = dt$, and $\varphi(t)$ is a continuous vector valued initial function defined on $[-\tau_d, 0]$.

The stochastic system controlled by a network is described in Figure 1.

To simplify the exposition, we make the following assumptions.

Assumption 1. All the system states variables of the controlled plant are available for a state-feedback control. The sensor is time-triggered with a constant sampling period h. The sampling sequence is described by the set $x(jh)$, $j = 1, 2, \ldots, \infty$.

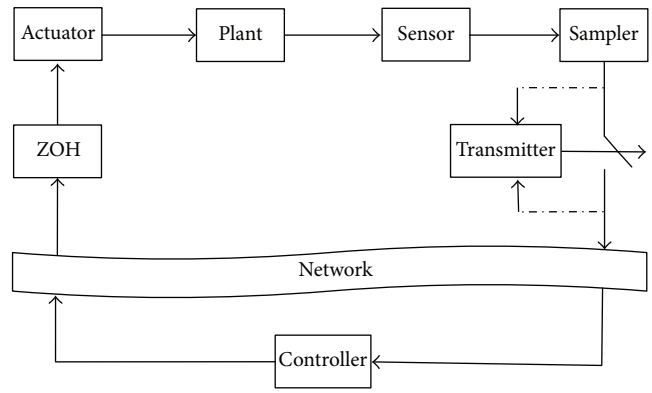


Figure 1: A framework of stochastic NCSs under an event-triggered communication scheme.

Assumption 2. The sampled data is transmitted in a single packet, and the packet losses do not occur in transmission. Whether the packets should be transmitted or not over network is determined by an event-triggered communication scheme. The successfully transmitted packet sequence is described by the set $x(t_k h)$, $k \in \{1, 2, \ldots, \infty\}$.

Assumption 3. The controllers and the actuators are event-triggered. The control input at the actuator is generated by a zero-order-holder (ZOH) with the holding time $t \in [t_k h + \tau_{t_k}, t_{k+1} h + \tau_{t_{k+1}}]$, where τ_{t_k} is the communication delay. Before the first control signal reaches the plant, the control input $u(t) = 0$.

The following state-feedback control law is employed for system (1):

$$u(t) = Kx(t). \tag{2}$$

Considering the effect of the transmission delay, (2) can be rewritten as

$$u(t) = Kx(t_k h), \quad t \in \left[t_k h + \tau_{t_k}, t_{k+1} h + \tau_{t_{k+1}}\right). \tag{3}$$

In order to improve the performance of the NCSs, the following event-triggered communication scheme is introduced to reduce the energy cost and the adverse effect of the transmission delay on performance of the NCSs:

$$\begin{aligned} &\left[x(t_k h) - x(t_k h + lh)\right]^T \Omega_2 \left[x(t_k h) - x(t_k h + lh)\right] \\ &\quad \geq \delta x^T(t_k h) \Omega_1 x(t_k h), \end{aligned} \tag{4}$$

where $l = 0, 1, \ldots, t_{k+1} - t_k - 1$, $0 \leq \delta < 1$, and Ω_1 and Ω_2 are the positive definite weighting matrices to be designed.

Remark 4. Only the sampled data satisfying inequality (4) will be transmitted since the transmitter in Figure 1 has a logic function to determine whether one should be transmitted or not; when $\delta = 0$, inequality (4) holds for all the sampled

state; hence, it shrinks to the periodic release case in [17, 18]. Obviously, the amount of transferred data can be reduced by set $\delta \neq 0$ in (4). If the amount of transferred data is reduced, the networked load will be reduced. Then, the energy cost used to transmit the unnecessary sampled data to the controller will be saved, and the adverse effect of the transmission delay on performance of the NCSs will be reduced since the transmission delay is mainly caused by the network traffic congestion.

Remark 5. A similar step to the one used in [16], two different weighting matrices Ω_1 and Ω_2 are introduced in event-triggering communication scheme (4) to obtain less conservative results. We further illustrate this in the following numerical example.

For a detailed timing analysis, we divide the holding interval of ZOH $t \in [t_k h + \tau_{t_k}, t_{k+1} h + \tau_{t_{k+1}}]$ into the following subsets:

$$\begin{aligned} &\left[t_k h + lh + \tau_{t_k}, t_k h + (l+1)h + \tau_{t_k}\right] \\ &\quad \cup \left[t_{k+1} h - h + \tau_{t_k}, t_{k+1} h + \tau_{t_{k+1}}\right], \\ &\qquad\qquad l = 0, 1, \ldots, t_{k+1} - t_k - 2. \end{aligned} \tag{5}$$

Define the following new variables:

$$\begin{aligned} &\tau(t) = t - t_k h - lh, \\ &e(t) = x(t_k h) - x(t_k h + lh), \\ &\qquad\qquad l = 0, 1, \ldots, t_{k+1} - t_k - 1. \end{aligned} \tag{6}$$

Then, we have

$$0 < \tau_1 \leq \tau(t) \leq h + \overline{\tau} \triangleq \tau_2. \tag{7}$$

The event-triggered algorithm (4) can be rewritten as

$$\begin{aligned} &e^T(t)\,\Omega_2 e(t) \\ &\quad \geq \delta\,(e(t) + x(t - \tau(t)))^T\,\Omega_1\,(e(t) + x(t - \tau(t))). \end{aligned} \tag{8}$$

Considering the effect of the stochastic actuator failure, the relationship between the control input $u(t)$ and the real actuator output $\widehat{u}(t)$ is described as follows:

$$\widehat{u}(t) = \Lambda u(t), \tag{9}$$

where

$$\Lambda = \operatorname{diag}\{\Lambda_1, \Lambda_2, \ldots, \Lambda_m\}, \quad \Lambda_i \in [0, \theta], \tag{10}$$

where $\theta \geq 1$ and Λ_i are the random variables which describe the relationship between the real executive amount of ith actuator and the control input for ith actuator. $\Lambda_i = 0$ represents that the ith actuator is completely fails, $\Lambda_i = 1$ represents that the ith actuator works normally, $0 < \Lambda_i < 1$ and $\Lambda_i > 1$ represent that the real actuator output is smaller and greater than the control input, respectively. The expectation and variance of Λ_i are $\overline{\Lambda}_i$ and β_i^2, respectively.

Define $e_i = \operatorname{diag}\{\underbrace{0, \ldots, 0}_{i-1}, I, \underbrace{0, \ldots, 0}_{m-i}\}$. Then

$$\begin{aligned} \overline{\Lambda} &= \operatorname{diag}\left\{\overline{\Lambda}_1, \overline{\Lambda}_2, \ldots, \overline{\Lambda}_m\right\} = \sum_{i=1}^{m} \overline{\Lambda}_i e_i, \\ \beta^2 &= \operatorname{diag}\left\{\beta_1^2, \beta_2^2, \ldots, \beta_m^2\right\} = \sum_{i=1}^{m} \beta_i^2 e_i. \end{aligned} \tag{11}$$

Based on the above analysis, (3) can be rewritten as

$$u(t) = B\Lambda K\,(e(t) + x(t - \tau(t))). \tag{12}$$

Substituting (12) into (1), the resultant closed-loop stochastic system with time delays can be expressed as follows:

$$\begin{aligned} dx(t) &= \left[Ax(t) + A_d x(t - \tau_d) + B\Lambda K x(t - \tau(t))\right. \\ &\quad \left. + B\Lambda K e(t)\right] dt + \left[Hx(t)\right. \\ &\quad \left. + H_d x(t - \tau_d)\right] d\varpi(t), \quad t > 0, \\ x(t) &= \varphi(t), \quad \forall t \in \left[-\max\{\tau_d, \tau_2\}, 0\right]. \end{aligned} \tag{13}$$

3. Main Results

This paper aims to develop a novel state-feedback controller for event-triggered stochastic NCSs with state delay, networked induce delay, and stochastic actuator failure. Before presenting the sufficient stabilization condition, we introduce the following lemmas, which are indispensable for the proofs in the sequel.

Lemma 6 (see [29]). *M_1, M_2, and M_3 are constant matrices with appropriate dimensions; then, a sufficient and necessary condition for*

$$\begin{aligned} &(\tau(t) - \tau_1) M_1 + (\tau_2 - \tau(t)) M_2 + M_3 < 0, \\ &\qquad\qquad \forall \tau(t) \in [\tau_1, \tau_2] \end{aligned} \tag{14}$$

is that the following inequalities hold simultaneously:

$$\begin{aligned} &(\tau_2 - \tau_1) M_2 + M_3 < 0, \\ &(\tau_2 - \tau_1) M_1 + M_3 < 0. \end{aligned} \tag{15}$$

Lemma 7 ((Schur complement) [30]). *Given matrices $Q = Q^T$, $R = R^T$, and S with appropriate dimension, the inequality*

$$\begin{bmatrix} Q & S \\ S^T & R \end{bmatrix} > 0 \tag{16}$$

is equivalent to $R > 0$, $Q - SR^{-1}S^T > 0$.

By employing delay-fraction technique, linear convex combination approach, and Lyapunov-Krasovskii stability theorem, the following sufficient stabilization condition for stochastic system (13) with time delay satisfying condition (7) is obtained.

Theorem 8. *For given scalars τ_d, τ_1, τ_2, and δ and a positive integer λ, the closed-loop stochastic system with time delays (13) and time-varying delay satisfying condition (7) is stochastically asymptotically stable if there exist $n \times n$ matrices $P = P^T > 0$, $\Omega_i = \Omega_i^T > 0$, $i = 1, 2$, $R_i = R_i^T > 0$, $i = 1, 2, \ldots, 6$, and $S_i = S_i^T > 0$, $i = 1, 2$, and $\lambda n \times \lambda n$ matrices $Q = Q^T > 0$ and $(\lambda + 5)n \times n$ matrices N_i, $i = 1, 2, 3, 4$, such that the following set of inequalities hold:*

$$\begin{bmatrix} \Xi_{11} & \Xi_{12} & \Xi_{13} & \Xi_{14} & \Xi_{15} & \Xi_{16} & \Xi_{17} & \Xi_{18}(l) \\ * & \Xi_{22} & 0_{4n\times 3n} & 0_{4n\times mn} & 0_{4n\times mn} & 0_{4n\times mn} & 0_{4n\times 6n} & 0_{4n\times n} \\ * & * & \Xi_{33} & 0_{3n\times mn} & 0_{3n\times mn} & 0_{3n\times mn} & 0_{3n\times 6n} & 0_{3n\times n} \\ * & * & * & \Xi_{44} & 0_{mn\times mn} & 0_{mn\times mn} & 0_{mn\times 6n} & 0_{mn\times n} \\ * & * & * & * & \Xi_{55} & 0_{mn\times mn} & 0_{mn\times 6n} & 0_{mn\times n} \\ * & * & * & * & * & \Xi_{66} & 0_{mn\times 6n} & 0_{mn\times n} \\ * & * & * & * & * & * & \Xi_{77} & 0_{6n\times n} \\ * & * & * & * & * & * & * & \Xi_{88} \end{bmatrix} < 0, \quad l = 1, 2, \tag{17}$$

where

$$\begin{aligned}
\Xi_{11} &= \mathrm{Sym}\left\{E_1 P \overline{f}_c\right\} + W_{\zeta_1}^T Q W_{\zeta_1} - W_{\zeta_2}^T Q W_{\zeta_2} \\
&\quad + E_1 S_1 E_1^T - E_{\lambda+2} S_1 E_{\lambda+2}^T + E_{\lambda+2} S_2 E_{\lambda+2}^T \\
&\quad - E_{\lambda+4} S_2 E_{\lambda+4}^T + \delta\left(E_{\lambda+3} + E_{\lambda+5}\right)\Omega_1\left(E_{\lambda+3}^T + E_{\lambda+5}^T\right) \\
&\quad - E_{\lambda+5}\Omega_2 E_{\lambda+5}^T + \mathrm{Sym}\left\{N_1\left(E_1^T - E_2^T\right)\right. \\
&\quad + N_2\left(E_1^T - E_{\lambda+2}^T\right) + N_3\left(E_{\lambda+2}^T - E_{\lambda+3}^T\right) \\
&\quad \left. + N_4\left(E_{\lambda+3}^T - E_{\lambda+4}^T\right)\right\}, \\
\Xi_{12} &= \left[g_c^T, \sqrt{\frac{\tau_d}{\lambda}} g_c^T, \sqrt{\tau_1} g_c^T, \sqrt{\tau_{12}} g_c^T\right], \\
\Xi_{22} &= \mathrm{diag}\left\{-P^{-1}, -R_2^{-1}, -R_4^{-1}, -R_6^{-1}\right\}, \\
\Xi_{13} &= \left[\sqrt{\frac{\tau_d}{\lambda}} f_c^T, \sqrt{\tau_1} f_c^T, \sqrt{\tau_{12}} f_c^T\right], \\
\Xi_{33} &= \mathrm{diag}\left\{-R_1^{-1}, -R_3^{-1}, -R_5^{-1}\right\}, \\
\Xi_{14} &= \sqrt{\frac{\tau_d}{\lambda}}\left[\beta_1 \widehat{f}_1^T, \beta_2 \widehat{f}_2^T, \ldots, \beta_m \widehat{f}_m^T\right], \\
\Xi_{44} &= \mathrm{diag}\left\{-R_1^{-1}, -R_1^{-1}, \ldots, -R_1^{-1}\right\}, \\
\Xi_{15} &= \sqrt{\tau_1}\left[\beta_1 \widehat{f}_1^T, \beta_2 \widehat{f}_2^T, \ldots, \beta_m \widehat{f}_m^T\right], \\
\Xi_{55} &= \mathrm{diag}\left\{-R_3^{-1}, -R_3^{-1}, \ldots, -R_3^{-1}\right\}, \\
\Xi_{16} &= \sqrt{\tau_{12}}\left[\beta_1 \widehat{f}_1^T, \beta_2 \widehat{f}_2^T, \ldots, \beta_m \widehat{f}_m^T\right], \\
\Xi_{66} &= \mathrm{diag}\left\{-R_5^{-1}, -R_5^{-1}, \ldots, -R_5^{-1}\right\}, \\
\Xi_{17} &= \left[\sqrt{\frac{\tau_d}{\lambda}} N_1, N_1, \sqrt{\tau_1} N_2, N_2, N_3, N_4\right], \\
\Xi_{77} &= \mathrm{diag}\left\{-R_1, -R_2, -R_3, -R_4, -R_6, -R_6\right\}, \\
\Xi_{18}(1) &= \sqrt{\tau_{12}} N_4, \\
\Xi_{18}(2) &= \sqrt{\tau_{12}} N_3, \\
\Xi_{88} &= -R_5, \\
W_{\zeta_1} &= \left[I_{\lambda n}, 0_{\lambda n\times 5n}\right], \\
W_{\zeta_2} &= \left[0_{\lambda, n\times n}, I_{\lambda n}, 0_{\lambda n\times 4n}\right], \\
\overline{f}_c &= A E_1^T + A_d E_{\lambda+1}^T + B\overline{\Lambda} K E_{\lambda+3}^T + B\overline{\Lambda} K E_{\lambda+5}^T, \\
\widehat{f}_i &= B e_i K E_{\lambda+3}^T + B e_i K E_{\lambda+5}^T, \\
g_c &= H E_1^T + H_d E_{\lambda+1}^T.
\end{aligned} \tag{18}$$

Proof. For the sake of notational convenience, let

$$\begin{aligned}
f(t) &= \overline{f}(t) + \widehat{f}(t), \\
g(t) &= Hx(t) + H_d x\left(t - \tau_d\right),
\end{aligned} \tag{19}$$

where

$$\begin{aligned}
\overline{f}(t) &= Ax(t) + A_d x\left(t - \tau_d\right) + B\overline{\Lambda} K x\left(t - \tau(t)\right) \\
&\quad + B\overline{\Lambda} K e(t), \\
\widehat{f}(t) &= B\left(\Lambda - \overline{\Lambda}\right) K x\left(t - \tau(t)\right) + B\left(\Lambda - \overline{\Lambda}\right) K e(t).
\end{aligned} \tag{20}$$

Then, we can recast (13) into following form:

$$d(t) = f(t)\, dt + g(t)\, d\varpi(t). \tag{21}$$

We divide the time-delay interval $[t - \tau_d, t]$ into λ uniform subintervals and define the following augmented variable:

$$\begin{aligned}
&\varsigma^T(t) \\
&\quad = \left[x^T(t), x^T\left(t - \frac{\tau_d}{\lambda}\right), \ldots, x^T\left(t - (\lambda - 1)\frac{\tau_d}{\lambda}\right)\right].
\end{aligned} \tag{22}$$

Choose the following simple functional candidate for system (21):

$$V(x(t), t) = \sum_{i=1}^{4} V_i(x(t), t), \tag{23}$$

where $x(t)$ denotes the function $x(s)$ defined on the interval $[t - \max\{\tau_d, \tau_2\}, t]$, and

$$\begin{aligned}
V_1(x(t),t) &= x^T(t) P x(t),\\
V_2(x(t),t) &= \int_{t-\tau_d/\lambda}^{t} \left[\varsigma^T(s) Q \varsigma(s) + \left(\frac{\tau_d}{\lambda} - t + s\right)\right.\\
&\quad \left.\cdot \left(f^T(s) R_1 f(s) + g^T(s) R_2 g(s)\right)\right] ds,\\
V_3(x(t),t) &= \int_{t-\tau_1}^{t} \left[x^T(s) S_1 x(s) + (\tau_1 - t + s)\right.\\
&\quad \left.\cdot \left(f^T(s) R_3 f(s) + g^T(s) R_4 g(s)\right)\right] ds,\\
V_4(x(t),t) &= \int_{t-\tau_1}^{t} \tau_{12} \left(f^T(s) R_5 f(s) + g^T(s)\right.\\
&\quad \left.\cdot R_6 g(s)\right) ds + \int_{t-\tau_2}^{t-\tau_1} \left[x^T(s) S_2 x(s)\right.\\
&\quad + (\tau_2 - t + s)\\
&\quad \left.\cdot \left(f^T(s) R_5 f(s) + g^T(s) R_6 g(s)\right)\right] ds.
\end{aligned} \tag{24}$$

For the sake of notational simplification, let

$$\begin{aligned}
\zeta_t^T &= \left[\varsigma^T(t), x^T(t-\tau_d), x^T(t-\tau_1), x^T(t-\tau(t)),\right.\\
&\quad \left. x^T(t-\tau_2), e^T(t)\right],\\
E_s &= \left[\underbrace{0_{n\times n}, \ldots, 0_{n\times n}}_{s-1}, I_{n\times n}, \underbrace{0_{n\times n}, \ldots, 0_{n\times n}}_{\lambda+5-s}\right]^T,\\
&\qquad\qquad s = 1, 2, \ldots, \lambda + 5,\\
\overline{f}_c &= AE_1^T + A_d E_{\lambda+1}^T + B\overline{\Lambda}KE_{\lambda+3}^T + B\overline{\Lambda}KE_{\lambda+5}^T,\\
\widehat{f}_i &= Be_iKE_{\lambda+3}^T + Be_iKE_{\lambda+5}^T,\\
g_c &= HE_1^T + H_dE_{\lambda+1}^T.
\end{aligned} \tag{25}$$

Taking the time derivatives of $V_i(x(t), t)$, $i = 1, 2, 3, 4$, along the trajectory of system (21) and taking expectation on it yield

$$\begin{aligned}
E\{dV_1(x(t),t)\} &= E\left\{2x^T(t) P f(t)\, dt\right.\\
&\quad \left. + g^T(t) P g(t)\, dt + 2x^T(t) P g(t)\, d\varpi(t)\right\}\\
&= \zeta_t^T \left\{2E_1 P \overline{f}_c dt + g_c^T P g_c dt\right\} \zeta_t,\\
E\left\{\dot{V}_2(x(t),t)\right\} &= \zeta_t^T \left\{W_{\varsigma_1}^T Q W_{\varsigma_1} - W_{\varsigma_2}^T Q W_{\varsigma_2}\right.\\
&\quad \left. + \frac{\tau_d}{\lambda} g_c^T R_2 g_c\right\} \zeta_t + E\left\{\frac{\tau_d}{\lambda} f^T(t) R_1 f(t)\right.\\
&\quad - \int_{t-\tau_d/\lambda}^{t} f^T(s) R_1 f(s)\, ds\\
&\quad \left. - \int_{t-\tau_d/\lambda}^{t} g^T(s) R_2 g(s)\, ds\right\},\\
E\left\{\dot{V}_3(x(t),t)\right\} &= \zeta_t^T \left\{E_1 S_1 E_1^T - E_{\lambda+2} S_1 E_{\lambda+2}^T\right.\\
&\quad \left. + \tau_1 g_c^T R_4 g_c\right\} \zeta_t + E\left\{\tau_1 f^T(t) R_3 f(t)\right.\\
&\quad \left. - \int_{t-\tau_1}^{t} f^T(s) R_3 f(s)\, ds - \int_{t-\tau_1}^{t} g^T(s) R_4 g(s)\, ds\right\},\\
E\left\{\dot{V}_4(x(t),t)\right\} &= \zeta_t^T \left\{E_{\lambda+2} S_2 E_{\lambda+2}^T - E_{\lambda+4} S_2 E_{\lambda+4}^T\right.\\
&\quad \left. + \tau_{12} g_c^T(t) R_6 g_c\right\} \zeta_t + E\left\{\tau_{12} f^T(t) R_5 f(t)\right.\\
&\quad \left. - \int_{t-\tau_2}^{t-\tau_1} f^T(s) R_5 f(s)\, ds - \int_{t-\tau_2}^{t-\tau_1} g^T(s) R_6 g(s)\, ds\right\},
\end{aligned} \tag{26}$$

where $\tau_{12} \triangleq \tau_2 - \tau_1$.

The calculation shows that

$$\begin{aligned}
&E\left\{\frac{\tau_d}{\lambda} f^T(t) R_1 f(t) + \tau_1 f^T(t) R_3 f(t) + \tau_{12} f^T(t)\right.\\
&\quad \left.\cdot R_5 f(t)\right\}\\
&= E\left\{\zeta_t^T \left\{\overline{f}_c^T \left(\frac{\tau_d}{\lambda} R_1 + \tau_1 R_3 + \tau_{12} R_5\right) \overline{f}_c\right.\right.\\
&\quad \left.\left. + \sum_{i=1}^{m} \beta_i^2 \widehat{f}_i^T \left(\frac{\tau_d}{\lambda} R_1 + \tau_1 R_3 + \tau_{12} R_5\right) \widehat{f}_i\right\} \zeta_t\right\}.
\end{aligned} \tag{27}$$

According to the Newton-Leibniz formula, the inner relationship among the terms in the Leibniz-Newton formula is revealed by introducing free weighting matrices N_i, $i = 1, 2, 3, 4$. Then, the following terms are true:

$$\begin{aligned}
&2\zeta_t^T N_1 \left(x(t) - x\left(t - \frac{\tau_d}{\lambda}\right) - \int_{t-\tau_d/\lambda}^{t} f(s)\, ds\right.\\
&\quad \left. - \int_{t-\tau_d/\lambda}^{t} g(s)\, d\varpi(s)\right) \equiv 0,\\
&2\zeta_t^T N_2 \left(x(t) - x(t-\tau_1) - \int_{t-\tau_1}^{t} f(s)\, ds\right.\\
&\quad \left. - \int_{t-\tau_1}^{t} g(s)\, d\varpi(s)\right) \equiv 0,\\
&2\zeta_t^T N_3 \left(x(t-\tau_1) - x(t-\tau(t)) - \int_{t-\tau(t)}^{t-\tau_1} f(s)\, ds\right.\\
&\quad \left. - \int_{t-\tau(t)}^{t-\tau_1} g(s)\, d\varpi(s)\right) \equiv 0,
\end{aligned}$$

$$2\zeta_t^T N_4\left(x(t-\tau(t))-x(t-\tau_2)-\int_{t-\tau_2}^{t-\tau(t)} f(s)\,ds - \int_{t-\tau_2}^{t-\tau(t)} g(s)\,d\omega(s)\right)\equiv 0. \tag{28}$$

Moreover, from the well-known fact, $-2a^Tb \le a^TR_i^{-1}a + b^TR_ib$, $i=1,2,\ldots,6$, for two appropriate dimensional vectors a and b, it can be easily checked that

$$\begin{aligned}
&2\zeta_t^T N_1\left(-\int_{t-\tau_d/\lambda}^{t} f(s)\,ds-\int_{t-\tau_d/\lambda}^{t} g(s)\,d\omega(s)\right)\\
&\le \zeta_t^T\left(\frac{\tau_d}{\lambda}N_1R_1^{-1}N_1^T+N_1R_2^{-1}N_1^T\right)\zeta_t\\
&+\int_{t-\tau_d/\lambda}^{t} f^T(s)R_1f(s)\,ds\\
&+\left(\int_{t-\tau_d/\lambda}^{t} g(s)\,d\omega(t)\right)^T R_2\left(\int_{t-\tau_d/\lambda}^{t} g(s)\,d\omega(t)\right),\\
&2\zeta_t^T N_2\left(-\int_{t-\tau_1}^{t} f(s)\,ds-\int_{t-\tau_1}^{t} g(s)\,d\omega(s)\right)\\
&\le \zeta_t^T\left(\tau_1N_2R_3^{-1}N_2^T+N_2R_4^{-1}N_2^T\right)\zeta_t\\
&+\int_{t-\tau_1}^{t} f^T(s)R_3f(s)\,ds+\left(\int_{t-\tau_1}^{t} g(s)\,d\omega(t)\right)^T\\
&\cdot R_4\left(\int_{t-\tau_1}^{t} g(s)\,d\omega(t)\right),\\
&2\zeta_t^T N_3\left(-\int_{t-\tau(t)}^{t-\tau_1} f(s)\,ds-\int_{t-\tau(t)}^{t-\tau_1} g(s)\,d\omega(s)\right)\\
&\le \zeta_t^T\left((\tau(t)-\tau_1)N_3R_5^{-1}N_3^T+N_3R_6^{-1}N_3^T\right)\zeta_t\\
&+\int_{t-\tau(t)}^{t-\tau_1} f^T(s)R_5f(s)\,ds+\left(\int_{t-\tau(t)}^{t-\tau_1} g(s)\,d\omega(t)\right)^T\\
&\cdot R_6\left(\int_{t-\tau(t)}^{t-\tau_1} g(s)\,d\omega(t)\right),\\
&2\zeta_t^T N_4\left(-\int_{t-\tau_2}^{t-\tau(t)} f(s)\,ds-\int_{t-\tau_2}^{t-\tau(t)} g(s)\,d\omega(s)\right)\\
&\le \zeta_t^T\left((\tau_2-\tau(t))N_4R_5^{-1}N_4^T+N_4R_6^{-1}N_4^T\right)\zeta_t\\
&+\int_{t-\tau(t)}^{t-\tau_1} f^T(s)R_5f(s)\,ds+\left(\int_{t-\tau_2}^{t-\tau(t)} g(s)\,d\omega(t)\right)^T\\
&\cdot R_6\left(\int_{t-\tau_2}^{t-\tau(t)} g(s)\,d\omega(t)\right),
\end{aligned} \tag{29}$$

as

$$\begin{aligned}
&E\left\{\left(\int_{t-\tau_d/\lambda}^{t} g(s)\,d\omega(t)\right)^T R_2\left(\int_{t-\tau_d/\lambda}^{t} g(s)\,d\omega(t)\right)\right\}\\
&\quad=E\left\{\int_{t-\tau_d/\lambda}^{t} g^T(s)R_2g(s)\,ds\right\},\\
&E\left\{\left(\int_{t-\tau_1}^{t} g(s)\,d\omega(t)\right)^T R_4\left(\int_{t-\tau_1}^{t} g(s)\,d\omega(t)\right)\right\}\\
&\quad=E\left\{\int_{t-\tau_1}^{t} g^T(s)R_4g(s)\,ds\right\},\\
&E\left\{\left(\int_{t-\tau(t)}^{t-\tau_1} g(s)\,d\omega(t)\right)^T R_6\left(\int_{t-\tau(t)}^{t-\tau_1} g(s)\,d\omega(t)\right)\right\}\\
&\quad=E\left\{\int_{t-\tau(t)}^{t-\tau_1} g^T(s)R_6g(s)\,ds\right\},\\
&E\left\{\left(\int_{t-\tau_2}^{t-\tau(t)} g(s)\,d\omega(t)\right)^T R_6\left(\int_{t-\tau_2}^{t-\tau(t)} g(s)\,d\omega(t)\right)\right\}\\
&\quad=E\left\{\int_{t-\tau_2}^{t-\tau(t)} g^T(s)R_6g(s)\,ds\right\}.
\end{aligned} \tag{30}$$

Combining (26)–(30) yields

$$\begin{aligned}
\dot V(x(t),t)\le\zeta_t^T\Bigg\{&2E_1P\overline{f}_c+g_c^TPg_c+W_{\varsigma_1}^TQW_{\varsigma_1}\\
&-W_{\varsigma_2}^TQW_{\varsigma_2}+\frac{\tau_d}{\lambda}g_c^TR_2g_c+\frac{\tau_d}{\lambda}\overline{f}_c^TR_1\overline{f}_c\\
&+\frac{\tau_d}{\lambda}\sum_{i=1}^{m}\beta_i^2\widehat{f}_i^TR_1\widehat{f}_i+E_1S_1E_1^T-E_{\lambda+2}S_1E_{\lambda+2}^T\\
&+\tau_1g_c^TR_4g_c+\tau_1\overline{f}_c^TR_3\overline{f}_c+\tau_1\sum_{i=1}^{m}\beta_i^2\widehat{f}_i^TR_3\widehat{f}_i\\
&+E_{\lambda+2}S_2E_{\lambda+2}^T-E_{\lambda+4}S_2E_{\lambda+4}^T+\tau_{12}g_c^TR_6g_c\\
&+\tau_{12}\overline{f}_c^TR_5\overline{f}_c+\tau_{12}\sum_{i=1}^{m}\beta_i^2\widehat{f}_i^TR_5\widehat{f}_i\\
&+\delta\left(E_{\lambda+3}+E_{\lambda+5}\right)\Omega_1\left(E_{\lambda+3}+E_{\lambda+5}\right)^T\\
&-E_{\lambda+5}\Omega_2E_{\lambda+5}^T+2N_1\left(E_1^T-E_2^T\right)\\
&+2N_2\left(E_1^T-E_{\lambda+2}^T\right)+2N_3\left(E_{\lambda+2}^T-E_{\lambda+3}^T\right)\\
&+2N_4\left(E_{\lambda+3}^T-E_{\lambda+4}^T\right)+\frac{\tau_d}{\lambda}N_1R_1^{-1}N_1^T+N_1R_2^{-1}N_1^T\\
&+\tau_1N_2R_3^{-1}N_2^T+N_2R_4^{-1}N_2^T\\
&+(\tau(t)-\tau_1)N_3R_5^{-1}N_3^T+N_3R_6^{-1}N_3^T\\
&+(\tau_2-\tau(t))N_4R_5^{-1}N_4^T+N_4R_6^{-1}N_4^T\Bigg\}\zeta_t+e^T(t)
\end{aligned}$$

$$\cdot \Omega_2 e(t) - \delta (e(t) + x(t-\tau(t)))^T \Omega_1 (e(t) + x(t-\tau(t))) = \zeta_t^T \Xi(\tau(t)) \zeta_t + e^T(t) \Omega_2 e(t) - \delta (e(t) + x(t-\tau(t)))^T \Omega_1 (e(t) + x(t-\tau(t))), \tag{31}$$

where

$$\Xi(\tau(t)) = \Xi_{11} + \Xi_{12}(\tau(t)), \quad \tau(t) \in [\tau_1, \tau_2]. \tag{32}$$

Ξ_{11} is defined in Theorem 8, and

$$\begin{aligned}\Xi_{12}(\tau(t)) = {} & g_c^T P g_c + \frac{\tau_d}{\lambda} g_c^T R_2 g_c + \frac{\tau_d}{\lambda} \overline{f}_c^T R_1 \overline{f}_c \\ & + \frac{\tau_d}{\lambda} \sum_{i=1}^{m} \beta_i^2 \widehat{f}_i^T R_1 \widehat{f}_i + \tau_1 g_c^T R_4 g_c \\ & + \tau_1 \overline{f}_c^T R_3 \overline{f}_c + \tau_1 \sum_{i=1}^{m} \beta_i^2 \widehat{f}_i^T R_3 \widehat{f}_i \\ & + \tau_{12} g_c^T R_6 g_c + \tau_{12} \overline{f}_c^T R_5 \overline{f}_c \\ & + \tau_{12} \sum_{i=1}^{m} \beta_i^2 \widehat{f}_i^T R_5 \widehat{f}_i + \frac{\tau_d}{\lambda} N_1 R_1^{-1} N_1^T \\ & + N_1 R_2^{-1} N_1^T + \tau_1 N_2 R_3^{-1} N_2^T + N_2 R_4^{-1} N_2^T \\ & + (\tau(t) - \tau_1) N_3 R_5^{-1} N_3^T + N_3 R_6^{-1} N_3^T \\ & + (\tau_2 - \tau(t)) N_4 R_5^{-1} N_4^T + N_4 R_6^{-1} N_4^T. \end{aligned} \tag{33}$$

From (31)–(33), it is not difficult to see that $\zeta_t^T \Xi(\tau(t)) \zeta_t$ is a first-order function on $\tau(t)$, and the first-order coefficient is $N_3 R_5^{-1} N_3^T - N_4 R_5^{-1} N_4^T$. According to Lemma 6, the condition $\Xi(\tau(t)) < 0$ is equivalent to

$$\begin{aligned} \Xi_{11} + \Xi_{12}(\tau_1) < 0, \\ \Xi_{11} + \Xi_{12}(\tau_2) < 0. \end{aligned} \tag{34}$$

Consequently, further resorting to the Schur complement lemma [30], inequalities (17) can be obtained.

In the following, we will show that stochastic system (21) satisfies event-triggered algorithm (4). Integrate on both sides of inequality (31) from 0 to t, and according to $V(0) = 0$, $V(t) \geq 0$, we can learn that if inequalities (17) are feasible, then

$$\begin{aligned} & e^T(t) \Omega_2 e(t) \\ & \quad \geq \delta (e(t) + x(t-\tau(t)))^T \Omega_1 (e(t) + x(t-\tau(t))) \end{aligned} \tag{35}$$

holds for arbitrary $t \in [t_k h + \tau_{t_k}, t_{k+1} h + \tau_{t_{k+1}}]$.

This completes the proof. □

The sufficient condition that stochastic system (13) with time-delay satisfying condition (7) is stochastically asymptotically stable is given in Theorem 8 with coupled matrix variables, but the controller gain matrix cannot be obtained directly. So, based on Theorem 8, we next prove a new theorem for the event-triggered controller design in terms of linear matrix inequalities through congruent transformation.

Theorem 9. *For given scalars τ_d, τ_1, τ_2, and δ and a positive integer λ, the closed-loop stochastic system with time delays (13) and time-varying delay satisfying condition (7) is stochastically asymptotically stable if there exist $n \times n$ matrices $X = X^T > 0$, $\widetilde{\Omega}_i = \widetilde{\Omega}_i^T > 0$, $i = 1, 2$, $\widetilde{R}_i = \widetilde{R}_i^T > 0$, $i = 1, 2, \ldots, 6$, and $\widetilde{S}_i = \widetilde{S}_i^T > 0$, $i = 1, 2$, and $\lambda n \times \lambda n$ matrices $\widetilde{Q} = \widetilde{Q}^T > 0$ and $(\lambda + 5)n \times n$ matrices $\widetilde{N}_i$, $i = 1, 2, 3, 4$, such that the following set of inequalities hold:*

$$\begin{bmatrix} \widetilde{\Xi}_{11} & \widetilde{\Xi}_{12} & \widetilde{\Xi}_{13} & \widetilde{\Xi}_{14} & \widetilde{\Xi}_{15} & \widetilde{\Xi}_{16} & \widetilde{\Xi}_{17} & \widetilde{\Xi}_{18}(l) \\ * & \widetilde{\Xi}_{22} & 0_{4n\times 3n} & 0_{4n\times mn} & 0_{4n\times mn} & 0_{4n\times mn} & 0_{4n\times 6n} & 0_{4n\times n} \\ * & * & \widetilde{\Xi}_{33} & 0_{3n\times mn} & 0_{3n\times mn} & 0_{3n\times mn} & 0_{3n\times 6n} & 0_{3n\times n} \\ * & * & * & \widetilde{\Xi}_{44} & 0_{mn\times mn} & 0_{mn\times mn} & 0_{mn\times 6n} & 0_{mn\times n} \\ * & * & * & * & \widetilde{\Xi}_{55} & 0_{mn\times mn} & 0_{mn\times 6n} & 0_{mn\times n} \\ * & * & * & * & * & \widetilde{\Xi}_{66} & 0_{mn\times 6n} & 0_{mn\times n} \\ * & * & * & * & * & * & \widetilde{\Xi}_{77} & 0_{6n\times n} \\ * & * & * & * & * & * & * & \widetilde{\Xi}_{88} \end{bmatrix} < 0, \quad l = 1, 2, \tag{36}$$

where

$$\begin{aligned} \widetilde{\Xi}_{11} = {} & \mathrm{Sym}\{E_1 \widetilde{f}_c\} + W_{\zeta_1}^T \widetilde{Q} W_{\zeta_1} - W_{\zeta_2}^T \widetilde{Q} W_{\zeta_2} + E_1 \widetilde{S}_1 E_1^T \\ & - E_{\lambda+2} \widetilde{S}_1 E_{\lambda+2}^T + E_{\lambda+2} \widetilde{S}_2 E_{\lambda+2}^T - E_{\lambda+4} \widetilde{S}_2 E_{\lambda+4}^T \\ & + \delta (E_{\lambda+3} + E_{\lambda+5}) \widetilde{\Omega}_1 (E_{\lambda+3}^T + E_{\lambda+5}^T) - E_{\lambda+5} \widetilde{\Omega}_2 E_{\lambda+5}^T \\ & + \mathrm{Sym}\{\widetilde{N}_1 (E_1^T - E_2^T) + \widetilde{N}_2 (E_1^T - E_{\lambda+2}^T) \\ & + \widetilde{N}_3 (E_{\lambda+2}^T - E_{\lambda+3}^T) + \widetilde{N}_4 (E_{\lambda+3}^T - E_{\lambda+4}^T)\}, \\ \widetilde{\Xi}_{12} = {} & \left[\widetilde{g}_c^T, \sqrt{\frac{\tau_d}{\lambda}} \widetilde{g}_c^T, \sqrt{\tau_1} \widetilde{g}_c^T, \sqrt{\tau_{12}} \widetilde{g}_c^T \right], \\ \widetilde{\Xi}_{22} = {} & \mathrm{diag}\{-X, -2X + \widetilde{R}_2, -2X + \widetilde{R}_4, -2X + \widetilde{R}_6\}, \\ \widetilde{\Xi}_{13} = {} & \left[\sqrt{\frac{\tau_d}{\lambda}} \widetilde{f}_c^T, \sqrt{\tau_1} \widetilde{f}_c^T, \sqrt{\tau_{12}} \widetilde{f}_c^T \right], \\ \widetilde{\Xi}_{33} = {} & \mathrm{diag}\{-2X + \widetilde{R}_1, -2X + \widetilde{R}_3, -2X + \widetilde{R}_5\}, \\ \widetilde{\Xi}_{14} = {} & \sqrt{\frac{\tau_d}{\lambda}} \left[\beta_1 \check{f}_1^T, \beta_2 \check{f}_2^T, \ldots, \beta_m \check{f}_m^T \right], \\ \widetilde{\Xi}_{44} = {} & \mathrm{diag}\{-2X + \widetilde{R}_1, -2X + \widetilde{R}_1, \ldots, -2X + \widetilde{R}_1\}, \\ \widetilde{\Xi}_{15} = {} & \sqrt{\tau_1} \left[\beta_1 \check{f}_1^T, \beta_2 \check{f}_2^T, \ldots, \beta_m \check{f}_m^T \right], \\ \widetilde{\Xi}_{55} = {} & \mathrm{diag}\{-2X + \widetilde{R}_3, -2X + \widetilde{R}_3, \ldots, -2X + \widetilde{R}_3\}, \\ \widetilde{\Xi}_{16} = {} & \sqrt{\tau_{12}} \left[\beta_1 \check{f}_1^T, \beta_2 \check{f}_2^T, \ldots, \beta_m \check{f}_m^T \right], \end{aligned}$$

$$\begin{aligned}
\widetilde{\Xi}_{66} &= \operatorname{diag}\left\{-2X + \widetilde{R}_5, -2X + \widetilde{R}_5, \ldots, -2X + \widetilde{R}_5\right\},\\
\widetilde{\Xi}_{17} &= \left[\sqrt{\frac{\tau_d}{\lambda}}\widetilde{N}_1, \widetilde{N}1, \sqrt{\tau_1}\widetilde{N}_2, \widetilde{N}_2, \widetilde{N}_3, \widetilde{N}_4\right],\\
\widetilde{\Xi}_{77} &= \operatorname{diag}\left\{-\widetilde{R}_1, -\widetilde{R}_2, -\widetilde{R}_3, -\widetilde{R}_4, -\widetilde{R}_6, -\widetilde{R}_6\right\},\\
\widetilde{\Xi}_{18}(1) &= \sqrt{\tau_{12}}\widetilde{N}_4,\\
\widetilde{\Xi}_{18}(2) &= \sqrt{\tau_{12}}\widetilde{N}_3,\\
\widetilde{\Xi}_{88} &= -\widetilde{R}_5,\\
W_{\zeta_1} &= \left[I_{\lambda n}, 0_{\lambda n\times 5n}\right],\\
W_{\zeta_2} &= \left[0_{\lambda n\times n}, I_{\lambda n}, 0_{\lambda n\times 4n}\right],\\
\widetilde{f}_c &= AXE_1^T + A_dXE_{\lambda+1}^T + B\overline{\Lambda}\widehat{K}E_{\lambda+3}^T + B\overline{\Lambda}\widehat{K}E_{\lambda+5}^T,\\
\breve{f}_i &= Be_i\widehat{K}E_{\lambda+3}^T + Be_i\widehat{K}E_{\lambda+5}^T,\\
\widetilde{g}_c &= HXE_1^T + H_dXE_{\lambda+1}^T
\end{aligned} \tag{37}$$

and a desired controller gain matrix in (2) can be chosen as

$$K = \widehat{K}X^{-1}. \tag{38}$$

Proof. Define

$$J = \operatorname{diag}\left\{J_1, J_2, J_3, J_4, J_4, J_4, J_5, P^{-1}\right\}, \tag{39}$$

where

$$\begin{aligned}
J_1 &= \operatorname{diag}\left\{P^{-1}, P^{-1}, \ldots, P^{-1}\right\} \in \mathbb{R}^{(\lambda+5)n},\\
J_2 &= \operatorname{diag}\{I, I, I, I\},\\
J_3 &= \operatorname{diag}\{I, I, I\},\\
J_4 &= \operatorname{diag}\{I, I, \ldots, I\} \in \mathbb{R}^{mn},\\
J_5 &= \operatorname{diag}\left\{P^{-1}, P^{-1}, P^{-1}, P^{-1}, P^{-1}, P^{-1}\right\}.
\end{aligned} \tag{40}$$

Pre- and postmultiplying (17) with J^T and J, noticing that

$$R_i^{-1} - 2X + XR_iX = \left(R_i^{-1} - X\right)R_i\left(R_i^{-1} - X\right) \geq 0, \quad i = 1, 2, \ldots, 6 \tag{41}$$

is equivalent to

$$-R_i^{-1} \leq -2X + XR_iX, \quad i = 1, 2, \ldots, 6, \tag{42}$$

we introduce the following new matrix variables:

$$\begin{aligned}
X &= P^{-1};\\
\widetilde{R}_i &= XR_i^TX, \quad i = 1, 2, \ldots, 6;\\
\widetilde{S}_i &= XS_i^TX, \quad i = 1, 2,\\
\widetilde{Q} &= J_6^TQJ_6, \quad J_6 = \operatorname{diag}\{X, X, \ldots, X\} \in \mathbb{R}^{\lambda n};\\
\widetilde{\Omega}_i &= X\Omega_i^TX, \quad i = 1, 2,\\
\widehat{K} &= KX;\\
\widetilde{N}_i &= J_1N_iX, \quad i = 1, 2, 3, 4,\\
\widetilde{f}_c &= AXE_1^T + A_dXE_{\lambda+1}^T + B\overline{\Lambda}\widehat{K}E_{\lambda+3}^T + B\overline{\Lambda}\widehat{K}E_{\lambda+5}^T,\\
\breve{f}_i &= Be_i\widehat{K}E_{\lambda+3}^T + Be_i\widehat{K}E_{\lambda+5}^T,\\
\widetilde{g}_c &= HXE_1^T + H_dXE_{\lambda+1}^T,
\end{aligned} \tag{43}$$

and then, we can learn that (36) is equivalent to (17).

This completes the proof. □

Remark 10. It is worth mentioning that Theorem 9 is also applicable to the event-triggered faults tolerant control for system (1) without state-delay and stochastic perturbations. The proof follows a similar line to Theorems 8 and 9. Now we formally present this result as the following corollary.

Corollary 11. *For given scalars τ_1 and τ_2 and $0 \leq \delta < 1$, the closed-loop system with time delays (13) and time-varying delay satisfying condition (7) is asymptotically stable if there exist $n\times n$ matrices $X = X^T > 0$, $\widetilde{\Omega}_i = \widetilde{\Omega}_i^T > 0$, $i = 1, 2$, $\widetilde{R}_i = \widetilde{R}_i^T > 0$, $i = 3, 5$, and $\widetilde{S}_i = \widetilde{S}_i^T > 0$, $i = 1, 2$, and $(\lambda+5)n\times n$ matrices $\widetilde{N}_i$, $i = 1, 2, 3$, such that the following set of inequalities hold:*

$$\begin{bmatrix}
\widetilde{\Xi}_{11} & \widetilde{\Xi}_{12} & \widetilde{\Xi}_{13} & \widetilde{\Xi}_{14} & \widetilde{\Xi}_{15} & \widetilde{\Xi}_{16}(l)\\
* & \widetilde{\Xi}_{22} & 0_{2n\times 2n} & 0_{2n\times n} & 0_{2n\times n} & 0_{2n\times n}\\
* & * & \widetilde{\Xi}_{33} & 0_{2n\times n} & 0_{2n\times n} & 0_{2n\times n}\\
* & * & * & \widetilde{\Xi}_{44} & 0_{2n\times n} & 0_{2n\times n}\\
* & * & * & * & \widetilde{\Xi}_{55} & 0_{n\times n}\\
* & * & * & * & * & \widetilde{\Xi}_{66}
\end{bmatrix} < 0, \quad l = 1, 2, \tag{44}$$

where

$$\begin{aligned}
\widetilde{\Xi}_{11} &= \operatorname{Sym}\left\{E_1\widetilde{A}_c\right\} + E_1\widetilde{S}_1E_1^T - E_2\widetilde{S}_1E_2^T + E_2\widetilde{S}_2E_2^T\\
&\quad - E_4\widetilde{S}_2E_4^T + \delta\left(E_3 + E_5\right)\widetilde{\Omega}_1\left(E_3^T + E_5^T\right) - E_5\widetilde{\Omega}_2E_5^T\\
&\quad + \operatorname{Sym}\left\{\widetilde{N}_1\left(E_1^T - E_2^T\right) + \widetilde{N}_2\left(E_2^T - E_3^T\right)\right.\\
&\quad \left. + \widetilde{N}_3\left(E_3^T - E_4^T\right)\right\},
\end{aligned}$$

$$\begin{aligned}
\widetilde{\Xi}_{12} &= \left[\sqrt{\tau_1}\widetilde{A}_c^T, \sqrt{\tau_{12}}\widetilde{A}_c^T\right],\\
\widetilde{\Xi}_{22} &= \operatorname{diag}\left\{-2X + \widetilde{R}_3, -2X + \widetilde{R}_5\right\},\\
\widetilde{\Xi}_{13} &= \sqrt{\tau_1}\left[\beta_1\check{A}_{c1}^T, \beta_2\check{A}_{c2}^T, \ldots, \beta_m\check{A}_{cm}^T\right],\\
\widetilde{\Xi}_{33} &= \operatorname{diag}\left\{-2X + \widetilde{R}_3, -2X + \widetilde{R}_3, \ldots, -2X + \widetilde{R}_3\right\},\\
\widetilde{\Xi}_{14} &= \sqrt{\tau_{12}}\left[\beta_1\check{A}_{c1}^T, \beta_2\check{A}_{c2}^T, \ldots, \beta_m\check{A}_{cm}^T\right],\\
\widetilde{\Xi}_{44} &= \operatorname{diag}\left\{-2X + \widetilde{R}_5, -2X + \widetilde{R}_5, \ldots, -2X + \widetilde{R}_5\right\},\\
\widetilde{\Xi}_{15} &= \sqrt{\tau_1}\widetilde{N}_1,\\
\widetilde{\Xi}_{55} &= -\widetilde{R}_3,\\
\widetilde{\Xi}_{16}(1) &= \sqrt{\tau_{12}}\widetilde{N}_3,\\
\widetilde{\Xi}_{16}(2) &= \sqrt{\tau_{12}}\widetilde{N}_2,\\
\widetilde{\Xi}_{66} &= -\widetilde{R}_5,\\
\widetilde{A}_c &= AXE_1^T + B\overline{\Lambda}\widehat{K}E_3^T + B\overline{\Lambda}\widehat{K}E_5^T,\\
\check{A}_{ci} &= Be_i\widehat{K}E_3^T + Be_i\widehat{K}E_5^T
\end{aligned} \tag{45}$$

and a desired controller gain matrix in (2) can be chosen as

$$K = \widehat{K}X^{-1}. \tag{46}$$

4. A Numerical Example

In this section, a numerical example is used to illustrate the effectiveness and improvement of the proposed approach in this paper.

Consider a simplified model of an unstable batch reactor with parameters:

$$A = \begin{bmatrix} 1.38 & -0.2077 & 6.715 & -5.676\\ -0.5814 & -4.29 & 0 & 0.675\\ 1.067 & 4.273 & -6.654 & 5.893\\ 0.048 & 4.273 & 1.343 & -2.104 \end{bmatrix},$$
$$B = \begin{bmatrix} 0 & 0\\ 5.679 & 0\\ 1.136 & -3.146\\ 1.136 & 0 \end{bmatrix}. \tag{47}$$

The eigenvalues of A are 1.9910, 0.0635, −5.0566, and −8.6659; it is obvious that the unforced system (47) is unstable. The unstable system is controlled under a network which is described in Figure 1, and the event-triggered communication scheme (4) is introduced to optimize the performance of the NCSs. Considering the effect of the transmission delay, it is assumed that the lower bound of the network-induced delay $\tau_1 = 0.01$ s. The purpose of the design is to make this upper bound, τ_2, as large as possible and the sampled data transmitted as low as possible so that the system can work under actuator failures conditions.

Two cases of probability distributions are considered.

Case 1. Actuator failures Λ_i are Bernoulli distributed white sequences with parameters:

$$\begin{aligned}
E\{\Lambda_i = 1\} &= \overline{\Lambda}_i \geq 0,\\
E\{\Lambda_i = 0\} &= 1 - \overline{\Lambda}_i.
\end{aligned} \tag{48}$$

It is not difficult to verify that

$$\beta_i^2 = \overline{\Lambda}_i\left(1 - \overline{\Lambda}_i\right) \tag{49}$$

when $\delta = 0$ and $\Omega_1 \neq \Omega_2$. We conduct the simulation studies with different Λ_i. The maximal allowable delays τ_2 and fault tolerant controller parameter K from Corollary 11 and those in [23, 24] are listed in Table 1. From Table 1, it is clear that Corollary 11 provides the largest delay; that is, Corollary 11 is less conservative than all the other list results.

Case 2. Actuator failures Λ_i are Gaussian distributed sequences with parameters.

Case 2.1. We have the following:

$$\begin{aligned}
\overline{\Lambda}_1 &= 0.2,\\
\overline{\Lambda}_2 &= 0.3,\\
\beta_1 &= 0.5,\\
\beta_2 &= 0.6.
\end{aligned} \tag{50}$$

Case 2.2. We have the following:

$$\begin{aligned}
\overline{\Lambda}_1 &= 0.2,\\
\overline{\Lambda}_2 &= 0.3,\\
\beta_1 &= 0.2,\\
\beta_2 &= 0.3
\end{aligned} \tag{51}$$

when $\delta = 0$ and $\Omega_1 \neq \Omega_2$. Table 2 lists the maximal allowable delays τ_2 and fault tolerant controller K obtained from Corollary 11 and those in [17, 18]. From Table 2, it can be concluded that our results are less conservative than the results from [17, 18].

In the following, under actuators failures Λ_i in Case 2.2 with $\overline{\Lambda}_1 = 0.2$, $\overline{\Lambda}_2 = 0.3$, $\beta_1 = 0.2$, and $\beta_2 = 0.3$, we will demonstrate that the fault tolerant state-feedback controller also optimizes the performance of the NCSs, that is, considerable reduction of the network resource occupancy:

(1) When $\delta = 0.1$ and $\Omega_1 \neq \Omega_2$, applying Corollary 11, we can solve the maximum $\tau_2 = 0.23$ s, and the

Table 1: The maximal allowable delays τ_2 and fault tolerant controller K for different Λ_i in Case 1.

Case 1	Maximal allowable delays	$\overline{\Lambda}_1 = 0.4$ and $\overline{\Lambda}_2 = 0.5$
[17]	$\tau_2 = 0.11$	$K = \begin{bmatrix} 0.4961 & -0.1158 & 0.3942 & -0.3413 \\ 3.7924 & 0.7272 & 2.9007 & -1.4987 \end{bmatrix}$
[18]	$\tau_2 = 0.12$	$K = \begin{bmatrix} -0.7941 & -0.3913 & -0.6128 & 0.0998 \\ 1.7764 & 0.3258 & 1.2664 & -0.6207 \end{bmatrix}$
Corollary 11	$\tau_2 = 0.23$	$K = \begin{bmatrix} -1.5767 & -0.4505 & -1.1426 & 0.4351 \\ 1.7901 & 0.4449 & 1.3089 & -0.5460 \end{bmatrix}$
Case 1	Maximal allowable delays	$\overline{\Lambda}_1 = 0.2$ and $\overline{\Lambda}_2 = 0.3$
[17]	$\tau_2 = 0.08$	$K = \begin{bmatrix} 0.9457 & -0.2368 & 0.7506 & -0.6604 \\ 6.5698 & 1.2461 & 5.0072 & -2.5903 \end{bmatrix}$
[18]	$\tau_2 = 0.10$	$K = \begin{bmatrix} -1.9127 & -0.7303 & -1.4330 & 0.3972 \\ 2.8661 & 0.5948 & 2.0576 & -0.9473 \end{bmatrix}$
Corollary 11	$\tau_2 = 0.17$	$K = \begin{bmatrix} -3.2189 & -0.9113 & -2.3315 & 0.8898 \\ 3.4078 & 0.8480 & 2.4877 & -1.0380 \end{bmatrix}$
Case 1	Maximal allowable delays	$\overline{\Lambda}_1 = 0.1$ and $\overline{\Lambda}_2 = 0.2$
[17]	$\tau_2 =$ —	$K =$ —
[18]	$\tau_2 = 0.08$	$K = \begin{bmatrix} -2.9434 & -1.2942 & -2.2716 & 0.4601 \\ 4.8397 & 0.9686 & 2.4877 & -1.0380 \end{bmatrix}$
Corollary 11	$\tau_2 = 0.12$	$K = \begin{bmatrix} -5.2739 & -1.4717 & -3.8744 & 1.4775 \\ 5.3644 & 1.3453 & 3.9057 & -1.6241 \end{bmatrix}$

Table 2: The maximal allowable delays τ_2 and fault tolerant controller K for different Λ_i in Case 2.

Case 2.1	Maximal allowable delays	$\overline{\Lambda}_1 = 0.2$, $\overline{\Lambda}_2 = 0.3$, $\beta_1 = 0.5$, and $\beta_2 = 0.6$
[17]	$\tau_2 = 0.06$	$K = \begin{bmatrix} 0.9892 & -0.2256 & 0.8052 & -0.6907 \\ 6.5727 & 1.2294 & 5.0305 & -2.5708 \end{bmatrix}$
[18]	$\tau_2 = 0.09$	$K = \begin{bmatrix} -2.4224 & -0.7447 & -1.7829 & 0.6406 \\ 2.5588 & 0.6054 & 1.8586 & -0.7902 \end{bmatrix}$
Corollary 11	$\tau_2 = 0.13$	$K = \begin{bmatrix} -3.6116 & -0.9842 & -2.6450 & 1.0301 \\ 3.3758 & 0.8475 & 2.4587 & -1.0213 \end{bmatrix}$
Case 2.2	Maximal allowable delays	$\overline{\Lambda}_1 = 0.2$, $\overline{\Lambda}_2 = 0.3$, $\beta_1 = 0.2$, and $\beta_2 = 0.3$
[17]	$\tau_2 =$ —	$K =$ —
[18]	$\tau_2 = 0.13$	$K = \begin{bmatrix} -2.8578 & -0.9068 & -2.0906 & 0.7377 \\ 2.1139 & 0.4710 & 1.5436 & -0.6720 \end{bmatrix}$
Corollary 11	$\tau_2 = 0.23$	$K = \begin{bmatrix} -3.8191 & -1.1124 & -2.7404 & 1.0449 \\ 2.6036 & 0.6281 & 1.9122 & -0.8087 \end{bmatrix}$

corresponding controller K and the trigger parameter Ω_1 and Ω_2 are given by

$$K = \begin{bmatrix} -3.8157 & -1.1116 & -2.7398 & 1.0445 \\ 2.5989 & 0.6263 & 1.9078 & -0.8073 \end{bmatrix},$$

$$\Omega_1 = \begin{bmatrix} 0.0558 & 0.0173 & 0.0401 & -0.0150 \\ 0.0173 & 0.0086 & 0.0113 & -0.0042 \\ 0.0401 & 0.0113 & 0.0341 & -0.0157 \\ -0.0150 & -0.0042 & -0.0157 & 0.0087 \end{bmatrix}, \quad (52)$$

$$\Omega_2 = \begin{bmatrix} 171.9090 & 43.1360 & 125.6951 & -52.4983 \\ 43.1360 & 10.8542 & 31.5284 & -13.1503 \\ 125.6951 & 31.5284 & 91.9150 & -38.3995 \\ -52.4983 & -13.1503 & -38.3995 & 16.0563 \end{bmatrix}.$$

(2) When $\delta = 0.1$ and $\Omega_1 = \Omega_2$, applying Corollary 11, we can solve the maximum $\tau_2 = 0.15$ s, and the corresponding controller K and the trigger parameter Ω are given by

$$K = \begin{bmatrix} -3.9951 & -1.3160 & -2.7469 & 0.9501 \\ 3.0396 & 0.6144 & 2.2942 & -1.0344 \end{bmatrix}, \quad (53)$$

$$\Omega = \begin{bmatrix} 3.3177 & 0.8513 & 2.4087 & -0.9850 \\ 0.8513 & 0.2276 & 0.6132 & -0.2482 \\ 2.0487 & 0.6132 & 1.7600 & -0.7259 \\ -0.9850 & -0.2482 & -0.7259 & 0.3034 \end{bmatrix}. \quad (54)$$

Remark 12. From Tables 1 and 2, it can be concluded that when $\delta = 0$, that is, in the case of time-triggered communication scheme, our results are less conservative than existing ones. Moreover, when $\delta = 0.1$, the maximal allowable delay τ_2 obtained from Corollary 11 under condition $\Omega_1 \neq \Omega_2$ ($\tau_2 = 0.23$ s) is bigger than the case of $\Omega_1 = \Omega_2$ ($\tau_2 = 0.15$ s); that is, two different weighting matrices Ω_1 and Ω_2 are introduced in (4) which guaranteed less conservative results.

Select the sampling period $h = 0.10$ s. According to (7), we can see that the allowable maximum transmission delay can be 0.05 s. The state response of system (47) with controller K in (53) and Ω in (54) is shown in Figure 2. Figure 3 plots the release instants and release interval.

From Figures 2 and 3, we can see that system (47) remains asymptotically stable under the feedback controller K in (53) and trigger matrix Ω in (54); the simulation results for $t \in [0, 30]$ show that only 58 sampled signals need to be transmitted to the controller through the transmitter, which takes 19.33% of the whole sample signals. Thus, the fault tolerant state-feedback controller presented in this paper also optimizes the performance of the NCSs by reducing the networked load.

Remark 13. The simulation results show that the amount of transferred data is reduced by introducing the event-triggering communication scheme. In addition, the maximum

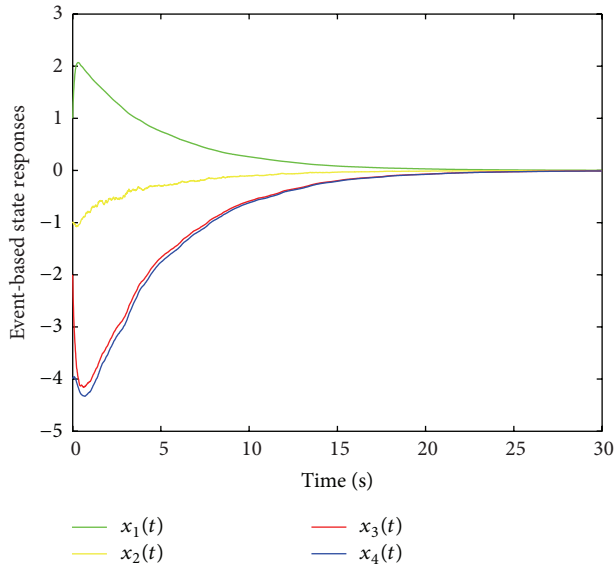


Figure 2: The state responses of system (47) under the feedback controller K in (53) and trigger matrix Ω in (54).

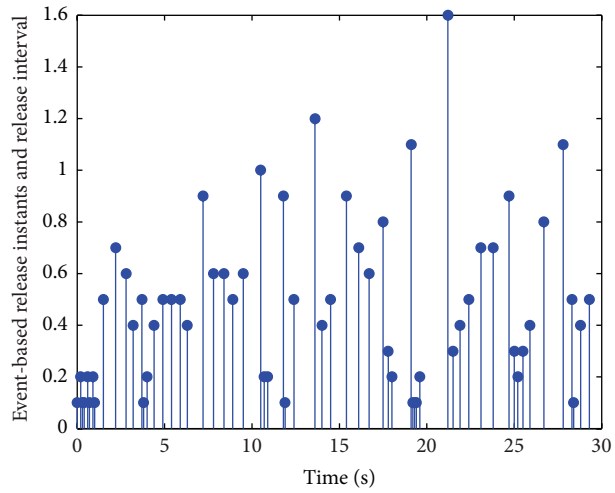


Figure 3: The release instants and release interval under the feedback controller K in (53) and trigger matrix Ω in (54).

value, τ_2, in this paper is bigger than the value, τ_2, in [17, 18]. Therefore, the proposed controller design method in the paper not only is less conservative but also has less transferred data than [17, 18].

5. Conclusion

The fault tolerant state-feedback controller design problem for stochastic NCSs has been investigated in the paper. Improved state-feedback controller design method with less conservation and less sampled data transmission has been proposed by employing event-triggered communication scheme, the delay-decomposition technique, and convex combination approach. A practical numerical example for the stabilization of batch reactor has been given to demonstrate

that the theory represented in this paper is less conservative than existing ones, and introducing event-triggered communication scheme can optimize the performance of stochastic NCSs. Therefore, the proposed controller design method in the paper is an improvement over existing ones.

Competing Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

This work was supported in part by the National Natural Science Funds for Young Scholar of China under Grant 61304046, Natural Science Foundation of Heilongjiang Province of China (Grant no. QC 2015064), and Scientific Research Foundation for the Returned Overseas Chinese Scholars, Heilongjiang Province of China (Grant no. LC 2016027).

References

[1] R.-A. Gupta and M.-Y. Chow, "Networked control system: overview and research trends," *IEEE Transactions on Industrial Electronics*, vol. 57, no. 7, pp. 2527–2535, 2010.

[2] L. Zhang, H. Gao, and O. Kaynak, "Network-induced constraints in networked control systems-a survey," *IEEE Transactions on Industrial Informatics*, vol. 9, no. 1, pp. 403–416, 2013.

[3] F. Li, P. Shi, L. Wu, M. V. Basin, and C.-C. Lim, "Quantized control design for cognitive radio networks modeled as nonlinear semi-Markovian jump systems," *IEEE Transactions on Industrial Electronics*, vol. 62, no. 4, pp. 2330–2340, 2015.

[4] Y. Shi and B. Yu, "Robust mixed H_2/H_∞ control of networked control systems with random time delays in both forward and backward communication links," *Automatica*, vol. 47, no. 4, pp. 754–760, 2011.

[5] R. Yang, P. Shi, and G.-P. Liu, "Filtering for discrete-time networked nonlinear systems with mixed random delays and packet dropouts," *IEEE Transactions on Automatic Control*, vol. 56, no. 11, pp. 2655–2660, 2011.

[6] E. Tian and D. Yue, "Reliable H_∞ filter design for T–S fuzzy model-based networked control systems with random sensor failure," *International Journal of Robust and Nonlinear Control*, vol. 23, no. 1, pp. 15–32, 2013.

[7] S. H. Kim, "$\mathscr{H}_\infty$ output-feedback tracking control for networked control systems," *Mathematical Problems in Engineering*, vol. 2015, Article ID 724389, 10 pages, 2015.

[8] J. Mazo and P. Tabuada, "Decentralized event-triggered control over wireless sensor/actuator networks," *IEEE Transactions on Automatic Control*, vol. 56, no. 10, pp. 2456–2461, 2011.

[9] Q. Liu, Z. Wang, X. He, and D. Zhou, "A survey of event-based strategies on control and estimation," *Systems Science and Control Engineering*, vol. 2, no. 1, pp. 90–97, 2014.

[10] C. Peng and T. C. Yang, "Event-triggered communication and H_∞ control co-design for networked control systems," *Automatica*, vol. 49, no. 5, pp. 1326–1332, 2013.

[11] H. Yan, S. Yan, H. Zhang, and H. Shi, "Event-triggered H_∞control for networked control systems with time-varying delay," *Mathematical Problems in Engineering*, vol. 2014, Article ID 874823, 7 pages, 2014.

[12] S. Hu, X. Yin, Y. Zhang, and E. G. Tian, "Event-triggered guaranteed cost control for uncertain discrete-time networked control systems with time-varying transmission delays," *IET Control Theory and Applications*, vol. 6, no. 18, pp. 2793–2804, 2012.

[13] X.-M. Zhang and Q.-L. Han, "Event-triggered dynamic output feedback control for networked control systems," *IET Control Theory & Applications*, vol. 8, no. 4, pp. 226–234, 2014.

[14] D. Zhang, Q.-L. Han, and X. Jia, "Network-based output tracking control for T-S fuzzy systems using an event-triggered communication scheme," *Fuzzy Sets and Systems*, vol. 273, pp. 26–48, 2015.

[15] H. Wang, P. Shi, and J. Zhang, "Event-triggered fuzzy filtering for a class of nonlinear networked control systems," *Signal Processing*, vol. 113, pp. 159–168, 2015.

[16] S. Hu, D. Yue, X. Xie, and Z. Du, "Event-triggered H_∞ stabilization for networked stochastic systems with multiplicative noise and network-induced delays," *Information Sciences*, vol. 299, pp. 178–197, 2015.

[17] C. Peng, T. C. Yang, and E. G. Tian, "Brief paper: robust fault-tolerant control of networked control systems with stochastic actuator failure," *IET Control Theory & Applications*, vol. 4, no. 12, pp. 3003–3011, 2010.

[18] S. Wang, J. Feng, and H. Zhang, "Robust fault tolerant control for a class of networked control systems with state delay and stochastic actuator failures," *International Journal of Adaptive Control and Signal Processing*, vol. 28, no. 9, pp. 798–811, 2014.

[19] Y. Zhai, R. Yan, H. Liu, and J. Liu, "Event-triggered reliable control in networked control systems with probabilistic actuator faults," *Mathematical Problems in Engineering*, vol. 2013, Article ID 131942, 9 pages, 2013.

[20] J. Liu, "Event-based reliable H_∞ control for networked control system with probabilistic actuator faults," *Chinese Journal of Electronics*, vol. 22, no. 4, pp. 843–848, 2013.

[21] J. Liu and D. Yue, "Event-triggering in networked systems with probabilistic sensor and actuator faults," *Information Sciences*, vol. 240, pp. 145–160, 2013.

[22] X. Su, P. Shi, L. Wu, and Y. Song, "Fault detection filtering for nonlinear switched stochastic systems," *IEEE Transactions on Automatic Control*, vol. 61, no. 5, pp. 1310–1315, 2015.

[23] N. Touijer and S. Kamoun, "Design of robust self-tuning control schemes for stochastic systems described by input-output mathematical models," *International Journal of Innovative Computing, Information and Control*, vol. 11, no. 3, pp. 1101–1112, 2015.

[24] X. Su, L. Wu, P. Shi, and C. L. Philip Chen, "Model approximation for fuzzy switched systems with stochastic perturbation," *IEEE Transactions on Fuzzy Systems*, vol. 23, no. 5, pp. 1458–1473, 2015.

[25] F. Li, L. Wu, P. Shi, and C.-C. Lim, "State estimation and sliding mode control for semi-Markovian jump systems with mismatched uncertainties," *Automatica*, vol. 51, pp. 385–393, 2015.

[26] F. Li, P. Shi, L. Wu, and X. Zhang, "Fuzzy-model-based D-stability and nonfragile control for discrete-time descriptor systems with multiple delays," *IEEE Transactions on Fuzzy Systems*, vol. 22, no. 4, pp. 1019–1025, 2014.

[27] B. Song, J. H. Park, Z.-G. Wu, and X. Li, "New results on delay-dependent stability analysis and stabilization for stochastic time-delay systems," *International Journal of Robust and Nonlinear Control*, vol. 24, no. 16, pp. 2546–2559, 2014.

[28] L. Huang, X. Xie, Y. Zhao, and Z. Hu, "Disturbance attenuation observer design for fuzzy time-delay system with fault," *Journal of Intelligent and Fuzzy Systems*, vol. 30, no. 2, pp. 811–819, 2016.

[29] D. Yue, E. Tian, Y. Zhang, and C. Peng, "Delay-distribution-dependent stability and stabilization of T–S fuzzy systems with probabilistic interval delay," *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, vol. 39, no. 2, pp. 503–516, 2009.

[30] S. Boyd, L. Ghaoui, E. Feron, and V. Balakrishnan, *Linear Matrix Inequalities in System and Control Theory*, Society for Industrial and Applied Mathematics Press (SIAM), Philadelphia, Pa, USA, 1994.

Hindawi Publishing Corporation
Scientific Programming
Volume 2016, Article ID 7340859, 11 pages
http://dx.doi.org/10.1155/2016/7340859

Research Article

Identifying and Analyzing Strong Components of an Industrial Network Based on Cycle Degree

Zhiying Zhang,[1] Xiaozhen Chen,[2] Wenwen Xiao,[1] and Guijie Qi[1]

[1]*School of Management, Shandong University, Jinan, Shandong 250100, China*
[2]*Business School, Shandong Normal University, Jinan, Shandong 250014, China*

Correspondence should be addressed to Xiaozhen Chen; cxz005566@sdnu.edu.cn

Received 13 March 2016; Accepted 17 April 2016

Academic Editor: Guo Chen

In the era of big data and cloud computing, data research focuses not only on describing the individual characteristics but also on depicting the relationships among individuals. Studying dependence and constraint relationships among industries has aroused significant interest in the academic field. From the network perspective, this paper tries to analyze industrial relational structures based on cycle degree. The cycle degree of a vertex, that is, the number of cycles through a vertex in an industrial network, can describe the roles of the vertices of strong components in industrial circulation. In most cases, different vertices in a strong component have different cycle degrees, and the one with a larger cycle degree plays more important roles. However, the concept of cycle degree does not involve the lengths of the cycles, which are also important for circulations. The more indirect the relationship between two industries is, the weaker it is. In order to analyze strong components thoroughly, this paper proposes the concept of circular centrality taking into consideration the influence by two factors: the lengths and the numbers of cycles through a vertex. Exemplification indicates that a profound analysis of strong components in an industrial network can reveal the features of an economy.

1. Introduction

With the rapid development of the Internet, Internet of Things, and cloud computing technology, data has the potential for an explosive growth. The big-data era, which depends on cloud computing and cloud storage, has arrived. Large scale, diversity, and fast processing speed are the major characteristics of big data. Currently, data research focuses not only on describing individual characteristics but also on depicting the relationships among individuals [1, 2]. Relational data has been explored in the research of the economic management sector; for example, Acemoglu et al. used a degree sequence of a relational indicator to study the influences of the relationships among different departments on the fluctuation of the total output [3]; Hidalgo et al. studied an industry (product) upgrading issue by studying internet relational stricter among industries (products) [4]; Zhao et al. studied the linkage structure effect of blue economy under the perspective of an industrial network [5]; McNerney et al. studied the structure of interindustry relationships using networks of money flows between industries in 45 national economies [6]. On the basis of an analysis of industrial relational structures using relational data, this paper proposes the concept of circular centrality to detect unknown and potentially useful circular relational data hidden behind Input-Output Tables and study the topological properties of industrial circulation relationships.

In a connected component of an undirected graph, all the vertices are reachable from each other, but this is not necessarily true for a digraph. When there is no path from one vertex to another in a digraph, it is not reachable from the vertex. If there are mutually reachable paths between two vertices, they must be in the same subdigraph, a strong component. A strong component of a digraph, also known as a strongly connected subdigraph, is a maximally induced subdigraph of the digraph. Because any two vertices in a strong component can reach each other, a strong component has features that are different from other components. These make strong components meaningful to a digraph, especially to the one with practical significance such as an industrial network.

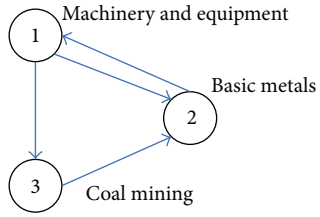


Figure 1: Industry circular linkages.

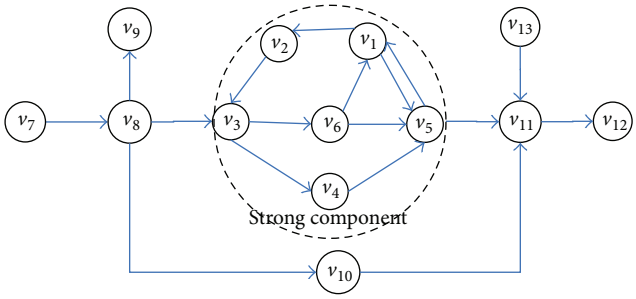


Figure 2: An industry network.

With vertices being the sectors and arcs representing significant industrial linkages, an economy can be abstracted into a network—an industrial network. The strong components of an industrial network depict industrial circulation linkages in an economy, where the former sector provides the input for the latter while the output of the latter comes back to the former. As shown in Figure 1, the sectors of coal mining, basic metals, and machinery and the linkages among them make up a strong component, and there are industrial circulation linkages among them. As fundamental linkages, industrial circulation linkages appear in various economic phenomena such as industrial clusters and circular economy [7].

In an industrial network, the general properties of strong components make them an important factor influencing the economy. Since any two sectors in a strong component can reach each other, any change of a sector can be felt by the others, and feedback effects also can influence the initial sector. The effects circulate with diminishing strength, as is shown in Figure 2.

Figure 2 describes an industrial network representing product flows. It contains 13 vertices, and 6 of them are in a strong component, forming vertices set V_c. If any sector of V_c undergoes some changes, such as shrink or expansion, the changes would pass on to all sectors of V_c and pass it on to some vertices outside the strong component. For instance, if sector v_4 alters, all sectors of V_c would change accordingly, and the effects circulate again and again over. As for the sectors outside, such as v_{11}, v_4 would push it along the directed path $v_4 \to v_5 \to v_{11}$. When the influence circulates within the strong component and reaches v_4 again, the effects on v_{11} would repeat again with lessened strength.

Any changes of a strong component can influence the circulation of the network. For instance, if v_1 becomes isolated in Figure 2, there would be no circulation, and any changes of a sector in the network would pass on along unidirectional paths and stop.

Consequently, strong components play significant roles in an economy. Campbell [8, 9] recognized the importance and proposed that a strong component could be regarded as a vertex to build the condensation digraph. Some scholars followed this method, such as Morillas et al. [10]. This approach would do a lot to clarify the relationship structure of the whole network but did not involve the internal structure of a strong component. In order to do further researches on strong components, Zhao et al. [11, 12] developed the concepts of cycle degree and cycle length distribution of sectors and testified their applicability to show the structure of strong components.

Suppose that N is an industrial network, V is the set of all sectors in N, and sector $v \in V$. An industry cycle through v is defined as a directed cycle containing vertex v, where no two vertices are the same. It is obvious that an industry cycle, the fundamental circular unit of an industrial network, is a closed path through v. Any sector in a strong component is in one or more industry cycles. An industry cycle contains two sectors at least (regardless of loops—arcs beginning and ending at the same vertex) and all sectors of an industrial network at most. In Figure 2, there are four industry cycles through v_1, that is, $v_1 \to v_5 \to v_1$, $v_1 \to v_2 \to v_3 \to v_6 \to v_1$, $v_1 \to v_2 \to v_3 \to v_4 \to v_5 \to v_1$, and $v_1 \to v_2 \to v_3 \to v_6 \to v_5 \to v_1$.

The cycle degree of a sector in an industrial network, denoted by d_c, means the number of industry cycles through the sector, describing the number of circular linkages between the sector and the others. As a sector with bigger d_c can influence more industry cycles, it has greater driving circulating power. In Figure 2, the cycle degrees of sectors v_1, v_2, v_3, v_4, v_5, and v_6 are 4, 3, 3, 1, 3, and 2, respectively. If v_1 becomes isolated, there would be no cycles in the network. It can be seen that the sectors with big cycle degrees have more influences on the strong component. Now that the sectors outside the strong component do not exist in any cycles, their cycle degrees are all 0.

Differing from the degree, the cycle degree of a sector depicts the structure information from the whole network. A vertex with small degree is not necessarily with small cycle degree. With in-degree and out-degree both being 1, maybe a sector is only adjacent to two sectors, but the three sectors are all in multiple cycles, so that the cycle degree of the sector is large still. In Figure 2, with the in-degree and out-degree being 1, $d_c(v_2)$ is 3, while $d_c(v_4)$ is 1.

The concept of cycle degree can depict some features of the strong component, but it is not enough. For a vertex, its impact on circulation is related to its cycle degree, but also to the steps in circulations, which is reflected in the lengths of cycles. The effects circulating in a shorter cycle are stronger than in a longer one. If an indicator can capture the two aspects, it is more accurate. For this reason, the concept of circular centrality is proposed in this paper. In addition, in order to better analyze the strong component of an industrial network, the cycle degree of a strong component is presented too.

In order to analyze the effectiveness of these concepts, some practical calculations are made. With more than 30

years of sustained rapid growth, China's economy has become one of the fastest growing economies in the world. This paper analyzes China's strong components and compares them to those of the US and Japan. The result shows the differences between China and the other two countries.

The organization of this paper is as follows. Section 2 describes the indicators to research strong components based on cycle degree. Section 3 briefly explains the methodology we employed in this study. The next part presents the main empirical results and analyses. Finally, some conclusions are offered.

2. Relevant Indicators to Analyze Strong Component Based on Cycle Degree

Strong components of an industrial network can be analyzed from a vertex or from the whole network. From a vertex perspective, as different vertices play different roles in circulations, it is important to analyze these differences. From the whole network, the strong components with different numbers of cycles create different circular effects. In addition, the distribution of sectors with large cycle degree can also show the characteristic of the economy.

2.1. Circular Centrality. The cycle degree of a vertex depicts the number of closed paths through the vertex in circulations. For a vertex, cycle degree can show its effect on circulation, but that is not wholly so. As is seen in Figure 1, there are two cycles through vertex ①, ① → ② → ① and ① → ③ → ② → ①, but they have difference. Vertex ① gets to itself through two steps along the first cycle, but three steps along the second one. The length of the path affects the linkage between the two sectors. As the path becomes longer, the linkage would be weaker. To a closed directed path, it is obvious that the cycle's length is a factor which can affect the roles of a vertex in circulation. If a vertex is contained in a special long cycle, the effect of the sector on itself would be slow and weak.

In order to express the effect of a sector on circulation, two factors should be taken into consideration, the cycle degree and lengths of cycles through the sector.

Freeman [13] proposed the concept of closeness centrality and suggested that the shorter the distance between a vertex and the others is, the higher closeness centrality it has. A vertex with big cycle degree and short cycles has a high circulation. Considering the two factors, we define the circular centrality of a vertex, which involves the concept of length distribution of cycles [11].

Length distribution of cycles depicts the numbers of cycles with different lengths through a sector. Regardless of loop, the length of the shortest cycle is 2. As the vertices in a cycle are different, the length of the longest cycle is no more than the number of vertices in the strong component, here assumed to be m. Supposing $d_c^k(v)$ denote the number of k cycles through sector v, we get length distribution of cycles $\begin{pmatrix} 2 & \cdots & k & \cdots & m \\ d_c^2(v) & & d_c^k(v) & & d_c^m(v) \end{pmatrix}$. In Figure 2, there are 4 cycles through vertex v_1, with two cycles of the length of 5 and the others of 2 and 4. So, the length distribution of industry cycles through v_1 is $\begin{pmatrix} 2 & 4 & 5 \\ 1 & 1 & 2 \end{pmatrix}$.

Length distribution of cycles involves the lengths of cycles and the corresponding cycle degrees at each length through a vertex. To a certain length, the ratio of cycle degree to its length shows the circular effect. The sum of all ratios depicts the role of the sector in circulation, which is defined as the circular centrality of a sector here. Denote the circular centrality of sector v by cc_v, and then

$$cc_v = \sum_{k=2}^{m} \frac{d_c^k(v)}{k}. \tag{1}$$

The percentage is applicable when the vertices within strong component are compared with each other. The coefficient of circular centrality of sector v, denoted by p_{cc_v}, is the percentage of circular centrality of sector v in those of all sectors in the strong component, and then

$$p_{cc_v} = \frac{cc_v}{\sum_{v=1}^{m} cc_v} \times 100\%. \tag{2}$$

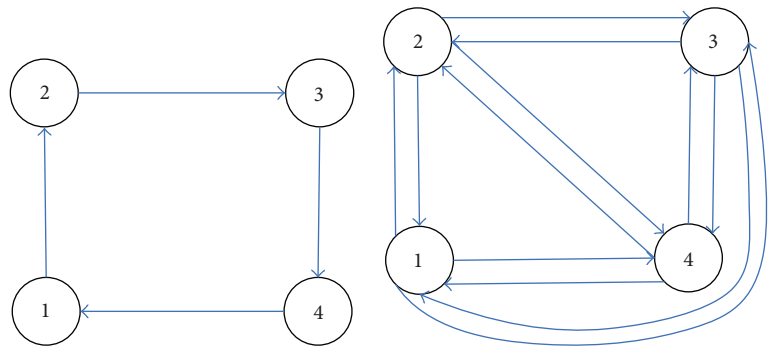
The circular centrality of a vertex is a concept based on cycle degree but involves the lengths of cycles. A vertex with big coefficient of circular centrality plays an important role in circulation.

2.2. The Cycle Degree of a Strong Component. Cycles are the fundamental units of a strong component. There must be some cycles in a strong component, at least one. Complete strong components contain the most cycles. The more cycles a strong component contains, the stronger the circularity is. The number of cycles in a strong component is related to the number of vertices and the features of arcs.

The cycle degree of a strong component, denoted by d_t, is the number of cycles it contains. As the number of vertices in a strong component is constant, it is highly correlated to the arcs there. In general, the cycle degree of a strong component is highly positively correlated to the number of arcs. Sometimes, it is also related to the directions of arcs, and so forth. In Figure 2, if the arc $v_1 \to v_2$ becomes $v_2 \to v_1$, there would remain only one cycle.

As a cycle including several vertices, the cycle degree of the strong component is not the sum of cycle degrees of all vertices there. If there are k vertices in a cycle, the cycle is counted k times, for it is considered in calculating the cycle degree of each vertex in the cycle. In order to calculate it, we can classify all cycles according to their lengths and sum up the cycle degree at each length. Suppose there are $d_c^k(v)$ k cycles through sector v and m sectors in a strong component, $v = 1, 2, \ldots, m$. As the number of vertices is that of arcs in a cycle, the number of k cycles in the strong component is $\sum_{v=1}^{m} d_c^k(v)/k$.

Regardless of loops, arcs beginning and ending at the same vertex, as an industry cycle contains two sectors at least, the length of the shortest cycle is 2. And that of the longest cycle is m, for m sectors in the strong component. So, the cycle



(a) A strong component (b) A complete strong component

Figure 3: Strong component and the corresponding complete subdigraph.

degree of a strong component, denoted by d_t, is the sum at each length. That is,

$$d_t = \sum_{k=2}^{m} \frac{\sum_{v=1}^{m} d_c^k(v)}{k}. \tag{3}$$

The cycle degree of a strong component depicts network connectivity of the strong component. The bigger it is, the stronger the connectivity is.

3. The Methodology to Identify Strong Components and Calculate the Cycle Degree of a Vertex

In order to study strong components, we need to identify them from an industrial network first. In addition, calculating the cycle degree of a vertex is the basis to study the strong component. Here, we employ the methodology proposed by Zhao et al. [11] to do these.

3.1. Identifying Strong Components of an Industrial Network. A strong component should be identified from its unique features. All sectors in a strong component can reach one another along directed paths. If arcs are drawn along all directed paths, a strong component would become a complete subdigraph, where there are arcs between any two sectors, as is shown in Figure 3. In the adjacent matrix attached to a complete subdigraph, the entries off the principal dialogue are all "1." According to this, we can distinguish complete subdigraphs from others and identify the strong components accordingly.

In order to identify the strong component of an industrial network, we draw arcs along all directed paths and get a new network, which is called the expansion network here. In the expansion network, the strong components of the primary network become complete strong components. Suppose that the adjacent matrixes of an industrial network and its expansion network are $\mathbf{U}$ and $\mathbf{U}^*$, respectively. In the industrial network, if there are p directed walks, length of k from i to j, the entries at the intersection of row i and column j of $\mathbf{U}^k$ would be p [14]. Suppose that an industrial network contains n vertices; then the length of the longest path (without cycles) of the industrial network will be not more than $n-1$. The sum of the power sequence of the $\mathbf{U}^k$ ($k = 1, \ldots, n-1$) matrices gives all walks no more than k steps. When arcs are drawn along all paths, the entries of the corresponding adjacent matrix will become 1 at all positive entries of the sum matrix. So we can get the adjacent matrix $\mathbf{U}^*$ from the Boolean summation (i.e., 1 + (#) 1 = 1) of the power sequence matrices; that is,

$$\mathbf{U}^* = \#\left(\mathbf{U} + \mathbf{U}^2 + \cdots + \mathbf{U}^{n-1}\right), \tag{4}$$

where # denotes Boolean summation.

From $\mathbf{U}^*$, we search for all the appropriate j and k for satisfying $u^*(k, j) = 1$ and $u^*(j, k) = 1$. All the vertexes j and k are in one complete subdigraph of the industry linkage expansion network; that is, they all belong to one strong component in the industrial network.

3.2. Calculating the Cycle Degree of a Sector. In a strong component, cycle degree of sector i means the number of all cycles through sector i, denoted by $d_c(i)$. In order to calculate, divide the cycles through sector i into groups according to their length and denote the cycle degree of length k by $d_c^{(k)}(i)$. It is obvious that the sum of cycle degrees of all lengths cycles is the cycle degree through sector i. Consider

$$d_c(i) = \sum_k d_c^{(k)}(i). \tag{5}$$

Ignoring the loops, the shortest cycle is 2 cycles (length of 2), and the longest is no more than the number of vertices in the strong component.

From the adjacency matrix of an industrial network, $\mathbf{U}$, we can get the liaisons of any two sectors. If $u_{ij} = 1$, there is a direct linkage from sector i to j; if $u_{iv_1} = u_{v_1v_2} = \cdots = u_{v_{k-1}j} = 1$ ($i, v_1, v_2, \ldots, j$ all differ from one another, so that there is a directed path length of k from sector i to j), there is an indirect linkage from sector i to j, as is shown in Figure 4; otherwise, there is no linkage. When sectors i and j are the same one, the directed path becomes a directed cycle. In other words, there is a k cycle through sector i.

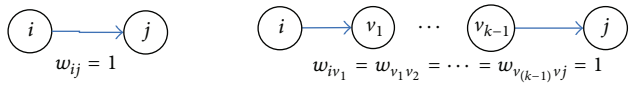


Figure 4: Directed path from sector i to sector j.

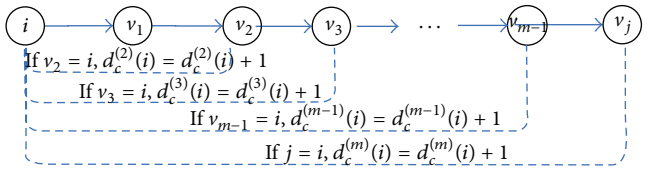


Figure 5: Process to calculate cycle degree through sector i.

In order to calculate $d_c^{(k)}(i)$, we search for the directed closed paths $i \to v_1 \to v_2 \to v_3 \to \cdots \to v_{k-2} \to v_{k-1} \to i$ ($i, v_1, v_2, \ldots$ all differ from one another) first and then count. The number of all these paths is $d_c^{(k)}(i)$, and the sum of all lengths of cycles is $d_c(i)$.

As vertices $i, v_1, v_2, \ldots, v_{k-1}$ are different from each other, the searching process is complex, especially to larger k. But an industrial network has its unique characteristics, which make the calculation possible. Firstly, the number of vertices in an industrial network is clear and small, and so is that of a strong component. Secondly, as industrial networks are sparse, if we only calculate unit entries and omit nil entries of the adjacent matrixes, the calculation will be reduced greatly. The algorithm is as follows.

Suppose that there are m sectors in a strong component, and $\mathbf{W}$ is the adjacent matrix attached to the strong component. Let $d_c^{(k)}(i)$ be equal to 0 for sector i first. As is shown in Figure 5, the steps are as follows.

Step 1. Search for the suitable v_1 for $w_{iv_1} = 1$.

Step 2. Search for the suitable v_2 for $w_{v_1v_2} = 1$. As no loop is in a strong component, it is certain that $v_2 \neq v_1$. If $v_2 = i$, there must be a 2-cycle; then $d_c^{(2)}(i) \leftarrow d_c^{(2)}(i) + 1$; else, go to next step.

Step 3. Search for the suitable v_3 for $w_{v_2v_3} = 1$, and $v_3 \neq v_1$. If $v_3 = i, d_c^{(3)}(i) \leftarrow d_c^{(3)}(i) + 1$; else, go to next step.

$\vdots$

Step $m-1$. Search for the suitable v_{m-1} for $w_{v_{m-2}v_{m-1}} = 1$, and $v_{m-1} \neq v_1, v_2, \ldots, v_{m-3}$. If $v_{m-1} = i, d_c^{(m-1)}(i) \leftarrow d_c^{(m-1)}(i) + 1$; else, go to next step.

Step m. Search for the suitable v_m for $w_{v_{m-1}v_m} = 1$, and $v_m \neq v_1, v_2, \ldots, v_{m-2}$. If $v_m = i, d_c^{(m)}(i) \leftarrow d_c^{(m)}(i) + 1$; else, go to next step.

Step $m+1$. Sum up $d_c^k(i)$ ($k = 2, \ldots, m$); then we get $d_c(i)$.

Applying the procedure to all vertices of a strong component, we obtain all cycle degrees.

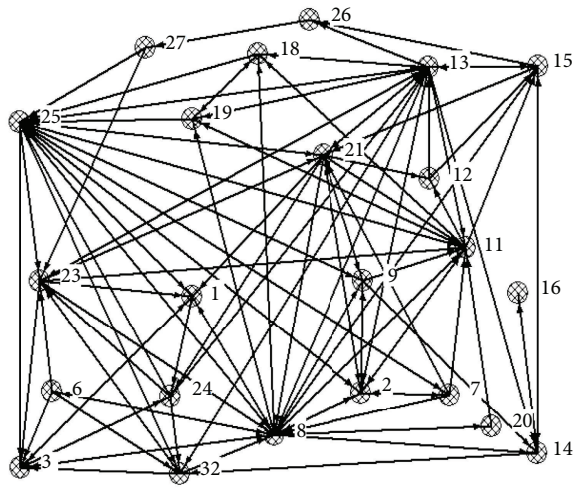


Figure 6: Strong component of China's industrial network.

4. Empirical Calculations and Analyses

In order to study the effectiveness of the concepts mentioned, some practical calculations are made. China's rapid development has intrigued a great deal of economists to try to understand the process [15–18]. So we try to analyze China's strong components and compare them to those of the US and Japan.

For comparison purposes, the data employed are derived from the OECD Input-Output Database, which is a part of the structural analysis (STAN) database. Data are used from Domestic Input-Output Tables 2005 of China, the US, and Japan (the 2005 Input-Output Tables were the last ones available). The tables all have 37 homogeneous sectors, as presented in Appendix C, making it possible to provide internationally comparable data for research. In order to find the fundamental relationships from the impacts on total production, we calculate the corrected influence coefficients of industrial linkage (Chen and Zhao [19], see Appendix A), on the basis of which industrial networks are constructed (Zhao et al., 2011 [20], see Appendix B).

4.1. Identifying the Strong Component of China's Industrial Network. Based on the methodology of Section 3, we obtain China's strong component. There is only one strong component, containing 23 sectors and 144 arcs, as is shown in Figure 6. The cycle degree of the strong component of China is 1263356.

4.2. Cycle Degrees of Sectors in China's Strong Component. The lengths distribution of cycles and cycle degrees of sectors in China are calculated, as is shown in Table 1. It can be seen that, to one sector, when the lengths of cycles are short, as the length is growing, the corresponding cycle degrees rise gradually. When the cycle degree reaches the maximum value, it starts to fall until reaching zero with the increase of the length.

To represent length distribution of industry cycles graphically, with abscissa being the lengths of cycles and ordinate

Table 1: The cycle degree and length contribution of the industry cycle in China.

SC	CL																				CD
	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	
1	2	4	25	107	380	1313	4026	10899	25554	51547	88434	127442	151844	146632	112795	67276	29412	8739	1572	108	828111
2	4	11	40	178	650	2052	5808	14740	32929	63913	106706	150591	176137	167981	127753	74793	31866	9209	1606	108	967075
3	2	6	24	99	351	1243	3906	10886	26291	54313	94775	138083	165730	160852	123523	72857	31344	9141	1596	108	895130
6	1	3	8	24	95	332	980	2530	5762	11366	19176	27400	32520	31793	25109	15151	6547	1931	306	12	181046
7	2	4	18	82	300	989	2884	7511	17328	34788	60037	87754	106841	107042	86807	54980	25545	8039	1522	108	602581
8	5	19	91	360	1239	3794	10271	24579	51524	93889	147017	194780	214946	194481	141537	80104	33293	9447	1622	108	1203106
9	0	1	11	51	197	689	2156	6122	15218	32672	60053	92626	117231	119530	96764	60793	28003	8639	1586	108	642450
11	2	14	64	251	883	2780	7880	19866	43791	83147	134135	181398	203212	186175	136908	78136	32705	9373	1622	108	1122450
12	0	3	12	47	170	560	1692	4590	10896	22583	41000	63044	79701	81707	67046	42191	18606	5413	1138	108	440507
13	7	26	112	412	1351	3976	10489	24674	51229	92888	145253	192679	213173	193435	140944	79708	33142	9435	1622	108	1194663
14	2	3	12	50	197	715	2353	6901	17620	38534	71472	110377	138890	139031	108388	64369	28007	8417	1570	108	737016
15	2	7	30	125	456	1501	4466	11827	27481	55934	98438	145565	176637	172832	133298	78096	33000	9431	1622	108	950856
16	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
18	1	3	26	113	398	1278	3625	9072	19997	38282	62701	86852	100837	97134	76051	47225	22776	7897	1568	108	575944
19	1	3	26	113	398	1278	3625	9072	19997	38282	62701	86852	100837	97134	76051	47225	22776	7897	1568	108	575944
20	0	0	2	12	58	191	529	1324	2894	5510	9071	12772	15411	15806	13578	9556	5267	2138	606	96	94821
21	2	16	66	252	892	2817	7928	19746	43046	81291	131103	177581	199263	182709	134497	76911	32279	9267	1610	108	1101384
23	1	9	45	179	639	2116	6286	16450	37559	73706	122878	171619	198136	186101	139081	79856	33349	9459	1622	108	1079199
24	0	8	37	136	495	1708	5228	14041	32515	64157	107018	149683	173631	164338	124043	72031	30474	8859	1594	108	950104
25	2	15	87	343	1181	3635	9938	24070	51183	94355	148777	197784	218558	197656	143437	80798	33425	9459	1622	108	1216433
26	0	0	2	14	75	277	913	2658	6953	16234	33021	56729	80801	94058	87068	60555	29345	9157	1622	108	479590
27	0	0	2	14	75	277	913	2658	6953	16234	33021	56729	80801	94058	87068	60555	29345	9157	1622	108	479590
32	1	7	32	133	482	1626	5000	13526	31830	64073	108425	152058	175267	164745	124286	72746	31288	9217	1622	108	956472

Note. SC: sector code; LC: cycles' length; CD: cycle degree.

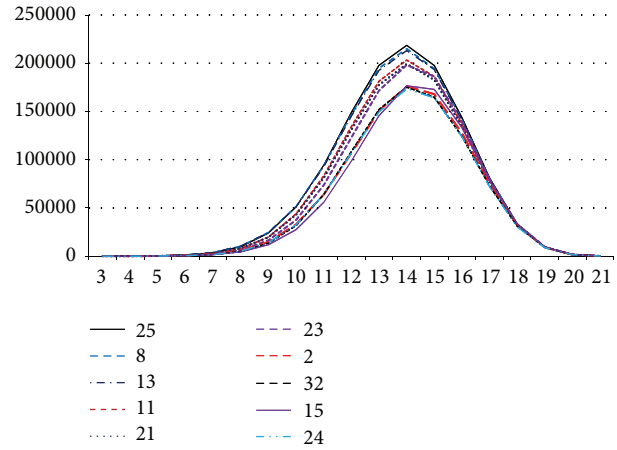


Figure 7: Length distribution of cycles of the top ten sectors.

being cycle degrees, we get the line charts of the top ten sectors, as is shown in Figure 7.

From Figure 7, it appears that the curves look like bells, increasing first and then decreasing. In the declining process, several curves overlap gradually. Given closer analysis it can be seen that as cycle becomes longer, there are more sectors in one cycle, so the cycle degrees of the top sectors increase simultaneously.

Dividing curves into halves in the middle parts, it can be seen that the differences of cycle degrees mainly come from the shorter cycles. Since shorter cycles circulate faster than the longer ones, the sectors with larger cycle degrees play more important roles and show greater competitiveness in an economy.

The cycle degree through a sector is an absolute value describing all closed paths through the sector. When we analyze how much a sector impacts the whole structure, relative cycle degree is needed. The relative cycle degree through sector i is the percentage of the cycle degree through sector i in that of the strong component, denoted by $p_{d_c(i)}$; then

$$p_{d_c(i)} = \frac{d_c(i)}{d_t} \times 100\%. \quad (6)$$

The relative cycle degrees of sectors in China are calculated, and the top ten ones are taken out, as is shown in Table 2.

4.3. Strong Component Comparisons among China, the US, and Japan. The cycle degree of the strong component of China is 1263356, while those of the US and Japan are 72712187 and 94706, respectively. It shows that, within an economy, the economic circulation of China is in the middle of the US and Japan. Relatively, Japan is the most dependent on foreign economy, and the US is the weakest.

For comparison, the coefficients of circular centrality and the relative cycle degrees of sectors are calculated and ordered in the three economies. Listing the top ten sectors, we get Table 2.

From Table 2, the top ten cycle degree sectors in the three economies basically share the same rankings with their coefficients of circular centrality. It is thus evident that the sectors with bigger cycle degrees have higher circular centrality generally.

Comparing the ten largest cycle degree sectors in the three economies, we can find that there exist some similarities but more differences in the three economies, as is shown in Figure 8. Overall, there are three common sectors, sectors 23, 25, and 32. They all belong to the service sector.

The distributions of the top ten sectors in the three economies are largely in accordance with their characteristics.

In China, the top ten sectors are distributed in three industries, one belonging to the resource industry, five to the manufacturing industry, and four to the service industry. Sector 2 (mining and quarrying) ranks in the top ten only in China. Of the five sectors belonging to the manufacturing industry, four are in the top five. It is evident that the manufacturing industry is the predominant part in China. The US shows its unique characteristics. Of the top ten sectors, nine belong to the service sector. Obviously, the service industry plays a dominant part in the US. Among them, sectors 28 (real estate activities), 31 (research and development), 36 (other community, social, and personal services), and 3 (food products, beverages, and tobacco) are in the top ten sectors only in the US, and sectors 28, 31, and 36 are even in the top five. Japan is somewhere between the US and China. The top ten sectors are both evenly distributed in the manufacturing and the service industry.

5. Conclusions and Further Research

Strong components are important components of a digraph for the circulations among vertices. Studying the strong components of an industrial network makes it easier to further identify the structural characteristics of the vertices and the whole network. To describe the strong components of an industrial network more accurately, this paper proposes the concept of circular centrality of a vertex and cycle degree of a strong component on the basis of cycle degree. As circular centrality of a vertex takes into consideration the influence by the lengths and the numbers of cycles through the vertex, it can better describe the real world. Using these indicators, the features of the strong components of China are analyzed and compared with those of the US and Japan.

We can analyze the strong components of an industrial network from two aspects: a vertex and strong component. From the viewpoint of a vertex, circular centrality is a concept involving the cycle degree and lengths of the cycles. Generally, a sector with a higher circular centrality has a stronger circulation and a larger influence on the economy. The study on the strong components of China and the other two economies shows that a vertex with a large cycle degree, with a few exceptions, has a high coefficient of circular centrality. From the viewpoint of a strong component, its cycle degree describes the number of circular paths within it. With a constant vertex number, a strong component with a larger cycle degree circulates stronger.

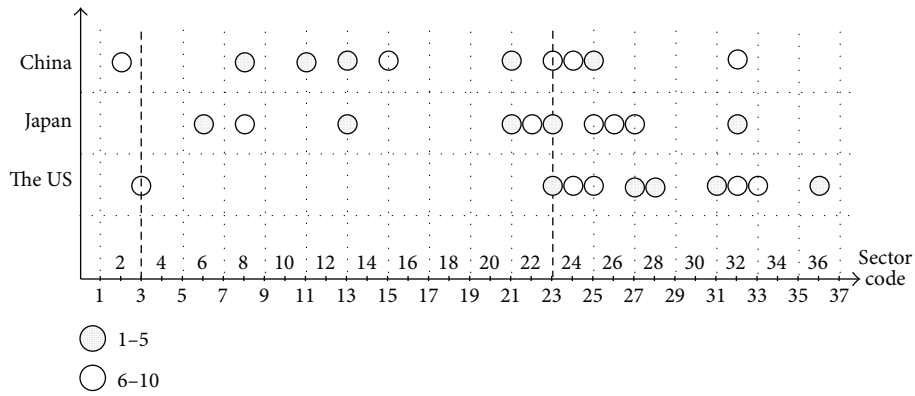


Figure 8: The top ten largest cycle degree sectors of the three economies.

Table 2: The top ten sectors of relative cycle degree and coefficients of circular centrality.

China					Japan					The US				
SC	RCD		CCC		SC	RCD		CCC		SC	RCD		CCC	
	(%)	RA	(%)	RA		(%)	RA	(%)	RA		(%)	RA	(%)	RA
25	96.3	1	7.20	1	23	98.3	1	8.0	1	23	23	1	6.24	1
8	95.2	2	7.13	2	21	89.7	2	7.17	2	36	36	2	6.23	2
13	94.6	3	7.08	3	32	89.2	3	7.12	3	31	31	3	6.17	3
11	88.8	4	6.60	4	13	88.5	4	7.04	4	28	28	4	6.01	4
21	87.2	5	6.47	5	6	80.1	5	6.41	5	27	27	5	5.85	5
23	85.4	6	6.28	6	27	78.4	6	6.20	6	33	33	6	5.79	6
2	76.5	7	5.61	7	25	76.8	7	6.09	7	25	25	7	5.1	7
32	75.7	8	5.55	8	22	76	8	5.99	8	32	32	8	4.97	8
15	75.3	9	5.45	10	26	70.8	9	5.52	9	53	53	9	4.53	9
24	75.2	10	5.51	9	8	67.8	10	5.3	10	24	24	10	4.32	10

Note. SC: sector code; RCD: relative cycle degree; RA: ranking; CCC: coefficients of circular centrality.

Although the concept of circular centrality of a vertex can describe strong components thoroughly, the calculation methods also have some limitations. Since the calculation process involves circulation, it may take a long time when there are many vertices and arcs in a strong component. As the intermediate inputs of industries are uneven (see Appendix B), the industrial networks are relatively sparse. This makes the method effective in most cases.

Generally, a sector with higher circular centrality has more competitiveness. In order to improve the competitiveness of an economy, we can try to change some linkages through policy guidance to enhance the circular level of the sector and the whole network. It is meaningful to find these linkages and take appropriate action to change them in future research. In addition, there are probably some vertices with more intimate relationships than others within a strong component. Finding these groups will be useful to further analyze the network.

Appendix

A. The Corrected Influence Coefficients of Industrial Linkage

From the perspective of the total output of an economy, corrected influence coefficients of industrial linkage delineate the effects of the linkages between two sectors on the total output with the unchanged final demand. The hypothetical extraction method (HEM) is a suitable method for quantitatively calculating the coefficients.

The original idea of the HEM is to extract a sector hypothetically from an economy and examine the impact of the extraction on the economy (Schultz [21]; Cella [22]; Clements [23]; Dietzenbacher and Van Der Linden [24]; Miller and Lahr [25]; Cai and Leung [26]). The differences of total output show effect of the linkages.

Similarly, to assess the linkage effect between two specific sectors, we assume that sector j would purchase import

goods to substitute for the inputs of sector i completely. Thus, by comparing the total outputs before and after the hypothetical extraction, the linkage between sectors i and j can be calculated. Based on the Leontief model, the industry linkage effect is studied from backward linkages.

Formally, denote the technical coefficient matrix of an economy with n sectors by $\mathbf{A}$, $\mathbf{A} = (a_{ij})_{n\times n}$, and the total output vector by $\mathbf{X}$, $\mathbf{X} = (x_i)_{n\times 1}$, the final total demand vector by $\mathbf{F}$, $\mathbf{F} = (f_i)_{n\times 1}$. The impact of a sector on its own is not considered.

In accordance with the Leontief model, the basic balance equation is

$$\mathbf{X} = \mathbf{AX} + \mathbf{F}. \tag{A.1}$$

According to HEM, it is assumed that the technical coefficient matrix $\mathbf{A}$ is partitioned into two groups: group one (g_1) being the two sectors to be studied (sectors i and j) and group two (g_2, $g_1 + g_2 = n$) being the remaining sectors. $\mathbf{X}$ and $\mathbf{F}$ are divided accordingly. Consider

$$\mathbf{A} = \begin{bmatrix} \mathbf{A_{11}} & \mathbf{A_{12}} \\ \mathbf{A_{21}} & \mathbf{A_{22}} \end{bmatrix},$$

$$\mathbf{X} = \begin{bmatrix} \mathbf{X_1} \\ \mathbf{X_2} \end{bmatrix}, \tag{A.2}$$

$$\mathbf{F} = \begin{bmatrix} \mathbf{F_1} \\ \mathbf{F_2} \end{bmatrix},$$

where

$$\mathbf{A_{11}} = \begin{bmatrix} a_{ii} & a_{ij} \\ a_{ji} & a_{jj} \end{bmatrix}_{2\times 2};$$

$$\mathbf{A_{12}} = \begin{bmatrix} a_{il} & \cdots & a_{ik} \\ a_{jl} & \cdots & a_{jk} \end{bmatrix}_{2\times(n-2)};$$

$$\mathbf{A_{21}} = \begin{bmatrix} a_{li} & a_{lj} \\ \cdots & \cdots \\ a_{ki} & a_{kj} \end{bmatrix}_{(n-2)\times 2}; \tag{A.3}$$

$$\mathbf{A_{22}} = \begin{bmatrix} a_{lm} & \cdots & a_{ln} \\ \cdots & \cdots & \cdots \\ a_{km} & \cdots & a_{kn} \end{bmatrix}_{(n-2)\times(n-2)}$$

with $l, k, m, n \neq i, j$.

Equation (A.1) can be rewritten as

$$\begin{bmatrix} \mathbf{X_1} \\ \mathbf{X_2} \end{bmatrix} = \begin{bmatrix} \mathbf{A_{11}} & \mathbf{A_{12}} \\ \mathbf{A_{21}} & \mathbf{A_{22}} \end{bmatrix} \times \begin{bmatrix} \mathbf{X_1} \\ \mathbf{X_2} \end{bmatrix} + \begin{bmatrix} \mathbf{F_1} \\ \mathbf{F_2} \end{bmatrix}. \tag{A.4}$$

Solve (A.4):

$$\begin{aligned} \mathbf{X_1} &= \mathbf{HA_{12}B_{22}F_2} + \mathbf{HF_1} \\ \mathbf{X_2} &= \mathbf{B_{22}F_2} + \mathbf{B_{22}A_{21}HA_{12}B_{22}F_2} + \mathbf{B_{22}A_{21}HF_1}, \end{aligned} \tag{A.5}$$

where $\mathbf{B_{22}} = [\mathbf{I} - \mathbf{A_{22}}]^{-1}$, $\mathbf{H} = [\mathbf{I} - \mathbf{A_{11}} - \mathbf{A_{12}B_{22}A_{21}}]^{-1}$, and $\mathbf{I}$ denotes the identity matrix.

Now, suppose that the input from sector i to sector j is extracted from the economy; that is, $a_{ij} = 0$. $\mathbf{A_{11}}$ becomes $\mathbf{A'_{11}}$, $\mathbf{A'_{11}} = \begin{bmatrix} a_{ii} & 0 \\ a_{ji} & a_{jj} \end{bmatrix}_{2\times 2}$, and $\mathbf{X}$ becomes $\mathbf{X'}$. Equation (A.4) can be expressed as

$$\begin{bmatrix} \mathbf{X'_1} \\ \mathbf{X'_2} \end{bmatrix} = \begin{bmatrix} \mathbf{A'_{11}} & \mathbf{A_{12}} \\ \mathbf{A_{21}} & \mathbf{A_{22}} \end{bmatrix} \times \begin{bmatrix} \mathbf{X'_1} \\ \mathbf{X'_2} \end{bmatrix} + \begin{bmatrix} \mathbf{F_1} \\ \mathbf{F_2} \end{bmatrix}. \tag{A.6}$$

Solve (A.6):

$$\begin{aligned} \mathbf{X'_1} &= \mathbf{H'A_{12}B_{22}F_2} + \mathbf{H'F_1} \\ \mathbf{X'_2} &= \mathbf{B_{22}F_2} + \mathbf{B_{22}A_{21}H'A_{12}B_{22}F_2} + \mathbf{B_{22}A_{21}H'F_1}, \end{aligned} \tag{A.7}$$

where $\mathbf{H'} = [\mathbf{I} - \mathbf{A'_{11}} - \mathbf{A_{12}B_{22}A_{21}}]^{-1}$.

The difference between (A.5) and (A.7), denoted by ILE_{ij}, means the industry linkage effect of sectors i and j. Consider

$$\begin{aligned} \mathrm{ILE}_{ij} = \mathbf{k}\left(\mathbf{X} - \mathbf{X'}\right) &= \left[\mathbf{k_1}\left(\mathbf{H} - \mathbf{H'}\right)\right. \\ &+ \left.\mathbf{k_2B_{22}A_{21}}\left(\mathbf{H} - \mathbf{H'}\right)\right]\mathbf{F_1} + \left[\mathbf{k_1}\left(\mathbf{H} - \mathbf{H'}\right)\mathbf{A_{12}B_{22}}\right. \\ &+ \left.\mathbf{k_2B_{22}A_{21}}\left(\mathbf{H} - \mathbf{H'}\right)\mathbf{A_{12}B_{22}}\right]\mathbf{F_2}, \end{aligned} \tag{A.8}$$

where $\mathbf{k}$, $\mathbf{k_1}$, and $\mathbf{k_2}$ are row summation vectors for $\mathbf{g}$, $\mathbf{g}_1$, and $\mathbf{g}_2$.

The corrected influence coefficient between sectors i and j, $ilcic_{ij}$, is the percentage of ILE_{ij} in the total output. Consider

$$ilcic_{ij} = \frac{\mathrm{ILE}_{ij}}{\mathbf{kX}} \times 100\%. \tag{A.9}$$

B. Constructing the Industrial Network

The industrial network is a fundamental model to analyze economic structures. With information in Input-Output Tables, the industrial network can be constructed with sectors being vertices, while arcs (directed edges) represent industrial linkages. The numeric values matrices, such as input-output quantitative transactions, or coefficients calculated from them, such as corrected influence coefficients, depict various relationships and determine the existence of arcs between sectors. In principle, with a defined threshold value, a numeric values matrix can be transferred into an adjacency matrix attached to a directed graph, with all positive entries larger than or equal to the filter to unity and the rest of the entries to zero.

It can be observed that the intermediate inputs of any sector are uneven. In the technical coefficient matrix $\mathbf{A}$ of any economy, only few a_{ij} values are large. For instance, in Chinese IO tables for the year 2005 (37×37), the 53 highest intermediate flows comprise 50% of the total intermediate consumption, while the remaining ones (1316) account for the other 50%. Nonuniform inputs make the industrial networks determined by several important relations.

Considering that the inputs have obvious inflection point, Zhao et al. [20] proposed the method to establish networks with Weaver-Thomas index.

Table 3: OECD sector classification.

Code	Sector
1	Agriculture, hunting, forestry, and fishing
2	Mining and quarrying
3	Food products, beverages, and tobacco
4	Textiles, textile products, leather, and footwear
5	Wood and products of wood and cork
6	Pulp, paper, and paper
7	Coke, refined petroleum products, and nuclear fuel
8	Chemicals and chemical products
9	Rubber and plastics products
10	Other nonmetallic mineral
11	Basic metals
12	Fabricated metal products except machinery and equipment
13	Machinery and equipment n.e.c
14	Office, accounting, and computing machinery
15	Electrical machinery and apparatus n.e.c
16	Radio, television, and communication equipment
17	Medical, precision, and optical instruments
18	Motor vehicles, trailers, and semitrailers
19	Other transport equipment
20	Manufacturing n.e.c; recycling
21	Electricity, gas, and water supply
22	Construction
23	Wholesale and retail trade; repairs
24	Hotels and restaurants
25	Transport and storage
26	Post and telecommunications
27	Finance and insurance
28	Real estate activities
29	Renting of machinery and equipment
30	Computer and related activities
31	Research and development
32	Other business activities
33	Public admin and defense; compulsory social security
34	Education
35	Health and social work
36	Other community, social, and personal services
37	Private households with employed persons

Source: This sector classification uses the latest version of the OECD I-O tables: 2005 Edition. http://stats.oecd.org/ to obtain the 2005 edition of OECD input-output tables for free.

The Weaver-Thomas index is an effective tool for finding the significant index developed by Weaver first and improved by O. Thomas later. Comparing an observed distribution with an assumed one, the closest approximation distribution is established to identify the key elements in numerical sequences. For the availability of finding crucial factors, it is widely used in regional economics.

Let $\mathbf{C}$ denote the coefficient matrix of an economy with n sectors (it can be technical coefficient matrix, or other matrixes); $C = (c_{ij})_{n\times n}$. According to the sequence from big to small order, rearrange $c_{11}, c_{12}, \ldots, c_{nn}$ and get the vector $\mathbf{C}^*$. Denote the lth element of $\mathbf{C}^*$ by c_{ij}; then the Weaver index is

$$w(l) = \sum_{k=1}^{n^2} \left[s(k,l) - 100 \times \frac{c_{ij}}{\sum_{j=1}^{n}\sum_{i=1}^{n} c_{ij}} \right]^2, \tag{B.1}$$

where

$$s(k,l) = \begin{cases} \dfrac{100}{l} & (k \le l) \\ 0 & (k > l). \end{cases} \tag{B.2}$$

Let $w(l^*) = \min\{w(1), w(2), \ldots, w(n \times n)\}$; the l^*th element of $\mathbf{C}^*$, c^*, is the threshold value. If $c_{ij} \ge c^*$, there would be an arc from sector i to sector j in the industrial network with the direction in keeping with the flow. Denote the adjacency matrix by $\mathbf{W}$; then

$$w_{ij} = \begin{cases} 1, & \text{if } c_{ij} \ge c^* \\ 0, & \text{if } c_{ij} < c^*. \end{cases} \tag{B.3}$$

C. OECD Sector Classification

See Table 3.

Competing Interests

The authors declare that they have no competing interests.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (no. 71371108).

References

[1] X. Su, P. Shi, L. Wu, and Y.-D. Song, "Fault detection filtering for nonlinear switched stochastic systems," *IEEE Transactions on Automatic Control*, vol. 61, no. 5, pp. 1310–1315, 2015.

[2] X. Su, L. Wu, P. Shi, and C. L. Chen, "Model approximation for fuzzy switched systems with stochastic perturbation," *IEEE Transactions on Fuzzy Systems*, vol. 23, no. 5, pp. 1458–1473, 2015.

[3] D. Acemoglu, V. Carvalho, A. Ozdaglar et al., "The network origins of aggregate fluctuations," *Econometrica*, vol. 80, no. 5, pp. 1977–2016, 2012.

[4] C. A. Hidalgo, B. Klinger, A.-L. Barabási, and R. Hausmann, "The product space conditions the development of nations," *Science*, vol. 317, no. 5837, pp. 482–487, 2007.

[5] B. Zhao, W. Xiao, R. Tong et al., "Blue economic connotation in industrial network perspective and its correlation structure effect—a case study of Shandong Province," *China Soft Science*, no. 07, pp. 135–147, 2015 (Chinese).

[6] J. McNerney, B. D. Fath, and G. Silverberg, "Network structure of inter-industry flows," *Physica A: Statistical Mechanics and its Applications*, vol. 392, no. 24, pp. 6427–6441, 2013.

[7] M. Fischer-Kowalski, "On the history of industrial metabolism," in *Perspectives on Industrial Ecology*, D. Bourg and S. Erkman, Eds., Greenleaf Publishing, Sheffield, UK, 2003.

[8] J. Campbell, *The relevance of input—output analysis and digraph concepts to growth pole theory [Ph.D. thesis]*, Department of Geography University of Washington, 1970.

[9] J. Campbell, "Growth pole theory, digraph analysis and interindustry relationships," *Tijdschrift voor Economische en Sociale Geografie*, vol. 63, no. 2, pp. 79–87, 1972.

[10] A. Morillas, L. Robles, and B. Diaz, "I-O coefficients importance: a fuzzy logic approach," *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, vol. 19, no. 6, pp. 1013–1031, 2011.

[11] B. Zhao, X. Chen, and J. Zhang, "Cycle degree of an industry and its algorithm," *System Engineering—Theory & Practice*, vol. 36, no. 6, pp. 1388–1397, 2014 (Chinese).

[12] B. Zhao and J. Zhang, *Introduction to Industrial Network Theory*, Economic Science Press, Beijing, China, 2013 (Chinese).

[13] L. C. Freeman, "Centrality in social networks conceptual clarification," *Social Networks*, vol. 1, no. 3, pp. 215–239, 1978.

[14] F. Harary, *Graph Theory*, Addison Wesley, London, UK, 1969.

[15] Y. F. Lin, F. Cai, and Z. Li, *China's Miracle: The Development Strategy and the Economic Reform*, Shanghai People's Publishing House, Shanghai, China, 1994 (Chinese).

[16] H. Fehr, S. Jokisch, and L. J. Kotlikoff, "Will China eat our lunch or take us out to dinner? Simulating the transition paths of the US, EU, Japan, and China," NBER Working Paper 11668, NBER, 2005.

[17] J. Ren, W. Li, Y. Wang, and L. Zhou, "Graph-mine: a key behavior path mining algorithm in complex software executing network," *International Journal of Innovative Computing, Information and Control*, vol. 11, no. 2, pp. 541–553, 2015.

[18] H. He, J. Wang, and J. Ren, "Measuring the importance of functions in software execution network based on complex network," *International Journal of Innovative Computing, Information and Control*, vol. 11, no. 2, pp. 719–731, 2015.

[19] X. Chen and B. Zhao, "Research on industry linkage correction influence coefficients based on hypothetical extraction method," *Management Review*, vol. 26, no. 6, pp. 23–32, 2014 (Chinese).

[20] B. Zhao, C. Yin, and J. Zhang, "A study on industry complex network and its modeling—on the example of Shandong province," *Economic Management*, no. 7, pp. 139–148, 2011 (Chinese).

[21] S. Schultz, "Approaches to identifying key sectors empirically by means of input–output analysis," *The Journal of Development Studies*, vol. 14, no. 1, pp. 77–96, 2007.

[22] G. Cella, "The input-output measurement of interindustry linkages," *Oxford Bulletin of Economics and Statistics*, vol. 46, no. 1, pp. 73–84, 1984.

[23] B. J. Clements, "On the decomposition and normalization of interindustry linkages," *Economics Letters*, vol. 33, no. 4, pp. 337–340, 1990.

[24] E. Dietzenbacher and J. A. Van Der Linden, "Sectoral and spatial linkages in the EC production structure," *Journal of Regional Science*, vol. 37, no. 2, pp. 235–257, 1997.

[25] R. E. Miller and M. L. Lahr, "A taxonomy of extractions," in *Regional Science Perspectives in Economic Analysis*, Elsevier, Amsterdam, The Netherlands, 2001.

[26] J. Cai and P. Leung, "Linkage measures: a revisit and a suggested alternative," *Economic Systems Research*, vol. 16, no. 1, pp. 65–86, 2004.

Hindawi Publishing Corporation
Scientific Programming
Volume 2016, Article ID 8038045, 10 pages
http://dx.doi.org/10.1155/2016/8038045

Research Article

Acquisition Pricing and Inventory Decisions on Dual-Source Spare-Part System with Final Production and Remanufacturing

Yancong Zhou,[1] Xudong Guo,[2] and Xiaochen Sun[3]

[1]*School of Information Engineering, Tianjin University of Commerce, Tianjin 300134, China*
[2]*Tourism and Historical Culture College, Zhaoqing University, Zhaoqing 526061, China*
[3]*Department of Mathematics, Tianjin University, Tianjin 300072, China*

Correspondence should be addressed to Yancong Zhou; zycong78@126.com

Received 27 March 2016; Accepted 10 May 2016

Academic Editor: Guo Chen

The life spans of durable goods are longer than their warranty periods. To satisfy the service demand of spare parts and keep the market competition advantage, enterprises have to maintain the longer inventory planning of spare parts. However, how to obtain a valid number of spare parts is difficult for those enterprises. In this paper, we consider a spare-part inventory problem, where the inventory can be replenished by two ways including the final production order and the remanufacturing way. Especially for the remanufacturing way, we consider the acquisition management problem of used products concerning an acquisition pricing decision. In a multiperiod setting, we formulate the problem into a dynamic optimization problem, where the system decisions include the final production order and acquisition price of used products at each period. By stochastic dynamic programming, we obtain the optimal policy of the acquisition pricing at each period and give the optimal policy structure of the optimization problem at the first period. Then, a recursion algorithm is designed to calculate the optimal decisions and the critical points in the policy. Finally, the numerical analyses show the effects of demand information and customer's sensitive degree on the related decisions and the optimal cost.

1. Introduction

With the improvement of market competition, customers have larger difficulty in distinguishing the products with similar functions. In addition to the price, they also have to consider more other factors. In those factors, after-sale service is the most important indicator, especially for durable products, such as automobile, computer, household appliances, and engineering equipment.

Current enterprises concern more the after-sale service within the regular warranty period. However, for durable products, the warranty period is much less than the product life span. For example, the warranty period of refrigerator is less than three years, but the product life is more than 10 years. And the probability of the product failure outside the warranty period is larger. When a failure occurs, although the product is outside the warranty period, the user still hopes the enterprise provides related services. However, if the enterprise cannot effectively help the user to solve the problem, it will have a negative impression of their products and directly affect the consumers' choice at the next time purchase option. For some personalized products, users are more dependent on the services provided by the manufacturer, ignoring the after-sale service outside the regular warranty period due to more serious results, which is showed by Nagler [1]: "you buy a car and eight years later cannot get it serviced for a reasonable price because the manufacturer has discontinued a particular part, you will remember it when you buy your next vehicle". Besides, for the necessity of keeping the competitive advantage of enterprises, some countries also have obliged enterprises to provide the service at outside the regular warranty period, for example, for 15 years in case of automobiles in Germany. Therefore, providing the extended after-sale service will be a trend practical issue, and the related research is important.

In current actual operations, the shortage of spare parts is the most difficult problem for providing after-sale service at outside the regular warranty period. This especially holds for durable products, because the production line has been closed. The shortage of spare parts will improve the service cost and brings a long waiting time, which results in a lower user's satisfaction and vanishes the action of providing after-sale service outside the regular warranty period. Because of no valid ways of replenishing inventory, the higher inventory or operation cost may be the major reason that enterprises do not want to provide the after-sale service outside the regular warranty period. Thus, how to replenish the inventory of spare part is the key problem for the extended after-sale service outside the regular warranty period.

The inventory replenishment problem of spare parts had became a concerned problem in the field of management and engineering sciences (Armstrong and Atkins [2], Bruggeman and van Dierdonck [3], Dhakar et al. [4], and Díaz and Fu [5]). The traditional replenishment mainly includes two ways: (i) producing enough spare parts before closing the regular production line and (ii) extra producing or ordering when the inventory is not enough. However, the first way has the high inventory cost and risk because of a very long service period; and the latter does not need to hold a high inventory, but it is discontinuous and has a high setup cost. The two ways are not enough to support the after-sale service outside the regular warranty period. Therefore, a kind of way with lower cost and continuous replenishment is necessary for the after-sale service outside the regular warranty period. Remanufacturing has been proved to be a good resource reusing way. For after-sale service, remanufacturing used products also can provide the necessary spare parts, and the related cost is lower and the replenishment process is also continuous. Fleischmann et al. [6] showed that product recovery can be as a major source of spare parts. Teunter and Fortuin [7] provided a case study, where spare parts can be replenished by recovering used products. However, remanufacturing also have shortcomings, such as higher uncertainty, and are constrained by the history of selling quantity of the product, so the traditional ways are still to be valid and necessary. Thus, how to integrate multiways to replenish the inventory of spare parts is important.

On the optimal acquisition problem of after-sale spare parts, Fortuin [8] firstly studied the size of the final production lot for the regular production. Teunter and Fortuin [9] extended the approach in Fortuin to a more complicated version. Cattani and Souza [10] considered the time of starting the final lot. van Kooten and Tan [11] provided a numerical method for determining the final order by Markov chain. The above papers only considered the final lot decision. Teunter and Klein Haneveld [12] consider a combination of final order and extra production order and gave an order-up-to policy. Inderfurth and Mukherjee [13] considered the combination of three ways of the spare parts' acquisition, that is, (i) the final lot of regular production, (ii) performing extra production runs until the end of service, and (iii) using remanufacturing to gain spare parts from used products. The decision is to find out the optimal combination of these three options. The model is modeled by decision tree and solved by a heuristic method based on stochastic dynamic programming. Pourakbar et al. [14] considered the determination problem of the optimal time to switch, where there are multioptions including the final order and repair policy, and the repairing component can be substituted by a new product. However, few papers consider the acquisition management problem of used products in a spare-part inventory system. The problem is important for ensuring the stability of the remanufacturing process of used products, and it is considered in this paper.

In a spare-part inventory system, the accurate demand information is important, which can reduce safety stock and thus might reduce cost without reducing service levels. A manager may need to focus on the development of information technology for effective inventory management. We know that, under a communication link technology, typically electronic data interchange (EDI), a vendor-managed inventory (VMI) is more effective to plan inventory and place orders. For example, Bourland et al. [15] investigated the operational problem of the tradeoff between communication (i.e., EDI) and inventory. They assumed a fixed cost of acquiring timely demand information and examined how the firm to adjust its inventory behavior. And information technology has enabled some researches to be more proactive and obtain advance demand information in addition to improving demand forecast, such as Gallego and Özer (2001 [16], 2003 [17]) and Özer and Wei (2004 [18]); they incorporated advance demand information into periodic-review inventory control problems. Although the main methods for forecasting demand are still statistical methods, such as moving average and exponential smoothing (also see Axsater, 2006 [19]), under more advanced information technologies (such as Radio-Frequency Identification (RFID), cloud computing, and Internet of Things), firms might obtain more in advance and more accurate information, such as the real-time status of one product, historical sales data, customers' position, the correlation between the historical service data, the using environment of the product, and customers' types. These possibilities also enable firms to change the operations policy, such as making a better and longer spare-part inventory planning. Therefore, we can consider a multiperiod spare-part inventory problem under the existing of remanufacturing.

In this paper, we will consider the inventory decisions problem in a multisource spare-part system, where we adopt the final production order and remanufacturing ways to replenish spare parts. The final production order provides the initial inventory of spare parts; then, the inventory is continuously replenished by collecting and remanufacturing used products. Under the setting, the manager needs to determine the optimal final production order and acquisition price of collecting used products.

The rest of this paper is organized as follows. In Section 2, we give the problem description and formulation. Section 3 provides the optimal acquisition pricing policy. Furthermore, we analyze the optimal size of the final production order and the acquisition pricing at the first period in Section 4. In Section 5, we propose an algorithm for the parameters in optimal policy structures. Numerical examples are provided in Section 6. Finally, we conclude our paper in Section 7.

2. Problem Description and Formulation

We consider an inventory system with single spare part and two replenishment ways, including the final production order and remanufacturing. The final production order only needs to be determined at the time of closing production line, which maintains the demand of the regular warranty period. To satisfy the demand in an extended warranty period, especially for outside the regular warranty period, the enterprise starts an acquisition and remanufacturing planning of used products, that is, by collecting and remanufacturing the used products into serviceable components and furthermore satisfying the demand in the service period. For controlling the return quantity of used products, the enterprise will provide an incentive for users which may be a cash payment. We call it acquisition price. A finite period setting will be considered in the following. Without loss of generality, the length of a period is assumed to be one, and the periods are numbered by $k = 1, 2, \ldots, N$.

Let τ_k denote the acquisition price at period k and $\xi_k(\tau_k)$ denote the returned quantity when the acquisition price is τ_k; it is defined as the following form:

$$\xi_k(\tau_k) = \mu_k(\tau_k) + \varepsilon_k, \tag{1}$$

where $\mu_k(\tau_k) = \alpha + \beta\tau_k$ and ε_k is a random factor with support $[a_k, b_k]$ and zero mean value. $\mu_k(\tau_k)$ is the expected quantity of returned products under the given acquisition price τ_k. This linear form is similar to the linear demand function, and the latter is frequently used in many studies (e.g., Federgruen and Heching [20], Simchi-Levi et al. [21], and Sun et al. [22]). We assume that the supply capacity of returned products is unlimited, and all returned products can be remanufactured into serviceable parts. And the remanufacturing process has no production loss, and the unit returned product will yield the unit serviceable part.

All demands of serviceable parts at different periods are stochastic and independent of each other. Because demands and supplies are stochastic, both the inventory shortage and surplus status are possible. The unsatisfied demands need to be backlogged and incur the inventory shortage cost. Further, all surplus will be transferred into the next period and incur the inventory holding cost.

The sequence of system events is given as follows. First, the firm determines the size of the final production order and acquisition price at the beginning of the planning horizon. From the second period to the Nth period, the firm only determines the acquisition price level at the period. Then, within each period, the firm acquires used products from users and remanufactures returned products into serviceable parts and reviews the inventory state of the serviceable part and calculates the related costs. The holding inventory cost can be obtained by the multiplication of the unit holding cost and the quantity of serviceable parts of entering the inventory, which also means that even if one spare part is put into the inventory and taken away in the same period, its cost is still calculated.

The optimization problem is to find the optimal final production order and the optimal acquisition price policies of different periods so that the total inventory and acquisition costs can be minimized.

From the sequence of system events, we have the following state equation of the dynamic system:

$$I_{k+1} = \begin{cases} I_k + \xi_k(\tau_{k+1}) - D_{k+1}, & k \geq 1, \\ i_0 + Q + \xi_1(\tau_1) - D_1, & k = 0. \end{cases} \tag{2}$$

For period k, $k \geq 2$, when the initial inventory level is I_{k-1} and the acquisition price is τ_k, the cost function at the period is given as follows:

$$\begin{aligned} g_k(I_{k-1}, \tau_k) = {} & \tau_k \xi_k(\tau_k) + h\left((I_{k-1})^+ + \xi_k(\tau_k)\right) \\ & + q\left(D_k - I_{k-1} - \xi_k(\tau_k)\right)^+, \end{aligned} \tag{3}$$

where $(x)^+ = \max\{x, 0\}$. The first term of the above equation is the acquisition cost, and the second term is the inventory holding cost, and the third term is the penalty cost of the shortage parts. Especially for the first period, there is the final production order and the cost function is given as follows:

$$\begin{aligned} g_1(i_0, Q, \tau_1) = {} & \tau_1 \xi_1(\tau_1) + h\left(i_0 + Q + \xi_1(\tau_1)\right) \\ & + q\left(D_1 - i_0 - Q - \xi_1(\tau_1)\right)^+. \end{aligned} \tag{4}$$

Let π denote a feasible policy and be defined by $\pi = \{(Q, \tau_1, \ldots, \tau_N) \mid Q \geq 0,\ \tau_k \in [0, \tau_u]\}$. Therefore, for a given initial inventory level i_0 and a feasible policy π, the total cost of the whole system is

$$W(i_0, \pi) = g_1(i_0, Q, \tau_1) + \sum_{k=2}^{N} g_k(I_{k-1}, \tau_k), \tag{5}$$

where we assume the end cost at the planning horizon to be zero.

Let Π denote the set including all feasible policies, so there is $\pi \in \Pi$. The optimization problem in this paper is as follows: for a given initial inventory level i_0, to find an optimal policy $\pi^* = (Q^*, \tau_1^*, \ldots, \tau_N^*)$ so that the expectation value of the total cost in (5) is minimum; that is,

$$W(i_0, \pi^*) = \min_{\pi \in \Pi} \{E[W(i_0, \pi)]\}. \tag{6}$$

We call Q^* the optimal final production order and τ_k^* the optimal acquisition price at the period k. In the following, we will give the dynamic programming equation of solving the optimization problem.

Let i_{k-1} denote the realized inventory state at the beginning of the period k. And for $k \geq 2$, let

$$W_k(i_{k-1}, \pi_k) = E\left[\sum_{j=k}^{N} g_j\left(I_{j-1}, \tau_j\right) \mid I_{k-1} = i_{k-1}\right], \tag{7}$$

and for $k = 1$, let

$$W_1(i_0, \pi_1) = E[g_1(i_0, Q, \tau_1)] + E\left[\sum_{j=2}^{N} g_j\left(I_{j-1}, \tau_j\right)\right]. \tag{8}$$

Furthermore, define the following optimal value function:

$$W_k\left(i_{k-1}\right) = \min_{\pi_k}\left\{W_k\left(i_{k-1}, \pi_k\right)\right\}, \quad 1 \le k \le N+1, \tag{9}$$

so $W_1(i_0) = W(i_0, \pi^*)$. In the above equations, $\pi_k = (\tau_k, \ldots, \tau_N)$ for $k \ge 2$ and $\pi_1 = (Q, \tau_1, \ldots, \tau_N)$ for $k = 1$. Therefore, the dynamic programming equations of the optimization problem in (6) are given as follows:

$$\begin{aligned} W_1\left(i_0\right) &= \min_{Q\ge 0, \tau_1 \in [0,\tau^u]}\left\{E\left[g_1\left(i_0, Q, \tau_1\right)\right]\right. \\ &\quad \left. + E\left[W_2\left(i_0 + Q + \xi_1\left(\tau_1\right) - D_1\right)\right]\right\}, \\ W_k\left(i_{k-1}\right) &= \min_{\tau_k \in [0,\tau^u]}\left\{E\left[g_k\left(i_{k-1}, \tau_k\right)\right]\right. \\ &\quad \left. + E\left[W_{k+1}\left(i_{k-1} + \xi_k\left(\tau_k\right) - D_k\right)\right]\right\}, \quad 2 \le k \le N \\ W_{N+1}\left(i_N\right) &= 0. \end{aligned} \tag{10}$$

In the following section, we will analyze the optimal decisions.

3. Optimal Acquisition Pricing for $k \ge 2$

In this section, we will give the optimal acquisition pricing decision τ_k^* for $k \ge 2$. From the second equation in (10), we define

$$\begin{aligned} H_k\left(i_{k-1}, \tau_k\right) &= E\left[g_k\left(i_{k-1}, \tau_k\right)\right] \\ &\quad + E\left[W_{k+1}\left(i_{k-1} + \xi_k\left(\tau_k\right) - D_k\right)\right]. \end{aligned} \tag{11}$$

To analyze the properties of $H_k(i_{k-1}, \tau_k)$, we firstly give the following lemmas.

Lemma 1. *If the function $h(x)$ is convex in x, then the function $g(x, y) = E[h(x + ay - D)]$ is jointly convex in x and y, where D is a stochastic variable with the cumulative distribution function $F(\cdot)$.*

Proof. For any given $z_1 = (x_1, y_1)$ and $z_2 = (x_2, y_2)$ and $\theta \in [0, 1]$, we have

$$\begin{aligned} g\left(\theta z_1 + (1-\theta) z_2\right) &= g\left(\theta x_1 + (1-\theta) x_2, \theta y_1 + (1\right. \\ &\quad \left. - \theta) y_2\right) = E\left[h\left(\theta x_1 + (1-\theta) x_2\right.\right. \\ &\quad \left.\left. + a\left(\theta y_1 + (1-\theta) y_2\right) - D\right)\right] = \int_{-\infty}^{+\infty} h\left(\theta x_1\right. \\ &\quad \left. + (1-\theta) x_2 + a\left(\theta y_1 + (1-\theta) y_2\right)\right. \\ &\quad \left. - z\right) dF\left(z\right) \\ &= \int_{-\infty}^{+\infty} h\left(\theta\left(x_1 + a y_1 - z\right)\right. \\ &\quad \left. + (1-\theta)\left(x_2 + a y_2 - z\right)\right) dF\left(z\right) \\ &\le \theta \int_{-\infty}^{+\infty} h\left(x_1 + a y_1 - z\right) dF\left(z\right) + (1-\theta) \\ &\quad \cdot \int_{-\infty}^{+\infty} h\left(x_2 + a y_2 - z\right) dF\left(z\right) = \theta E\left[h\left(x_1\right.\right. \\ &\quad \left.\left. + a y_1 - z\right)\right] + (1-\theta) E\left[h\left(x_2 + a y_2 - z\right)\right] \\ &= \theta g\left(z_1\right) + (1-\theta) g\left(z_2\right). \end{aligned} \tag{12}$$

Therefore, $g(x, y)$ is convex in x and y. □

Lemma 2. *If $f(x, y)$ is jointly convex in x and y and C is a convex set, then $g(x) = \min_{y\in C} f(x, y)$ is also convex in x.*

Proof. For the proof of the lemma, please see page 84 in Boyd and Vandenberghe [23]. □

In the following, we analyze the properties of $E[g_k(i_{k-1}, \tau_k)]$.

Theorem 3. *The period cost function $E[g_k(i_{k-1}, \tau_k)]$ in (3) has the following properties: (a) $E[g_k(i_{k-1}, \tau_k)]$ is convex in i_{k-1} and τ_k; (b) $E[g_k(i_{k-1}, \tau_k)]$ is supermodular in i_{k-1} and τ_k.*

Proof. We firstly prove part (a). It is obvious that the function $\max\{-x, 0\}$ is convex. From (1), there is $\xi_k(\tau_k) = \rho_k\alpha + \beta\rho_k\tau_k$. And by Lemma 1, we easily know that $E[\max\{D_k - x - \xi_k(\tau_k), 0\}]$ is convex in x and τ_k. Other parts in $E[g_k(i_{k-1}, \tau_k)]$ are convex obviously. So part (b) holds.

For part (b), for any $x_1 \ge x_2$ and $\tau_k^1 \ge \tau_k^2$, we have

$$\begin{aligned} &E\left[g_k\left(x_1, \tau_k^1\right)\right] + E\left[g_k\left(x_2, \tau_k^2\right)\right] - E\left[g_k\left(x_1, \tau_k^2\right)\right] \\ &\quad - E\left[g_k\left(x_2, \tau_k^1\right)\right] = q\left(E\left[\left(D_k - x_1 - \xi_k\left(\tau_k^1\right)\right)^+\right]\right. \\ &\quad \left. + E\left[\left(D_k - x_2 - \xi_k\left(\tau_k^2\right)\right)^+\right]\right. \\ &\quad \left. - E\left[\left(D_k - x_1 - \xi_k\left(\tau_k^2\right)\right)^+\right]\right. \\ &\quad \left. - E\left[\left(D_k - x_2 - \xi_k\left(\tau_k^1\right)\right)^+\right]\right). \end{aligned} \tag{13}$$

Let $g(y, d) = (d - y - x_1)^+ - (d - y - x_2)^+$, easily knowing that $g(y, d)$ is nondecreasing in y. From $\tau_k^1 \ge \tau_k^2$, there is $\xi_k(\tau_k^1) \ge_{st} \xi_k(\tau_k^2)$. And from Ross [24], we have $E[g(\xi_k(\tau_k^1, D_k))] \ge E[g(\xi_k(\tau_k^2, D_k))]$, so

$$\begin{aligned} &E\left[g_k\left(x_1, \tau_k^1\right)\right] + E\left[g_k\left(x_2, \tau_k^2\right)\right] - E\left[g_k\left(x_1, \tau_k^2\right)\right] \\ &\quad - E\left[g_k\left(x_2, \tau_k^1\right)\right] \ge 0. \end{aligned} \tag{14}$$

Therefore, $E[g_k(x_1, \tau_k)]$ is supermodular in x_1 and τ_k. □

Furthermore, the following theorem gives the properties of the optimal value function and the objective function.

Theorem 4. *The optimal value function $W_k(i_{k-1})$ in (9) and the objective function $H_k(i_{k-1}, \tau_k)$ in (11) have the following properties:*

(a) *$W_{k+1}(i_k)$ is convex in i_k for $k = 1, 2, \ldots, N$.*

(b) *$H_k(i_{k-1}, \tau_k)$ is convex in i_{k-1} and τ_k for $k = 2, 3, \ldots, N$.*

(c) *$H_k(i_{k-1}, \tau_k)$ is supermodular in i_{k-1} and τ_k for $k = 2, 3, \ldots, N$.*

Proof. Apply backward recursive method.

When $k = N$, there is always $W_{N+1}(i_N) = 0$ for any i_N, so part (a) holds for $k = N + 1$. From (11), there is

$$H_k(i_{k-1}, \tau_k) = E[g_k(i_{k-1}, \tau_k)], \tag{15}$$

and by Theorem 3, it is obvious that $H_k(i_{k-1}, \tau_k)$ is convex and supermodular in i_{k-1} and τ_k.

Assume that the theorem holds for $k = n + 1$; that is,

(a) $W_{n+2}(i_{n+1})$ is convex in i_{n+1};

(b) $H_{n+1}(i_n, \tau_{n+1})$ is convex in i_n and τ_{n+1};

(c) $H_{n+1}(i_n, \tau_{n+1})$ is super-modular in i_n and τ_{n+1}.

We need to prove that the theorem still holds for $k = n$. From (10) and (11), we have

$$W_{n+1}(i_n) = \min_{\tau_{n+1} \in [0, \tau^u]} \{H_{n+1}(i_n, \tau_{n+1})\}. \tag{16}$$

From inductive assumption (b), $H_{n+1}(i_n, \tau_{n+1})$ is convex in i_n and τ_{n+1}, and $[0, \tau^u]$ is a convex set, from Lemma 2, so $W_{n+1}(i_n)$ is convex in i_n.

From (11), we have

$$\begin{aligned} H_k(i_{n-1}, \tau_n) &= E[g_n(i_{n-1}, \tau_n)] \\ &\quad + E[W_{n+1}(i_{n-1} + \xi_n(\tau_n) - D_n)]. \end{aligned} \tag{17}$$

Because $W_{n+1}(i_n)$ is convex in i_n, from Lemma 1, we have the notion that $E[W_{n+1}(i_{n-1} + \xi_n(\tau_n) - D_n)]$ is convex in i_{n-1} and τ_n. And from Theorem 3, we have the notion that $E[g_n(i_{n-1}, \tau_n)]$ is convex in i_{n-1} and τ_n. Therefore, $H_n(i_{n-1}, \tau_n)$ is convex in i_{n-1} and τ_n.

For the supermodular property in part (c), we firstly prove $E[W_{n+1}(i_{n-1} + \xi_n(\tau_n) - D_n)]$ to be supermodular in i_{n-1} and τ_n. Set $H(y) = E[W_{k+1}(x_2 + y - D_k)] - E[W_{k+1}(x_1 + y - D_k)]$. For any $x_1 \le x_2$, we have

$$\begin{aligned} &H(y + x_2 - x_1) - H(y) \\ &\quad = (E[W_{k+1}(2x_2 + y - x_1 - D_k)] \\ &\quad - E[W_{k+1}(x_2 + y - D_k)]) \\ &\quad - (E[W_{k+1}(x_2 + y - D_k)] \\ &\quad - E[W_{k+1}(x_1 + y - D_k)]) \\ &\quad = E[W_{k+1}(2x_2 + y - x_1 - D_k)] \\ &\quad - 2E[W_{k+1}(x_2 + y - D_k)] \\ &\quad + E[W_{k+1}(x_1 + y - D_k)]. \end{aligned} \tag{18}$$

Because $W_{n+1}(i_n)$ is convex in i_n, $E[W_{k+1}(i_n - D_k)]$ is also convex. Therefore, $H(y + x_2 - x_1) - H(y) \ge 0$; that is, $H(y)$ is increasing with respect to y. For $\tau_k^1 \le \tau_k^2$, we easily know that $\xi_k(\tau_k^1) \le_{st} \xi_k(\tau_k^2)$, so there is $E[H(\xi_k(\tau_k^1))] \le E[H(\xi_k(\tau_k^2))]$; that is,

$$\begin{aligned} &E[W_{k+1}(x_2 + \xi_k(\tau_k^1) - D_k)] \\ &\quad - E[W_{k+1}(x_1 + \xi_k(\tau_k^1) - D_k)] \\ &\le E[W_{k+1}(x_2 + \xi_k(\tau_k^2) - D_k)] \\ &\quad - E[W_{k+1}(x_1 + \xi_k(\tau_k^2) - D_k)]. \end{aligned} \tag{19}$$

Therefore, $E[W_{n+1}(i_{n-1} + \xi_n(\tau_n) - D_n)]$ is supermodular in i_{n-1} and τ_n. And from part (b) in Theorem 3 and (11), $H_n(i_{n-1}, \tau_n)$ is supermodular in i_{n-1} and τ_n. In summary, the induction method has been completed. □

From part (b) in Theorem 4, we know that the optimal solution of $\min_{\tau_k \in [0, \tau^u]} \{H_k(i_{k-1}, \tau_k)\}$ is unique, so we can make the following definition:

$$\tau_k(i_{k-1}) = \arg \min_{\tau_k \in [0, \tau^u]} \{H_k(i_{k-1}, \tau_k)\}. \tag{20}$$

For $\tau_k(i_{k-1})$, we have the following property.

Theorem 5. *The optimal solution $\tau_k(i_{k-1})$ in (20) is decreasing with respect to i_{k-1}.*

Proof. From part (c) in Theorem 4, we have the notion that $H_k(i_{k-1}, \tau_k)$ is supermodular in i_{k-1} and τ_k. And according to Puterman [25] (Lemma 4.7.2 in page 94), $\tau_k(i_{k-1})$ is decreasing with respect to i_{k-1}. □

Theorem 6. *There exists X_k^u with the following definition:*

$$X_k^u = \inf\{i_{k-1} \mid \tau_k(i_{k-1}) = 0\}. \tag{21}$$

Proof. The first-order derivative of $H_k(i_{k-1}, \tau_k)$ with respect to τ_k is given as follows:

$$\begin{aligned} \frac{\partial H_k(i_{k-1}, \tau_k)}{\partial \tau_k} &= \alpha + 2\beta\tau_k + \beta h - q\beta \\ &\quad + q\beta F(i_{k-1} + \alpha + \beta\tau_k) \\ &\quad + \beta \frac{\partial E[W_{k+1}(i_{k-1} + \xi_k(\tau_k) - D_k)]}{\partial \tau_k}. \end{aligned} \tag{22}$$

For a finite period problem, when the inventory level i_{k-1} at the period $k - 1$ is large enough, the shortage cost can be ignored in the following periods, so there is $\partial E[W_{k+1}(i_{k-1} + \xi_k(\tau_k) - D_k)]/\partial \tau_k = h$. Therefore, $\partial H_k(i_{k-1}, \tau_k)/\partial \tau_k > 0$; that is, the optimal solution is $\tau_k(i_{k-1}) = 0$. □

Corollary 7. *The optimal acquisition price decision obeys the following decision rule:*

$$\tau_k^* = \begin{cases} \tau_k(i_{k-1}), & i_{k-1} < X_k^u, \\ 0, & i_{k-1} \ge X_k^u. \end{cases} \tag{23}$$

And if there exists X_k^l with the following definition

$$X_k^l = \sup\left\{i_{k-1} \mid \tau_k\left(i_{k-1}\right) = \tau^u\right\}, \tag{24}$$

then the optimal acquisition price is

$$\tau_k^* = \begin{cases} \tau^u, & i_{k-1} \le X_k^l, \\ \tau_k\left(i_{k-1}\right), & X_k^l < i_{k-1} < X_k^u, \\ 0, & i_{k-1} \ge X_k^u, \end{cases} \tag{25}$$

where $\tau_k(i_{k-1})$ is the solution of $\partial H_k(i_{k-1}, \tau_k)/\partial\tau_k = 0$. X_k^l is defined in (24), and X_k^u is defined in (21).

Proof. From the monotonicity of $\tau_k(i_{k-1})$ in Theorem 5 and the definitions of X_k^l in (24) and X_k^u in (21), the corollary is obvious. □

4. Optimal Acquisition Pricing and Final Order

In this section, we will give the optimal decisions Q^* and τ_k^* for $k = 1$. From the first equation in (10), we define

$$\begin{aligned} H_1\left(i_0, Q, \tau_1\right) &= E\left[g_1\left(i_0, Q, \tau_1\right)\right] \\ &\quad + E\left[W_2\left(i_0 + Q + \xi_1\left(\tau_1\right) - D_1\right)\right]. \end{aligned} \tag{26}$$

Let $y_1 = i_0 + Q_1$ denote the inventory level of serviceable parts after replenishing. Therefore, we have $E[g_1(0, y_1, \tau_1)] = E[g_1(i_0, Q, \tau_1)]$, and, further, there is

$$\begin{aligned} H_1\left(i_0, Q, \tau_1\right) &= H_1\left(0, y_1, \tau_1\right) \\ &= E\left[g_1\left(0, y_1, \tau_1\right)\right] \\ &\quad + E\left[W_2\left(y_1 + \xi_1\left(\tau_1\right) - D_1\right)\right] \\ &= \tau_1\xi_1\left(\tau_1\right) + h\left(y_1 + \xi_1\left(\tau_1\right)\right) \\ &\quad + q\left(D_1 - y_1 - \xi_1\left(\tau_1\right)\right)^+ \\ &\quad + E\left[W_2\left(y_1 + \xi_1\left(\tau_1\right) - D_1\right)\right]. \end{aligned} \tag{27}$$

For convenience, let $H(y_1, \tau_1) = H_1(0, y_1, \tau_1)$; that is,

$$\begin{aligned} H\left(y_1, \tau_1\right) &= \tau_1\xi_1\left(\tau_1\right) + h\left(y_1 + \xi_1\left(\tau_1\right)\right) \\ &\quad + q\left(D_1 - y_1 - \xi_1\left(\tau_1\right)\right)^+ \\ &\quad + E\left[W_2\left(y_1 + \xi_1\left(\tau_1\right) - D_1\right)\right]. \end{aligned} \tag{28}$$

And we have the following theorem.

Theorem 8. *$H(y_1, \tau_1)$ in (28) is jointly convex in y_1 and τ_1.*

Proof. From part (a) in Theorem 4, we know that $W_2(i_1)$ is convex in i_1. And from Lemma 1, $E[W_2(y_1 + \xi_1(\tau_1) - D_1)]$ is jointly convex in y_1 and τ_1.

From Theorem 3, the cost function $E[g_k(i_{k-1}, \tau_k)]$ in (3) is convex in i_{k-1} and τ_k, so the function $E[g_1(0, y_1, \tau_1)]$ in (4) is also jointly convex in y_1 and τ_1. Therefore, $H(y_1, \tau_1)$ in (28) is jointly convex in y_1 and τ_1. □

From Theorem 8, we can give the following form of the optimal policy at the first period.

Theorem 9. *For the jointly inventory and the acquisition price decision problem at the first period, a policy (S, τ) is optimal.*

Proof. Since $H(y_1, \tau_1)$ is jointly concave in y_1 and τ_1, we have

$$\begin{aligned} &\min_{y_1 \ge i_0, \tau_1 \in [0, \tau^u]} \left\{H\left(y_1, \tau_1\right)\right\} \\ &\quad = \min_{y_1 \ge i_0} \left\{ \min_{\tau_1 \in [0, \tau^u]} \left\{H\left(y_1, \tau_1\right)\right\} \right\}. \end{aligned} \tag{29}$$

Obviously, there exists a unique $\tau_1(y_1)$ such that

$$\tau_1\left(y_1\right) = \arg\min_{\tau_1 \in [0, \tau^u]} \left\{H\left(y_1, \tau_1\right)\right\}. \tag{30}$$

From Lemma 2, we know that $\min_{\tau_1 \in [0, \tau^u]}\{H_1(y_1, \tau_1)\}$ is still convex in y_1. Therefore, the optimal final production order decision obeys a base-stock policy. And, furthermore, we can find the corresponding optimal inventory level S such that $(S, \tau_1(S))$ can minimize $H(y_1, \tau_1)$. □

Theorem 9 shows that the optimal decisions at the first period follow a simple policy with two parameters. Executing the policy, the firm firstly reviews its initial inventory i_0 at the beginning of the planning horizon. Then, it determines the optimal base-stock level S. If the initial inventory level is less than the optimal basic-stock level S, then it replenishes the inventory of spare parts to S and sets the optimal acquisition price level to be $\tau_1(S)$. If the initial inventory level is larger than the optimal basic-stock level S, then the firm does not need to replenish the spare-part inventory and set the optimal acquisition price level to be $\tau_1(i_0)$. For $\tau_1(S)$, we have the following property.

Theorem 10. *$\tau_1(S)$ is decreasing with respect to the optimal inventory level S.*

Proof. From the proof of part (c) in Theorem 4, we know that if $W_{k+1}(i_k)$ is convex, then $E[W_{k+1}(i_{k-1} + \xi_k(\tau_k) - D_k)]$ is supermodular in i_{k-1} and τ_k. From part (a) in Theorem 4, $W_2(i_1)$ is convex, so $E[W_2(y_1 + \xi_1(\tau_1) - D_1)]$ is supermodular in y_1 and τ_1.

From part (b) in Theorem 3, $E[g_1(0, y_1, \tau_1)]$ is supermodular in y_1 and τ_1. Furthermore, from (28), $H(y_1, \tau_1)$ is also supermodular in y_1 and τ_1. Therefore, $\tau_1(y_1)$ is decreasing with respect to y_1. And $\tau_1(S)$ is decreasing with respect to S. □

5. Algorithm for Optimal Decisions

Although the policy structure of the optimal acquisition pricing and the final production order are given in the above sections, the analytical forms of the related parameters in the policy structure are not obtained. In this section, we will give an algorithm to calculate the optimal decisions at each period. For convenience, we provide an algorithm description for the case of discrete state variables.

Algorithm 11.

Step 1. Set $k = N + 1$ and $W_{N+1}(i_N) = 0$ for all i_N.

Step 2. Set $k \leftarrow k - 1$.

Step 3. Find X_k^u in (21) using the following judgement condition by a search algorithm (e.g., binary search method):

$$\begin{aligned}\frac{\partial H_k(i_{k-1}, \tau_k)}{\partial \tau_k} &= \beta h - q\beta + q\beta F(i_{k-1} + \alpha + \beta\tau_k) + \alpha \\ &\quad + 2\beta\tau_k \\ &\quad + \beta \frac{\partial E[W_{k+1}(i_{k-1} + \xi_k(\tau_k) - D_k)]}{\partial \tau_k} \\ &> 0.\end{aligned} \tag{31}$$

If $k = 1$, then go to Step 4; otherwise, go to Step 3.1.

Step 3.1. Set $i_{k-1} = X_k^u$.

Step 3.2. Set $i_{k-1} \leftarrow i_{k-1} - 1$, and find the solution τ_k' of $\partial H_k(i_{k-1}, \tau_k)/\partial \tau_k = 0$. If the solution $\tau_k' \le \tau^u$, then transfer to Step 3.3; otherwise, go to Step 3.4.

Step 3.3. Set $\tau_k^*(i_{k-1}) = \tau_k'$, calculate $W_k(i_{k-1})$ by the following equation, and then go to Step 3.2:

$$W_k(i_{k-1}) = H_k(i_{k-1}, \tau_k^*(i_{k-1})). \tag{32}$$

Step 3.4. Set $\tau_k^*(i_{k-1}) = \tau^u$ and $X_k^l = i_{k-1}$, and go to the next step.

Step 3.5. Calculate $W_k(i_{k-1})$. If $i_{k-1} = \text{Min}\, i_{k-1}$ (the minimum value of system state), go to Step 2; otherwise, go to the next step.

Step 3.6. Set $i_{k-1} \leftarrow i_{k-1} - 1$, set $\tau_k^*(i_{k-1}) = \tau^u$, and go to Step 3.5.

Step 4. Set $y_1 = X_1^u$ and calculate $W_1(y_1)$; let H Min Val $= W_1(y_1)$ and $S = y_1$.

Step 4.1. Set $y_1 \leftarrow y_1 - 1$, and find the solution τ_1' of $\partial H_1(i_0, \tau_1)/\partial \tau_1 = 0$. If the solution $\tau_1' \le \tau^u$, then go to Step 4.2; otherwise, go to Step 4.3.

Step 4.2. Set $\tau_1^*(y_1) = \tau_1'$, and calculate $W_1(y_1)$. If $W_1(y_1) < H$ Min Val, then set H Min Val $= W_1(y_1)$ and $S = y_1$, and go to Step 4.1.

Step 4.3. Set $\tau_1^*(y_1) = \tau^u$ and $X_1^l = y_1$, and go to Step 4.4.

Step 4.4. Calculate $W_1(y_1)$. If $W_1(y_1) < H$ Min Val, then set H Min Val $= W_1(y_1)$ and $S = y_1$, and go to Step 4.5.

Step 4.5. Set $y_1 \leftarrow y_1 - 1$ and $\tau_1^*(y_1) = \tau^u$, and go to Step 4.4.

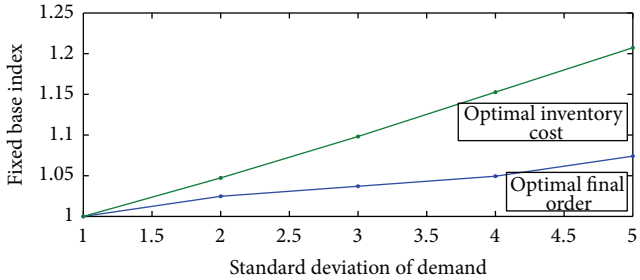


Figure 1: Effects on optimal final order and inventory cost.

6. Numerical Study

By some advanced information technologies (such as Wireless Communication Networks, Radio-Frequency Identification (RFID), cloud computing, etc.), one enterprise might obtain more information on the using products or the customers of holding used products. To analyze the effect of those information on the inventory policy parameters (including X_k^l, X_k^u, and S) and the optimal operation cost of the spare-part system, we will adjust parameters to obtain those trends by numerical ways in this section. We mainly consider the effects of two parameters: one is the standard deviation of the uncertain demand at each period, and the other is the customer's sensitive degree on the acquisition price.

We give a basic setting of related parameters. The demand is assumed to follow a normal distribution with the mean 20 and the standard deviation 2, and the disturbance variable ε of the return quantity follows a normal distribution with zero mean and the standard deviation 0.2. The other parameters are given as follows: $N = 5$, $\alpha = 4$, $\beta = 1$, $h = 1$, and $q = 30$.

6.1. Effect of the Demand Uncertainty. Firstly, we investigate the effects of the uncertainty of the period demand on the optimal final production order, pricing bounds in the above pricing policy structure, and the optimal inventory cost. Let the standard deviation of the uncertain period demand vary from 1 to 5; the fixed base indexes of optimal final production orders and the system optimal costs can be shown in Figure 1, where the fixed bases are the optimal production final orders and the system optimal cost in the first period.

From Figure 1, we know that, as the increasing of the standard deviation, both the optimal final order S and the system optimal cost are also increasing. Because the improvement of the demand uncertainty will make the increasing of the shortage cost larger than the increasing of the inventory holding cost, therefore, the optimal inventory level will be improved, and the corresponding optimal cost of the system is also increasing.

From Figure 2, we know that the bounds X_k^u and X_k^l are decreasing with respect to the increasing of the period number k. The upper bound means the critical point which the system does not need to acquire used products, and the lower bound means the critical point which the system needs to acquire used products as possible. If the enterprise

Table 1: Period bounds under different standard deviations.

Sta. dev.	Bounds	Period 2	Period 3	Period 4	Period 5
1	Lower bound	44	33	23	11
	Upper bound	63	48	33	17
2	Lower bound	45	34	23	11
	Upper bound	63	49	34	18
3	Lower bound	47	34	23	11
	Upper bound	64	49	35	19
4	Lower bound	47	35	24	11
	Upper bound	65	50	36	20
5	Lower bound	47	35	25	12
	Upper bound	66	51	37	21

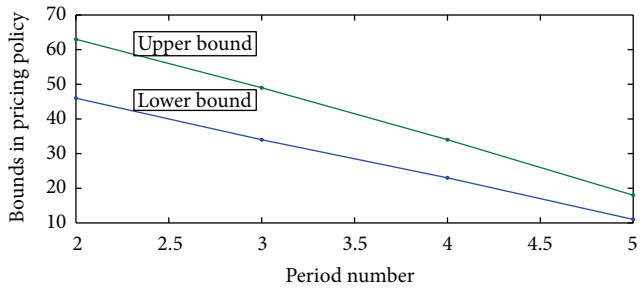


Figure 2: Period index on bounds in pricing policy.

Table 2: Period bounds under different return sensitivities.

Sensitivity degree	Bounds	Period 2	Period 3	Period 4	Period 5
1	Lower bound	46	34	23	11
	Upper bound	63	49	34	18
1.5	Lower bound	33	24	16	8
	Upper bound	60	48	33	19
2	Lower bound	27	20	13	6
	Upper bound	60	47	33	19
2.5	Lower bound	17	12	8	4
	Upper bound	58	46	33	19
3	Lower bound	9	7	5	1
	Upper bound	57	46	33	19

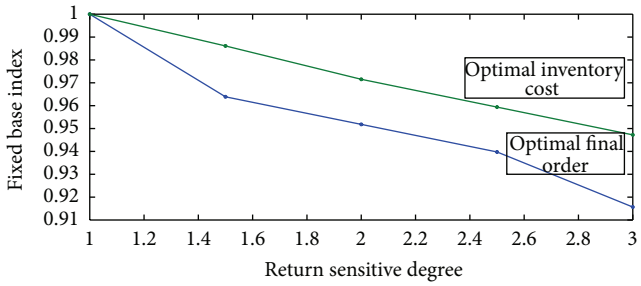


Figure 3: Effects on optimal final order and inventory cost.

faces a larger uncertainty within a longer future planning horizon, which will make the increasing of the shortage cost larger than the increasing of the inventory holding cost, the whole inventory holding level should be increased. When the number k is little, the enterprise needs to pay more attention for dealing with the future longer and larger uncertain demand, so both critical points of acquiring and not acquiring used products should be increased. And with respect to the decreasing of the planning length, the fluctuation between upper bound and lower bound is decreasing because the fluctuation of the future demand is decreasing.

From Table 1, we know that, as the increasing of the standard deviation, both bounds X_k^u and X_k^l are increasing. The results are similar to the above properties: the critical point of holding inventory level also will be increased when the enterprise faces the larger uncertainty.

6.2. Effect of the Return Sensitivity. In this subsection, we will investigate the effects of the return sensitivity.

From Figure 3, as the increasing of the return sensitivity degree, the optimal final production order S is decreasing, and the optimal inventory cost is also decreasing. The same acquisition price will acquire more used products than that for the case with little return sensitivity, and the enterprise does not need to hold a higher safety inventory if the product acquisition is more easy, so there are the results in Figure 3.

From Table 2, as the increasing of the return sensitivity degree, both bounds X_k^u and X_k^l are decreasing. The reasons of the results are similar to the reason for Figure 3: when the acquisition way is more effective, the enterprise does not need to hold a higher inventory. The critical points of acquiring and not acquiring used products can be decreased.

7. Conclusion

In order to gain a competitive advantage in an increasing competitive market, many enterprises have improved the concerning for the after-sale service in a regular warranty period. However, the service within the warranty period is not enough for maintaining the customer's satisfaction, which is showed by Nagler [1]: "you buy a car and eight years later cannot get it serviced for a reasonable price because the manufacturer has discontinued a particular part, you will remember it when you buy your next vehicle". Therefore, in addition to the regular warranty period, the service outside the warranty period is also important. The traditional ways of replenishing spare parts within the warranty period include the final production order and the additional production, but those are higher cost or setup cost, especially for durable products outside the warranty period.

In this paper, we consider two ways including the final production order and remanufacturing. the remanufacturing way has been concerned in recent years, but few papers consider the acquisition management problem of used products in an inventory system with remanufacturing. We assume that the returning process is controlled by the acquisition price of used products and consider a multiperiod dynamic optimization problem, where the system decisions include

the optimal final production batch size and the optimal acquisition price at each period. In order to obtain the optimal policy, we divided our analysis into two stages. First, we study the optimal acquisition pricing from the second period to final period. By stochastic dynamic programming, we prove the structural properties of the objective function and give an optimal monotonous pricing with key thresholds. Then, we analyze the optimal decision problem at the first period, which is jointly an inventory and pricing problem, and we prove a simple policy (S, τ) to be optimal. Furthermore, we propose a recursion algorithm and make some numerical examples.

In the future, the combination of more replenishment ways can be considered in one framework, which may include the final production order, extra production, substitution, and remanufacturing. Certainly, the acquisition pricing should be considered. The other direction is to consider a closed-loop structure; that is, the return of used products is dependent on the historic demands, which will be a more attractive perspective.

Notations

N: Total number of all running periods
h: Inventory holding cost of per unit component in one period
q: Penalty cost of per unit inventory shortage
τ^u: Maximum acquisition price
τ_k: Acquisition price at the period k, $1 \le k \le N$, $\tau_k \in [0, \tau^u]$
D_k: Demand of the spare part at period k, $1 \le k \le N$, which is a stochastic variable
$\xi_k(\tau_k)$: Returned quantity when the acquisition price is τ_k, $1 \le k \le N$, which is a stochastic variable with the form $\xi_k(\tau_k) = \mu_k(\tau_k) + \varepsilon_k$, where ε_k is a random factor with support $[a_k, b_k]$ and zero mean value and the probability density function $f(\cdot)$ and the cumulative distribution function $F(\cdot)$
I_k: Inventory level at the end of period k
i_0: Initial inventory level at the beginning of the planning horizon.

Competing Interests

The authors declare that there are no competing interests regarding the publication of this paper.

Acknowledgments

This research is supported by National Natural Science Foundation of China (NSFC), Research Fund no. 71302112.

References

[1] B. Nagler, "Reintroducing remanufacturing," *Machine Design*, vol. 71, no. 22, pp. 78–84, 1999.

[2] M. J. Armstrong and D. R. Atkins, "Joint optimization of maintenance and inventory policies for a simple system," *IIE Transactions*, vol. 28, no. 5, pp. 415–424, 1996.

[3] W. Bruggeman and R. van Dierdonck, "Maintenance resource planning an integrative approach," *Engineering Costs and Production Economics*, vol. 9, no. 1–3, pp. 147–154, 1985.

[4] T. S. Dhakar, C. P. Schmidt, and D. M. Miller, "Base stock level determination for high cost low demand critical repairable spares," *Computers & Operations Research*, vol. 21, no. 4, pp. 411–420, 1994.

[5] A. Díaz and M. C. Fu, "Models for multi-echelon repairable item inventory systems with limited repair capacity," *European Journal of Operational Research*, vol. 97, no. 3, pp. 480–492, 1997.

[6] M. Fleischmann, J. A. E. E. van Nunen, and B. Gräve, "Integrating closed-loop supply chains and spare-parts management at IBM," *Interfaces*, vol. 33, no. 6, pp. 44–56, 2003.

[7] R. H. Teunter and L. Fortuin, "End-of-life service: a case study," *European Journal of Operational Research*, vol. 107, no. 1, pp. 19–34, 1998.

[8] L. Fortuin, "The all-time requirement of spare parts for service after salestheoretical analysis and practical results," *International Journal of Operations & Production Management*, vol. 1, no. 1, pp. 59–70, 1980.

[9] R. H. Teunter and L. Fortuin, "End-of-life service," *International Journal of Production Economics*, vol. 59, no. 1, pp. 487–497, 1999.

[10] K. D. Cattani and G. C. Souza, "Good buy? Delaying end-of-life purchases," *European Journal of Operational Research*, vol. 146, no. 1, pp. 216–228, 2003.

[11] J. P. J. van Kooten and T. Tan, "The final order problem for repairable spare parts under condemnation," *Journal of the Operational Research Society*, vol. 60, no. 10, pp. 1449–1461, 2009.

[12] R. H. Teunter and W. K. Klein Haneveld, "Inventory control of service parts in the final phase," *European Journal of Operational Research*, vol. 137, no. 3, pp. 497–511, 2002.

[13] K. Inderfurth and K. Mukherjee, "Decision support for spare parts acquisition in post product life cycle," *Central European Journal of Operations Research*, vol. 16, no. 1, pp. 17–42, 2008.

[14] M. Pourakbar, J. B. G. Frenk, and R. Dekker, "End-of-life inventory decisions for consumer electronics service parts," Work Paper, Erasmus School of Economics, Erasmus University Rotterdam, Rotterdam, The Netherlands, 2009.

[15] K. E. Bourland, S. G. Powell, and D. F. Pyke, "Exploiting timely demand information to reduce inventories," *European Journal of Operational Research*, vol. 92, no. 2, pp. 239–253, 1996.

[16] G. Gallego and Ö. Özer, "Integrating replenishment decisions with advance demand information," *Management Science*, vol. 47, no. 10, pp. 1344–1360, 2001.

[17] G. Gallego and Ö. Özer, "Optimal replenishment policies for multiechelon inventory problems under advance demand information," *Manufacturing and Service Operations Management*, vol. 5, no. 2, pp. 157–175, 2003.

[18] Ö. Özer and W. Wei, "Inventory control with limited capacity and advance demand information," *Operations Research*, vol. 52, no. 6, pp. 988–1000, 2004.

[19] S. Axsater, *Inventory Control*, Springer, Berlin, Germany, 2006.

[20] A. Federgruen and A. Heching, "Combined pricing and inventory control under uncertainty," *Operations Research*, vol. 47, no. 3, pp. 454–475, 1999.

[21] D. Simchi-Levi, X. Chen, and J. Bramel, *The Logic of Logistics: Theory, Algorithms and Applications for Logistics and Supply Chain Management*, Springer, New York, NY, USA, 2nd edition, 2004.

[22] X. Sun, Y. Li, K. Govindan, and Y. Zhou, "Integrating dynamic acquisition pricing and remanufacturing decisions under random price-sensitive returns," *International Journal of Advanced Manufacturing Technology*, vol. 68, no. 1–4, pp. 933–947, 2013.

[23] S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge University Press, Cambridge, Mass, USA, 2004.

[24] S. M. Ross, *Stochastic Processes*, John Wiley & Sons, New York, NY, USA, 2nd edition, 1996.

[25] M. L. Puterman, *Markov Decision Processes: Discrete Stochastic Dynamic Programming*, John Wiley & Sons, New York, NY, USA, 1994.

Hindawi Publishing Corporation
Scientific Programming
Volume 2016, Article ID 8461857, 12 pages
http://dx.doi.org/10.1155/2016/8461857

Research Article

A Heuristics-Based Parthenogenetic Algorithm for the VRP with Potential Demands and Time Windows

Chenghua Shi,[1] Tonglei Li,[2] Yu Bai,[1] and Fei Zhao[3]

[1]*School of Economics and Management, Hebei University of Engineering, No. 199, Guangming South Street, Handan 056038, China*
[2]*Department of Information and Technology, Hebei Vocational College for Correctional Police, No. 515, Zhonghua North Street, Handan 056038, China*
[3]*Northeastern University at Qinhuangdao, No. 143, Taishan Road, Qinhuangdao 066004, China*

Correspondence should be addressed to Fei Zhao; zhaofei@neuq.edu.cn

Received 14 January 2016; Accepted 12 April 2016

Academic Editor: Chengyan Yue

We present the vehicle routing problem with potential demands and time windows (VRP-PDTW), which is a variation of the classical VRP. A homogenous fleet of vehicles originated in a central depot serves customers with soft time windows and deliveries from/to their locations, and split delivery is considered. Also, besides the initial demand in the order contract, the potential demand caused by conformity consuming behavior is also integrated and modeled in our problem. The objective of minimizing the cost traveled by the vehicles and penalized cost due to violating time windows is then constructed. We propose a heuristics-based parthenogenetic algorithm (HPGA) for successfully solving optimal solutions to the problem, in which heuristics is introduced to generate the initial solution. Computational experiments are reported for instances and the proposed algorithm is compared with genetic algorithm (GA) and heuristics-based genetic algorithm (HGA) from the literature. The comparison results show that our algorithm is quite competitive by considering the quality of solutions and computation time.

1. Introduction

The vehicle routing problem (VRP) is a well-known combinatorial optimization problem arising in transportation logistic, which was first introduced by Dantzig and Ramser in their study on the truck dispatching problem [1]. The classical VRP (CVRP) is generally defined by a graph $G = (V, E)$, where $V = \{0, 1, \ldots, N\}$ is the set of vertices comprised of a central depot 0 and N customers and $E = \{(i, j) : i, j \in V, i \neq j\}$ is the arc set. The CVRP aims to determine a set of routes for K identical and capacitated vehicles by minimizing the total routing cost (distance, time, etc.), such that each customer is visited exactly once. It mainly focuses on deterministic demand, which means customer demands are fixed and known in advance, and there is no requirement for the service time.

Numerous variants of CVRP have been put forward and extensively studied which capture the complications of real-world problems [2–4], for example, VRP with time windows (VRPTW) [5–21], in which customers have to be visited within a predefined time interval which can be described by the earliest arrival time and the latest arrival; VRP with multiple depots (VRPMD) [20, 22–25]; VRP with simultaneous delivery and pickup (VRPSDP) [10, 26], where customers require simultaneous pickup and delivery service; VRP with stochastic demands (VRPSD) [19, 27–35], which involves demands that are random variables with known distributions; VRP with split delivery, where the demand of a customer can be split and delivered by multiple vehicles [36]; and the heterogeneous fleet VRP with several vehicle types [37].

Besides, lots of surveys were dedicated to studying solution methods of VRPs which can be classified into exact algorithms, heuristics, and metaheuristics [3, 4, 22]. We refer the interested reader to the most common exact algorithms, namely, a branch-and-cut by Hà et al. [38]; a branch-and-price by Christiansen and Lysgaard [28] and Hadjar and Soumis [5]; a branch-cut-and-price by Gauvin et al. [29] and

Lysgaard and Wøhlk [39]; Lagrangian relaxation by Kallehauge et al. [6]; Dantzig-Wolfe decomposition by Qureshi et al. [7] and Chabrier [40]; and a branch-and-bound by Carpaneto et al. [23]. However, exact algorithms are usually designed for the static and deterministic VRPs, where all input is known and keeps unchanged once vehicles are in service. Since VRP is a typical NP-hard problem, many heuristics-based studies are proposed for various VRPs. The work of Pang [8] and Figliozzi [9] applied the improved route construction heuristics to solve VRPTM, and Liu et al. [24] and Z. Wang and Z. Wang [41] raised the novel heuristics based on the traditional two-phase method and the results demonstrated the effectiveness of the proposed heuristics. Most of the recent researches for VRPs paid extensive attention to the development of metaheuristics. This can be summarized as (1) genetic algorithms, (2) Ant Colony Optimization (ACO), (3) simulated annealing (SA), (4) Variable Neighborhood Search (VNS), (5) tabu search (TS), (6) local search algorithm, (7) Artificial Bee Colony (ABC), and (8) Particle Swarm Optimization (PSO). For further details on VRP metaheuristics, we refer the reader to the work of [3, 22, 42].

Conformity behavior was firstly introduced by M. Sherif, a social psychologist, and proposed to study the how peers' behavior affects an individual's behavior in a group [43]. Many studies have focused on herd behavior [44, 45]. For example, Jiang and Du analyzed the homebuyers' conformity behavior [46] and Cipriani and Guarino studied herd behavior in financial market [47]. However, there is currently very little methodology in the literature to model and optimize the vehicle routing problem with conformity behavior. Our VRP is motivated by a real-life problem, where the depot receives orders from a set of customers and then assigns vehicles to deliver products after signing the contract. The customer demand is generally confirmed in the contract, which is confirmed as the initial demand. However, once a customer knows the initial demand of other customers, they could make some adjustments for their initial demand due to conformity behavior. Thus, customer's ordering behavior is prone to be affected by other customers [47], which is named as conformity consuming behavior in this paper. This may lead to generation of new orders in the delivery, which is consistent with the rush order; the reduction or canceling of the initial demand is also possible. In the following sections, we call it the potential demand to differ it from the initial demand in the contract. Once the potential demand occurs, the manager has to reschedule the routes of vehicles. Henceforth, how to assign vehicles and give the routing plan is an issue that needs to be solved, which is what our study concerned. To the best of our knowledge, there is no literature devoted to the VRP with potential demands and time windows.

Our paper has three main contributions, which are as follows: the first is the establishment of the vehicle routing problem with potential demands and time windows, in which the concept of conformity consuming behavior is taken into consideration and the route of vehicles is constructed by multiple cycles; the second is the use of parthenogenetic algorithm for saving the computation time caused by searching/repairing the invalid solutions after crossover operator in GA, in which a novel coding technique using 2D vector is proposed to address our problem; and then heuristics is designed for the generation of the initial population for the parthenogenetic algorithm (PGA) such that a hybrid algorithm is proposed, which is our third contribution. The remaining outline of this work is organized as follows. Section 2 reviews the literature; Section 3 introduces the modeling assumptions and the mathematic model of the VRP-PDTW; in Sections 4 and 5, HPGA by combining parthenogenetic algorithm with heuristics is designed and developed to solve this problem; this is followed by computational results and comparison in Section 6; Section 7 concludes the paper and the further research is also given.

2. Literature Review

In the deterministic capacitated vehicle routing problem (DCVPR), as the variation of the CVRP, the demand of each customer is deterministic and there is no limit of time windows for visiting customers [39]. The objective in the work of Lysgaard and Wøhlk is to minimize the sum of arrival times at the customer which is referred to as cumulative VRP, and then a set partitioning formulation and a vehicle flow formulation are proposed and combined to apply into a branch-and-cut-and-price algorithm [39]. De Oliveira et al. considered a multidepot DCVPR (MD-DCVRP) where the maximum route duration time is set, and a coevolutionary approach is introduced through decomposing the MD-DCVRP into a number of single-depot DCVRPs [25]. Tasan and Gen developed a genetic algorithm based approach for the DCVRP with simultaneous delivery and pickup (DCVRPSDP) [26]. Wang and Chen modeled for the DCVRPSDP with time windows and in order to solve this problem, the cheapest insertion method is introduced into the genetic algorithm to speed up the solution procedure [10]. Kallehauge [11], Baños et al. [12], Qi et al. [13], Yu et al. [14], Vidal et al. [15], and Koç et al. [16] considered the case of the DCVPR with hard time windows, in which the vehicles have to arrive at the customers before the latest arrival time; otherwise, the vehicles may wait at a customer location if they arrive earlier than the earliest arrival time. In the literature [14–16], neighborhood search is also introduced to enhance the local search ability. A Pareto-based hybrid algorithm that combines evolutionary computation and simulated annealing is reported by Baños et al. for solving the multiobjective formulations of the DCVRPTW [12]. The DCVPR with soft time windows is investigated by Ghoseiri and Ghannadpour [17], in which the immediate service from late arrival vehicles is accepted by customers but a penalized cost is imposed on the latter arrival vehicles. An adaptive genetic algorithm incorporating push-forward insertion heuristic and λ-interchange mechanism is then constructed to search for better routing solutions. Related works in the soft time windows area are also proposed by Lau et al. [18] and Qureshi et al. [7], in which an exact algorithm arising from the Dantzig-Wolfe decomposition method and a tabu search approach are presented, respectively.

However, in most real applications, one or more parameters (time, demand, etc.) of the CVRP tend to be random or stochastic, giving rise to the stochastic vehicle routing problem (SVRP). We found that most researches mainly focused on VRPSD, where customer demands are known only when the vehicle arrives to location [28–35]. In the following, some studies will be mentioned. Christiansen and Lysgaard developed a branch-and-price algorithm [28] for the minimal distribution cost. Marinakis et al. [30] proposed a new hybrid algorithmic approach, which combined the PSO algorithm with the 2-opt and 3-opt local search algorithms and with the path relinking strategy. A paired cooperative optimization strategy is developed for the VRPSD with split delivery [31]. But, the constraint for the visiting time of customers is not considered in these studies [28–34]. However, to be closer to the practice, Tan et al. [35] and Lei et al. [19] applied a soft time windows constraint for the VRPSD, which means that vehicles may be violated if the arrival time at the location of customers is beyond time windows. Tan et al. [35] defined the time constrained VRPSD as a multiobjective optimization problem such that an evolutionary algorithm with enhanced genetic operators, Pareto fitness ranking, and local search heuristics is presented. The simulation results illustrated that the solutions obtained from the proposed algorithm are robust to the stochastic nature of the problem. The VRPSD is modeled as a stochastic program with recourse by Lei et al. and for its solution an adaptive large neighborhood search heuristic is designed, in which several remove-insert operations are combined to efficiently explore the solution space [19].

There are some researches on the fuzzy VRP (FVRP) which arises whenever some elements of the problem are uncertain, subjective, and vague. Common examples are fuzzy travel times and fuzzy demands [20, 21, 48]. Generally, the fuzzy variable is firstly handled using the possibility measure theory. For example, the fuzzy demand on each customer and the travel times between nodes and time windows are often described as triangular fuzzy numbers. Tang et al. considered the vehicle routing problem with fuzzy time windows, in which the deviation of the servicing time from the customer's time window is assumed to be associated with the level of customer's satisfaction, and proposed a two-stage algorithm to obtain a Pareto solution [20]. Asl et al. extended work [20] to multidepot and multilevel VRP with the fuzzy time windows and sequentially a three-section algorithm is proposed [21]. Cao and Lai designed a fuzzy chance constrained program model based on fuzzy credibility theory and a hybrid intelligent algorithm with stochastic simulation and differential evolution algorithm is presented to solve this problem [48].

3. Model Formulation

3.1. Problem Description and Assumptions. We focused on the vehicle routing problem with a central depot, denoted by "0," and a set of customers $\{1, 2, \ldots, N\}$ to be serviced, which is defined by a directed graph $G = (V, E)$. $V = \{0, 1, 2, \ldots, N\}$ is the node set including the central depot "0" and N customers and E is the set of arcs (i, j), $i, j \in V$. Arc (i, j) represents the possibility of traveling from i to j with an associated distance, duration, or cost. The customers are serviced by κ homogeneous vehicles with a limited capacity $Q_{\max}$ and the same speed, and all vehicles are located at the central depot such that they have to depart from the depot and return to the depot after visiting some customers. In the following, some assumptions are given for the modeling.

(1) Consider a single type of products stored at the central depot, which is delivered by vehicles and ordered by customers.

(2) The duration for a vehicle from departing from depot to ending at depot after going through several customers is defined as a cycle.

(3) The split delivery for the customer demands is allowed; thus each customer may be visited more than once and serviced by different vehicles. This implies that the route of a vehicle may be subject to multiple cycles. Taking the route of a vehicle "$0 \to 3 \to 2 \to 6 \to 0 \to 1 \to 4 \to 0$" as an example, there are two cycles "$0 \to 3 \to 2 \to 6 \to 0$" and "$0 \to 1 \to 4 \to 0$." Particularly, the number of the cycle in a route of vehicle k ($k = 1, 2, \ldots, \kappa$) is represented by the maximum delivery frequency L_k in our following model.

(4) The initial demand of customer $j, j \in V \setminus 0$, is d_j ($d_j \geq 0$), but his/her potential demand d_{pj} may occur before a vehicle visits him/her for the first time. d_{pj} relies on the difference between the initial demand of customer j and that of other customers' h ($\forall h \in H, H = V \setminus j$). Thus, the potential demand is defined as $d_{pj} = G_{h \in H}(d_j, d_h)$, in which it has a wide choice of function forms for G including linear function, logarithmic function, and semilogarithmic function.

(5) Time windows $[E_i, L_i]$, in which we let E_i denote the earliest arrival time and let L_i be the latest arrival time that customer i can be serviced by a vehicle, are imposed on customer i. These are soft bounds that can be violated, so the penalty cost incurred by waiting for servicing and late servicing at customers arises. However, there is no constraint of time windows for the central depot.

(6) The service time at node i is f_i, and the travel time t_{ij} from customer i to customer j with the travel cost per unit time c is associated with each arc in the directed graph.

(7) The product is always available for the central depot, that is, no shortage.

3.2. Mathematical Model. The objective function of the proposed VRP-PDTW is to minimize the total cost including the travel cost caused by the delivery of vehicles and the penalty cost due to violating the time windows, which is given by the following:

$$\text{minimize} \quad \text{TC} = \sum_{k=1}^{K}\sum_{l=1}^{L_k}\sum_{i=0}^{N}\sum_{j=0}^{N} x_{ijl}^{k} \cdot t_{ij} \cdot c_{ij} + \sum_{k=1}^{K}\sum_{l=1}^{L_k}\sum_{j=1}^{N} \left[p \max\left(0, E_j - t_{rjl}^{k}\right) + q \max\left(0, t_{rjl}^{k} - L_j\right)\right] \tag{1}$$

$$\text{subject to} \quad \sum_{j=1}^{N} x_{0jl}^{k} = 1, \quad \forall k \in \kappa,\ \forall l \in \Omega_k \tag{2}$$

$$\sum_{i=1}^{N} x_{i0l}^{k} = 1, \quad \forall k \in \kappa,\ \forall l \in \Omega_k \tag{3}$$

$$\sum_{i=0(i\neq j)}^{N} x_{ijl}^{k} = 1, \quad \forall k \in \kappa,\ \forall l \in \Omega_k,\ \forall j \in V \setminus 0 \tag{4}$$

$$\sum_{i\in V} x_{ihl}^{k} - \sum_{j\in V\setminus i} x_{hjl}^{k} = 0, \quad \forall k \in \kappa,\ \forall l \in \Omega_k,\ h \in V \setminus (i \cup j) \tag{5}$$

$$\sum_{i=0}^{N}\sum_{j=1}^{N} x_{ijl}^{k} \cdot b_{jl}^{k} \leq Q_{\max}, \quad \forall k \in \kappa,\ \forall l \in \Omega_k \tag{6}$$

$$\sum_{k=1}^{K}\sum_{l=1}^{L_k}\sum_{i=0}^{N} x_{ijl}^{k} \cdot b_{jl}^{k} = d_j + d_{pj}, \quad \forall j \in V \setminus 0 \tag{7}$$

$$t_{ril}^{k} + f_i + t_{ij} \leq \sum_{j\in V\setminus i} x_{ijl}^{k} \cdot t_{rjl}^{k}, \quad \forall k \in \kappa,\ \forall l \in \Omega_k,\ \forall i \in V \setminus 0, \tag{8}$$

where κ and Ω_k are set of all vehicles and the number of cycles for the route of vehicle k, respectively. Let t_{rjl}^{k} denote the arrival time of a vehicle k in the lth cycle at customer j, and $\max(a, b)$ returns the maximum value from a and b. There are two decision variables, that is, x_{ijl}^{k} $(i \neq j; \forall i, j \in V)$ and b_{jl}^{k}. $x_{ijl}^{k} = 1$ if vehicle k travels from customer i to customer j in the lth cycle; otherwise, $x_{ijl}^{k} = 0$; and b_{jl}^{k} is the demand of customer j which is satisfied by vehicle k in the lth cycle.

In the objective function (1), p and q are the penalty cost per unit time as related to violating time windows, respectively. Constraints (2) and (3) represent that a vehicle k starts its route at the depot, visits a number of customers, and returns to the depot in the lth cycle. Constraint (4) imposes the notion that, in the lth cycle, a vehicle k visits customer i only once. The sequence of a route "$i \rightarrow h \rightarrow j$" for vehicle k in the lth cycle is imposed by constraint (5). The demand of all customers met by the vehicle k in the lth cycle does not exceed the maximum capacity, as mentioned in constraint (6). Constraint (7) indicates that both the initial and potential demands of customer j are satisfied. The time constraint is controlled by (8) to ensure the time of a vehicle k in the lth cycle reached to customer j must be not earlier than $t_{ril}^{k} + f_i + t_{ij}$, in which the equality holds if t_{ril}^{k} fails to the interval $[E_i, +\infty)$; otherwise, $t_{ril}^{k} + f_i + t_{ij} < \sum_{j\in V\setminus i} x_{ijl}^{k} \cdot t_{rjl}^{k}$.

The VRP with potential demands and time windows is a typical problem in combinatorial optimization, which is also an NP-hard problem [49–53]. It is theoretically and practically impossible to use exact methods for the large-scale VRP; thus a hybrid parthenogenetic algorithm mixed with heuristics is designed in our view to find approximate solution, which is described in the following section.

4. Framework of the Heuristics-Based Parthenogenetic Algorithm

The genetic algorithm based on the drawing concept of evolution, survival of the fittest, which was proposed by Holland in 1975, has been extensively adapted to solve the NP-hard problems. It and the improved algorithms based on GA have been widely implemented on VRP and its variants over the past decades (see [10, 15–17, 26]). In GA, the candidate solutions of solving problem are represented as a population of chromosomes (individuals) and each chromosome is comprised of a string of genes. The search starts from the initial population, which is either randomly generated or generated using heuristics, and runs in generations where the individuals evaluated by the fitness function keep improving by applying genetic operators. Commonly, genetic algorithm iterates over generations till a predefined stop criterion is met. The steps of the genetic algorithm are summarized as follows.

Step 1 (problem representation). In GA the problem is in the form of a genotype representation by using a chromosome regarding a solution. The genes in chromosomes can be encoded in binary for bit string representation or represented as a string of real numbers.

Step 2 (initialization). Randomly generate the initial population of solutions with the effective population size to cover the search space.

Step 3 (fitness evaluation). The fitness value of each chromosome in the population is calculated and ranked by the fitness function.

Step 4 (genetic operators). Evolution mechanisms subject to selection, crossover, and mutation are used to create fitter individuals for the next generation.

Step 5. Repeat Steps 3 and 4 until terminating condition is met, in which the stop criterion can be either the maximum number of generations or a satisfactory fitness level.

The crossover operator, as the major genetic operator, is used to generate new offspring by mixing two parents such that some characteristics from the parents can be inherited by newer individuals. The traditional crossover operators, that is, one-point crossover, two-point crossover, scattered crossover, and a partially mapped crossover, can be found in the VRPs, and some improved crossover operators are also proposed depending on the various problems [3, 4]. The crossover operator is intended to generate solutions in the whole search space. However, how to design a feasible crossover operator for a special VRP is still a key challenge since invalid offspring could be obtained through crossover and it is time-consuming to repair them.

Parthenogenetic algorithm is fortunately proposed [54] to deal with the above issue by the removal of the crossover operator, which greatly simplifies the procedure of GA and improves the effectiveness and performance of searching optimal solutions. This is because in PGA there is the shift operator only performed on a single chromosome. It avoids the offspring obtained by crossover operator jump to the invalid solution area. As we mentioned in the previous section, that is, (1), it is impossible or difficult to create the initial solution randomly that constitutes the initial population of our problem in PGA. A heuristic algorithm is therefore designed to construct the initial population, which will be shown in the following section. Figure 1 shows the framework of the heuristics-based parthenogenetic algorithm.

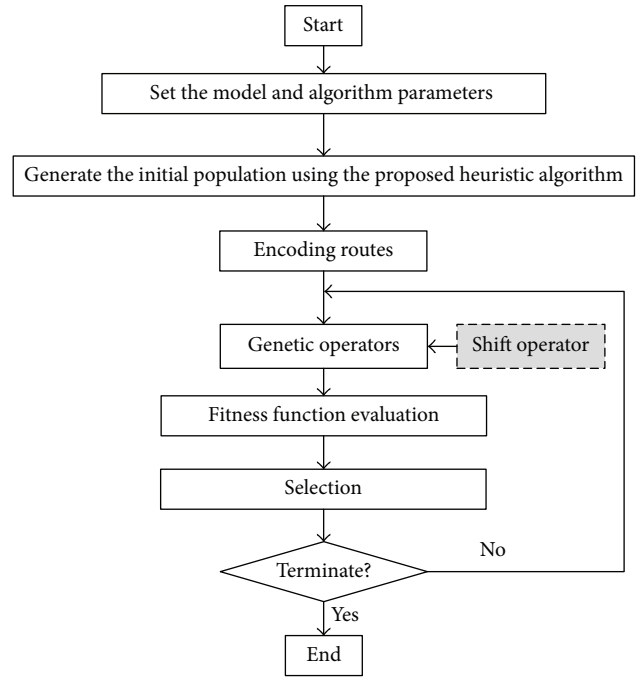


Figure 1: Heuristics-based parthenogenetic algorithm procedure for the VRP-PDTW.

5. Design of the Heuristics-Based Parthenogenetic Algorithm

5.1. Heuristic for Initial Solutions. From the problem description mentioned in Section 3, we design a heuristic algorithm for the generation of initial solutions, as described in Figure 2, in which two rules are firstly developed to perform search. For more details, the procedure can be found in Appendix.

Rule 1. For a set of customers with the initial demand $d_i > 0$, a vehicle k chooses the nearest customer to provide service. However, the customer with the largest demand is visited by vehicle prior to others once there are two or more customers with the same traveling distance.

Table 1: An initial solution.

Vehicle	Route
1	$0 \to 4 \to 9 \to 7 \to 0 \to 2 \to 6 \to 0$
2	$0 \to 5 \to 7 \to 1 \to 4 \to 0$
3	$0 \to 2 \to 9 \to 0 \to 3 \to 8 \to 0 \to 1 \to 6 \to 0$

Rule 2. The initial demand for each customer is always considered to be preferentially satisfied against his potential demand.

The quality of the initial population size is also a key issue since it directly affects the performance and efficiency of HPGA. It is commonly preset as an integer depending on the problem to be solved. Without exception, we give the initial population size M as an integer which is associated with the number of vehicles and customers, that is, $M = \varphi(K, N)$.

5.2. Solution Coding. The important step for HPGA is to determine the solution coding technique to represent chromosome. Binary representation, permutation representation, integer representation, and sector representation are commonly applied in the literature on VRPs. Assume that if we have obtained an initial solution for our problem, which is described in Table 1, it is impossible to code it using the coding approaches mentioned above. It is because for a vehicle there may be multiple cycles, rather than delivery only once. So, we propose here a novel permutation representation based on a 2D vector $[k, i]$, where k and i are the indices of vehicle and node, respectively, such that $[k, i]$ is defined as the vehicle k visits node i. The generalizable chromosome representation is then described in Figure 3.

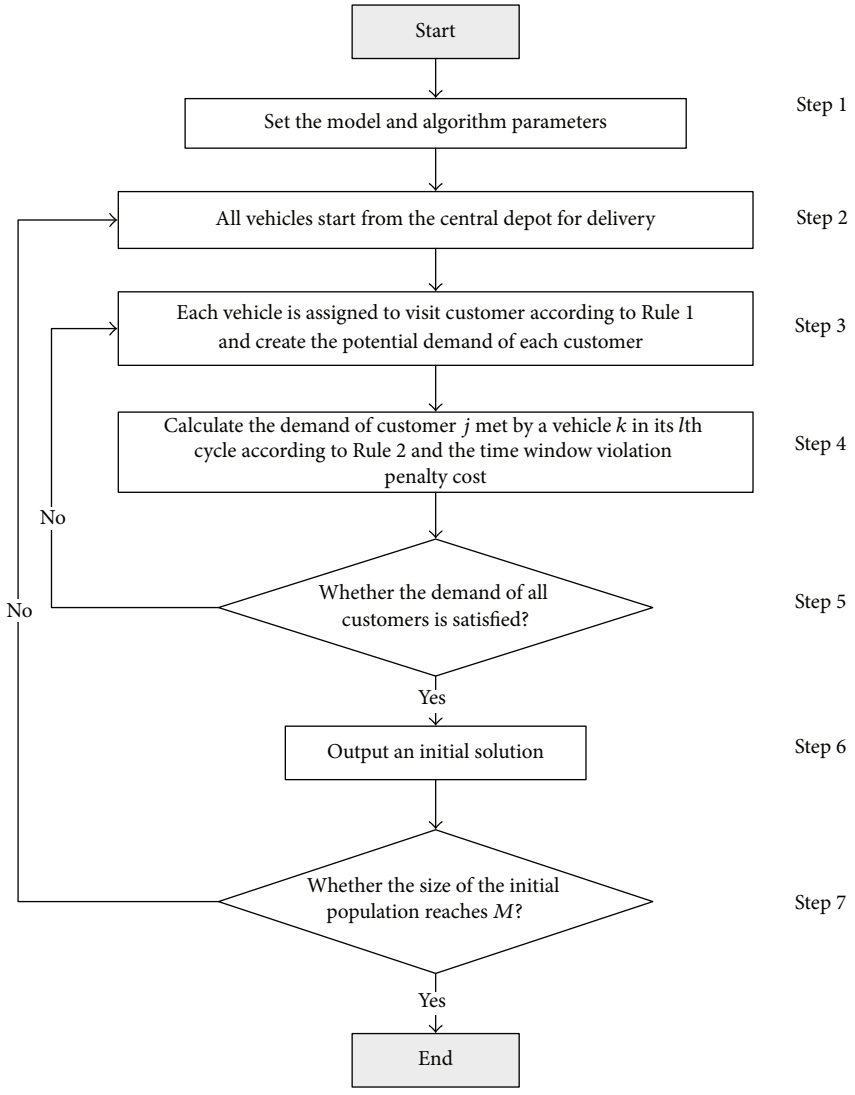


Figure 2: Heuristics algorithm for the generation of initial solutions.

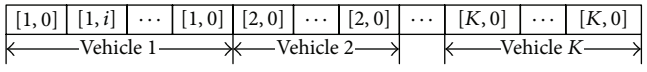


Figure 3: Chromosome representation.

It can be seen from Figure 3 that each vehicle k ($\forall k \in \kappa$) leaves the central depot 0 and comes back to the depot through some customers after completion of its delivery. Using the solution coding technique in Figure 3, the initial solution in Table 1 can be expressed as

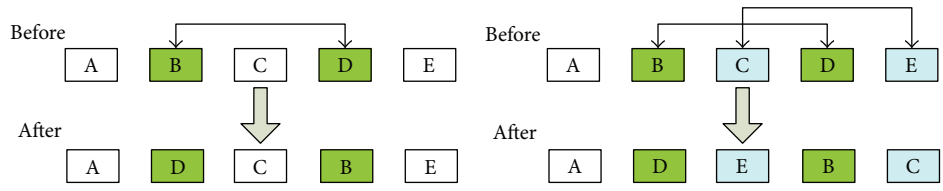
$$\left\{ \begin{array}{c} [1,0]\,[1,4]\,[1,9]\,[1,7]\,[1,0]\,[1,2]\,[1,6]\,[1,0] \\ [2,0]\,[2,5]\,[2,7]\,[2,1]\,[2,4]\,[2,0] \\ [3,0]\,[3,2]\,[3,9]\,[3,0]\,[3,3]\,[3,8]\,[3,0]\,[3,1]\,[3,6]\,[3,0] \end{array} \right\}. \quad (9)$$

Obviously, there are multiple cycles for vehicles 1 and 3.

5.3. Shift Operator. Different from genetic operators in GA, shift operator is usually applied to generate offspring in PGA, as defined in the following.

Definition 1. Shift operator exchanges some genes with a certain probability and works on a chromosome, in which the location of genes exchanged is randomly chosen.

Considering the length difference of each cycle in a vehicle route, that is, the number of customers C_N visited by vehicle in a cycle, different scenarios could be applied based on the int $[C_N/2]$-point exchange policy, in which int $[C_N/2]$ rounds a number down to a nearest integer. It implies that there are two different shift operators subject to single-point exchange and multipoint exchange, as shown in Figures 4(a) and 4(b), respectively. Single-point exchange corresponds to the location of a pair of genes that mutually interchange, referred to as genes B and D in Figure 4(a).



(a) Single-point exchange (b) Two-point exchange

Figure 4: Shift operator.

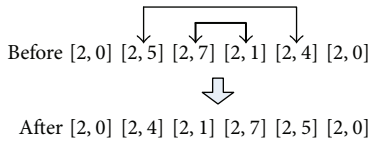


Figure 5: Process of shift operator.

However, multipoint exchange is defined as two or more pairs of genes are interchanged. From Figure 4(b), it can be obviously seen that there are two pairs of genes exchanged, that is, {B, D} and {C, E}.

Shift operator in Definition 1 is utilized to produce newer individual, but in our context, it is restricted to a cycle of each vehicle. This is because exchange across cycles will result in invalid individuals and cost more repair time. For the chromosome in Table 1, Figure 5 gives the process of shift operator that is used to the route of vehicle 2. Similarly, it is also implemented on the routes of vehicles 1 and 3.

5.4. Evaluation and Selection. Using the objective function shown in (1), the fitness is defined as the reciprocal of TC; that is, Fitness = 1/TC. Thus, the fitness value of each individual from the initial population and its offspring can be calculated to scale the performance of each individual. The roulette wheel selection is used to select fitter individuals, as defined in (10). It implies that the probability of individuals with lower cost to be selected for the next generation is higher. Consider

$$P\,(\text{individual } z \text{ to be selected}) = \frac{\text{fitness}\,(z)}{\sum_{z=1}^{2M} \text{fitness}\,(z)}, \quad (10)$$

where $2M$ is the number of individuals to be selected.

6. Experimental Results and Comparisons

We coded the above procedure using MATLAB R2012b and ran it on a PC with 3.60 GHz CPU and 1 T memory, and all computing times are reported in CPU seconds. This section presents comparison results and analyses for results carried out to investigate the effectiveness of the proposed HPGA through a number of computational experiments. We first apply the proposed HPGA to a real-world instance of a distribution company in Handan to verify its applicability in real life, and then simulation experiments are conducted to assess the performance of HPGA by comparing with the GA and HGA.

6.1. Parameter Settings. There are 12 customer nodes and they are serviced by 4 homogeneous vehicles which are located at the central depot and the vehicle capacity is 80 units. The traveling time t_{ij} between customers is set in terms of the corresponding Euclidean distance between them, and some parameters values used in the mathematical model are also given in Table 2. The potential demand d_{pi} in Table 2 is obtained by simplifying assumption (4), in which only one related customer of i is considered in this simulation and a positive/negative number indicates the increase/decrease of the initial demand, respectively. The following cost parameter values are used in our experiments: $c = 100$, $p = 100$, and $q = 150$.

Also, the proposed algorithm HPGA is applied with population size $M = C_{N'}^{K}/2$, where we let N' denote the number of customers with nonzero initial demand, that is, $d_i > 0$. In our experimental studies, the run is terminated if the fitness is not improved; that is, the solution value is within a preset percentage of the optimum 10^{-4}. The shift probability on shift operator is set to 0.7 by a pilot study.

6.2. Computation Results. Depending on the parameters in Section 6.1 and the proposed algorithm, the optimal routes and work time for vehicles are obtained, as shown in Table 3. More details on the arrival/departure time (AT, DT) of vehicles at each node are given in Table 4, and the corresponding quantity unloaded at nodes is also output. From Tables 3 and 4, it can be seen that vehicles 1 and 3 firstly end their first cycle at 54 and 58, respectively, and then start their second cycle to service customers after uploading products at the central depot "0." Furthermore, Figure 6 shows the best objective function value through the evolution process in this instance, which reveals the optimal solution yields when the proposed algorithm runs for 250 CPU seconds approximately. However, it should be noted that different results could be obtained for other parameters or the regeneration of potential demands.

6.3. Comparative Analyses. We here perform comparative analyses of the results of the HPGA with the GA and HGA to evaluate the efficiency between them, using the previous modeling parameter settings. Heuristics proposed in Section 5.1 is further integrated into the GA to create the initial solution. We expect HGA can save computational time since it is not easy to obtain feasible solution by random generation for our VRP with potential demands and time windows. Note that in GA and heuristics-based GA some

Table 2: Parameter values.

Customer i	Customer j	d_i	d_{pi}	$[E_i, L_i]$	f_i	t_{ij} 0	1	2	3	4	5	6	7	8	9	10	11	12
0	—	—	—	—	2	0	25	15	30	24	10	26	11	12	35	28	20	18
1	4	40	−4	[35, 50]	4	—	0	12	8	10	15	7	5	11	9	15	12	10
2	6	20	+35	[20, 80]	3	—	—	0	23	9	12	14	13	15	10	9	21	13
3	5	0	+16	[40, 110]	4	—	—	—	0	11	14	13	13	10	14	10	13	7
4	11	35	+2	[30, 90]	2	—	—	—	—	0	9	6	4	11	17	15	8	20
5	3	25	−14	[25, 60]	1	—	—	—	—	—	0	8	25	10	18	12	10	14
6	12	70	−5	[15, 70]	5	—	—	—	—	—	—	0	6	8	10	17	12	15
7	1	0	+30	[40, 90]	3	—	—	—	—	—	—	—	0	17	13	5	18	13
8	9	45	−14	[20, 80]	3	—	—	—	—	—	—	—	—	0	10	9	16	9
9	7	0	0	[30, 95]	4	—	—	—	—	—	—	—	—	—	0	7	12	8
10	2	50	−7	[25, 85]	2	—	—	—	—	—	—	—	—	—	—	0	10	9
11	10	40	+6	[30, 80]	1	—	—	—	—	—	—	—	—	—	—	—	0	12
12	8	60	−10	[20, 115]	4	—	—	—	—	—	—	—	—	—	—	—	—	0

Table 3: Optimal routes of vehicles.

Vehicle	Route	Work time
1	$0 \to 6 \to 7 \to 0 \to 5 \to 8 \to 10 \to 0$	119
2	$0 \to 12 \to 3 \to 1 \to 0$	81
3	$0 \to 2 \to 4 \to 0 \to 1 \to 10 \to 0$	134
4	$0 \to 11 \to 4 \to 7 \to 1 \to 0$	82

Table 4: Timetable of routes.

(a) Timetable of route 1

Node	0	6	7	0	5	8	10	0
AT	—	26	37	54	66	77	89	119
DT	0	31	43	56	67	80	91	—
b_{jl}^k	—	65	15	—	11	31	38	—

(b) Timetable of route 2

Node	0	12	3	1	0
AT	—	18	31	52	81
DT	0	24	44	56	—
b_{jl}^k	—	50	16	14	—

(c) Timetable of route 3

Node	0	2	4	0	1	10	0
AT	—	15	32	58	85	104	134
DT	0	23	34	60	89	106	—
b_{jl}^k	—	55	25	—	15	5	—

(d) Timetable of route 4

Node	0	11	4	7	1	0
AT	—	20	39	45	53	82
DT	0	31	41	48	57	—
b_{jl}^k	—	46	12	15	7	—

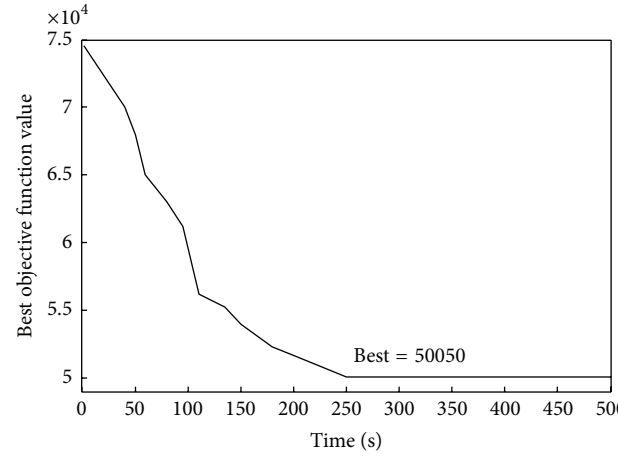


Figure 6: Best objective function values through the evolution process.

initial parameters including population size, the crossover and mutation probabilities, and the termination condition should be first set. To remove any factor that could impact the performance, the population size and the termination condition are the same in GA, heuristics-based GA, and HPGA. Since what is concerned is the effect of the crossover probability in GA and HGA on the performance of solution, the crossover probability p_c varies from 0 to 1 with a step of 0.2. However, the mutation probability p_m is set to 0.3 in our instances by using the proposed method by Greenwell et al. [55].

The comparison results are summarized in Table 5, where the proposed algorithm is tested in six instances and in the first column each instance is numbered. These instances are available by randomly generating potential demands, keeping other parameters unchanged. The efficiency of the algorithm is measured by the best objective function value and the computation time, which are denoted by Obj and Time in

Table 5: Comparison results.

Instance	p_c	GA		HGA		HPGA		Gap	
		Obj	Time	Obj	Time	Obj	Time	GA	HGA
1	0.2	71584	8059	70152	3750	58291	271	0.23	0.20
	0.4	70514	5041	69841	2351			0.21	0.20
	0.6	**67415**	2580	**66410**	1225			**0.16**	**0.14**
	0.8	69045	1102	68125	952			0.18	0.17
2	0.2	71201	—	68410	4059	56124	258	0.27	0.22
	0.4	64151	4245	65124	2584			0.14	0.16
	0.6	61490	1751	60155	1354			0.10	0.07
	0.8	**59510**	1105	**57742**	852			**0.06**	**0.03**
3	0.2	80159	8156	68411	4594	61025	312	0.31	0.12
	0.4	75512	5894	75142	3154			0.24	0.23
	0.6	64589	3015	63142	1981			0.06	0.03
	0.8	**62015**	1705	**62144**	1101			**0.02**	**0.02**
4	0.2	68150	9425	65124	5154	55410	215	0.23	0.18
	0.4	65058	4059	62042	2129			0.17	0.12
	0.6	**61254**	1759	61458	905			**0.11**	0.11
	0.8	62548	851	**60458**	638			0.13	**0.09**

Table 5, respectively. Also, the gap is calculated in terms of the relative deviation from the best known solution to measure the performance of solutions with the following equation:

$$\text{gap}_{\#} = \frac{\text{Obj}_{\#} - \text{Obj}_{\text{HPGA}}}{\text{Obj}_{\text{HPGA}}}, \tag{11}$$

where $\text{Obj}_{\#}$ is the value of one of the other methods (# is either GA or HGA) and Obj_{HPGA} is the solution value of HPGA. In Table 5, the optimal solutions for each instance are denoted with bold face letters, and the corresponding gap values are also marked in bold.

As Table 5 demonstrates, our proposed algorithm HPGA is more effective and very competitive compared to GA and HGA. It is noted that, for each instance, there is only one best solution in HPGA. This can be explained as the crossover operator is not used in HPGA, so the change of p_c does not work on HPGA. We can see that the computational time of GA and HGA decreases with the increase of p_c, which results from the fact that a larger p_c speeds up the convergence. For the scenario $p_c = 0.2$ in instance 2, the computation time of GA is represented as "—," which implies that it runs beyond the preset time threshold, 3 hours. Obviously, HPGA searches the optimal solution with the smallest time while GA costs the longest time, HGA between them. This is probably because heuristics improves the search speed of HGA and HPGA by generating the reasonable initial solutions as soon as possible. It may be also because the crossover operator in GA and HGA increases the time of searching the optimal solution and repairing the invalid solution.

In Table 5, it can be concluded from the results of Obj and gap that the quality of our proposed algorithm outperforms those of GA and HGA, which is what we expected. The gap values obtained from these instances are positive, varying between 0.31 and 0.02. This provides full and strong support to illustrate the solution quality of HPGA. Additionally, it can be found that the optimal solution for different instances is obtained with different crossover probabilities, which is out of the scope of this study. However, the proposed algorithm is not proved to be the most effective compared to all other algorithms ever published in the literature.

7. Conclusions

In this paper, we investigated the vehicle routing problem with potential demands and time windows inspired by a real-world problem, which considered the initial demand from customers and the possible potential demand induced by conformity consuming behavior. A fleet of vehicles originated in a central depot serves customers and the split delivery is allowed in our problem. The VRP-PDTW is formulated as a single objective model to minimize the total cost and find the optimal routes of vehicles. The VRP-PDTW is also an NP-hard problem; thus it is difficult and time-consuming to obtain the solutions using exact algorithms. Then, a heuristics-based parthenogenetic algorithm for solving VRP-PDTW was proposed in this study, in which the initial solution can be created using heuristics. The shift operator is only performed on a chromosome in our proposed algorithm, which differs from that in GA. Also, the coding technique based on a 2D vector is presented and applied in our algorithm. Finally, some experiments are carried out to evaluate the performance of the proposed algorithm. The results show that our proposed algorithm produces competitive solutions and is the fastest when compared with the best results of GA and HGA. The benefit of HPGA comes from its ability to produce the initial solution by heuristics and avoid the occurrence of invalid offspring which is generated by the crossover operator.

Step 1: Set the model and algorithm parameters $\kappa, c, p, q, t_{ij}, Q_{\max}, [E_i, L_i], d_i, d_{pj}, b_{il}^k$ and initialize some transient variables for heuristics $l = 1, s_i = 0, Q_{il}^k, t_{ril}^k, t_{dil}^k, d_i' = d_i, C_p = 0$.

We let $s_i = 0$ if customer i has not been visited by vehicles yet, otherwise, $s_i = 1$; Q_{il}^k is defined as the remaining quantity of products loaded by a vehicle k when it departs from customer i in its lth cycle; C_p is the cost due to the time windows is violated, that is, the second term of the objective function, (1).

Step 2: Initially, each vehicle k starts from the central depot "0" for delivery, and update $t_{dil}^k = 0$ and $Q_{il}^k = Q_{\max}$ for simplicity where $i = 0$ and $l = 1$;

Step 3: Vehicle k ($\forall k \in \kappa$) is assigned to service customer j in terms of Rule 1, and update $x_{ijl}^k = 1$ and $t_{rjl}^k = t_{dil}^k + t_{ij}$;

If $s_j = 0$

Generate the potential demand of customer j, update $s_j = 1$, and go to Step 4;

Else go to Step 4;

End;

Step 4: Calculate the decision variable b_{il}^k in terms of Rule 2, that is, the demand of customer j met by a vehicle k in its lth cycle, and the penalty cost induced by the time windows.

(a) **If** $d_{pj} > 0$

If $Q_{il}^k > d_j'$ then $b_{jl}^k = d_j' + \min\{d_{pj}, Q_{il}^k - d_j'\}$ and update

$d_j' = 0, d_{pj} = \max\{0, d_{pj} - (Q_{il}^k - d_j')\}, Q_{jl}^k = Q_{il}^k - b_{jl}^k$;

Else $b_{jl}^k = Q_{il}^k$ and update $d_j' = d_j' - Q_{il}^k, Q_{jl}^k = 0$;

End;

Else

If $Q_{il}^k > d_j' + d_{pj}$ then $b_{jl}^k = d_j' + d_{pj}$ and update $d_j' = 0, Q_{jl}^k = Q_{il}^k - b_{jl}^k$;

Else $b_{jl}^k = Q_{il}^k$ and update $d_j' = d_j' + d_{pj} - b_{jl}^k, Q_{jl}^k = 0$;

End;

End;

(b) It can be concluded that from assumption (5):

If $t_{rjl}^k < E_j$ then $t_{djl}^k = E_j + f_j, C_p = C_p + p \cdot (E_j - t_{rjl}^k)$;

Elseif $t_{rjl}^k \in [E_j, L_j]$ then $t_{djl}^k = t_{rjl}^k + f_j$;

Else $t_{djl}^k = t_{rjl}^k + f_j, C_p = C_p + q \cdot (t_{djl}^k - L_j)$;

End;

Step 5: Decision whether to continue to delivery or return to the central depot.

If $\sum_{\forall w \in V \setminus 0}[(d_w' \mid s_w = 0) + (d_w' + d_{pw} \mid s_w = 1 \bigcap d_{pw} > 0) + (d_w' \mid s_w = 1 \bigcap d_{pw} \le 0)] = 0$

$L_k = l, x_{j0L_k}^k = 1$, go to Step 6;

Else

If $Q_{jl}^k = 0$ then $x_{j0l}^k = 1, t_{r0l}^k = t_{djl}^k + t_{0j}$ and update $l = l + 1, Q_{0l}^k = Q_{\max}$;

Go to Step 3;

Else go to Step 3;

End;

End;

Step 6: Output the route of all vehicles according to $x_{ijl}^k = 1$, and the decision variable b_{jl}^k such that a solution for VRP is created.

Step 7: Repeat Steps 1–6 and terminate until M individuals are generated.

Algorithm 1

Our further studies will be focused on other real-world routing problems, for example, considering multiple types of products, the vehicle availability and the shock from environment; and relaxing the fixed number of routes will be further optimized.

Appendix

See Algorithm 1.

Competing Interests

The authors declare that they have no competing interests.

Acknowledgments

This research is partially supported by the Natural Science Foundation of Hebei Province (no. G2014402027), Science and Technology Project of Hebei Province (nos. 15454707D and 16457402D), Humanities and Social Sciences Research Project for Universities in Hebei Province (no. 2014068), Three Modernization and Collaborative Development Base Project of Hebei Province (no. SH2015QR01), and High Level Talent Support Project of Hebei Province (no. A201500112).

References

[1] G. B. Dantzig and J. H. Ramser, "The truck dispatching problem," *Management Science*, vol. 6, no. 1, pp. 80–91, 1959.

[2] V. Pillac, M. Gendreau, C. Guéret, and A. L. Medaglia, "A review of dynamic vehicle routing problems," *European Journal of Operational Research*, vol. 225, no. 1, pp. 1–11, 2013.

[3] B. Eksioglu, A. V. Vural, and A. Reisman, "The vehicle routing problem: a taxonomic review," *Computers & Industrial Engineering*, vol. 57, no. 4, pp. 1472–1483, 2009.

[4] J. F. Cordeau, G. Laporte, M. W. P. Savelsbergh et al., "Vehicle routing," in *Handbooks in Operations Research & Management Science*, vol. 14, chapter 6, pp. 367–428, Elsevier, 2007.

[5] A. Hadjar and F. Soumis, "Dynamic window reduction for the multiple depot vehicle scheduling problem with time windows," *Computers & Operations Research*, vol. 36, no. 7, pp. 2160–2172, 2009.

[6] B. Kallehauge, J. Larsen, and O. B. G. Madsen, "Lagrangian duality applied to the vehicle routing problem with time windows," *Computers & Operations Research*, vol. 33, no. 5, pp. 1464–1487, 2006.

[7] A. G. Qureshi, E. Taniguchi, and T. Yamada, "An exact solution approach for vehicle routing and scheduling problems with soft time windows," *Transportation Research Part E: Logistics & Transportation Review*, vol. 45, no. 6, pp. 960–977, 2009.

[8] K.-W. Pang, "An adaptive parallel route construction heuristic for the vehicle routing problem with time windows constraints," *Expert Systems with Applications*, vol. 38, no. 9, pp. 11939–11946, 2011.

[9] M. A. Figliozzi, "An iterative route construction and improvement algorithm for the vehicle routing problem with soft time windows," *Transportation Research Part C: Emerging Technologies*, vol. 18, no. 5, pp. 668–679, 2010.

[10] H.-F. Wang and Y.-Y. Chen, "A genetic algorithm for the simultaneous delivery and pickup problems with time window," *Computers & Industrial Engineering*, vol. 62, no. 1, pp. 84–95, 2012.

[11] B. Kallehauge, "Formulations and exact algorithms for the vehicle routing problem with time windows," *Computers & Operations Research*, vol. 35, no. 7, pp. 2307–2330, 2008.

[12] R. Baños, J. Ortega, C. Gil, A. L. Márquez, and F. De Toro, "A hybrid meta-heuristic for multi-objective vehicle routing problems with time windows," *Computers & Industrial Engineering*, vol. 65, no. 2, pp. 286–296, 2013.

[13] Y. Qi, Z. Hou, H. Li, J. Huang, and X. Li, "A decomposition based memetic algorithm for multi-objective vehicle routing problem with time windows," *Computers & Operations Research*, vol. 62, pp. 61–77, 2015.

[14] B. Yu, Z. Z. Yang, and B. Z. Yao, "A hybrid algorithm for vehicle routing problem with time windows," *Expert Systems with Applications*, vol. 38, no. 1, pp. 435–441, 2011.

[15] T. Vidal, T. G. Crainic, M. Gendreau, and C. Prins, "A hybrid genetic algorithm with adaptive diversity management for a large class of vehicle routing problems with time-windows," *Computers & Operations Research*, vol. 40, no. 1, pp. 475–489, 2013.

[16] Ç. Koç, T. Bektaş, O. Jabali, and G. Laporte, "A hybrid evolutionary algorithm for heterogeneous fleet vehicle routing problems with time windows," *Computers & Operations Research*, vol. 64, pp. 11–27, 2015.

[17] K. Ghoseiri and S. F. Ghannadpour, "Multi-objective vehicle routing problem with time windows using goal programming and genetic algorithm," *Applied Soft Computing*, vol. 10, no. 4, pp. 1096–1107, 2010.

[18] H. C. Lau, M. Sim, and K. M. Teo, "Vehicle routing problem with time windows and a limited number of vehicles," *European Journal of Operational Research*, vol. 148, no. 3, pp. 559–569, 2003.

[19] H. Lei, G. Laporte, and B. Guo, "The capacitated vehicle routing problem with stochastic demands and time windows," *Computers & Operations Research*, vol. 38, no. 12, pp. 1775–1783, 2011.

[20] J. Tang, Z. Pan, R. Y. K. Fung, and H. Lau, "Vehicle routing problem with fuzzy time windows," *Fuzzy Sets and Systems*, vol. 160, no. 5, pp. 683–695, 2009.

[21] V. M. Asl, S. A. Sadeghi, and S. Fathi, "A mathematical model and solving method for multi-depot and multi-level vehicle routing problem with fuzzy time windows," *Advances in Intelligent Transportation Systems*, vol. 1, no. 1, 2012.

[22] J. R. Montoya-Torres, J. L. Franco, S. N. Isaza, H. Felizzola Jiménez, and N. Herazo-Padilla, "A literature review on the vehicle routing problem with multiple depots," *Computers & Industrial Engineering*, vol. 79, pp. 115–129, 2015.

[23] G. Carpaneto, M. Dell'Amico, M. Fischetti, and P. Toth, "A branch and bound algorithm for the multiple depot vehicle scheduling problem," *Networks*, vol. 19, no. 5, pp. 531–548, 1989.

[24] R. Liu, Z. Jiang, R. Y. K. Fung, F. Chen, and X. Liu, "Two-phase heuristic algorithms for full truckloads multi-depot capacitated vehicle routing problem in carrier collaboration," *Computers & Operations Research*, vol. 37, no. 5, pp. 950–959, 2010.

[25] F. B. De Oliveira, R. Enayatifar, H. J. Sadaei, F. G. Guimarães, and J.-Y. Potvin, "A cooperative coevolutionary algorithm for the Multi-Depot Vehicle Routing Problem," *Expert Systems with Applications*, vol. 43, pp. 117–130, 2016.

[26] A. S. Tasan and M. Gen, "A genetic algorithm based approach to vehicle routing problem with simultaneous pick-up and deliveries," *Computers & Industrial Engineering*, vol. 62, no. 3, pp. 755–761, 2012.

[27] A. Gupta and R. Ravi, "Approximation algorithms for VRP with stochastic demands," *Operations Research*, vol. 60, no. 1, pp. 123–127, 2012.

[28] C. H. Christiansen and J. Lysgaard, "A branch-and-price algorithm for the capacitated vehicle routing problem with stochastic demands," *Operations Research Letters*, vol. 35, no. 6, pp. 773–781, 2007.

[29] C. Gauvin, G. Desaulniers, and M. Gendreau, "A branch-cut-and-price algorithm for the vehicle routing problem with stochastic demands," *Computers & Operations Research*, vol. 50, pp. 141–153, 2014.

[30] Y. Marinakis, G.-R. Iordanidou, and M. Marinaki, "Particle swarm optimization for the vehicle routing problem with stochastic demands," *Applied Soft Computing Journal*, vol. 13, no. 4, pp. 1693–1704, 2013.

[31] L. Zhu, L.-M. Rousseau, W. Rei, and B. Li, "Paired cooperative reoptimization strategy for the vehicle routing problem with stochastic demands," *Computers & Operations Research*, vol. 50, pp. 1–13, 2014.

[32] C. Novoa and R. Storer, "An approximate dynamic programming approach for the vehicle routing problem with stochastic demands," *European Journal of Operational Research*, vol. 196, no. 2, pp. 509–515, 2009.

[33] J. C. Goodson, "A priori policy evaluation and cyclic-order-based simulated annealing for the multi-compartment vehicle routing problem with stochastic demands," *European Journal of Operational Research*, vol. 241, no. 2, pp. 361–369, 2015.

[34] D. G. Pandelis, E. G. Kyriakidis, and T. D. Dimitrakos, "Single vehicle routing problems with a predefined customer sequence, compartmentalized load and stochastic demands," *European Journal of Operational Research*, vol. 217, no. 2, pp. 324–332, 2012.

[35] K. C. Tan, C. Y. Cheong, and C. K. Goh, "Solving multiobjective vehicle routing problem with stochastic demand via evolutionary computation," *European Journal of Operational Research*, vol. 177, no. 2, pp. 813–839, 2007.

[36] C. Archetti, M. G. Speranza, and A. Hertz, "A tabu search algorithm for the split delivery vehicle routing problem," *Transportation Science*, vol. 40, no. 1, pp. 64–73, 2006.

[37] M. Yousefikhoshbakht, F. Didehvar, and F. Rahmati, "Solving the heterogeneous fixed fleet open vehicle routing problem by a combined metaheuristic algorithm," *International Journal of Production Research*, vol. 52, no. 9, pp. 2565–2575, 2014.

[38] M. H. Hà, N. Bostel, A. Langevin, and L.-M. Rousseau, "An exact algorithm and a metaheuristic for the generalized vehicle routing problem with flexible fleet size," *Computers & Operations Research*, vol. 43, no. 1, pp. 9–19, 2014.

[39] J. Lysgaard and S. Wøhlk, "A branch-and-cut-and-price algorithm for the cumulative capacitated vehicle routing problem," *European Journal of Operational Research*, vol. 236, no. 3, pp. 800–810, 2014.

[40] A. Chabrier, "Vehicle Routing Problem with elementary shortest path based column generation," *Computers & Operations Research*, vol. 33, no. 10, pp. 2972–2990, 2006.

[41] Z. Wang and Z. Wang, "A novel two-phase heuristic method for vehicle routing problem with backhauls," *Computers & Mathematics with Applications*, vol. 57, no. 11-12, pp. 1923–1928, 2009.

[42] G. Laporte, M. Gendreau, J.-Y. Potvin, and F. Semet, "Classical and modern heuristics for the vehicle routing problem," *International Transactions in Operational Research*, vol. 7, no. 4-5, pp. 285–300, 2000.

[43] S. Bikhchandani, D. Hirshleifer, and I. Welch, "Learning from the behavior of others: conformity, fads, and informational cascades," *Journal of Economic Perspectives*, vol. 12, no. 3, pp. 151–170, 1998.

[44] M. David, *Exploring Social Psychology*, McGraw-Hill, New York, NY, USA, 2009.

[45] A. Mehrabian and S. Ksionzky, "Models for affiliative and conformity behavior," *Psychological Bulletin*, vol. 74, no. 2, pp. 110–126, 1970.

[46] X. Jiang and P. Du, "Analysis of homebuyers' conformity behavior based on herd effect," Eastern Academic Forum, pp. 430–434, 2010.

[47] M. Cipriani and A. Guarino, "Herd behavior in financial markets: an experiment with financial market professionals," *Journal of the European Economic Association*, vol. 7, no. 1, pp. 206–233, 2009.

[48] E. Cao and M. Lai, "A hybrid differential evolution algorithm to vehicle routing problem with fuzzy demands," *Journal of Computational and Applied Mathematics*, vol. 231, no. 1, pp. 302–310, 2009.

[49] M. N. A. Khalid and U. K. Yusof, "Optimizing distributed production scheduling problem in flexible manufacturing system subjects to machine maintenance: a modified chemical reaction approach," *International Journal of Innovative Computing, Information and Control*, vol. 11, no. 1, pp. 213–228, 2015.

[50] N. B. Mulyono and Y. Ishida, "Spatial strategy of disaster relief inventory," *International Journal of Innovative Computing, Information and Control*, vol. 11, no. 2, pp. 391–406, 2015.

[51] Y. Jou, B. Po-Tsang, and C. Lo, "Consideration of an imperfect production process in the supply chain of an integrated inventory model with polynomial present value and dependent crashing cost," *International Journal of Innovative Computing, Information and Control*, vol. 11, no. 3, pp. 775–786, 2015.

[52] J. Ruan, P. Shi, C.-C. Lim, and X. Wang, "Relief supplies allocation and optimization by interval and fuzzy number approaches," *Information Sciences*, vol. 303, pp. 15–32, 2015.

[53] J. Ruan, X. Wang, and Y. Shi, "A two-stage approach for medical supplies intermodal transportation in large-scale disaster responses," *International Journal of Environmental Research and Public Health*, vol. 11, no. 11, pp. 11081–11109, 2014.

[54] J. Wang, W. Huang, G. Ma, and S. Chen, "An improved partheno genetic algorithm for multi-objective economic dispatch in cascaded hydropower systems," *International Journal of Electrical Power & Energy Systems*, vol. 67, pp. 591–597, 2015.

[55] R. N. Greenwell, J. E. Angus, and M. Finck, "Optimal mutation probability for genetic algorithms," *Mathematical & Computer Modelling*, vol. 21, no. 8, pp. 1–11, 1995.