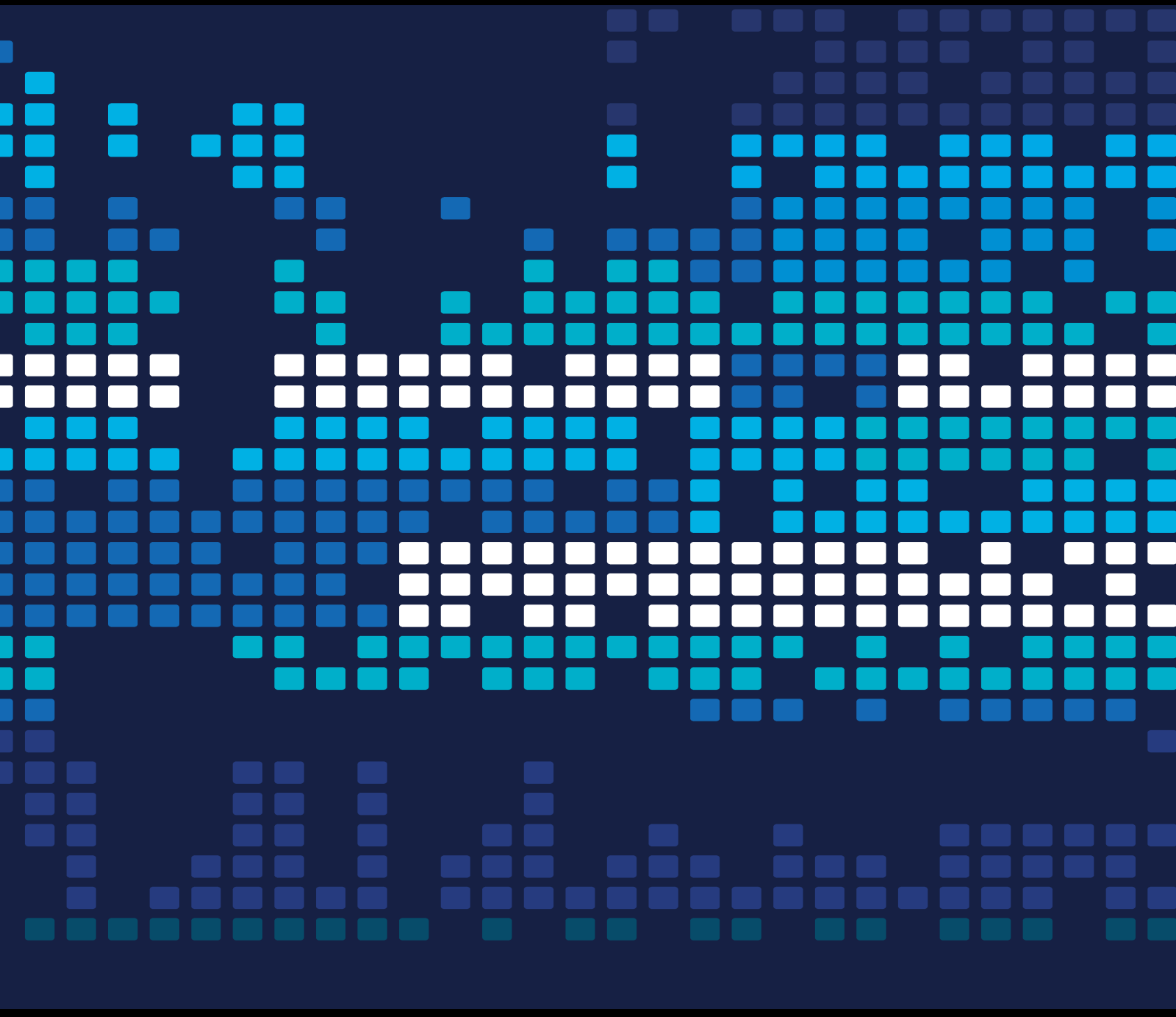
Scientific Programming

Data-Driven Decision Making in Business Intelligence

Lead Guest Editor: Xiaofeng Xu
Guest Editors: Dan Ralescu and Hongguang Ma



Data-Driven Decision Making in Business Intelligence

Scientific Programming

Data-Driven Decision Making in Business Intelligence

Lead Guest Editor: Xiaofeng Xu
Guest Editors: Dan Ralescu and Hongguang Ma

Contents

Hindawi
Scientific Programming
Volume 2023, Article ID 1606642, 15 pages
https://doi.org/10.1155/2023/1606642

Research Article

A Distributionally Robust Fuzzy Optimization Method for Single-Period Inventory Management Problems

Zhaozhuang Guo, Yuefang Sun, Shengnan Tian, and Zikun Li

School of Liberal Arts and Sciences, North China Institute of Aerospace Engineering, Langfang 065000, Hebei, China

Correspondence should be addressed to Yuefang Sun; syuefang@163.com

Received 12 October 2022; Revised 28 October 2022; Accepted 25 January 2023; Published 18 February 2023

Academic Editor: Hongguang Ma

This paper investigates single-period inventory management problems with uncertain market demand, where the exact possibility distribution of demand is unavailable. In this condition, it is important to order a reliable quantity which can immunize against distribution uncertainty. To model this type of single-period inventory management problem, this paper characterizes the uncertain demand by generalized interval-valued possibility distributions. We present a novel concept about an uncertain distribution set to describe distribution perturbation characterization. First, we introduce a lambda selection of the interval-valued fuzzy variable, and the uncertain distribution set is a collection of all generalized possibility distributions of lambda selection variables. According to the uncertain distribution set, a new distributionally robust fuzzy optimization method is developed for single-period inventory management problems. Under mild assumptions, the robust counterpart of the proposed fuzzy single-period inventory management model is formulated, which is an optimization program with certain linear objectives and infinitely many integral constraints. We discuss the computational issue of integral constraints and reformulate equivalently the robust counterpart as three deterministic inventory submodels under generalized interval-valued trapezoidal possibility distributions. According to the characteristics of three submodels, a domain decomposition method is designed to find the robust optimal solution that can immunize against uncertainty in our single-period inventory management problem. Finally, some computational results demonstrate the efficiency of the proposed distributionally robust fuzzy optimization method.

1. Introduction

The single-period inventory management problem (newsvendor problem) is a classical problem in the literature on inventory management [1]. The essential characteristic of the problem is that only one period is relevant and that there is no chance to place any subsequent orders during the period. In our real life, many products have such characteristics like seasonal products, sports goods, and fashion items, so the single-period inventory management problem provides a very useful framework to make decision about the optimal order quantity. The single-period inventory management problem has been studied since the 18th century, and it has been widely used to analyze supply chains with perishable and fashionable products. Khouja [2] reviewed the extensive contributions to the single-period inventory problem. Qin et al. [1] reviewed additional contributions and extended the prior reviews by considering several specific extensions. Most of the extensions have been made in a probabilistic framework and focused on the case that market demand was assumed to be random and characterized by random variables. However, many products today have shorter and shorter life cycles due to rapid technology upgrades. That is to say, growing innovation rates and shorter product life cycles make the market demand extremely variable. In this case, decision makers do have not enough historical data to determine the exact probability distribution of demand. Under incomplete information about the probability distribution of demand, some interesting research studies have been documented in the literature [3–5]. The stochastic approach seems to be less conservative than the worst-case-oriented robust optimization approach. This is so if indeed

the uncertain data are of a stochastic nature and if decision makers are able to provide the associated probability distribution.

However, the above two if's are too restrictive in some practical inventory management problems, in which the probability distribution of the market demand is unavailable due to the lack of the related historical data or information. At the same time, the demand in these situations can be approximately estimated based on the experts' experiences or subjective judgments. In this respect, there are some early applications of fuzzy set theory to inventory management problems in the literature [6–8]. Since then, there has been growing interest in the study of the single-period inventory model under fuzzy uncertainty [9–12]. In numerous uncertain inventory management problems, optimal order quantities often depend heavily on the distributions of uncertain market demands. The work mentioned above studied inventory management problems under the assumption that the exact possibility distributions of uncertain demands were available, which motivates us to address uncertain inventory management problems in a more advanced setting.

In this paper, we develop a new distributionally robust optimization method for the single-period inventory management problem, in which the uncertain demand is characterized by uncertain distribution sets. Compared with the existing literature, the novelties of this study include the following several aspects:

(i) When the distribution of uncertain market demand is only partially known, this paper presents a new method to model the uncertain demand by uncertain distribution sets.

(ii) Based on uncertain distribution sets, this paper investigates a new robust modeling framework for a single-period inventory management problem that incorporates the robust credibilistic optimization method and the risk-neutral criterion. Under mild assumptions, the robust counterpart of the original inventory management optimization model is formulated.

(iii) Under generalized interval-valued trapezoidal possibility distributions of uncertain market demand, theoretical analysis demonstrates that the robust counterpart of original inventory problems is equivalent to three submodels with different subregions. As a result, we design a domain decomposition method to find the robust optimal solution of our single-period inventory management problem.

(iv) The computational results demonstrate that the proposed distributionally robust optimization method can help the retailer order a reliable quantity to immunize against the uncertain demand in our single-period inventory management problem.

The rest of this paper is organized as follows: Section 2 gives an overview of related works. Section 3 first introduces generalized interval-valued trapezoidal fuzzy variables. Then, a new concept of an uncertain distribution set is proposed for a given parametric interval-valued possibility distribution. Section 4 develops a robust modeling framework for single-period inventory management problems, in which the possibility distributions of lambda selections vary in a given uncertain distribution set. Under mild assumptions, the robust counterpart of the original uncertain inventory problem is also formulated. Section 5 discusses the computational issue of infinitely many integral constraints and turns the robust counterpart problem into its equivalent parametric programming submodels. Section 6 designs a domain decomposition method to solve the obtained three deterministic parametric programming submodels. Section 7 performs some numerical experiments to demonstrate our new robust modeling idea. Section 8 gives our conclusions.

2. Literature Review

The single-period inventory management problem has been studied since the 18th century in the economic literature. Starting from the 1950s, the single-period inventory management problem has been extended to model a great variety of real-life problems. In the following, we classify the literature of single-period inventory management problems into three streams. The first stream is the random market demand with exact probability distributions. The second stream is the incomplete information about probability distributions of demand, and the third stream is the study of the single-period inventory model under fuzzy uncertainty.

2.1. Random Market Demand. The classical single-period inventory management problem was studied under random market demand, where the market demand was treated as an exogenous parameter. Based on the fact that the retailer may adjust the price in order to reduce or increase market demand, Lau and Lau [13] studied the extension of the classical single-period inventory management problem with a stochastic price-demand. For the simplest price-demand relationship, analytical solutions to the extended single-period inventory problem were obtained. For other cases, they developed numerical solution procedures. Based on the fact that market demand could be influenced by many marketing activities, such as advertising and sales calls, Kraiselburd et al. [14] studied the effect of marketing efforts on market demand, where the mean market demand was modeled as a nondecreasing function of the marketing effort. In addition, some marketing researchers believe that quantity stock has a positive impact on market demand. Balakrishnan et al. [15] generalized the single-period inventory problem to incorporate the stochastic and initial stock-level-dependent demand. In order to capture the effect of the stock level on demand, they considered a general random demand model via an inverse fractile function. Khouja [2] reviewed the extensive contributions to the single-period inventory management problem, such as different news-vendor pricing policies, discounting structures, and different states of information about demand. Qin et al. [1] reviewed

additional contributions and extended the prior reviews by considering several specific extensions such as analyzing the impact of the price, marketing effort, and stocking quantity on market demand.

2.2. Incomplete Information about Probability Distribution of Demand. Some researchers considered the single-period inventory management problem where the market demand did not satisfy the assumption of having a specific probability distribution of demand. In this case, some interesting research studies have been documented in the literature. Scarf [16] addressed the single-period inventory management problem where only the mean and variance of market demand was known. He modeled the problem by maximizing the expected profit against the worst possible distribution of the demand. Gallego and Moon [17] presented a simpler proof of Scarf's ordering rule and provided four extensions to the distribution-free single-period inventory management problem. Hill [18] applied a Bayesian methodology to the single-period inventory management problem where random demands followed known distributions with unknown but fixed parameters. Ridder et al. [19] studied the effects of demand variability on the profit. That is, higher demand variability resulted in larger variance and smaller profit. When the distribution of demand had known support, mean, and variance, Kamburowski [3] studied the single-period inventory management problem and derived the closed form formulas for the worst-case and best-case order quantities. Qin and Shang [4] and Wang et al. [5] applied a robust optimization approach to stochastic inventory management problems. So far, the robust optimization approach [20] has become an important research direction.

2.3. Fuzzy Market Demand. In fuzzy decision systems, several researchers have studied the inventory management problem based on fuzzy set theory. Petrović et al. [8] proposed two fuzzy models for single-period inventory management problems and discussed the effects of changing the membership function shapes of fuzzy inventory data on the optimal order quantity. Ishii and Konno [21] considered the stochastic demand with fuzzy shortage cost in the single-period inventory management problem. Kao and Hsu [22] proposed a single-period inventory model to find the optimal order quantity of the newsboy, in which the demand was described by subjectively determined membership functions. Li et al. [23] considered a single-period fuzzy inventory model, in which the optimal order quantity was achieved through fuzzy ordering of fuzzy numbers with respect to their total integral values. Dutta et al. [24] formulated a single-period inventory model with reordering opportunities under fuzzy demand. Chen and Ho [11] considered the optimal inventory policy for the single-period inventory problem with quantity discount. Xu and Zhai [25] expressed the demand as an L-R type fuzzy number. Under perfect coordination and in contrast with the noncoordination case, they investigated the optimization of the vertically integrated two-stage supply chain. Yu and Jin [9] developed the return policy in a supply chain with symmetric channel information and asymmetric channel information, respectively. Yu et al. [10] proposed a single-period inventory model with fuzzy price-dependent demand and discussed the conditions to determine the optimal pricing and inventory decisions jointly so that the expected profit could be maximized. Sang [12] considered a supply chain model with two competitive manufacturers and a common retailer, where the parameters of demand function and manufacturing costs were treated as fuzzy variables. For a single-vendor multiretailer supply chain, Sadeghi et al. [26] developed a constrained vendor-managed inventory model with trapezoidal fuzzy demand. Based on credibility measures, Guo [27] proposed two optimization models where uncertain demands were characterized by discrete and continuous possibility distributions, respectively. The analytical expressions of the optimal order quantity were derived in the above cases. Under variable possibility distributions of uncertain demand, Guo et al. [28] studied a multiproduct single-period inventory management problem.

Most of the existing literature studied inventory management problems under the assumption that the exact membership function or possibility distribution of fuzzy variables was available. Based on a given uncertainty distribution set, Guo and Liu [29] studied a distributionally robust optimization method for the single-period inventory management problem, which motivates us to study inventory management problems from a new perspective.

3. Uncertain Distribution Set

Fuzzy possibility theory has been introduced in the literature [30, 31]. For a detailed overview of the relationship between interval-valued fuzzy variables and interval type-2 fuzzy variables, we refer to Pagola et al. [32] and Bustince et al. [33].

In order to characterize the perturbation of the possibility distribution in some practical inventory management problems, we first introduce the representation method for the interval-valued distribution of fuzzy variables [34]. If the secondary possibility distribution of a type-2 fuzzy variable ξ is the following subinterval,

$$\left[\frac{x-r_1}{r_2-r_1}-\theta_l\frac{x-r_1}{r_2-r_1},\frac{x-r_1}{r_2-r_1}+\theta_r\frac{r_2-x}{r_2-r_1}\right], \tag{1}$$

of [0,1] for $x \in [r_1, r_2]$, the subinterval $[1-\theta_l, 1]$ of [0,1] for $x \in [r_2, r_3]$, and the following subinterval,

$$\left[\frac{r_4-x}{r_4-r_3}-\theta_l\frac{r_4-x}{r_4-r_3},\frac{r_4-x}{r_4-r_3}+\theta_r\frac{x-r_3}{r_4-r_3}\right], \tag{2}$$

of [0,1] for $x \in [r_3, r_4]$, then the fuzzy variable ξ is called a generalized parametric interval-valued trapezoidal fuzzy variable, where $\theta_l, \theta_r \in [0,1]$ are two parameters characterizing the degree of uncertainty that ξ takes on the value x. For simplicity, we denote the generalized parametric interval-valued trapezoidal fuzzy variable ξ by $\mathrm{Tra}(r_1, r_2, r_3, r_4; \theta)$ with $\theta = (\theta_l, \theta_r)$. The secondary

possibility distribution is called the nominal possibility distribution of ξ when $\theta_l = \theta_r = 0$, and the fuzzy variable characterized by the nominal possibility distribution is denoted by ξ^n.

In order to address the fuzzy variable with the parametric interval-valued possibility distribution, we next introduce its selection variable, which is different from the existing literature [35, 36]. Suppose ξ is a generalized parametric interval-valued fuzzy variable with the secondary possibility distribution $\widetilde{\mu}_\xi(x) = [\mu_{\xi^L}(x;\theta_l), \mu_{\xi^U}(x;\theta_r)]$. The fuzzy variable ξ^L is called the lower selection variable of ξ if ξ^L has the generalized parametric possibility distribution $\mu_{\xi^L}(x;\theta_l)$. The fuzzy variable ξ^U is called the upper selection variable of ξ if ξ^U has the generalized parametric possibility distribution $\mu_{\xi^U}(x;\theta_r)$. For any $\lambda \in [0,1]$, a fuzzy variable ξ^λ is called a lambda selection of ξ if ξ^λ has the following generalized parametric possibility distribution:

$$\mu_{\xi^\lambda}(x;\theta) = \lambda\mu_{\xi^U}(x;\theta_r) + (1-\lambda)\mu_{\xi^L}(x;\theta_l). \tag{3}$$

Theorem 1. *Let $\xi = Tra(r_1, r_2, r_3, r_4; \theta)$ be a generalized parametric interval-valued trapezoidal fuzzy variable and ξ^λ be its lambda selection. Then, for any real r, the credibility of $\{\xi^\lambda \le r\}$ is computed by*

$$Cr\{\xi^\lambda \le r\} = \begin{cases} 0, r \in (-\infty, r_1), \\ \dfrac{1}{2}\lambda\theta_r + \dfrac{(r-r_1)[1-(1-\lambda)\theta_l - \lambda\theta_r]}{2(r_2-r_1)}, r \in [r_1, r_2), \\ \dfrac{1}{2} - \dfrac{(1-\lambda)\theta_l}{2}, r \in [r_2, r_3), \\ 1-(1-\lambda)\theta_l, r \in [r_3, r_4), \\ -\dfrac{1}{2}\left\{\lambda\theta_r + \dfrac{(r_4-r)[1-(1-\lambda)\theta_l - \lambda\theta_r]}{r_4-r_3}\right\}, \\ 1-(1-\lambda)\theta_l, r \in [r_4, +\infty). \end{cases} \tag{4}$$

Proof. According to the definition of the generalized parametric interval-valued trapezoidal fuzzy variable, if $\xi = \mathrm{Tra}(r_1, r_2, r_3, r_4; \theta)$, then the generalized possibility distributions of the lower selection variable ξ^L and the upper selection variable ξ^U can be determined [34].

By the definition of ξ^λ, the generalized possibility distribution of ξ^λ is

$$\mu_{\xi^\lambda}(r;\theta) = \begin{cases} \lambda\theta_r + \dfrac{(r-r_1)[1-(1-\lambda)\theta_l - \lambda\theta_r]}{r_2-r_1}, & r \in [r_1, r_2], \\ 1-(1-\lambda)\theta_l, & r \in [r_2, r_3], \\ \lambda\theta_r + \dfrac{(r_4-r)[1-(1-\lambda)\theta_l - \lambda\theta_r]}{r_4-r_3}, & r \in [r_3, r_4]. \end{cases} \tag{5}$$

According to the definition of the credibility measure [37], the credibility of $\{\xi^\lambda \le r\}$ is computed by

$$\begin{aligned} \mathrm{Cr}\{\xi^\lambda \le r\} &= \frac{1}{2}\left(\sup_{t\in\Re}\mu_{\xi^\lambda}(t;\theta) + \sup_{t\le r}\mu_{\xi^\lambda}(t;\theta) - \sup_{t>r}\mu_{\xi^\lambda}(t;\theta)\right) \\ &= \begin{cases} 0, r \in (-\infty, r_1), \\ \dfrac{1}{2}\lambda\theta_r + \dfrac{(r-r_1)[1-(1-\lambda)\theta_l - \lambda\theta_r]}{2(r_2-r_1)}, r \in [r_1, r_2), \\ \dfrac{1}{2} - \dfrac{(1-\lambda)\theta_l}{2}, r \in [r_2, r_3), \\ 1-(1-\lambda)\theta_l, r \in [r_3, r_4), \\ -\dfrac{1}{2}\left\{\lambda\theta_r + \dfrac{(r_4-r)[1-(1-\lambda)\theta_l - \lambda\theta_r]}{r_4-r_3}\right\}, \\ 1-(1-\lambda)\theta_l, r \in [r_4 + \infty). \end{cases} \end{aligned} \tag{6}$$

The proof of the theorem is complete.

For a generalized parametric interval-valued trapezoidal fuzzy variable $\xi = \mathrm{Tra}(r_1, r_2, r_3, r_4; \theta)$, the generalized credibility $\mathrm{Cr}\{\xi^\lambda \le r\}$ with respect to r is plotted in Figure 1.

Given an interval-valued fuzzy variable ξ, we next define the uncertain distribution set to describe the distribution perturbation of the interval-valued possibility distribution $[\mu_{\xi^L}(x;\theta_l), \mu_{\xi^U}(x;\theta_r)]$.

Definition 1. Assume that ξ is a generalized parametric interval-valued fuzzy variable with the secondary possibility distribution $\widetilde{\mu}_\xi(\mathrm{x}) = [\mu_{\xi^L}(\mathrm{x};\theta_l), \mu_{\xi^U}(\mathrm{x};\theta_r)]$. For any $\lambda \in [0,1]$, the generalized possibility distribution of the lambda selection ξ^λ is denoted as $\mu_{\xi^\lambda}(\mathrm{x};\theta)$. Then, the uncertain distribution set $\mathscr{U}$ of ξ is defined as a collection of all generalized possibility distributions $\mu_{\xi^\lambda}(\mathrm{x};\theta)$ of lambda selections ξ^λ, i.e.,

$$\begin{aligned} \mathscr{U} = \Big\{\mu_{\xi^\lambda}(x;\theta) \,\Big|\, &\mu_{\xi^\lambda}(x;\theta) = \lambda\mu_{\xi^U}(x;\theta_r) \\ &+ (1-\lambda)\mu_{\xi^L}(x;\theta_l), \lambda \in [0,1]\Big\}. \end{aligned} \tag{7}$$

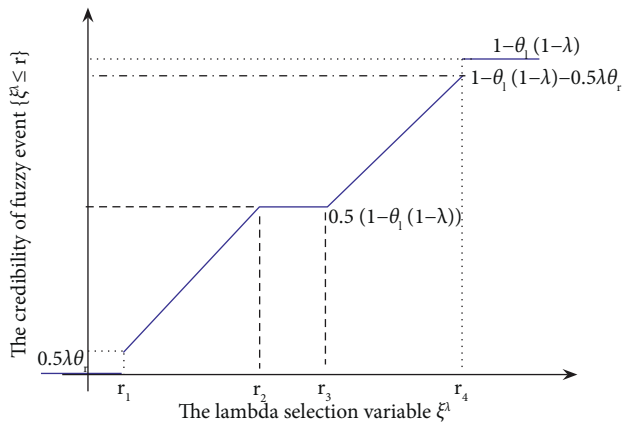


FIGURE 1: The credibility of the fuzzy event $\{\xi^{\lambda} \le r\}$.

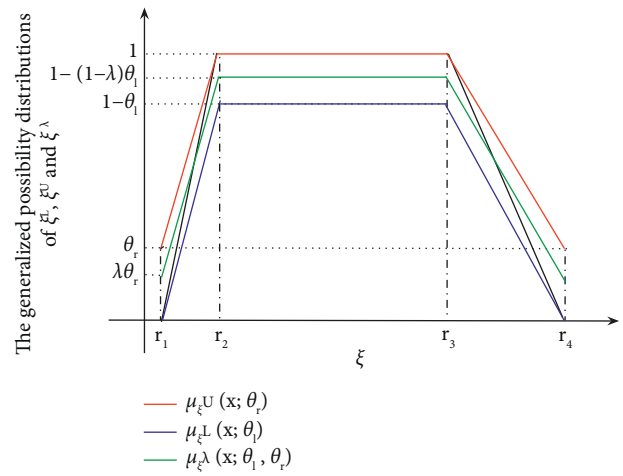


FIGURE 2: The generalized possibility distributions of ξ^{L}, ξ^{U}, and ξ^{λ}.

For a generalized parametric interval-valued trapezoidal fuzzy variable $\xi = \mathrm{Tra}(r_1, r_2, r_3, r_4; \theta)$, the generalized possibility distributions of ξ^{L}, ξ^{U}, and ξ^{λ} are plotted in Figure 2.

In the next section, we develop a robust modeling method for single-period inventory management problems under the given uncertain distribution set in (7).

4. Formulation of Robust Inventory Models

4.1. Problem Description and Notations. Consider a single-period inventory management problem consisting of one supplier and one retailer, where the supplier sells products to the retailer and the retailer faces the uncertain market demand. Suppose that products are sold only in one period and that the retailer has no chance to place a second order. After the supplier sets a price for his products, the retailer makes the decision to maximize his profit according to holding costs, goodwill costs for shortages, and his estimate on the demand. In practice, the difficulty faced by the retailer is to forecast the demand during the decision-making process. Because the uncertain market demand is nonnegative and bounded variable in the light of the actual conditions, the trapezoidal fuzzy demand variable is a common variable to characterize this demand. In order to build our distributionally robust optimization model, we next give some necessary notations and model parameters.

Fixed parameters are as follows.

c_r is the retailer's treatment cost of a unit product.

c is the total cost of a unit product and $c = c_r + w_s$.

s is the salvage value of a unit residual product.

g is the retailer's goodwill cost for unit unmet demand.

p is the retailer's sales price of a unit product.

w_s is the wholesale price of a unit product charged by suppliers.

The decision variable is as follows.

Q is the retailer's order quantity.

Uncertain parameters are as follows.

ξ is the uncertain market demand $\xi = \mathrm{Tra}(r_1, r_2, r_3, r_4; \theta)$.

ξ^{λ} is the lambda selection of the uncertain demand ξ.

In the following discussion, we assume $p > c > s$ to avoid trivial problems. Consider a lambda selection ξ^{λ} of a demand ξ, where $\mu_{\xi^{\lambda}}(x; \theta) \in \mathscr{U}$ is the generalized parametric possibility distribution of ξ^{λ}. When a retailer determines to order Q units of products, the sales volume, holding, and shortage quantity for the retailer are denoted as $\min(\xi^{\lambda}, Q)$, $\max(Q - \xi^{\lambda}, 0)$, and $\max(\xi^{\lambda} - Q, 0)$, respectively. Under this condition, the total profit for the retailer can be represented as

$$\pi\left(Q, \xi^{\lambda}\right) = p \min\left(\xi^{\lambda}, Q\right) + s \max\left(Q - \xi^{\lambda}, 0\right) - g \max\left(\xi^{\lambda} - Q, 0\right) - (w_s + c_r)Q. \tag{8}$$

The objective of a retailer is to maximize the total profit. When the demand ξ is known exactly in advance, the best decision is to order exactly quantity ξ. In the next subsection, we build a novel robust inventory optimization model, where the uncertain demand is characterized by a given uncertain distribution set. For a detailed overview of the robust optimization framework, we refer to Ben-Tal et al. [38], Bertsimas et al. [39], and Gorissen et al. [40].

4.2. Development of the Single-Period Inventory Model in the Fuzzy Decision System. In the fuzzy decision system, the single-period inventory problem has been studied in [8, 22, 23], in which the basic model is built as

$$\max_{Q} \; I[\pi(Q, \xi)] \quad \text{s.t. } r_1 \le Q \le r_4, \tag{9}$$

where ξ is a fuzzy demand and I is some defuzzification method for the associated fuzzy profit $\pi(Q, \xi)$. The work mentioned above studied the single inventory management problem under the assumption that the exact membership function or possibility distribution of uncertain demand ξ was available.

In the present paper, we study the single-period inventory management problem from a new perspective. When the information of the uncertain demand ξ is partially known, we characterize it by the parametric interval-valued possibility distribution $\tilde{\mu}_{\xi}(x) = [\mu_{\xi^L}(x; \theta_l), \mu_{\xi^U}(x; \theta_r)]$. After

that, we introduce the lambda selection ξ^{λ} of the interval-valued demand and define the uncertain distribution set as in (7).

Given the generalized possibility distribution $\mu_{\xi^{\lambda}}(x;\theta)$ of the lambda selection ξ^{λ}, the total profit $\pi(Q,\xi^{\lambda})$ for the retailer is represented by (8). Based on L–S integral [41], the mean profit of $\pi(Q,\xi^{\lambda})$ is computed by

$$\int_{[r_1,r_4]} \pi(Q,r)\mathrm{dCr}\left\{\xi^{\lambda} \le r\right\}, \tag{10}$$

where the measure is induced by the nondecreasing function $\mathrm{Cr}\left\{\xi^{\lambda} \le r\right\}$. It is obvious that the mean profit (10) depends on the generalized parametric possibility distribution $\mu_{\xi^{\lambda}}(x;\theta)$ of the lambda selection ξ^{λ} of the uncertain demand ξ.

Finally, we develop the following robust inventory optimization model for a single-period inventory management problem:

$$\left\{\max_{Q}\left\{\int_{[r_1,r_4]} \pi(Q,r)\mathrm{dCr}\left\{\xi^{\lambda} \le r\right\}\colon r_1 \le Q \le r_4\right\}\right\}_{\mu_{\xi^{\lambda}}(x;\theta)\in\mathscr{U}}, \tag{11}$$

which is a collection of optimization models.

$$\begin{aligned} &\max_{Q} \int_{[r_1,r_4]} \pi(Q,r)\mathrm{dCr}\left\{\xi^{\lambda} \le r\right\}, \\ &\text{s.t. } r_1 \le Q \le r_4, \end{aligned} \tag{12}$$

where possibility distribution $\mu_{\xi^{\lambda}}(x;\theta)$ varies in $\mathscr{U}$.

It is evident that our robust inventory optimization model (11) is totally different from model (9). In addition, the solution concept to model (11) is different from that to model (9), which is addressed in details in the next subsection.

4.3. Robust Counterpart of the Proposed Inventory Optimization Model. In this subsection, we deal with the robust counterpart of distributionally robust single-period inventory problem (11). In contrast to fuzzy inventory optimization model (9), where the uncertain demand has a fixed membership function or possibility distribution, a collection of a single-period inventory optimization problem like (12) is not associated by itself with the concepts of feasible solutions, optimal solutions and optimal values. The answer to the question rests on some implicit assumptions on the underlying decision-making environment. In this paper, we focus on the environment with the following assumptions.

A1. The retailer's order quantity Q in problem (11) represents "here and now" decision; it should be assigned a specific numerical value as a result of solving the problem before the actual demand data ξ reveal themselves.

A2. The decision maker is fully responsible for the consequence of the decision to be made when and only when the parametric possibility distribution $\mu_{\xi^{\lambda}}(x;\theta)$ varies in the uncertain distribution set $\mathscr{U}$ specified by (7).

A3. The constraints in model (11) cannot be violated when the parametric possibility distribution $\mu_{\xi^{\lambda}}(x;\theta)$ varies in $\mathscr{U}$.

The three assumptions lead to the robust feasible solution to optimization problem (11). Considering all the generalized possibility distributions in $\mathscr{U}$, the objective function will be constructed with respect to the worst-case mean profit. Therefore, this leads to solving the distributionally robust optimization model. By **A1**, the meaningful feasible solution to the uncertain single-period inventory management problem should be a fixed variable; Based on the spirit of the worse-case-oriented assumptions, **A2** and **A3**, the robust mean value of the objective in (11) at a candidate order quantity Q is the smallest mean value $\int_{[r_1,r_4]}\pi(Q,r)\mathrm{dCr}\left\{\xi^{\lambda} \le r\right\}$ over all parametric possibility distribution $\mu_{\xi^{\lambda}}(x;\theta)$ from the uncertain distribution set $\mathscr{U}$, i.e.,

$$\inf_{\mu_{\xi^{\lambda}}(r;\theta)\in\mathscr{U}} \int_{[r_1,r_4]} \pi(Q,r)\mathrm{dCr}\left\{\xi^{\lambda} \le r\right\}. \tag{13}$$

The best robust feasible solution is the one that solves the following optimization problem:

$$\begin{aligned} &\max_{Q} \inf_{\mu_{\xi^{\lambda}}(x;\theta)\in\mathscr{U}} \int_{[r_1,r_4]} \pi(Q,r)\mathrm{dCr}\left\{\xi^{\lambda} \le r\right\}, \\ &\text{s.t. } r_1 \le Q \le r_4. \end{aligned} \tag{14}$$

Problem (14) is called the robust counterpart of problem (11). We should seek the maximal robust mean value of the objective among all robust feasible solutions to the uncertain single-period inventory problem. If we introduce an extra variable t and reformulate problem (14) as an optimization problem with a certain objective, then robust counterpart (14) can be rewritten equivalently as the following optimization problem:

$$\begin{aligned} &\max_{t,Q} \ t \\ &\text{s.t.} \int_{[r_1,r_4]} \pi(Q,r)\mathrm{dCr}\left\{\xi^{\lambda} \le r\right\} \ge t, \forall \mu_{\xi^{\lambda}}(x;\theta) \in \mathscr{U}, \\ &r_1 \le Q \le r_4. \end{aligned} \tag{15}$$

So far, we have obtained equivalent robust counterpart optimization problem (15) in variables Q and t, where the objective is not affected by uncertainty at all. The optimal solution and optimal value to robust counterpart (14) or (15) is called the robust optimal solution and the robust optimal value to problem (11), respectively. Note that robust counterpart (15) has a certain linear objective and infinitely many integral constraints, which are usually computationally intractable. In the next section, we discuss the equivalent deterministic programming model of problem (15) and design its solution algorithm.

5. The Equivalent Programming Models of the Robust Counterpart

5.1. The Analytical Representation of the Mean Profit. In order to solve robust counterpart (15), it is required to derive the analytical representation of the following mean profit:

$$\int_{[r_1,r_4]} \pi(Q,r)\mathrm{dCr}\left\{\xi^{\lambda} \le r\right\}, \tag{16}$$

where the generalized credibility $\mathrm{Cr}\{\xi \le r\}$ is computed by (4). We first give the following result about the analytical representation of the mean profit.

Theorem 2. *(Guo [27]). If ξ is a fuzzy demand variable with a finite expected value, then the analytical representation of the mean profit is*

$$\int_{[0,+\infty]} \pi(Q,r)\mathrm{dC}\, r\{\xi \le r\} = (p+g-w_s-c_r)hQ - (p+g-s)\int_0^Q Cr\{\xi \le r\}\mathrm{d}r - g\mu, \tag{17}$$

where $h = \lim_{r \longrightarrow +\infty} Cr\{\xi \le r\}$ and $\mu = \int_{[0,+\infty]} r dC\, r\{\xi \le r\}$.

If ξ^{λ} is the lambda selection of $\mathrm{Tra}(r_1, r_2, r_3, r_4; \theta)$, then $\pi(Q, \xi^{\lambda})$ is computed by

$$\int_{[r_1,r_4]} \pi(Q,r)\mathrm{dCr}\left\{\xi^{\lambda} \le r\right\} = (p+g-w_s-c_r)(1-(1-\lambda)\theta_l)Q - (p+g-s)\int_0^Q Cr\{\xi \le r\}dr - g\mu, \tag{18}$$

where $\mu = [1-(1-\lambda)\theta_l](r_1+r_2+r_3+r_4/4) + (1/4)\lambda\theta_r(r_1-r_2-r_3+r_4)$.

According to Theorem 1, one has

$$\int_{r_1}^{Q} \mathrm{Cr}\left\{\xi^{\lambda} \le r\right\}\mathrm{d}r = \begin{cases} \lambda\dfrac{\theta_r\left[-Q^2+2r_2Q-(2r_2-r_1)r_1\right]}{4(r_2-r_1)} + \dfrac{h(Q-r_1)^2}{4(r_2-r_1)}, & r_1 \le Q < r_2 \\ \lambda\dfrac{\theta_r(r_2-r_1)}{4} + \dfrac{h(2Q-r_1-r_2)}{4}, & r_2 \le Q < r_3 \\ \lambda\dfrac{\theta_r\left[-Q^2+2r_3Q+(r_2-r_1)(r_4-r_3)-r_3^2\right]}{4(r_4-r_3)} + \dfrac{h\left[Q^2+2(r_4-2r_3)Q\right]}{4(r_4-r_3)}, & r_3 \le Q \le r_4. \\ \dfrac{h\left[(r_1+r_2)(r_3-r_4)+r_3^2\right]}{4(r_4-r_3)}, \end{cases} \tag{19}$$

By (18) and (19), when $r_1 \le Q < r_2$, the mean profit has the following analytical representation:

$$\int_{[r_1,r_4]} \pi(Q,r)\mathrm{dCr}\left\{\xi^{\lambda} \le r\right\} = \begin{cases} \lambda\left\{-\dfrac{g[\theta_l(r_1+r_2+r_3+r_4)+\theta_r(r_1-r_2-r_3+r_4)]}{4}\right. \\ -\dfrac{Q\{2\theta_l r_1(p+g+s-2c)-2r_2[2\theta_l(p+g-c)-\theta_r(p+g-s)]\}}{4(r_2-r_1)} \\ -\dfrac{r_1(p+g-s)[r_1(\theta_r+\theta_l)-2r_2\theta_r]}{4(r_2-r_1)} \\ \left.+\dfrac{Q^2(p+g-s)(\theta_r-\theta_l)}{4(r_2-r_1)}\right\} \\ +Q(1-\theta_l)\left[p+g-c+\dfrac{r_1(p+g-s)}{2(r_2-r_1)}\right] \\ -\dfrac{g(1-\theta_l)(r_1+r_2+r_3+r_4)}{4} \\ -\dfrac{(1-\theta_l)r_1^2(p+g-s)}{4(r_2-r_1)} \\ -\dfrac{Q^2(1-\theta_l)(p+g-s)}{4(r_2-r_1)}. \end{cases} \tag{20}$$

Similarly, when $r_2 \le Q < r_3$, the mean profit has the following analytical representation:

$$\int_{[r_1,r_4]} \pi(Q,r)\mathrm{dCr}\left\{\xi^{\lambda} \le r\right\} = \begin{cases} \lambda\left\{\dfrac{Q\theta_l(p+g+s-2c)}{2}+\dfrac{(p+g-s)[(r_1-r_2)\theta_r+(r_1+r_2)\theta_l]}{4}\right. \\ \left.-\dfrac{g[\theta_l(r_1+r_2+r_3+r_4)+\theta_r(r_1-r_2-r_3+r_4)]}{4}\right\} \\ +\dfrac{Q(1-\theta_l)(p+g+s-2c)}{2}+\dfrac{(p+g-s)(1-\theta_l)(r_1+r_2)}{4} \\ -\dfrac{g(1-\theta_l)(r_1+r_2+r_3+r_4)}{4}. \end{cases} \tag{21}$$

Finally, when $r_3 \le Q \le r_4$, the mean profit has the following analytical representation:

$$\int_{[r_1,r_4]} \pi(Q,r)\mathrm{dCr}\left\{\xi^{\lambda} \le r\right\} = \begin{cases} \lambda\left\{\dfrac{Q\{r_3[4\theta_l(c-s)-2\theta_r(p+g-s)]+2\theta_l r_4(p+g+s-2c)\}}{4(r_4-r_3)}\right. \\ -\dfrac{g[\theta_l(r_1+r_2+r_3+r_4)+\theta_r(r_1-r_2-r_3+r_4)]}{4} \\ +\dfrac{\theta_r(p+g-s)\left[r_3^2-(r_2-r_1)(r_4-r_3)\right]}{4(r_4-r_3)} \\ -\dfrac{\theta_l(p+g-s)\left[(r_1+r_2)(r_3-r_4)+r_3^2\right]}{4(r_4-r_3)} \\ \left.+\dfrac{Q^2(p+g-s)(\theta_r-\theta_l)}{4(r_4-r_3)}\right\} \\ +Q(1-\theta_l)\left[\dfrac{2r_3-r_4}{2(r_4-r_3)}(p+g-s)+p+g-c\right] \\ -\dfrac{(1-\theta_l)(p+g-s)\left[(r_1+r_2)(r_3-r_4)+r_3^2\right]}{4(r_4-r_3)} \\ -\dfrac{g(1-\theta_l)(r_1+r_2+r_3+r_4)}{4} \\ -\dfrac{Q^2(1-\theta_l)(p+g-s)}{4(r_4-r_3)}. \end{cases} \tag{22}$$

5.2. The Equivalent Parametric Programming Submodels of the Robust Counterpart. In order to solve robust counterpart problem (15), we introduce the following functions:

$$\begin{aligned} f_1(Q) = &-\frac{Q\{2\theta_l r_1(p+g+s-2c)-2r_2[2\theta_l(p+g-c)-\theta_r(p+g-s)]\}}{4(r_2-r_1)} \\ &-\frac{g[\theta_l(r_1+r_2+r_3+r_4)+\theta_r(r_1-r_2-r_3+r_4)]}{4} \\ &-\frac{r_1(p+g-s)[r_1(\theta_r+\theta_l)-2r_2\theta_r]}{4(r_2-r_1)}+\frac{Q^2(p+g-s)(\theta_r-\theta_l)}{4(r_2-r_1)}, \\ g_1(Q) = &-\frac{Q^2(1-\theta_l)(p+g-s)}{4(r_2-r_1)}-\frac{(1-\theta_l)r_1^2(p+g-s)}{4(r_2-r_1)} \\ &+Q(1-\theta_l)\left[p+g-c+\frac{r_1(p+g-s)}{2(r_2-r_1)}\right]-\frac{g(1-\theta_l)(r_1+r_2+r_3+r_4)}{4}, \end{aligned}$$

$$
\begin{aligned}
f_2(Q) &= \frac{Q\theta_l(p+g+s-2c)}{2} + \frac{(p+g-s)[(r_1-r_2)\theta_r+(r_1+r_2)\theta_l]}{4} \\
&\quad - \frac{g[\theta_l(r_1+r_2+r_3+r_4)+\theta_r(r_1-r_2-r_3+r_4)]}{4}, \\
g_2(Q) &= \frac{Q(1-\theta_l)(p+g+s-2c)}{2} + \frac{(p+g-s)(1-\theta_l)(r_1+r_2)}{4} - \frac{g(1-\theta_l)(r_1+r_2+r_3+r_4)}{4}, \\
f_3(Q) &= \frac{Q\{r_3[4\theta_l(c-s)-2\theta_r(p+g-s)]+2\theta_l r_4(p+g+s-2c)\}}{4(r_4-r_3)} - \frac{g[\theta_l(r_1+r_2+r_3+r_4)+\theta_r(r_1-r_2-r_3+r_4)]}{4} \\
&\quad + \frac{\theta_r(p+g-s)\left[r_3^2-(r_2-r_1)(r_4-r_3)\right]}{4(r_4-r_3)} - \frac{\theta_l(p+g-s)\left[(r_1+r_2)(r_3-r_4)+r_3^2\right]}{4(r_4-r_3)} + \frac{Q^2(p+g-s)(\theta_r-\theta_l)}{4(r_4-r_3)}, \\
g_3(Q) &= Q(1-\theta_l)\left[\frac{2r_3-r_4}{2(r_4-r_3)}(p+g-s)+p+g-c\right] - \frac{Q^2(1-\theta_l)(p+g-s)}{4(r_4-r_3)} - \frac{g(1-\theta_l)(r_1+r_2+r_3+r_4)}{4} \\
&\quad - \frac{(1-\theta_l)(p+g-s)\left[(r_1+r_2)(r_3-r_4)+r_3^2\right]}{4(r_4-r_3)}.
\end{aligned}
\tag{23}
$$

Using the aforementioned notations, when $r_1 \le Q < r_2$, by the analytical representation of the mean profit, the robust mean value has the following equivalent representation:

$$
\inf_{\mu_{\xi^\lambda}(r;\theta)\in\mathcal{U}} \int_{[r_1,r_4]} \pi(Q,r)\mathrm{dCr}\left\{\xi^\lambda \le r\right\} = -\max\{-f_1(Q),0\} + g_1(Q). \tag{24}
$$

Similarly, when $r_2 \le Q < r_3$, by the analytical representation of the mean profit, the robust mean value has the following equivalent representation:

$$
\inf_{\mu_{\xi^\lambda}(r;\theta)\in\mathcal{U}} \int_{[r_1,r_4]} \pi(Q,r)\mathrm{dCr}\left\{\xi^\lambda \le r\right\} = -\max\{-f_2(Q),0\} + g_2(Q). \tag{25}
$$

Finally, when $r_3 \le Q \le r_4$, by the analytical representation of the mean profit, the robust mean value has the following equivalent representation:

$$
\inf_{\mu_{\xi^\lambda}(r;\theta)\in\mathcal{U}} \int_{[r_1,r_4]} \pi(Q,r)\mathrm{dCr}\left\{\xi^\lambda \le r\right\} = -\max\{-f_3(Q),0\} + g_3(Q). \tag{26}
$$

According to the above analytical representation of the robust mean value, we decompose the feasible region of robust counterpart (15) into three disjoint subregions according to the values of the decision Q.

When $r_1 \le Q < r_2$, robust counterpart (15) is equivalent to the quadratic programming submodel:

$$
\begin{aligned}
\max \quad & t_1 \\
\text{s.t.} \quad & g_1(Q) - u_1 \ge t_1, \\
& f_1(Q) + u_1 \ge 0, \\
& u_1 \ge 0, r_1 \le Q < r_2.
\end{aligned}
\tag{27}
$$

Similarly, when $r_2 \le Q < r_3$, robust counterpart (15) is equivalent to the linear programming submodel:

$$\begin{aligned} \max \quad & t_2 \\ \text{s.t.} \quad & g_2(Q) - u_2 \geq t_2, \\ & f_2(Q) + u_2 \geq 0, \\ & u_2 \geq 0, r_2 \leq Q < r_3. \end{aligned} \tag{28}$$

Finally, when $r_3 \leq Q \leq r_4$, robust counterpart (15) is equivalent to the quadratic programming submodel:

$$\begin{aligned} \max \quad & t_3 \\ \text{s.t.} \quad & g_3(Q) - u_3 \geq t_3, \\ & f_3(Q) + u_3 \geq 0, \\ & u_3 \geq 0, r_3 \leq Q \leq r_4. \end{aligned} \tag{29}$$

6. Domain Decomposition Method

We have decomposed robust counterpart problem (15) into three deterministic parametric programming submodels (27)–(29). Therefore, the feasible region of problem (15) is decomposed into three disjoint subregions according to the values of the decision Q. The three subregions are just the feasible regions of submodels (27)–(29). From this observation, we know that the global optimal solution of problem (15) can be obtained by solving submodels (27)–(29). By comparing the objective values of the obtained local optimal solutions, we can find the global optimal solution. This solution procedure is called the domain decomposition method.

Given the values of distribution parameters θ_l and θ_r, the solution process described above is summarized as follows.

Step 1. Parametric programming submodels (27)–(29) are solved by using LINGO software, and the obtained local optimal solutions are denoted as $(Q_i, u_i, t_i), i = 1, 2, 3$.

Step 2. The local objective values $t_i = g_i(Q_i) - u_i$ are compared at (Q_i, u_i, t_i) for $i = 1, 2, 3$, and the global optimal solution is obtained by the following formula:

$$t_k = \max_{1 \leq i \leq 3} t_i. \tag{30}$$

Step 3. Q_k is returned as the global optimal solution to robust counterpart (15) with the optimal value $g_k(Q_k) - u_k$.

The obtained Q_k value is called the robust optimal solution to model (11).

7. Numerical Experiments

7.1. Problem Statement. In this section, we consider a practical fanner inventory management problem during the summer. The retailer needs to order fanners before a selling season. Based on the knowledge of the retailer, the number of fanner demands ξ during the sales cycle is between 300 and 600, but the exact distribution on the interval $[300, 600]$ is unavailable. To model this situation, we characterize the uncertain demand ξ of the fanner by the generalized parametric interval-valued trapezoidal variable $\mathrm{Tra}(300, 450, 550, 600; \theta_l, \theta_r)$, where θ_l and θ_r represent the uncertainty degree of the market demand ξ in the interval $[300, 600]$. The unit wholesale price w_s of the fanner is set by the supplier, and the retailer orders the number of fanners based on holding costs, goodwill costs for shortages, and his estimate on demand during the summer. In this inventory management problem, the values of model parameters are set as follows. The unit wholesale price w_s is \$55, and the unit retail price p is \$130. The treatment cost c_r of the unit fanner is \$5, and the goodwill cost g for unit unmet demand is \$15. In order to avoid overstock, the retailer can hold a special clearance sale to sell all surplus fanners at the end of the sales cycle, and there is no initial inventory on hand. It is expected that any residual fanners could be sold at the price s = \$20.

7.2. Computational Results of the Robust Counterpart. In order to find the optimal order quantity of the fanner, we solve robust counterpart (15) with respect to the following uncertain distribution set:

$$\mathcal{U} = \left\{ \mu_{\xi^\lambda} \,\middle|\, \mu_{\xi^\lambda} \text{ is the possibility distribution of } \xi^\lambda \right\}, \tag{31}$$

where ξ^λ is the lambda selection of the uncertain demand $\xi = \mathrm{Tra}(300, 450, 550, 600; \theta_l, \theta_r)$.

Based on our designed domain decomposition method, we employ LINGO software to solve three equivalent submodels (27)–(29), where the values of parameters are $r_1 = 300$, $r_2 = 450$, $r_3 = 550$, and $r_4 = 600$.

To identify the influence of distribution parameters θ_r and θ_l on solution results, we first set the values of parameter θ_l as 0.05, 0.08, and 0.1, respectively, and observe the relationship between the robust optimal order quantity and the parameter θ_r and the relationship between the robust optimal mean profit and the parameter θ_r. The computational results are plotted in Figures 3 and 4, respectively, from which we find that the robust optimal order quantity Q^* is increasing with respect to the parameter θ_r, while the robust optimal mean profit of the retailer is decreasing with respect to the parameter θ_r. To further illustrate the impact of distribution parameters θ_r and θ_l on the robust optimal order quantity, the computational results with various values of θ_r and fixed $\theta_l = 0.02$ are reported in Table 1.

We next set the values of θ_r as 0.05, 0.1, 0.2, and 0.3, respectively, and observe the relationship between the robust optimal order quantity the and parameter θ_l and the relationship between the robust optimal mean profit and the parameter θ_l. The computational results are plotted in Figures 5 and 6, respectively, from which we find that the robust optimal order quantity Q^* is decreasing with respect to the parameter θ_l, and the robust optimal mean profit of the retailer is also decreasing with respect to the parameter θ_l. To further illustrate the impact of distribution parameters θ_r and θ_l on the robust optimal order quantity, the computational results with various values of θ_l and fixed $\theta_r = 0.28$ are reported in Table 2.

So far, we have discussed the relationship between the robust optimal order quantity and distribution parameters θ_r and θ_l. By the definition of distribution parameters θ_r and

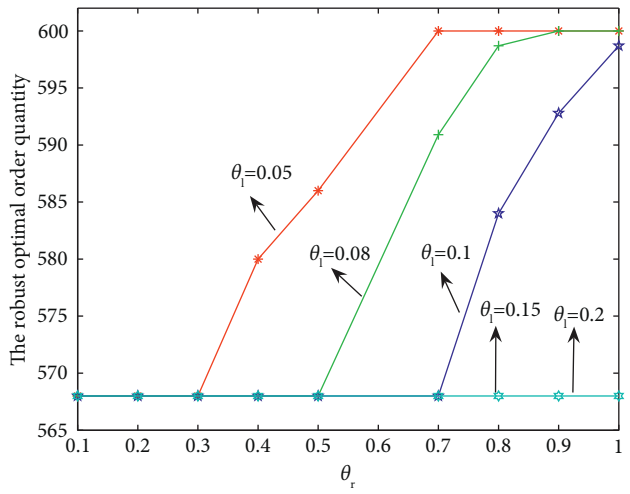


Figure 3: The relationship between the robust optimal order quantity and the parameter θ_r.

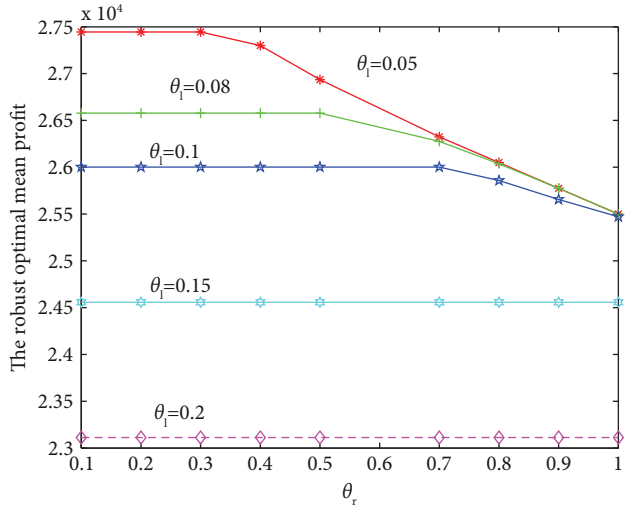


Figure 4: The relationship between the robust optimal mean profit and the parameter θ_r.

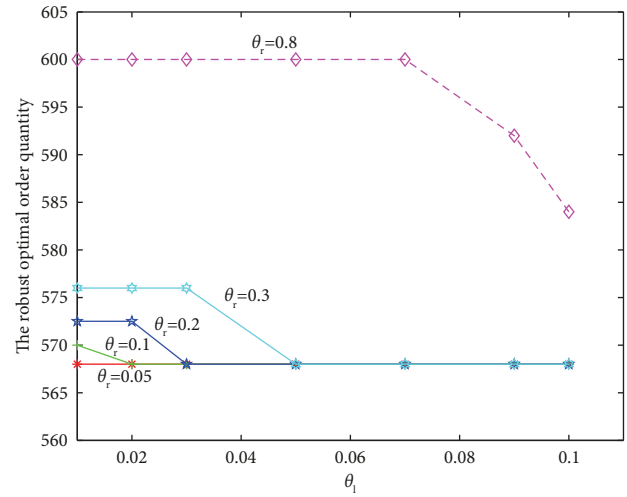


Figure 5: The relationship between the robust optimal order quantity and the parameter θ_l.

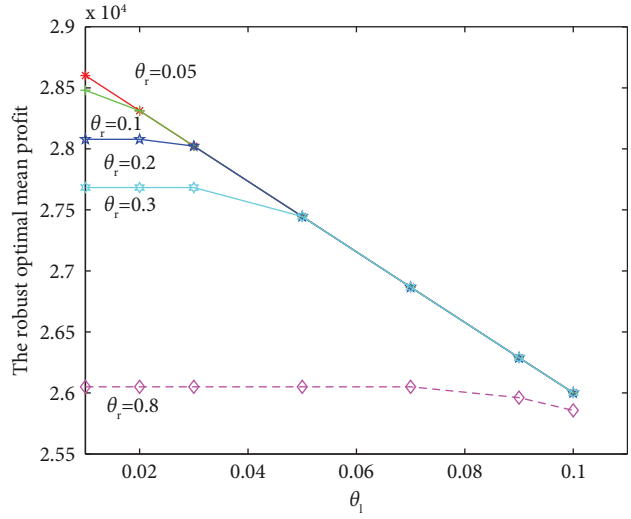


Figure 6: The relationship between the robust optimal mean profit and the parameter θ_l.

Table 1: The robust optimal order quantity and the mean profit with respect to θ_r.

θ_l	θ_r	Robust optimal order quantity	Robust optimal mean profit
0.02	0.15	571	28278.9
0.02	0.2	572	28078
0.02	0.22	573	27998.4
0.02	0.25	574	27879.4
0.02	0.28	575	27761.3
0.02	0.3	576	27683

Table 2: The robust optimal order quantity and the mean profit with respect to θ_l.

θ_l	θ_r	Robust optimal order quantity	Robust optimal mean profit
0.038	0.28	575	27761.2
0.0382	0.28	574	27760.8
0.0385	0.28	573	27759.5
0.0388	0.28	572	27757.2
0.039	0.28	571	27754.1
0.0395	0.28	570	27746.4

θ_l, the larger the distribution parameters, the larger the uncertainty degree of uncertain demand. The above computational results support our arguments. In the next section, we further compare the proposed distributionally robust optimization method with other optimization approaches to the single-period inventory management problem.

7.3. Comparison Study

7.3.1. Comparing with the Deterministic Optimization Method. We first compare our robust optimization method with the deterministic optimization method for the single-period inventory management problem, where the demand is known in advance. For the sake of comparison, we assume that the deterministic demand d is a mean value of 475 for the trapezoidal demand variable $\xi = \mathrm{Tra}(300, 450, 550, 600)$. In this situation, the optimization problem can be formulated as

$$\max_{r_1 \le Q \le r_4} \pi(Q), \tag{32}$$

where

$$\pi(Q) = \min\{(s - w_s - c_r)Q + 475(p - s), (p + g - w_s - c_r)Q - 475g\}. \tag{33}$$

Obviously, $\pi(Q)$ is a concave function with respect to Q. The total profit $\pi(Q)$ gets its maximum in $Q^* = d = 475$. That is, the optimal order quantity is 475 with a maximum profit of 33250. It is evident that the solution (475) is totally different from our robust optimal solutions reported in Tables 1 and 2. In fact, note that a value of 33250 is larger than all robust optimal values obtained in our numerical experiments, and the optimal solution (475) to model (32) is in the subregion [450,550]. Hence, there is no nonnegative number u_2 such that $(475, u_2, 33250)$ is a feasible solution to submodel (28), which implies $(475, u_2, 33250)$ is not feasible to robust counterpart problem (15).

7.3.2. Comparing with the Stochastic Optimization Method. We now compare our robust optimization method with the stochastic optimization method, in which the stochastic demand ξ follows a trapezoidal probability density function:

$$f(x) = \begin{cases} \dfrac{1}{30000}(x - 300), & 300 \le x \le 450, \\ \dfrac{1}{200}, & 450 \le x \le 550, \\ \dfrac{1}{10000}(600 - x), & 550 \le x \le 600. \end{cases} \tag{34}$$

According to the stochastic optimization method for the single-period inventory management problem [1], we know that the optimal order quantity is 511 with a maximum mean profit of 30007. In this case, the support of stochastic demand is $[300, 600]$, which is the same as the support of uncertain demand in our single-period inventory management problem. However, compared with our robust optimal solutions and robust optimal values reported in Tables 1 and 2, the optimal solution (511) to the stochastic model is not a feasible solution to our robust counterpart problem (15), which can be explained in the same way as in the deterministic optimization method.

7.3.3. Comparing with the Fuzzy Optimization Method. For the sake of comparison, suppose uncertain demand ξ follows a trapezoidal possibility distribution (300, 450, 550, 600), which is the nominal possibility distribution of interval-valued possibility distributions used in our numerical example. In this case, we consider the following optimization model:

$$\max_{r_1 \le Q \le r_4} E[\pi(Q, \xi)], \tag{35}$$

where $E[\pi(Q, \xi)] = \int_{[r_1, r_4]} \pi(Q, r)\mathrm{dCr}\{\xi \le r\}$. Model (35) corresponds to the case $\theta_r = \theta_l = 0$ in our robust counterpart problem (15). The optimal order quantity (568) to model (35) is called the nominal optimal order quantity, while a maximum mean profit of 28890 to model (35) is called the nominal maximal profit. The nominal maximal profit is larger than the robust optimal values obtained in Tables 1 and 2, which implies that the nominal optimal solution (568) in the subregion $[550, 600]$ is not a feasible solution to our robust counterpart problem (15). The price of robustness is the reduction from its nominal optimal value to its robust optimal value. From the computational results reported in Tables 1 and 2, we observe that the price of robustness is increasing with respect to θ_r or θ_l, which can be explained easily by the definition of parameters θ_r and θ_l. It is noted that with the meaning of price of robustness, the obtained robust order quantity is the best uncertainty-immunized solution for our inventory management problem.

From the above comparison study, we obtain the following observations:

(i) Some of data like market demands in inventory management do not exist and usually are replaced with retailer's forecasting. Deterministic inventory optimization is based on the assumption that future demands can be forecasted exactly. In the case that future demands cannot be forecasted exactly, one should not adopt the resulting optimal solution of model (32) to order the optimal quantity.

(ii) The stochastic optimization method for inventory management problems is based on the assumption that future demands are random variables and that decision makers are able to point out the associated probability distribution. In the case that the probability distributions of future demands cannot be determined exactly, the resulting optimal solution of the stochastic model cannot be used to order optimal quantity.

(iii) In the conventional fuzzy optimization method for inventory management problems, small data perturbation in the possibility distribution is usually ignored. The inventory problem is solved as if the given nominal possibility distributions were exact. The comparison study demonstrates that there exists a real need of a technique which is able to detect cases when data perturbation in possibility distributions can heavily affect the quality of nominal solutions. Applying the proposed distributionally robust optimization method, the resulting robust optimal order quantity is the best uncertainty-immunized solution we can associate with our uncertain single-period inventory management problem.

8. Conclusions

In this paper, we studied the single-period inventory management problem from a new perspective. The major new results include the following several aspects.

First, when only partial information about the distribution of market demand is available, we characterized the uncertain demand by an uncertain distribution set, which is a collection of all generalized possibility distributions of lambda selections.

Second, based on the proposed uncertain distribution set, a new robust fuzzy optimization method was developed for single-period inventory management problems. Under mild assumptions, we built the robust counterpart of the original inventory management problem.

Third, we discussed the computational issue of the robust counterpart. Based on the structural characteristics of the three submodels, a domain decomposition method was designed to find the robust optimal solution that can immunize against uncertainty in inventory management problems.

Finally, some computational results were provided to demonstrate the primary benefit of using the proposed robust fuzzy optimization method.

This study limits the consideration to the generalized parametric trapezoidal fuzzy variables with bounded possibility distributions. Extension to considering other types of uncertain distribution sets with the robust optimization method is another interesting research direction.

Data Availability

The data used to support the findings of this study are included within the article.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This work was supported by the Social Science Foundation of Hebei Province (No. HB22TJ005) and the Doctoral Science Foundation of North China Institute of Aerospace Engineering (No. BKY-2018-23).

References

[1] Y. Qin, R. X. Wang, A. J. Vakharia, Y. W. Chen, and M. M. Seref, "The newsvendor problem: review and directions for future research," *European Journal of Operational Research*, vol. 213, no. 2, pp. 361–374, 2011.

[2] M. Khouja, "The single period (news-vendor) problem: literature review and suggestions for future research," *Omega*, vol. 27, no. 5, pp. 537–553, 1999.

[3] J. Kamburowski, "The distribution-free newsboy problem under the worst-case and best-case scenarios," *European Journal of Operational Research*, vol. 237, no. 1, pp. 106–112, 2014.

[4] R. Qiu and J. Shang, "Robust optimisation for risk-averse multi-period inventory decision with partial demand distribution information," *International Journal of Production Research*, vol. 52, no. 24, pp. 7472–7495, 2014.

[5] C. X. Wang, S. Webster, and S. D. Zhang, "Robust price-setting newsvendor model with interval market size and consumer willingness-to-pay," *International Journal of Production Economics*, vol. 154, pp. 100–112, 2014.

[6] J. Kacprzyk and P. Stanieski, "Long-term inventory policy-making through fuzzy decision-making models," *Fuzzy Sets and Systems*, vol. 8, no. 2, pp. 117–132, 1982.

[7] H. J. Zimmermann, *Fuzzy Sets Theory and its Applications*, Kluwer Academic Publishers, London, UK, 2 edition, 1985.

[8] D. Petrović, R. Petrović, and M. Vujošević, "Fuzzy models for the newsboy problem," *International Journal of Production Economics*, vol. 45, no. 1-3, pp. 435–441, 1996.

[9] Y. Yu and T. D. Jin, "The return policy model with fuzzy demands and asymmetric information," *Applied Soft Computing*, vol. 11, no. 2, pp. 1669–1678, 2011.

[10] Y. Yu, J. Zhu, and C. W. Wang, "A newsvendor model with fuzzy price-dependent demand," *Applied Mathematical Modelling*, vol. 37, no. 5, pp. 2644–2661, 2013.

[11] S. P. Chen and Y. H. Ho, "Optimal inventory policy for the fuzzy newsboy problem with quantity discounts," *Information Sciences*, vol. 228, pp. 75–89, 2013.

[12] S. J. Sang, "Price competition of manufacturers in supply chain under a fuzzy decision environment," *Fuzzy Optimization and Decision Making*, vol. 14, no. 3, pp. 335–363, 2015.

[13] A. H. L. Lau and H. S. Lau, "The newsboy problem with price-dependent demand distribution," *IIE Transactions*, vol. 20, no. 2, pp. 168–175, 1988.

[14] S. Kraiselburd, V. G. Narayanan, and A. Raman, "Contracting in a supply chain with stochastic demand and substitute products," *Production and Operations Management*, vol. 13, no. 1, pp. 46–62, 2009.

[15] A. Balakrishnan, M. S. Pangburn, and E. Stavrulaki, "Integrating the promotional and service roles of retail inventories," *Manufacturing and Service Operations Management*, vol. 10, no. 2, pp. 218–235, 2008.

[16] H. Scarf, "A min-max solution of an inventory problem," in *Studies in the Mathematical Theory of Inventory and Production*, K. Arrow, S. Karlin, and H. Scarf, Eds., pp. 201–209, Stanford University Press, Stanford, CA, USA, 1958.

[17] G. Gallego and I. Moon, "The distribution free newsboy problem: review and extensions," *Journal of the Operational Research Society*, vol. 44, no. 8, pp. 825–834, 1993.

[18] R. M. Hill, "Applying Bayesian methodology with a uniform prior to the single period inventory model," *European Journal of Operational Research*, vol. 98, no. 3, pp. 555–562, 1997.
[19] A. Ridder, E. van der Laan, and M. Salomon, "How larger demand variability may lead to lower costs in the newsvendor problem," *Operations Research*, vol. 46, no. 6, pp. 934–936, 1998.
[20] X. F. Xu, Z. R. Lin, X. Li, C. J. Shang, and Q. Shen, "Multi-objective robust optimisation model for MDVRPLS in refined oil distribution," *International Journal of Production Research*, vol. 60, no. 22, pp. 6772–6792, 2021.
[21] H. Ishii and T. Konno, "A stochastic inventory problem with fuzzy shortage cost," *European Journal of Operational Research*, vol. 106, no. 1, pp. 90–94, 1998.
[22] C. Kao and W. K. Hsu, "A single-period inventory model with fuzzy demand," *Computers & Mathematics with Applications*, vol. 43, no. 6-7, pp. 841–848, 2002.
[23] L. Li, S. N. Kabadi, and K. P. K. Nair, "Fuzzy models for single-period inventory problem," *Fuzzy Sets and Systems*, vol. 132, no. 3, pp. 273–289, 2002.
[24] P. Dutta, D. Chakraborty, and A. R. Roy, "An inventory model for single-period products with reordering opportunities under fuzzy demand," *Computers & Mathematics with Applications*, vol. 53, no. 10, pp. 1502–1517, 2007.
[25] R. Xu and X. Zhai, "Analysis of supply chain coordination under fuzzy demand in a two-stage supply chain," *Applied Mathematical Modelling*, vol. 34, no. 1, pp. 129–139, 2010.
[26] J. Sadeghi, S. M. Mousavi, and S. T. A. Niaki, "Optimizing an inventory model with fuzzy demand, backordering, and discount using a hybrid imperialist competitive algorithm," *Applied Mathematical Modelling*, vol. 40, no. 15-16, pp. 7318–7335, 2016.
[27] Z. Z. Guo, "Optimal inventory policy for single-period inventory management problem under equivalent value criterion," *Journal of Uncertain Systems*, vol. 10, no. 4, pp. 302–311, 2016.
[28] Z. Z. Guo, S. N. Tian, and Y. K. Liu, "A multi-product single-period inventory management problem under variable possibility distributions," *Mathematical Problems in Engineering*, vol. 2017, Article ID 2159281, 14 pages, 2017.
[29] Z. Z. Guo and Y. K. Liu, "Modelling single-period inventory problem by distributionally robust fuzzy optimization method," *Journal of Intelligent and Fuzzy Systems*, vol. 35, no. 1, pp. 1007–1019, 2018.
[30] Z. Q. Liu and Y. K. Liu, "Type-2 fuzzy variables and their arithmetic," *Soft Computing*, vol. 14, no. 7, pp. 729–747, 2010.
[31] Y. Liu and Y. K. Liu, "The lambda selections of parametric interval-valued fuzzy variables and their numerical characteristics," *Fuzzy Optimization and Decision Making*, vol. 15, no. 3, pp. 255–279, 2016.
[32] M. Pagola, C. Lopez-Molina, J. Fernandez, E. Barrenechea, and H. Bustince, "Interval type-2 fuzzy sets constructed from several membership functions. Application to the fuzzy thresholding algorithm," *IEEE Transactions on Fuzzy Systems*, vol. 21, no. 2, pp. 230–244, 2013.
[33] H. Bustince, J. Fernandez, H. Hagras, F. Herrera, M. Pagola, and E. Barrenechea, "Interval type-2 fuzzy sets are generalization of interval-valued fuzzy sets: towards a wider view on their relationship," *IEEE Transactions on Fuzzy Systems*, vol. 23, no. 5, pp. 1876–1882, 2015.
[34] Z. Z. Guo, Y. K. Liu, and Y. Liu, "Coordinating a three level supply chain under generalized parametric interval-valued distribution of uncertain demand," *Journal of Ambient Intelligence and Humanized Computing*, vol. 8, no. 5, pp. 677–694, 2017.
[35] X. Bai and Y. K. Liu, "Semideviations of reduced fuzzy variables: a possibility approach," *Fuzzy Optimization and Decision Making*, vol. 13, no. 2, pp. 173–196, 2014.
[36] X. Bai and Y. K. Liu, "CVaR reduced fuzzy variables and their second order moments," *Iranian Journal of Fuzzy Systems*, vol. 12, no. 5, pp. 45–75, 2015.
[37] B. D. Liu and Y. K. Liu, "Expected value of fuzzy variable and fuzzy expected value models," *IEEE Transactions on Fuzzy Systems*, vol. 10, no. 4, pp. 445–450, 2002.
[38] A. Ben-Tal, L. E. Ghaoui, and A. Nemirovski, *Robust Optimization*, Princeton University Press, New Jersey, NJ, USA, 2009.
[39] D. Bertsimas, D. B. Brown, and C. Caramanis, "Theory and applications of robust optimization," *Society of Indian Automobile Manufacturers Review*, vol. 53, no. 3, pp. 464–501, 2011.
[40] B. L. Gorissen, İ. Yanikoğlu, and D. Hertog, "A practical guide to robust optimization," *Omega*, vol. 53, pp. 124–137, 2015.
[41] M. Carter and B. V. Brunt, *The Lebesgue-Stieltjes Integral*, Spinger-Verlag, New York, NY, USA, 2000.

Hindawi
Scientific Programming
Volume 2023, Article ID 8555423, 9 pages
https://doi.org/10.1155/2023/8555423

Research Article

Prediction Model of Fault Block Reservoir Measure Index Based on 1DCNN-LightGBM

Bin Wang,[1] Dawei Wu,[2] Kai Zhang,[3] Huaqing Zhang,[4] and Chao Zhang[1]

[1]*Exploration and Development Research Institute of Shengli Oil Field, SINOPEC, Dongying 257100, Shandong, China*
[2]*China University of Petroleum East China-Qingdao Campus, School of Petroleum Engineering, Qingdao 266580, Shandong, China*
[3]*School of Civil Engineering, Qingdao University of Technology, Qingdao 266520, China*
[4]*China University of Petroleum East China-Qingdao Campus, College of Control Science and Engineering and College of Sciences, Qingdao 266580, Shandong, China*

Correspondence should be addressed to Kai Zhang; zhangkai@qut.edu.cn

Received 1 June 2022; Revised 31 July 2022; Accepted 11 October 2022; Published 3 February 2023

Academic Editor: Debo Cheng

In view of the shortcomings of the prediction method of future development measures and indicators of fault block reservoir in the current oilfield practical application, a prediction method of fault block reservoir measures and indicators based on the random forest method and LightGBM is proposed, which can help the oilfield make more effective decisions in the middle and later development. Firstly, using the advantages of random forest (RF) in dealing with high-dimensional data sets, the main controlling factors are selected by feature analysis. Then, the measure prediction model is established by using the 1DCNN-LightGBM algorithm. Firstly, 1DCNN processes the reservoir dynamic data and then trains the LightGBM model with the extracted time series characteristics and static data characteristics as input to predict the measure indexes of fault block reservoir. The evaluation results show that the prediction models proposed in this paper have good performance and can obtain more accurate prediction results and more stable prediction performance. It provides a basis for the future planning and optimization of the oilfield.

1. Introduction

With the continuous development of the petroleum industry, most fault block oilfields in China have entered the stage of medium and high water cut adjustment and development, and there are problems such as significant production decline and fast water cut rise. Timely evaluating the development effect of fault block oilfields and establishing a complete potential evaluation index prediction method are of great significance in achieving stable production and increasing production [1]. Compared with ordinary reservoirs, the structure of fault block reservoir is relatively more complex, faults are developed, and there is an independent oil-water system composed of multiple separate fault blocks [2]. Therefore, the relationship of this type of reservoir is more complex, and it is difficult to realize the fine modeling of fault block reservoir structure, which leads to the significant increase in development difficulty and the deviation of the overall development effect of the reservoir. Previously, based on practical experience and theoretical derivation, oilfield workers proposed a variety of traditional methods for development potential evaluation, mainly including well pattern density method, numerical simulation method, water drive curve method, production decline method, and prediction model method [3, 4]. However, in the practical application of the abovementioned methods, due to the complex boundary conditions and oil-water distribution of the fault block reservoir, there will be a large deviation between the prediction results and the real situation.

In recent years, machine learning technology has developed rapidly and has been gradually applied to various

industrial fields, which has greatly promoted industrial production. Oilfield researchers have also found the advantages of machine learning technology in prediction and applied it to the petroleum field [5]. Machine learning methods such as artificial neural network, support vector machine, and swarm intelligence have been widely used in the field of production performance analysis [6]. It also provides a new method for index prediction of old oil fields, which is different from the previous method of studying oil reservoirs based on experience and classical reservoir engineering theory. Sun et al. used the LSTM neural network method instead of the decline curve analysis method to predict oilfield production, which improved the prediction accuracy [7]. Cai used the least squares support vector machine method based on particle swarm optimization to predict shale gas reservoir production. The results show that this method has good convergence and prediction accuracy [8]. However, at present, the relevant machine learning methods are relatively simple and do not consider the physical properties of the reservoir, which cannot meet the needs for rapid and accurate prediction of the oilfield [9]. Aiming at the problems of poor accuracy and low reliability of fault block reservoir potential evaluation methods, a new high-precision prediction method based on the 1DCNN LightGBM neural network model is proposed in this paper, which provides a scientific and accurate judgment basis for fault block reservoir planning and optimization.

2. Prediction Model of Measure Index of Fault Block Reservoir

This paper presents a fault block reservoir measure index prediction model based on the 1DCNN LightGBM neural network. Its framework process mainly includes three steps, namely, data cleaning and data standardization, feature selection, and neural network model construction.

2.1. Data Cleaning and Data Standardization. In the data directly obtained from the oil field, some defective data are often produced due to format conversion, staff operation errors, instrument failures, and other reasons. If these data are not processed, the error of prediction results will become larger and even impossible to calculate. The cleaning of the original data of index prediction of old oil fields mainly includes the following: (1) eliminating the recorded false data, such as the annual value of permeability difference and formation crude oil is 0; (2) eliminating the reservoir data with missing data; for example, there are no records of liquid production rate and oil increase in the current year in the fault block data; (3) eliminate the wrong data of logical verification, etc.

At the same time, normalizing the original data can remove the dimension of different evaluation indexes and the influence of dimension units, solve the comparability problem between data indexes, and is suitable for comprehensive comparative evaluation. The specific normalization formula is

$$x_i^* = \frac{x_i - x_{\min}}{x_{\max} - x_{\min}}, \tag{1}$$

where x_i and x_i^* represent the values before and after normalization, respectively; $x_{\min}$ and $x_{\max}$ represent the minimum and maximum values in the sample data, respectively.

As shown in Figures 1 and 2, the data of the total number of wells is partially missing, and there should be no negative value in the annual remaining recoverable reserves. Considering that the total sample size available from the oilfield is large and the proportion of such data is small, this paper chooses to delete the data containing missing values and abnormal values to avoid affecting the subsequent model training. The normalized results are shown in Figure 3. After normalization, the two indicators are in the same order of magnitude, which is suitable for subsequent comprehensive comparative evaluation.

2.2. Selection of Main Control Factors of Fault Block Reservoir with Random Forest Method. There are many factors affecting the development effect of fault block reservoirs. This paper considers the incremental oil production in the current year after implementing the measures as the index to evaluate the future development effect of the reservoir and selects the main controlling factors for the prediction of fault block reservoir from two aspects of dynamic factors and static factors. Static factors mainly refer to the geological factors related to the physical properties of the reservoir itself, which can be obtained through comprehensive logging and other data, such as the reservoir subclass and the medium depth of the reservoir, which can largely determine the development system of the reservoir and affect the final actual development effect. Dynamic factors are constantly changing factors with the development of the oilfield, which is dynamically updated with the real-time development process of the reservoir, such as oil production rate, annual remaining recoverable reserves, and recovery degree of geological reserves. These indicators can reflect the current development effect and oil reservoir production status of the oilfield from different aspects.

Considering that the dynamic and static factors of the oil field may contain a large amount of redundant information and information irrelevant to the prediction index, which are used as inputs, it will increase the sample dimension, slow down the training speed of the model, and even affect the accuracy of the model. Therefore, feature selection is of great significance to reduce the number of features [10]. Compared with other feature selection methods, the random forest method can output the relative importance of each feature [11]. Selecting the features with high correlation can obtain the preferred features for model training to reduce the complexity of the model and avoid overfitting of the model. In addition, the random forest feature selection method has the characteristics of fast training speed, high precision, and no complex parameter adjustment [12]. Therefore, this paper uses the random forest to rank the importance of dynamic and static factors of high-dimensional fault block

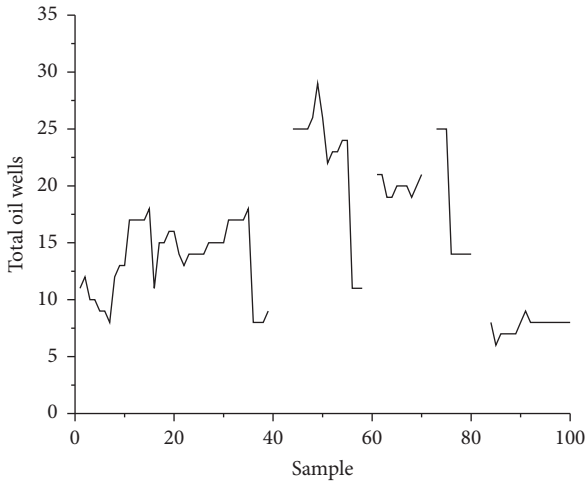


Figure 1: Initial data of total number of wells.

reservoirs and selects the main control factors for further machine learning model training.

2.3. 1DCNN-LightGBM Prediction Model. The research content of this paper is the prediction of measure indexes of fault block reservoirs. Due to the complex geological conditions of fault block reservoirs, traditional methods such as the water drive characteristic curve method and production attenuation method are difficult to accurately analyze the relationship between prediction indexes and various main control factors.

Firstly, the 1DCNN neural network is used to train the characteristics of dynamic data and deeply mine the time characteristics of dynamic data. Then, the integrated learning method LightGBM is used to learn the nonlinear mapping from static data and latent temporal data to incremental oil production. The prediction framework has higher prediction accuracy than the signal model.

2.3.1. 1DCNN Neural Network. The convolutional neural network imitates the construction of a biological visual perception mechanism, which has been applied in various fields. This method processes data by using the convolution layer and pooling layer rather than the full connection layer. Due to the structural characteristics such as local connection, weight sharing, and pooling, CNN is usually used for image classification, fault diagnosis, and image recognition [13].

The one-dimensional convolutional neural network (1DCNN) can be applied to a variety of one-dimensional signals. It has a strong processing ability for time series without complex feature extraction. The structure of 1DCNN can be roughly divided into two parts, namely, the convolution layer and the pooling layer, as shown in Figure 4.

As the core of DCNN, the convolution layer uses the same core to traverse the input with a fixed step size. At each traversal position, the convolution kernel and the neurons of the previous layer perform convolution operation and finally generate the feature map. The specific convolution layer calculation formula is

$$y^{(i,j)} = \omega^* x^{(i,j)} = \sum_m \sum_n x^{(i+m,j+n)} \omega^{(m,n)}, \tag{2}$$

where ω represents the weight of the convolution kernel, and $x^{(i,j)}$ represents the convolution region of the starting point (i, j).

The pooling layer reduces the number of parameters by compressing the extracted features. Therefore, adding a pooling layer can speed up the calculation and prevent overfitting. There are generally two ways of pooling, namely, maximum pooling and mean pooling. The maximum pool is used in this study, which can output the maximum value of neurons in the perception domain. Its calculation formula is shown as follows:

$$y_i = \max_{l \times l} (x_i), \tag{3}$$

where x_i and y_i are input features and output features, respectively, and l is the width scale of the pool area.

2.3.2. LightGBM Model. LightGBM is a new integrated learning model proposed by Microsoft in 2015. The principle of LightGBM is similar to XGBoost, which is based on Taylor expansion of loss function, so the residuals can be approximated [14]. The difference between the two is reflected in the following two aspects:

(1) By using the decision tree strategy of leaf wise (Figure 5), instead of the level-wise strategy used in XGBoost, only the node with the largest splitting gain is selected for splitting, avoiding the waste caused by the small gain of some nodes. However, the number of splits will also increase, making the final decision tree too large. Therefore, it is necessary to set a maximum depth for LightGBM.

(2) LightGBM uses a histogram-based decision tree algorithm. The basic idea is to discretize the continuous features into k integers and construct a histogram with width k. When traversing the data, the discrete value is used as the cumulative index of the statistical data in the histogram. After traversing the data once, the histogram accumulates the required statistical information and then traverses according to the discrete value of the histogram to find the optimal segmentation point. It uses the presorting algorithm used in the accurate algorithm, which reduces the use of memory and improves the training speed of the model.

2.3.3. 1DCNN- LightGBM Hybrid Model. There are two main factors affecting the production of fault block reservoirs, namely, dynamic data and static data. The traditional method only processes them according to the same data and cannot comprehensively extract the implicit relationship between feature data and index data, which often leads to poor prediction results. This paper proposes a fault block

Table 1: Comparison of the accuracy of each model on the training set.

Model	Accuracy (%)
1DCNN	62.8
LightGBM	65.7
1DCNN-LightGBM	**72.3**

The bold value indicates the best one in all results.

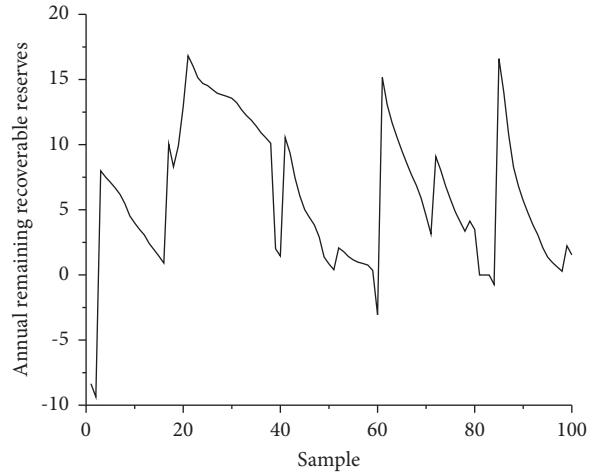


Figure 2: Initial data of annual remaining recoverable reserves.

index prediction method based on the 1DCNN-LightGBM hybrid model, and its model framework is shown in Figure 6.

The method mainly includes the following steps: Firstly, by preprocessing the data set of the fault block reservoir and using the random forest method to obtain the ranking of feature importance, the strong correlation characteristics of measure indicators are obtained. Secondly, the 1DCNN model is used as the feature extractor to extract the features of dynamic data, which are combined with the static features of the data and input into the LightGBM model for training. Finally, the evaluation index is calculated to evaluate the model.

Among them, 1DCNN is mainly used to adaptively extract features from input data. This layer mainly includes two convolution layers and two pooling layers. The pooling layer is the maximum pooling. The extracted feature vector is input to the LightGBM model in the form of time-series data from the last pool layer. At the same time, the static characteristics of the fault block reservoir are also input as the LightGBM model, so that the hybrid algorithm can integrate the dynamic and static information in production data and better predict the index of fault block reservoir.

3. Experimental Study

3.1. Construction of Dynamic and Static Database. In order to verify the effectiveness of the algorithm framework proposed in this paper, we use this algorithm to predict the actual data of an oilfield block. After data preprocessing and removing null values and outliers, a fault-block reservoir sample database was established, including 759 samples. There are 17 dimensions of sample feature variables in the database, and the label that needs to be predicted is the incremental oil production in the current year after implementing measures.

3.2. Feature Selection. The features are selected based on data collected in the dynamic and static databases. By using the random forest method, the relative importance of the initial factors to the labels can be calculated, and the importance between each feature and the output is shown in Figure 7. Here, we select the top 5 factors to construct the main control factor database of fault block oilfield index prediction. The static factors in the main control factors include producing geological reserves and recoverable reserves, and the dynamic factors include the total number of oil wells, the total number of measured wells, and the number of effective wells.

3.3. Model Training and Performance Analysis. The data set after feature selection is randomly divided into training set and test set according to the ratio of 7 : 3. The 5 features selected are input into 1DCNN, LightGBM, and 1DCNN-LightGBM models, respectively.

After training, the prediction results of each model can be obtained, as shown in Figures 8–10. Figure 8 shows the prediction results when the 1DCNN model is used, Figure 9 shows the prediction results of the LightGBM model using the ensemble learning method, and Figure 10 shows the prediction results of the 1DCNN-LightGBM model proposed in this paper. As can be seen from the figure, although the results of the 1DCNN model can predict the general trend, there are cases where the prediction of individual samples is inaccurate; the predicting results of the LightGBM

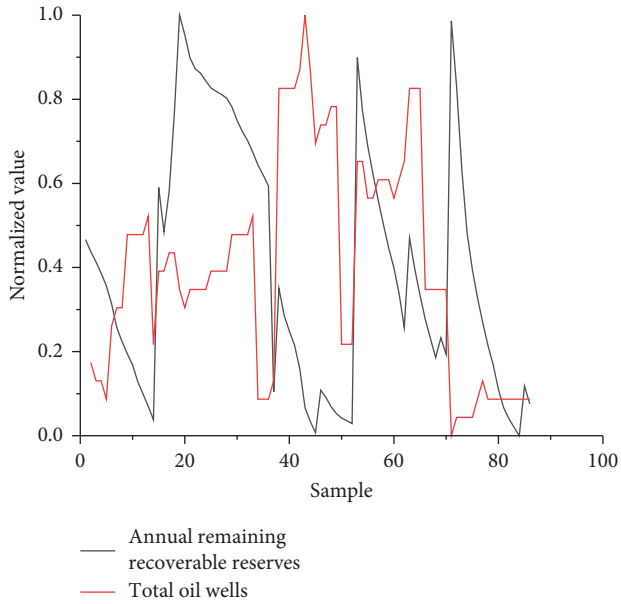


Figure 3: Annual remaining recoverable reserves and total number of oil wellheads after normalization.

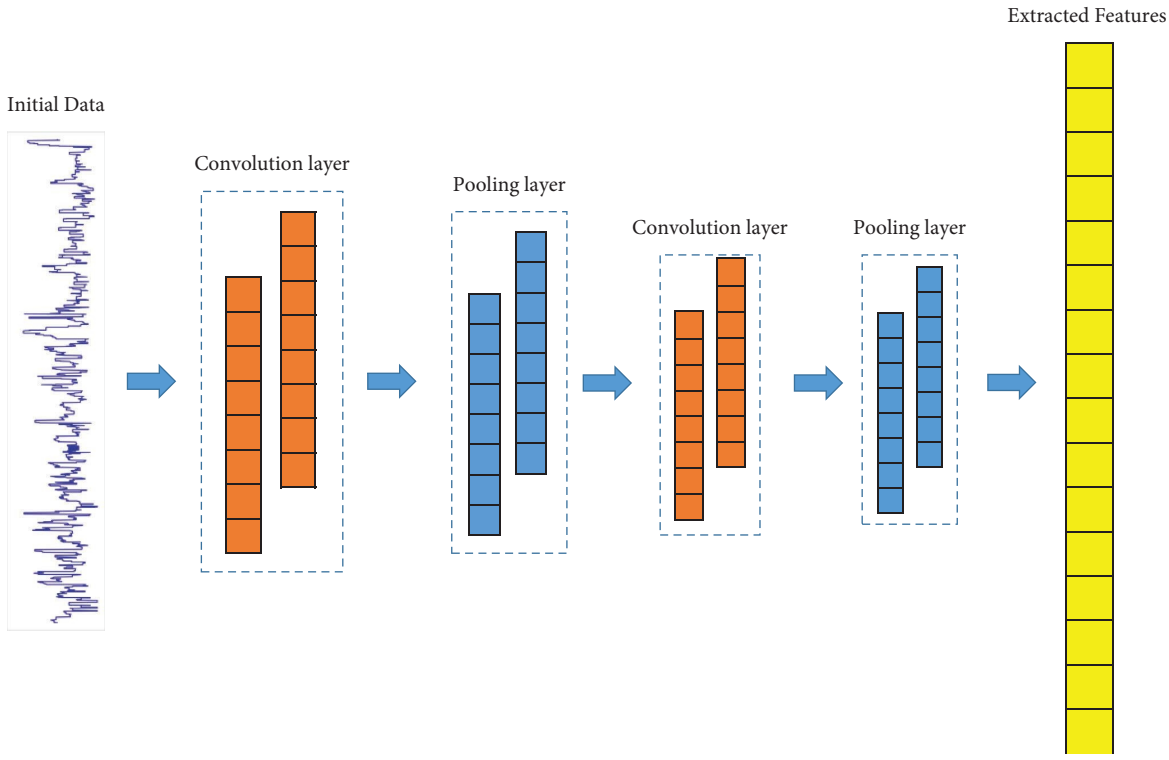


Figure 4: 1D convolutional neural network.

model are better than the results of the 1DCNN model, but the details are still not accurately predicted; in contrast, the prediction results of the 1DCNN-LightGBM model proposed in this paper are more in line with the actual value of the incremental oil production in the current year after implementing measures, showing stronger prediction reliability. In addition, compared with the standard 1DCNN model and the LightGBM model, the 1DCNN-LightGBM hybrid model has higher stability and better adaptability to samples with a wider range of input and output variables. The 1DCNN-LightGBM hybrid model utilizes 1DCNN to extract high-level time series features from dynamic production data and combines it with reservoir static data as the input of ensemble learning, fully excavating the time correlation and data space nonlinear correlation contained in the data. Therefore, compared with other models in the

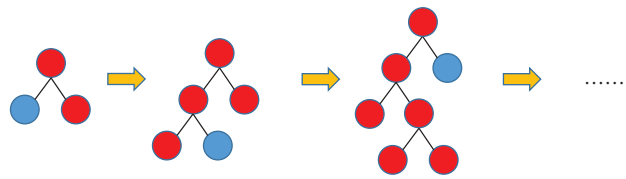


Figure 5: Decision tree strategy based on leaf growth of LightGBM.

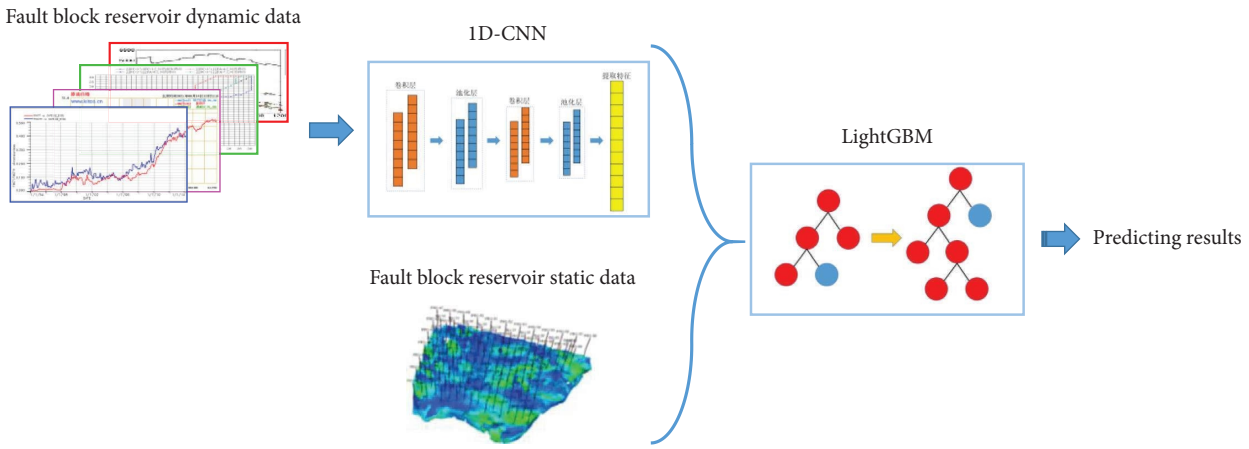


Figure 6: 1DCNN-LightGBM hybrid model structure.

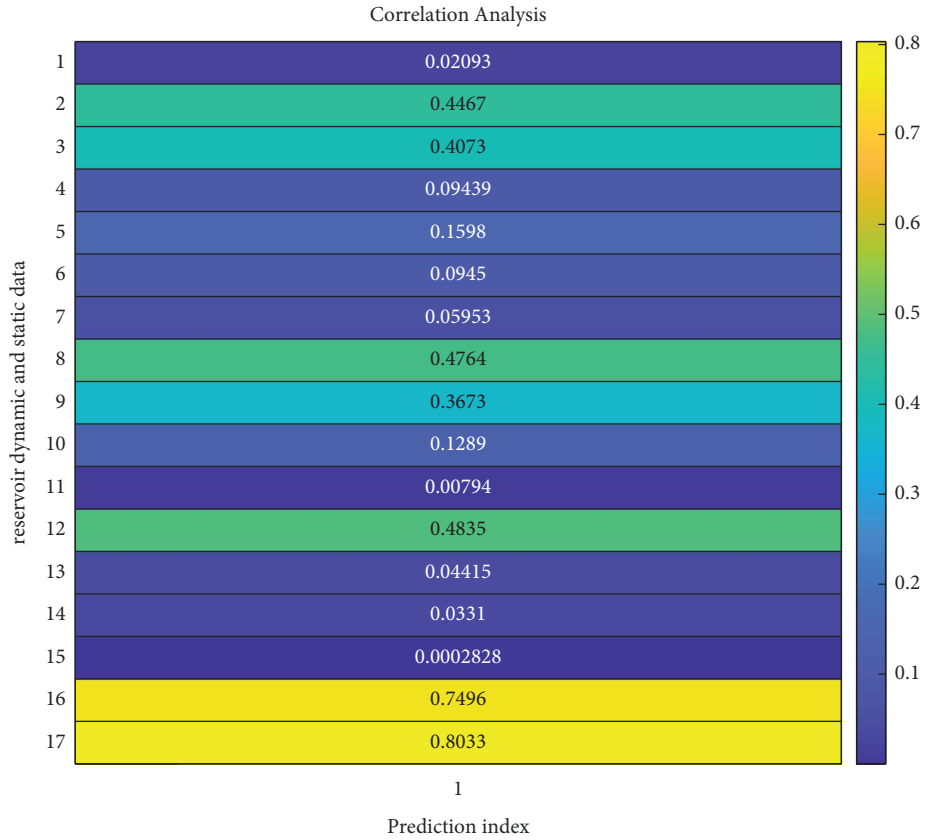


Figure 7: Correlation analysis between dynamic and static factors of fault block and measure index based on random forest method.

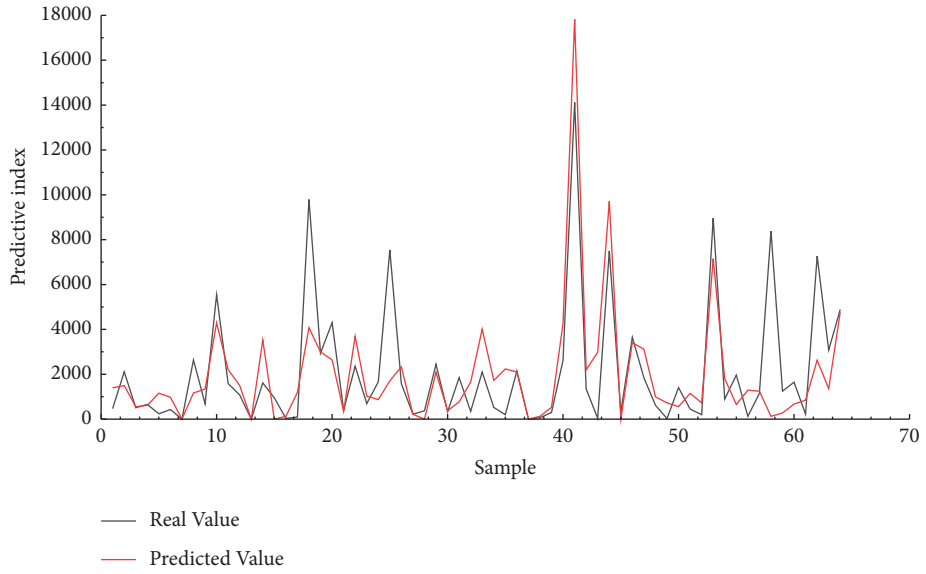


Figure 8: 1DCNN prediction results.

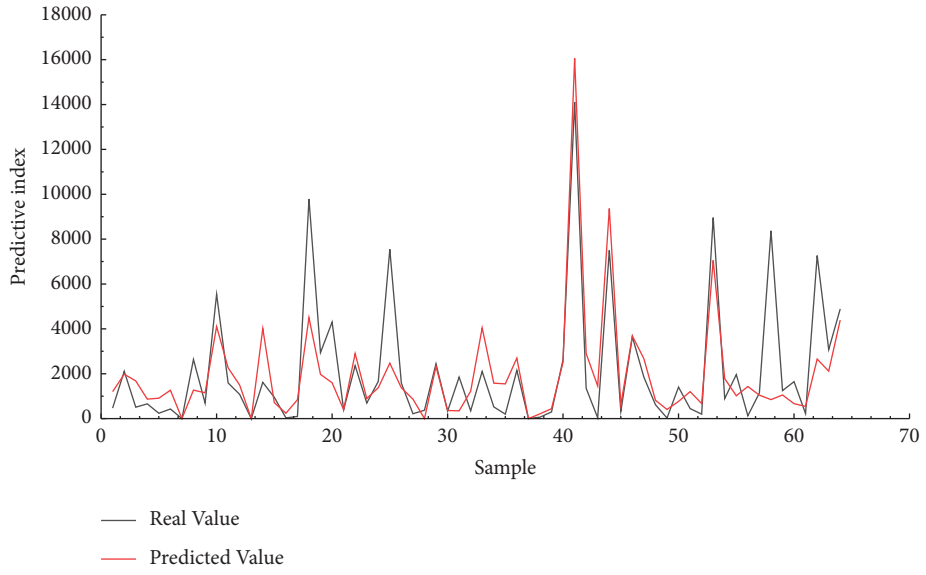


Figure 9: LightGBM prediction results.

combination, the prediction accuracy can be significantly improved.

In order to further verify the effectiveness of the model, each model is trained 10 times with the same data, and the average training accuracy is shown in Table 1. The accuracy rates of 1DCNN, LightGBM, and 1DCNN-LightGBM models are 62.8%, 65.7%, and 72.3%, respectively. The accuracy of 1DCNN-LightGBM is 6.6% higher than that of the LightGBM classification method and 11.5% higher than 1DCNN. It can be seen that the accuracy of the model proposed in this paper is higher than that of the single 1DCNN and LightGBM models. The experimental results can reflect the high potential application value of the 1DCNN-LightGBM algorithm in the prediction of fault-block reservoir indicators, which can extract deep-level information from dynamic and static data.

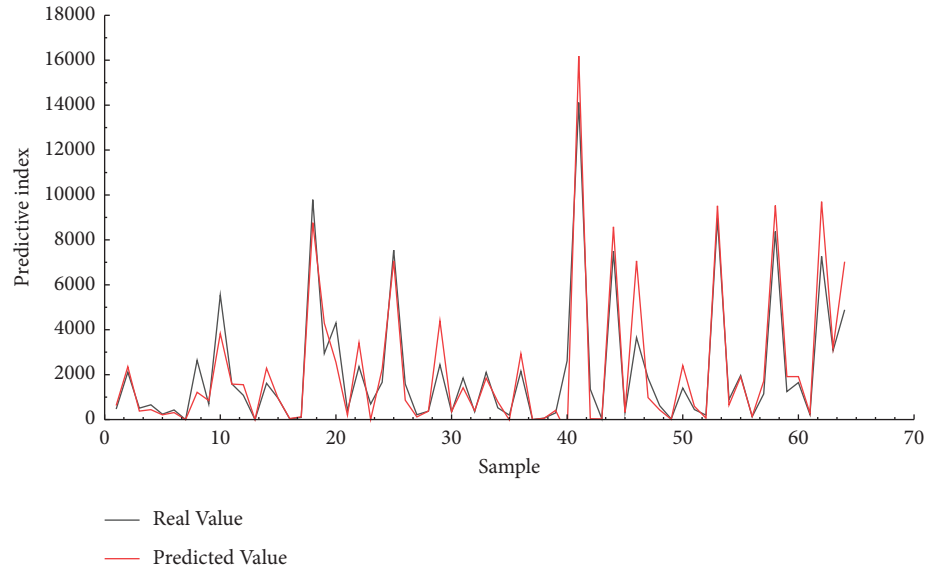


Figure 10: 1DCNN-LightGBM prediction results.

4. Conclusions and Discussion

Based on the characteristics of two 1DCNN and LightGBM, this paper proposes a hybrid prediction model combining LightGBM with 1DCNN. In this paper, the main controlling factor which has the highest correlation with the predictive index is selected by using the random forest method. After that, 1DCNN is adopted to extract time series features from dynamic production data which is then integrated with static data as model input to train LightGBM. The results show that the performance of the 1DCNN-LightGBM model is significantly improved compared with the 1DCNN and LightGBM models. It indicates that the mixed model has good prediction performance, provides a new way for the prediction research of reservoir measures index, and has guiding significance for the formulation of reservoir development plan.

Data Availability

Access to data is restricted as third-party rights.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported by the Research and Application of Old Oil Field Potential Evaluation and Planning Optimization Method of Sinopec (p20070-3).

References

[1] A. Mamghaderi and P. Pourafshary, "Water flooding performance prediction in layered reservoirs using improved capacitance-resistive model," *Journal of Petroleum Science and Engineering*, vol. 108, pp. 107–117, 2013.

[2] H. Zhou, X. Chang, and J. Hao, "Development technique and practice of horizontol wells for complex fault-block reservoirs in Jidong oilfield," in *Proceedings of the International Oil and Gas Conference and Exhibition in China*, Society of Petroleum Engineers, Beijing, China, December 2006.

[3] J. Du, "A method for calculating oil field relative permeability curve by using water drive characteristic curve in high water cut stage," *Journal of Geoscience and Environment Protection*, vol. 10, no. 02, pp. 47–54, 2022.

[4] P. Zhang, H. Chen, Q. Bai, and H. Wei, "New applications of low-frequency absorption attenuation gradient analysis for potential evaluation in oilfield development stage," *SEG International Exposition and Annual Meeting*, San Antonio, TX, USA, 2019.

[5] K. Zhang, X. Zhao, and L. Zhang, "Current status and prospect for the research and application of big data and intelligent optimization methods in oilfield development," *Journal of China University of Petroleum (Edition of Natural Science)*, vol. 44, no. 4, pp. 28–38, 2020.

[6] R. Zhang and H. Jia, "Production performance forecasting method based on multivariate time series and vector autoregressive machine learning model for waterflooding reservoirs," *Petroleum Exploration and Development*, vol. 48, no. 1, pp. 201–211, 2021.

[7] J. Sun, X. Ma, and M. Kazi, "Comparison of decline curve analysis dca with recursive neural networks rnn for production forecast of multiple wells," *SPE Western Regional Meeting*, Garden Grove, CA, USA, 2018.

[8] J. Cai, "Application of least squares support vector machines based on particle swarm optimization in production forecast of shale gas wells," *China Computer & Communication*, no. 13, pp. 35–41, 2019.

[9] B. Lin and J. Guo, "Discussion on current application of artificial intelligence in petroleum industry," *Petroleum Science Bulletin*, vol. 4, pp. 403–413, 2019.

[10] H. Zhang, J. Wang, Z. Sun, J. M. Zurada, and N. R. Pal, "Feature selection for neural networks using group lasso

regularization," *IEEE Transactions on Knowledge and Data Engineering*, vol. 32, no. 4, pp. 659–673, 2020.

[11] X. Li, W. Chen, Q. Zhang, and L. Wu, "Building Auto-Encoder Intrusion Detection System based on random forest feature selection," *Computers & Security*, vol. 95, Article ID 101851, 2020.

[12] Y. Ye, Q. Wu, J. Zhexue Huang, M. K. Ng, and X. Li, "Stratified sampling for feature subspace selection in random forests for high dimensional data," *Pattern Recognition*, vol. 46, no. 3, pp. 769–787, 2013.

[13] L. Li, R. Situ, and J. Gao, "A hybrid model combining convolutional neural network with XGBoost for predicting social media popularity," in *Proceedings of the 25th ACM International Conference on Multimedia*, pp. 1912–1917, New York, NY, USA, October 2017.

[14] C. Bentéjac, A. Csörgő, and G. Martínez-Muñoz, "A comparative analysis of gradient boosting algorithms," *Artificial Intelligence Review*, vol. 54, no. 3, pp. 1937–1967, 2020.

Hindawi
Scientific Programming
Volume 2022, Article ID 9343954, 14 pages
https://doi.org/10.1155/2022/9343954

Research Article

A Fuzzy Comprehensive Evaluation Method on the Construction of High-Level Universities Based on Data from Jiangsu Province

Suoming Huang,[1,2] Yuhao Wang,[2] and Tingqiang Chen[2]

[1]*School of Public Administration, Hohai University, Nanjing 211100, China*
[2]*School of Economics and Management, Nanjing Tech University, Nanjing 211816, China*

Correspondence should be addressed to Tingqiang Chen; tingqiang88888888@163.com

Received 29 May 2022; Accepted 12 July 2022; Published 28 August 2022

Academic Editor: Xiaofeng Xu

To respond to the call for the construction of a "double first-class university" in China and implement the "Focus on innovation, focus on enriching the people, and build a moderately prosperous society at a high level" development strategy. Jiangsu Province issued the first round of the "Jiangsu High-level University Construction Program" in 2016. Taking the deepening of the innovation-driven development strategy as the starting point, the main objective of promoting higher education is to strengthen its ability to support economic and social development. In this study, the fuzzy-ANP comprehensive evaluation model is used to create an index system for building high-level universities with Jiangsu Province features. Using this system, high-level university construction in Jiangsu Province is evaluated. The results of the study show that the first round of Jiangsu's high-level university construction has reached an excellent level and is the most effective in improving the quality of research. In addition, the participating universities are better able to promote socioeconomic development. However, the resources and social reputation of the participating universities are close to the level of excellence. Furthermore, due to the problem of prioritizing research above teaching, the teaching and internationalization levels of the participating universities are rather low.

1. Introduction

In 2021, Chinese Premier Li Keqiang stressed at a symposium on the reform and innovation in higher education that a high level of education is an important reflection of a country's comprehensive competitiveness. Jiangsu Province, a powerful province in China's higher education sector, has taken the lead in initiating the construction of high-level universities in 2016. Through the use of evaluation to promote construction, the first round of the construction of high-level universities in Jiangsu, including universities selected for the "double first-class" construction and provincial universities whose comprehensive operation level is among the top 100 in China, has been successfully completed in 2020. Therefore, an evaluation of the overall performance of the first round of high-level university construction in Jiangsu is not only valuable for Jiangsu Province to summarize its experience and carry out subsequent construction, but also sets a benchmark for the construction of high-level universities in China, and further contributes to the advancement of international higher education. Previous research on university evaluation has mainly focused on the selection of evaluation objects, the construction of evaluation dimensions, and the determination of evaluation criteria. In terms of the selection of evaluation objects, scholars have distinguished institutions of higher education into comprehensive, science, arts, and engineering universities according to the proportion of disciplines, or into research, research-teaching, teaching-research, and teaching universities according to the scale of university research [1]. In addition, there is a group of universities involved in government-led university construction projects. Scholars have selected a certain number of universities of the same type and in the same region to evaluate their discipline development, faculty teams, and research activities [2, 3], which is conducive to an

understanding of the development of universities as a whole. By contrast, others who use colleges within universities to evaluate the research management and financial situation of colleges [4], which can pinpoint the internal problems of universities and facilitate the development of specific optimization measures.

There are two main types of university evaluation dimensions constructed by existing studies: one is the evaluation of a single dimension such as research situation, teaching situation, faculty team, and social service situation [5–8]. The other is to construct a multidimensional evaluation model that starts with multiple dimensions and provides a comprehensive evaluation of the study population [9–11]. The evaluation of a single dimension is relevant and provides a concrete and intuitive reflection of the strengths and weaknesses of the university. But university development is often multidimensional at the same time, and the dimensions can influence each other. As a result, evaluating from a single dimension makes it impossible to account for the influence of other factors on the assessment objectives and to analyze the overall state of university development in a methodical manner. Further research can reveal that academics focus on performance evaluation research from the dimension of the research situation of universities but less on comprehensive evaluation of their teaching and research performance from the perspective of their functions and social contributions.

In terms of determining assessment criteria, the academic community tends to directly assess university performance in terms of quantifiable evaluation dimensions such as research and teaching output, using QS World University Rankings, Times Higher Education (THE) World University Rankings, Shanghai Ranking's Academic Ranking of World Universities (ARWU), and USNEWS rankings in the United States, which have great international influence [12]. The academic community has gradually changed from copying international university evaluation criteria to localizing international standards, adopting international standards in terms of student learning activities, faculty, and research, and integrating domestic evaluation criteria for teaching outputs and social services [13]. Scholars tend to highlight the quality of development, services, and contributions of universities, so different evaluation criteria are applied for different types of universities with different construction goals, and a multi-objective optimization approach can be used in the criteria making [14, 15].

In terms of research methods, scholars have mostly used quantitative analysis methods such as PCA methodology, DEA model, BSC method, MCDM method, QFD method, and Type-2 fuzzy set method to evaluate the performance of universities as a whole [5, 6, 16, 17]. However, at present, the domestic and international research on conducting comprehensive multidimensional evaluations of high-level university groups in provincial areas is rare. In fact, the construction of high-level universities is mostly coordinated by each province. Hence, it is especially important to construct evaluation standards adapted to each province's development needs and to make a comprehensive evaluation of the overall construction of high-level universities in each province. In addition, the construction of universities needs to be assessed in multiple dimensions, but not all dimensions have quantifiable evaluation criteria. Moreover, compared with universities in science and technology, it is difficult to quantify the progress of universities in liberal arts in terms of talent cultivation and teaching output [18].

In light of the above considerations, the index system of the first round of high-level university construction in Jiangsu Province is established in this study using the fuzzy-ANP comprehensive evaluation model and on the basis of High-level University Construction Program in Jiangsu Province and existing literature analysis. In this study, 19 experts and scholars in the field of higher education were invited to participate in the investigation and validation, and the first round of the overall construction of high-level universities in Jiangsu was evaluated on the basis of expert scoring. The rest of this paper is organized as follows. Section 2 describes the setting of the index system. Section 3 introduces the fuzzy-ANP comprehensive evaluation model and analyzes its applicability. Section 4 analyzes the evaluation results. Finally, Section 5 summarizes and elaborates the conclusions of the study.

2. Construction of Evaluation Index System

Educational resources, teaching level, research quality, social reputation, and internationalization level are the five important aspects of assessing high-level universities. For a high-level university, it is the development path for the university to continuously realize its value by improving teaching and internationalization, focusing on research quality, and forming a highly recognized social reputation based on the existing resources [19]. Therefore, in this paper, the secondary indicators involved in the construction of high-level universities in Jiangsu Province are described in five dimensions: educational resources, teaching level, research quality, social reputation, and internationalization level.

2.1. Educational Resources. Educational resources occupy a dominant position in the strategic development of universities. According to the resource dependence theory, the educational resources of universities can be briefly divided into physical, financial, and human resources [20]. Adequate educational resources are the cornerstone of daily teaching in universities [21, 22]. The construction of high-level universities requires sufficient material resources and teaching funds to guarantee basic teaching [23, 24]. At the same time, universities integrate high-quality faculty resources and form teamwork faculty group organizations to carry out teaching and research activities [25]. Therefore, the following three aspects can be considered in improving the educational resources.

2.1.1. Funding Input. The amount of funds invested by universities is the measure of whether the construction of high-level universities in Jiangsu can run efficiently and stably. According to "Provisional Measures for the Management of Comprehensive Award and Subsidy Funds for

the Construction of High-level Universities in Jiangsu Province," for specific scientific research projects, Jiangsu Province implements the form of matching funds subsidy for the construction of universities after the completion of the target requirements. The subsidy is canceled for the projects that do not reach the target. To a certain extent, this drives universities to turn pressure into motivation, on the one hand, to broaden the funding channels of universities and accumulate social donation income by enhancing the reputation of universities, and, on the other hand, to optimize the allocation of education expenditure and focus on the cost efficiency of funding input. According to the Financial System of Universities, this paper subdivides the index into four indicators, including the growth amount of total assets, the total growth amount of four expenses, the growth amount of education expenditure, and the ratio of social donation amount to total income.

2.1.2. Faculty Quality. In constructing high-level universities in Jiangsu, scientific research has gradually developed into a function that distinguishes universities from daily teaching [26, 27]. In this context, a scientific teacher-student ratio can provide solid faculty support for students [28]. In addition, faculty construction should be balanced between quantity and quality, and the percentage of teachers with senior faculty titles, the proportion of teachers with Ph.D. degrees, and the number of distinguished talents can reflect the construction of university faculty to some extent. Accordingly, this paper incorporates four indicators into the evaluation system; these indexes are the faculty-student ratio, the proportion of teachers with senior titles to the total number of teachers, the number of outstanding talents, and the proportion of teachers with doctoral degrees.

2.2. Teaching Level. "The Outline of National Medium and Long-term Education Reform and Development Plan (2010–2010)" states that universities should improve the quality of talent cultivation in all aspects. Additionally, teaching is the primary content of teachers' assessment, and quality teaching level is the guarantee for promoting high-quality development of higher education [29]. Therefore, this study reflects the teaching level in the following three aspects.

2.2.1. The Situation of University Discipline Construction. According to the "Implementation Plan of Jiangsu University Advantageous Discipline Construction Project," the participating universities of Jiangsu high-level university construction should strengthen the development of key disciplines and improve the overall teaching quality of the university through discipline development. The academic discipline construction results reflect how closely the high-level university construction is integrated with social development. Therefore, the participating universities of Jiangsu high-level university construction should realize the dynamic adjustment of degree points. Doctoral and master's degree points are flexibly increased or decreased according to the development prospect of related industries and the needs for social and economic development. The situation of university discipline construction should be subdivided into six indexes to evaluate high-level university construction in Jiangsu Province; these indexes are proportion of doctoral teachers, increase of doctoral points, increase of master's points, increase of national key disciplines, increase of provincial and ministerial key disciplines, and increase of disciplines ranked in the top 1% of ESI.

2.2.2. Quality of Student Source and Cultivation. Cultivating high-quality talents requires the support of high-quality students [30]. The quality of cultivation of high-quality students is a direct reflection of the effectiveness of high-level university construction. Jiangsu high-level university construction project requires participating universities to innovate talent cultivation mode, optimize, and reform curriculum and teaching methods. We encourage universities to pursue cultivating quality rather than cultivating quantity [31]. It seeks to cultivate high-level comprehensive talents at multiple levels through high-level university construction projects. Quality of student source and cultivation should be subdivided into seven indexes to evaluate high-level university construction in Jiangsu Province; these indexes are the average score of undergraduate freshmen, the ratio of graduate students to undergraduate students, the ratio of international students to undergraduate students, the number of national 100 outstanding doctoral dissertations, the number of provincial outstanding undergraduate dissertations and master's degree dissertations, the number of student innovation competition awards, and the number of international and domestic accredited undergraduate majors.

2.2.3. Teaching Output. The talent and intellectual advantages of an excellent teaching team are reflected through teaching output. Excellent teaching output can, on the one hand, reflect the solid teaching foundation function of universities and attract excellent students. On the other hand, it can concentrate resources and attract social and government resources to flow into discipline construction and professional construction, forming a virtuous circle and enhancing the core competitiveness of universities [32]. According to the high-quality teaching resources ensemble of the Ministry of Education, teaching output should be subdivided into six indexes to evaluate high-level university construction in Jiangsu Province; these indexes are the number of provincial and ministerial teaching achievement awards, the number of high-quality courses of the Ministry of Education, the number of provincial and ministerial teaching projects, the number of provincial and ministerial teaching platforms, the number of provincial and ministerial teaching courses, and the number of provincial and ministerial teaching materials.

2.3. Research Quality. According to the "National Medium- and Long-term Education Reform and Development Plan (2010–2010)," universities should improve scientific

research and are required to improve the innovation and quality-oriented scientific research evaluation mechanism. Despite the fact that certain parts of creating a top-tier institution have competing goals, the primary task of scientific research evaluation is to measure the scientific and social values of scientific research activities [7, 15]. Accordingly, this study improves the research quality from the following three aspects.

2.3.1. Scientific Research Input. Universities are resource-dependent in conducting scientific research activities and need to absorb external research resources continuously. Generally speaking, universities obtain research investment funds through government financial education funding income, tuition income from research grants, income from the transformation of research results, income from social donations, and fund investment and bank loans [7, 23, 33]. In recent years, under the policy background of the Chinese government requiring universities to expand their enrollment, the cost of running universities has increased, and the ceiling of scientific research investment determines the ceiling of scientific research achievements in universities to a certain extent. Therefore, scientific research input should be subdivided into three indexes to evaluate high-level university construction in Jiangsu Province; these indexes are the total amount of scientific research funds, the incremental number of entrusted scientific research funds undertaken by enterprises and institutions, and the incremental number of entrusted scientific research projects undertaken by enterprises and institutions.

2.3.2. Scientific Research Output. Under the rated research investment, the research results achieved by universities with coordinated funding and manpower allocation are measured by research output indicators. The number of authorized patents in universities is an important indicator to measure the degree of close integration between universities and industry [34, 35]. Universities strengthen the transformation of scientific research results into patent R&D, control the quality of patents, and improve the application rate of patents, which can ultimately improve the efficiency of university scientific research output in serving economic and social development. Therefore, scientific research output should be subdivided into six indexes to evaluate high-level university construction in Jiangsu Province; these indexes are the total number of authorized patents, the total number of international and domestic high-level scientific research papers, the total number of national- and provincial-level scientific research platforms, the total number of national- and provincial-level scientific research projects, the number of national- and provincial-level scientific research awards, and the number of international major awards.

2.3.3. Scientific Research Output Efficiency. On February 20, 2020, the Chinese Ministry of Education and Ministry of Science and Technology issued "Several Opinions on Regulating the Use of SCI Paper-related Indicators in Higher Education Institutions to Establish Correct Evaluation Guidance," which emphasized that scientific research activities in universities should be transformed from focusing on quantity to focusing on quality. In this background, participating universities should pay more attention to the practical application value of scientific research results. The number of citations of high-level theses is not only a sign of the transformation of scientific research output in the field but also a direct reflection of the quality of theses and the value of the contribution of the construction universities in each research field [36]. Moreover, according to "Measures for Sample Inspection of Undergraduate Theses (for trial implementation)," the results of sample inspection of dissertations directly affect the allocation of educational resources. In addition, the patent conversion rate reflects the practical application efficiency of the scientific research output of universities. Therefore, in this paper, scientific research output efficiency should be subdivided into three indexes to evaluate high-level university construction in Jiangsu Province; these indexes are the total number of citations of international and domestic high-level scientific research papers, passing rate of undergraduate and postgraduate dissertation sampling inspection, and patent conversion rate.

2.4. Social Reputation. Social reputation is one of the core competencies of high-level universities, which largely determines the future development of universities. Excellent social reputation not only attracts high-quality students and high-level teachers to universities but also expands the endowment income for universities. Therefore, this study is carried out through the following two aspects.

2.4.1. Service Contribution. Promoting social development is one of the construction goals of colleges and universities, and the construction of high-level universities requires colleges and universities to cultivate outstanding talents who can promote social development and serve as role models. The service contribution of outstanding alumni to the society can highlight the effectiveness of moral education work of colleges and universities, and the reputation effect brought by them is of great significance for colleges and universities to enhance social recognition [37]. Meanwhile, the approval of provincial and ministerial leaders and departments to adopt the results of university decision-making consultation is also a channel for universities to solve social problems and provide decision support for social development [38]. In addition, under the policy background of "stable employment" and "employment preservation" in Jiangsu, the construction of universities to deliver high-level talents to the society is also a reflection of their service contribution. Therefore, service contribution should be subdivided into three indexes to evaluate high-level university construction in Jiangsu Province; these indexes are the number of distinguished alumni, the number of instructions adopted by provincial and ministerial leaders and departments, and the proportion of graduates employed in Jiangsu.

2.4.2. Social Ranking. Universities vary significantly across the country due to their history of establishment and regional cultural traditions. Correspondingly, the rankings of universities in different countries and regions have their own focus and specificity [39–41]. However, internationally recognized ranking rankings can still reflect some of the problems of university development and have significant implications for building high-level universities [42]. This index system selects four highly recognized international rankings to evaluate high-level university construction in Jiangsu Province; these indexes are changes in Times Higher Education World University Rankings, changes in QS World University Rankings, changes in Softbank Academic Ranking of World Universities, and changes in USNEWS Rankings.

2.5. Internationalization Level. The internationalization level is an important way for universities to benchmark with the world's top universities and to enhance their international reputation. Moreover, the international exchange and cooperation of high-level universities is on the one hand conducive to the integration of resources among universities and the realization of complementary advantages. On the other hand, it can achieve both the localization of high-level university construction and high-quality development with an international vision. Therefore, international communication and cooperation is included as an evaluation index in this study.

2.5.1. International Cooperation and Communication. Promoting universities to improve their internationalization level and achieve world class is the goal of the construction of high-level universities in Jiangsu. The internationalization level of universities determines their influence and discourse in the international society [43]. The construction of Jiangsu high-level university should improve the internationalization collaborative development mechanism, integrate the internationalization concept and strategy into all elements such as departments and faculties as well as teachers, students, and managers within the university, and actively establish a smooth international exchange channel for teachers, students, and foreign introduced talents. For this reason, this index system selects five indexes to reflect the internationalization level comprehensively; these indexes are the number of international cooperation research platforms, the number of international conferences held, the internationalization level of faculty, the number of international joint training programs for students, and the proportion of students studying and visiting abroad.

By combining the experts' opinion (The Delphi expert group consists of 21 experts in the field of higher education. Three rounds of expert opinions were solicited, 21 questionnaires were distributed in each round, and 21 valid questionnaires were returned in each round. After multiple rounds of evaluation and scoring by experts, 54 tertiary indicators were identified and then divided into 12 secondary indicators and 5 primary indicators based on the correlation between indicators.) using the Delphi method, the index system of the first round of high-level university construction in Jiangsu Province is constructed and is shown in Table 1.

3. Evaluation Model

The ANP is a decision analysis method that combines quantitative and qualitative analysis. Based on the analytic hierarchy process (AHP), feedback and dependency relationships between levels and internal elements are considered [44], while the fuzzy comprehensive evaluation method, which is based on fuzzy mathematics' membership theory, effectively examines qualitative indexes [45, 46]. Therefore, the fuzzy-ANP comprehensive evaluation model provides numerous advantages in terms of evaluation and analysis. In the process of construction of Jiangsu high-level university, the construction subjects exhibit a clear hierarchy. In other words, all aspects of university construction present obvious dependency relationships. Therefore, the influencing factors of Jiangsu high-level university construction have typical hierarchical and dependency relationships. Therefore, its index system forms an organic whole with hierarchical network structure. Some indicators cannot be quantified because they present a fuzzy character in terms of values. Therefore, it is highly scientific and applicable to evaluate the construction of Jiangsu high-level universities by using the fuzzy-ANP comprehensive evaluation model.

3.1. Constructing the Network Structure of ANP. In the network structure of Jiangsu high-level university construction, the control layer contains the target and criteria; the target is A, and the criteria are the first-level indexes of the index system, including $B1$, $B2$, $B3$, $B4$, and $B5$. The network layer includes twelve sets of elements that correspond to second- and third-level indexes of the index system. These sets are $C01$, $C02$, $C03$, $C04$, $C05$, $C06$, $C07$, $C08$, $C09$, $C10$, $C11$, and $C12$. Among them, both the elements and the elements may interact and influence each other, according to which the network structure of ANP is constructed as in Figure 1.

3.2. Determination of the Index Weight. The weights of each indicator were determined with the help of ANP and Delphi expert scoring. The specific process is as follows.

Step 1. build a super matrix. The interactions between the network and control layers are evaluated, the internal relationships between the set of elements and the elements are determined, and the corresponding weights are assigned to the control layer guidelines. The importance of the indicators and guidelines is determined by fixed guidelines while ensuring the independence between the elements. As shown in Table 2, the importance levels among the indicators, using the scale method of 1–9, were compared, after which the AHP method was applied to obtain the weights.

Then, the unweighted supermatrix and the indicator element weights are obtained. The normalized eigenvector values are able to be obtained by the eigenroot method; based on the two-by-two judgment matrix, the super matrix W is obtained as follows:

The limit matrix of the weighted supermatrix $\overline{W}$ is obtained by taking infinite powers of the weighted

Table 1: Index system of the first round of high-level university construction in Jiangsu province.

Objective	First-level index	Second-level index	Third-level index
The first round of high-level university construction in Jiangsu Province (A)	Educational resources (B_1)	Material resources (C_{01})	School space per student (C_{011})
			Number of books per student (C_{012})
			Number of repositories and databases (C_{013})
		Funding input (C_{02})	Total asset growth (C_{021})
			Total increase of the four provisions (C_{022})
			Growth in education spending (C_{023})
			Social donation amount to total income ratio (C_{024})
		Faculty quality (C_{03})	Teacher-to-student ratio (C_{031})
			The proportion of senior title teachers to the total number of teachers (C_{032})
			Number of outstanding talents (C_{033})
			Percentage of faculty with doctoral degrees (C_{034})
	Teaching level (B_2)	The situation of university discipline construction (C_{04})	Increases in the number of doctoral sites (C_{041})
			Increase in the number of master's points (C_{042})
			Number of national key disciplines increased (C_{043})
			Increase in the number of key disciplines at provincial and ministerial level (C_{044})
			Incremental number of disciplines in the top 1% of the global ESI ranking (C_{045})
		Quality of student source and cultivation (C_{05})	Average undergraduate Entrance score (C_{051})
			Number of graduate students as a percentage of the number of undergraduate students (C_{052})
			Number of international students as a percentage of the number of undergraduate students (C_{053})
			Number of national 100 excellent doctoral theses (C_{054})
			Number of provincial excellent undergraduate thesis and master's thesis (C_{055})
			Number of student innovation competition awards (C_{056})
			Number of internationally and nationally accredited undergraduate programs (C_{057})
		Teaching output (C_{06})	Number of provincial and ministerial teaching Achievement awards for higher Education (C_{061})
			Number of high-quality courses of the Ministry of Education (C_{062})
			Number of provincial- and ministerial-level teaching projects (C_{063})
			Number of provincial- and ministerial-level teaching platforms (C_{064})
			Number of provincial- and ministerial-level teaching courses (C_{065})
			Number of teaching materials at provincial and ministerial level (C_{066})
	Research quality (B_3)	Scientific research input (C_{07})	Total scientific research funding (C_{071})
			Incremental amount of research funds entrusted by enterprises and institutions (C_{072})
			Increase in the number of research projects entrusted by enterprises and institutions (C_{073})
		Scientific research output (C_{08})	Total number of licensed patents (C_{081})
			Total number of international and domestic high-level scientific papers (C_{082})
			Total number of national-level and provincial- and ministerial-level research platforms (C_{083})
			Total number of scientific research projects at national and provincial levels (C_{084})
			Number of national and provincial scientific research awards (C_{085})
			Number of major international awards received (C_{086})
		Scientific research output efficiency (C_{09})	Total number of citations of international and domestic high-level scientific papers (C_{091})
			Undergraduate dissertation sample inspection pass rate (C_{092})
			Pass rate of graduate student dissertation sample inspection (C_{093})
			Authorized patent conversion rate (C_{094})
	Social reputation (B_4)	Service contribution (C_{10})	Number of outstanding alumni (C_{101})
			Percentage of graduates employed in Jiangsu (C_{102})
			The number of provincial and ministerial leaders and departmental instructions to adopt (C_{103})
		Social ranking (C_{11})	Progressing places in Times Higher Education World University Rankings (C_{111})
			Progressing places in QS World University Rankings (C_{112})
			Progressing places in Softbank Academic Ranking of World Universities (C_{113})
			Progressing places in NSNEWS Rankings (C_{114})
	Internationalization level (B_5)	International communication and cooperation (C_{12})	Number of international cooperative research platforms (C_{121})
			Number of international conferences held (C_{122})
			Internationalization level of the faculty (C_{123})
			Number of international joint training programs for students (C_{124})
			Percentage of students studying and visiting abroad (C_{125})

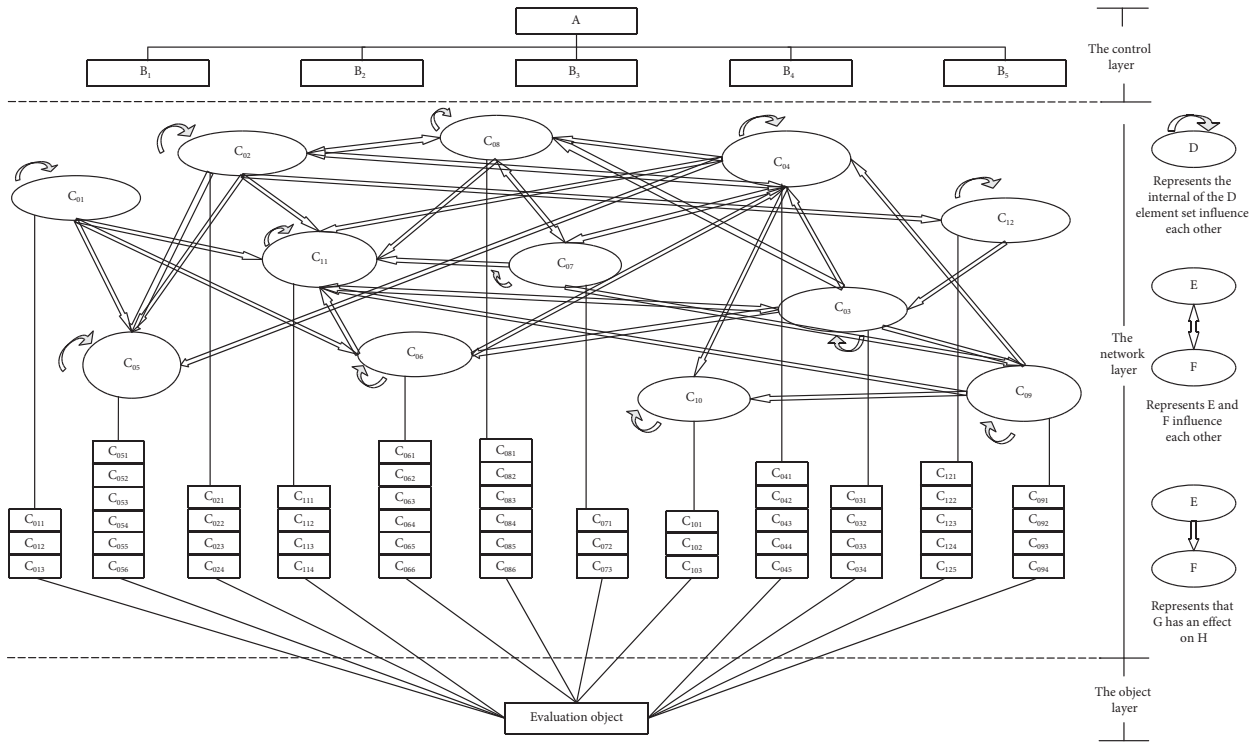


Figure 1: Network structure of ANP of Jiangsu high-level university construction.

Table 2: Scale method of 1–9 in comparison.

Scale	Definition
1	Element i is of equal importance to element j
3	Element i is slightly more important than element j
5	Element i is more important than element j
7	The i element is very important than the j element
9	The i element is definitely more important than the j element
2, 4, 6, 8	The middle value of the above adjacent judgments
Reciprocal	The importance scale of element j to element i

supermatrix W^{∞}. When $i \longrightarrow \infty$ and the limit converges uniquely, the column vectors in the limit matrix W^{∞} are the weights of each participant index.

Step 2. construct the weighted matrix and weighted super matrix. Under the criterion P_i, the importance of the relative criterion of element $C_j\,(j = 1, 2 \ldots, N)$ is compared to obtain a normalized row sequence vector $(a_{1j}, a_{2j}, \ldots, a_{nj})$ such as that a weighted matrix is obtained as follows:

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1N} \\ a_{21} & a_{22} & \cdots & a_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ a_{N1} & a_{N2} & \cdots & a_{NN} \end{bmatrix} \left(a_{ij} \in [0,1] \text{ and } \sum^{N} a_{ij} = 1 \right). \quad (1)$$

If the two elements have no interaction with each other, then $a_{ij} = 0$. The weighted super matrix is constructed as follows:

$$\overline{W} = \overline{W}_{ij} = A \times W = \left(a_{ij} \times W_{ij}\right) (i = 1, 2 \ldots, N, j = 1, 2 \ldots, N). \quad (2)$$

3.3. Determination of Evaluation Rating and Rules. The evaluation rating is assumed as follows: $v = (v_1, v_2, \ldots v_N)$ $(N = 1, 2, \ldots, n)$. The evaluation rating is divided into four grades: good, general, relatively poor, and poor.

The fuzzy relation matrix is defined as follows:

$$R = \left(r_{ij}\right)_{N \times M} = \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1M} \\ r_{21} & r_{22} & \cdots & r_{2M} \\ \vdots & \vdots & \ddots & \vdots \\ r_{N1} & r_{N2} & \cdots & r_{NM} \end{bmatrix}. \quad (3)$$

Among them, R is a member of the index number i of the level number j, r_{ij} = the number of the indicator number i to select the level v_i/the number of participating evaluation.

3.4. Determination of the Comprehensive Evaluation Level. A comprehensive evaluation vector is established by the fuzzy comprehensive operation of the fuzzy relation matrix of the set of weights.

$$S_i = W_i \times R = (w_{i1}, w_{i2}, \ldots, w_{iN}) \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1M} \\ r_{21} & r_{22} & \cdots & r_{2M} \\ \vdots & \vdots & \ddots & \vdots \\ r_{N1} & r_{N2} & \cdots & r_{NM} \end{bmatrix} \quad (4)$$

$$= (s_1, s_2, \ldots, s_M).$$

In the comprehensive evaluation vector of Jiangsu high-level university construction, it equals to 4. Therefore, the final score obtained by the weighted average method is $H = 1 \times s_1 + 0.8 \times s_2 + 0.6 \times s_3 + 0.2 \times s_4$.

4. Empirical Research

4.1. Weight Calculation. The weights of the indexes are calculated using the above weight calculation method and the super decision software, as shown in Table 3.

Table 3 shows that, among the first-level indicators, the indicator with the largest proportion of weight is research quality, with 58.8%. The next in order is 17.5% for educational resources, followed by 12.2% for teaching level, 6.8% for social reputation, and 4.6% for internationalization level. It can be seen that the index of research quality has the greatest influence on the performance evaluation of the construction of high-level universities in Jiangsu.

In addition, according to Table 3, it can be found that among the second-level indicators, the discipline construction situation has the highest weight of 19.9%, followed by research output of 16.5% and research input of 12%. The lowest weight is service contribution of 0.8%. This indicates that discipline construction, research output, and research input are the main factors affecting the performance evaluation of Jiangsu high-level university construction. By comparison, the weight of service contribution, international cooperation and exchange, social ranking, and material resources are less than 5%. This indicates that these indicators have a small influence on the performance evaluation. On the other hand, the low weight of service contribution indicates that universities can provide decision-making advice to government departments, but it has limited effect on the construction of universities themselves. Moreover, the low weight of international cooperation and exchange and social ranking reflects that the construction of high-level universities in Jiangsu Province is in line with the influential international standards but not limited by them. It also reflects that the construction of Jiangsu high-level universities refers to social ranking but does not blindly pursue the ranking.

Furthermore, based on the weighting of the three-level indicators in Table 4, it can be stated that the indicators with the highest weighting are primarily concerned with evaluating teaching quality and research quality. The number of disciplines in the top 1% of the global ESI ranking, with a weighted of 10.4 percent, has the largest impact on performance evaluation. The number of major international awards won (8.9%), the overall amount of research money (8.2%), the total number of citations of foreign and domestic high-level research publications (7.7%), and the number of national important disciplines (5.3%) are the next most influential indicators.

4.2. Fuzzy Comprehensive Calculation. Selecting a fuzzy measure set $H = \{\text{Good}(H_1), \text{General}(H_2), \text{Relatively poor}(H_3), \text{Poor}(H_4)\}$, its fuzzy measure benchmark value is $p(H) = \{p(H_1), p(H_2), p(H_3), p(H_4)\} = \{1, 0.8, 0.6, 0.2\}$.

Assuming that the decision maker's preference coefficient (Decision maker's preference coefficient reflects the degree of decision maker's preference for the upper and lower bound utility values of the indicator interval, representing whether the decision maker values the indicator more highly or less highly.) $\alpha = 0.9$, the normalized values of the first-level index weights are obtained by normalizing the indicators in Tables 3 and 4 as $\lambda = \{0.17526, 0.12209, 0.58827, 0.06832, 0.04606\}$. The normalized values of the second-level index weights are $\lambda_1 = \{0.08110, 0.34200, 0.57690\}$, $\lambda_2 = \{0.58155, 0.10945, 0.30900\}$, $\lambda_3 = \{0.07862, 0.26275, 0.65863\}$, $\lambda_4 = \{0.66667, 0.33333\}$, $\lambda_5 = \{1\}$.

After the data statistical results were validated by experts, the initial evidence credibility of the second-level indicators of the performance evaluation of the construction of high-level universities in Jiangsu was given. The results of the initial evidence credibility assignment of the second-level indexes are shown in Table 4. The elements in each row of Table 4 are first multiplied by the normalized weight corresponding to each second-level index to obtain the basic credibility of the second-level indexes. Then, the second-level indexes, first-level indexes, and the comprehensive evaluation score of the construction performance of Jiangsu high-level universities were further calculated, which are shown in Tables 5–7.

The final result of the comprehensive evaluation of the construction performance of Jiangsu high-level universities is $H = \sum_{q=1}^{4} m^q p(H_q) = 1 \times 0.61867 + 0.8 \times 0.28507 + 0.6 \times 0.09490 + 0.2 \times 0.00136 = 0.90394$.

Due to the result $0.8 < 0.90394 < 1$, the performance grade of Jiangsu high-level university construction is obtained between general and good; that is, the construction of Jiangsu high-level university basically reaches the good level on the whole.

The data results in Table 6 show that among the five dimensions of the construction of Jiangsu high-level universities, the evaluation score of research quality is the highest, and the construction results reach a good level. Next, in order, are educational resources, social reputation, teaching level, and internationalization level, and the construction results reach the general level. The scores in Table 6 show that the universities participating in the construction of Jiangsu high-level universities have achieved remarkable results in improving the quality of scientific research, although they have made promising achievements in educational resources, social reputation, teaching level, and internationalization level, but there is still some room for improvement. This necessitates a greater integration of resources, a faster pace of building, and an overall improvement in the overall effect of Jiangsu high-level university construction.

Table 3: Index weight.

First-level index	Weight	Second-level index	Weight	Third-level index	Weight
B_1	0.175	C_{01}	0.081	C_{011}	0.143
				C_{012}	0.286
				C_{013}	0.571
		C_{02}	0.342	C_{021}	0.066
				C_{022}	0.231
				C_{023}	0.582
				C_{024}	0.12
		C_{03}	0.577	C_{031}	0.118
				C_{032}	0.047
				C_{033}	0.513
				C_{034}	0.322
B_2	0.122	C_{04}	0.582	C_{041}	0.064
				C_{042}	0.048
				C_{043}	0.266
				C_{044}	0.098
				C_{045}	0.524
		C_{05}	0.109	C_{051}	0.040
				C_{052}	0.220
				C_{053}	0.066
				C_{054}	0.051
				C_{055}	0.257
				C_{056}	0.107
				C_{057}	0.259
		C_{06}	0.309	C_{061}	0.167
				C_{062}	0.093
				C_{063}	0.133
				C_{064}	0.301
				C_{065}	0.228
				C_{066}	0.079
B_3	0.588	C_{07}	0.079	C_{071}	0.745
				C_{072}	0.099
				C_{073}	0.156
		C_{08}	0.263	C_{081}	0.166
				C_{082}	0.109
				C_{083}	0.057
				C_{084}	0.041
				C_{085}	0.089
				C_{086}	0.538
		C_{09}	0.659	C_{091}	0.638
				C_{092}	0.178
				C_{093}	0.126
				C_{094}	0.058
B_4	0.068	C_{10}	0.667	C_{101}	0.717
				C_{102}	0.066
				C_{103}	0.217
		C_{11}	0.333	C_{111}	0.423
				C_{112}	0.174
				C_{113}	0.266
				C_{114}	0.137
B_5	0.046	C_{12}	1	C_{121}	0.412
				C_{122}	0.157
				C_{123}	0.261
				C_{124}	0.077
				C_{125}	0.093

As can be seen from Tables 4 and 5, the secondary indexes that reach the good level are research input, research output efficiency, research output, faculty quality, and funding input in order. The high score of scientific research input is mainly attributed to the expansion of the total amount of scientific research funds of the participating universities and the increase of research funds entrusted by enterprises and institutions. The total number of international and domestic high-level scientific research papers cited, the conversion rate of authorized patents, and the

Table 4: Initial evidence credibility results of the first-level index.

First-level index	Second-level index	Third-level index	Initial evidence credibility results of the first-level index			
			Good	General	Relatively poor	Poor
Educational resource	Material resources	C_{011}	0.42105	0.31579	0.21053	0.05263
		C_{012}	0.36842	0.42105	0.21053	0.00000
		C_{013}	0.68421	0.21053	0.10526	0.00000
	Funding input	C_{021}	0.47368	0.42105	0.10526	0.00000
		C_{022}	0.52632	0.42105	0.05263	0.00000
		C_{023}	0.63158	0.36842	0.00000	0.00000
		C_{024}	0.31579	0.36842	0.31579	0.00000
	Faculty quality	C_{031}	0.52632	0.42105	0.05263	0.00000
		C_{032}	0.52632	0.26316	0.21053	0.00000
		C_{033}	0.57895	0.42105	0.00000	0.00000
		C_{034}	0.52632	0.36842	0.10526	0.00000
Teaching level	The situation of university discipline construction	C_{041}	0.52632	0.26316	0.21053	0.00000
		C_{042}	0.47368	0.31579	0.21053	0.00000
		C_{043}	0.63158	0.21053	0.15789	0.00000
		C_{044}	0.52632	0.31579	0.15789	0.00000
		C_{045}	0.52632	0.36842	0.10526	0.00000
	Quality of student source and cultivation	C_{051}	0.52632	0.31579	0.10526	0.05263
		C_{052}	0.31579	0.21053	0.42105	0.05263
		C_{053}	0.05263	0.31579	0.47368	0.15789
		C_{054}	0.42105	0.36842	0.21053	0.00000
		C_{055}	0.52632	0.26316	0.21053	0.00000
		C_{056}	0.57895	0.26316	0.15789	0.00000
		C_{057}	0.47368	0.21053	0.31579	0.00000
	Teaching output	C_{061}	0.57895	0.26316	0.15789	0.00000
		C_{062}	0.52632	0.26316	0.21053	0.00000
		C_{063}	0.47368	0.26316	0.26316	0.00000
		C_{064}	0.42105	0.31579	0.26316	0.00000
		C_{065}	0.47368	0.26316	0.26316	0.00000
		C_{066}	0.52632	0.15789	0.31579	0.00000
Research quality	Scientific research input	C_{071}	0.73684	0.26316	0.00000	0.00000
		C_{072}	0.57895	0.36842	0.05263	0.00000
		C_{073}	0.42105	0.47368	0.10526	0.00000
	Scientific research output	C_{081}	0.52632	0.47368	0.00000	0.00000
		C_{082}	0.78947	0.05263	0.15789	0.00000
		C_{083}	0.73684	0.15789	0.10526	0.00000
		C_{084}	0.73684	0.10526	0.15789	0.00000
		C_{085}	0.68421	0.15789	0.15789	0.00000
		C_{086}	0.57895	0.31579	0.10526	0.00000
	Scientific research output efficiency	C_{091}	0.73684	0.21053	0.05263	0.00000
		C_{092}	0.63158	0.26316	0.10526	0.00000
		C_{093}	0.57895	0.31579	0.10526	0.00000
		C_{094}	0.73684	0.15789	0.10526	0.00000
Social reputation	Service contribution	C_{101}	0.57895	0.36842	0.05263	0.00000
		C_{102}	0.36842	0.31579	0.26316	0.05263
		C_{103}	0.47368	0.26316	0.26316	0.00000
	Social ranking	C_{111}	0.63158	0.21053	0.15789	0.00000
		C_{112}	0.57895	0.26316	0.15789	0.00000
		C_{113}	0.63158	0.26316	0.05263	0.05263
		C_{114}	0.52632	0.31579	0.15789	0.00000
Internationalization level	International communication and cooperation	C_{121}	0.57895	0.31579	0.10526	0.00000
		C_{122}	0.47368	0.31579	0.21053	0.00000
		C_{123}	0.47368	0.26316	0.26316	0.00000
		C_{124}	0.26316	0.36842	0.36842	0.00000
		C_{125}	0.31579	0.31579	0.26316	0.10526

Table 5: Basic credibility and scores of second-level index.

Second-level index	Basic credibility and scores of second-level index				Score
	Good	General	Relatively poor	Poor	
Material resources	0.55639	0.28571	0.15038	0.00752	0.87669
Funding input	0.55868	0.38409	0.05722	0.00000	0.90029
Faculty quality	0.55331	0.39670	0.04999	0.00000	0.90066
The situation of university discipline construction	0.55177	0.31196	0.13627	0.00000	0.88310
Quality of student source and cultivation	0.43526	0.24893	0.29168	0.02412	0.81424
Teaching output	0.48448	0.27071	0.24480	0.00000	0.84794
Scientific research input	0.67188	0.30647	0.02165	0.00000	0.93004
Scientific research output	0.61794	0.28172	0.10034	0.00000	0.90352
Scientific research output efficiency	0.69824	0.23007	0.07168	0.00000	0.92531
Service contribution	0.54224	0.34210	0.11220	0.00346	0.88393
Social ranking	0.60795	0.24813	0.12994	0.01398	0.88722
International communication and cooperation	0.48614	0.30613	0.19792	0.00981	0.85176

Table 6: Basic credibility and scores of second-level index.

First-level index	Basic credibility and scores of the second-level index				Score
	Good	General	Relatively poor	Poor	
Educational resources	0.55540	0.38339	0.06060	0.00061	0.89859
Teaching level	0.51822	0.29232	0.18682	0.00264	0.86470
Research quality	0.67507	0.24965	0.07528	0.00000	0.91996
Social reputation	0.56415	0.31078	0.11811	0.00697	0.88503
Internationalization level	0.48614	0.30613	0.19792	0.00981	0.85176

Table 7: Jiangsu high-level university construction performance evaluation index measurement level.

	Good	General	Relatively poor	Poor
Confidence	0.61867	0.28507	0.09490	0.00136

passing rate of undergraduate dissertation sampling contribute the most to the efficiency of scientific research output, which reflects that the scientific research achievements of participating universities are highly recognized in the world, and the conversion rate of scientific research achievements is high. Meanwhile, the participating universities have done a good job in guiding undergraduate students to standardize academic work. Among the three levels of indicators under scientific research output, except for the number of major international awards, which is close to but not yet at the good level, the other five indicators, including the total number of authorized patents, have reached a good level, which indicates that the overall scientific research output of the participating universities is fruitful, but they should further promote the high-quality development of scientific research and strive to reach the international first-class level. The excellent level of faculty mainly lies in the considerable number of outstanding talents in the participating universities. This reflects that the sufficient number of outstanding talents is the intellectual support to help the participating universities sprint to high level. In the funding input, the growth of education expenditure reaches an excellent level, which helps to provide a pulling effect for the participating universities to expand the quality students.

Further analysis of Tables 4 and 5 also shows that, except for the five second-level indexes mentioned above, all the other secondary indicators reach the general level, in which the social ranking, service contribution, discipline construction, and material resource score are close to a good level. Specific analysis shows that, firstly, the social ranking of the participating universities has been improved in the international and domestic authoritative rankings, but the overall ranking is still far from the international and domestic first-class level. Secondly, among the service contributions of participating universities, outstanding alumni bring a good accumulation of social reputation for universities, but universities have limited consulting suggestions adopted by provincial and ministerial leaders and departments. This reflects that colleges and universities should base on the platform of Jiangsu high-level university construction and provide more comprehensive and in-place decision-making consulting services for Jiangsu provincial and ministerial leaders and government departments as much as possible, so as to serve the economic and social development of Jiangsu Province indirectly. Moreover, the employment ratio of graduates staying in Jiangsu needs to be improved. This requires the joint efforts of the Jiangsu government and universities to provide more comprehensive employment support policies for graduates. After that,

in the construction of disciplines, the number of doctoral and master's degrees, national and provincial key disciplines, and ESI top 1% disciplines in participating universities have increased considerably, but the overall growth is relatively average. This requires universities to gather their resource advantages to the direction of university characteristic development and do a good job in the construction of characteristic disciplines. Finally, among the material resources, the area per capita of school buildings, the number of books per student, and the number of repositories and databases are relatively adequate, but the participating universities still need to further improve the material resources to further guarantee the learning, teaching and research activities of students and teachers.

In addition, among the second-level indexes that reach good level, the scores of student source and cultivation quality, teaching output, and international communication and cooperation are relatively low. This indicates that the participating universities have certain problems of emphasizing research and neglecting teaching. Therefore, the output of students' innovation competition awards, national 100 excellent doctoral dissertations, provincial excellent undergraduate, and master's dissertations needs to be improved; the output of teaching achievements, high-quality courses, teaching platforms, teaching projects, teaching courses, and teaching materials of the Ministry of Education and provincial ministries is slightly weak; the number of international joint training programs for students, the number of international conferences held, and the number of international cooperative research platforms needs to be increased. Moreover, the university should benchmark itself against world-class universities in terms of faculty construction and student training mode, and cultivate international talents based on Chinese soil.

In summary, the participating universities have achieved more excellent results in the first round of Jiangsu high-level university construction. However, there is still some room for improvement on the whole, especially the teaching level and internationalization level still need to be further strengthened.

5. Conclusion

Based on the first round of "Jiangsu High-level University Construction Program" and the existing literature, this study constructs a comprehensive evaluation index system for the construction of Jiangsu High-level University by using the fuzzy-ANP comprehensive evaluation model and Delphi expert scoring method. We have designed the evaluation index system to be able to comprehensively evaluate the construction of high-level universities in Jiangsu Province. Using this system, the first round of Jiangsu high-level university construction was evaluated regarding educational resources, teaching level, research quality, social reputation, and internationalization level. The empirical results are as follows:

(1) The first round of Jiangsu high-level university construction basically reached the good level on the whole. Among them, the best results in improving the quality of scientific research are mainly attributed to the steadily improving efficiency of scientific research output and sufficient investment in scientific research of the participating universities. This indicates that the Jiangsu government and Jiangsu enterprises and institutions have invested enough in providing economic support to the participating universities. Moreover, it provides guarantee for universities to absorb high-quality teachers and research teams and carry out research work. Meanwhile, the improvement of the overall scientific research output efficiency of the participating universities is mainly due to the improvement of the conversion rate of authorized patents in the participating universities. The ability of participating universities in the construction of high-level universities in Jiangsu Province to promote local social and economic development has been enhanced.

(2) The resources and social reputation of the participating universities in the first round of Jiangsu high-level university construction have not reached the excellent level. The main reason is that the educational resource, service contribution, and social ranking are not outstanding enough. The school space per student, number of books per student, and number of repositories and databases are still not sufficient compared with the increasing number of students. The participating universities have not done enough to provide consulting and advising services for provincial and ministerial leaders and departments. Although the number of key disciplines, and master's and doctoral programs has increased considerably, the participating universities as a whole are still lacking in the construction of special disciplines.

(3) In the first round of Jiangsu high-level university construction, the teaching and internationalization levels of the participating universities are rather low. There is a lack in student source and cultivation quality, teaching output, and international communication and cooperation. This is mainly attributed to the problem of emphasizing scientific research and neglecting teaching in the participating universities. The teaching achievements of the participating universities at the Ministry of Education and provincial levels, the number of students' innovation competition awards, and the output of excellent dissertations are not abundant. Moreover, the international research exchange activities of the participating universities need to be increased, and the internationalization level of faculty and student cultivation mode is not enough.

Data Availability

The methodology of this paper is the Delphi expert research method and the fuzzy-ANP comprehensive evaluation

method. The authors invited 21 experts and scholars in the field of higher education to participate in the survey and validation. Three rounds of expert opinion were solicited, with 21 questionnaires distributed in each round and more than 18 valid questionnaires returned in each round. After multiple rounds of evaluation and scoring by experts, 54 tertiary indicators were identified and then divided into 12 secondary indicators and 5 primary indicators based on the correlation between the indicators. The fuzzy-ANP comprehensive evaluation method is based on the principle of fuzzy relationship synthesis, which quantifies the influencing elements of the target problem and provides a comprehensive evaluation of the reality of the evaluation object from all angles. Based on the indicators determined by the Delphi expert research method, the authors quantitatively analyze the impact elements of the study on the evaluation of the construction performance of high-level universities in Jiangsu Province and complete the comprehensive evaluation. The data for this paper were obtained from 21 Delphi expert research questionnaires and from a special page on the construction of high-level universities on the official website of the Jiangsu Provincial Education Department (http://jyt.jiangsu.gov.cn/col/col38747/index.html).

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported by the Major Project of Philosophy and Social Science Research in Colleges and Universities in Jiangsu Province (no. 2019SJZDA035), the National Natural Science Foundation of China (71871115), Key Project of the Social Science Foundation of Jiangsu Province (22WTA-019), General Project of Social Science Foundation of Jiangsu Province (22GLB032), and Young and Middle-Aged Academic Leaders of Qinglan Project in Jiangsu Province.

References

[1] E. WardS. Buglione et al., "The Carnegie classification for community engagement," *University Engagement with Socially Excluded Communities*, Springer, Dordrecht, Netherlands, 2013.

[2] F. S. Alani, M. F. R. Khan, and D. F. Manuel, "University performance evaluation and strategic mapping using balanced scorecard (BSC): case study–Sohar University, Oman[J]," *International Journal of Educational Management*, vol. 32, no. 4, pp. 689–700, 2018.

[3] J. Johnes, "Data envelopment analysis and its application to the measurement of efficiency in higher education," *Economics of Education Review*, vol. 25, no. 3, pp. 273–288, 2006.

[4] J. Joanna, "The strategies of using a special kind of number patterns in different stages of education," *Educational Research and Reviews*, vol. 12, no. 12, pp. 643–652, 2017.

[5] S. Nazari-Shirkouhi, S. Mousakhani, M. Tavakoli, M. R. Dalvand, J. Saparauskas, and J. Antucheviciene, "Importance-performance analysis based balanced scorecard for performance evaluation in higher education institutions: an integrated fuzzy approach," *Journal of Business Economics and Management*, vol. 21, no. 3, pp. 647–678, 2020.

[6] X. Zhang and W. Shi, "Research about the university teaching performance evaluation under the data envelopment method," *Cognitive Systems Research*, vol. 56, pp. 108–115, 2019.

[7] L. Chen, V. Jagota, and A. Kumar, "Research on optimization of scientific research performance management based on BP neural network," *International Journal of System Assurance Engineering and Management*, pp. 1–10, 2021.

[8] M. K. Yousif and A. Shaout, "Fuzzy logic computational model for performance evaluation of Sudanese Universities and academic staff," *Journal of King Saud University-Computer and Information Sciences*, vol. 30, no. 1, pp. 80–119, 2018.

[9] S. El Gibari, T. Gómez, and F. Ruiz, "Evaluating university performance using reference point based composite indicators," *Journal of Informetrics*, vol. 12, no. 4, pp. 1235–1250, 2018.

[10] S. Cadez, V. Dimovski, and M. Zaman Groff, "Research, teaching and performance evaluation in academia: the salience of quality," *Studies in Higher Education*, vol. 42, no. 8, pp. 1455–1473, 2017.

[11] F. Ruiz, S. El Gibari, J. M. Cabello, and T Gomez, "MRP-WSCI: multiple reference point based weak and strong composite indicators," *Omega*, vol. 95, Article ID 102060, 2020.

[12] I. Alsmadi, Z. W. Taylor, and J. Childs, "US News & World Report Best Colleges rankings: which institutional metrics contribute to sustained stratification?" *Scientometrics*, vol. 124, no. 3, pp. 1851–1869, 2020.

[13] X. Huang and Y. Z. Chen, "Chinese characteristics and international standards: study on evaluation system of scientific and technological innovation capacity of "double first-class" universities," *Higher Education Studies*, vol. 9, no. 4, pp. 189–199, 2019.

[14] R. Aliyev, H. Temizkan, and R. Aliyev, "Fuzzy analytic hierarchy process-based multi-criteria decision making for universities ranking," *Symmetry*, vol. 12, no. 8, 2020.

[15] X. Xu, Z. Lin, X. Li, C. Shang, and Q. Shen, "Multi-objective robust optimisation model for MDVRPLS in refined oil distribution," *International Journal of Production Research*, pp. 1–21, 2021.

[16] J. M. Arenas Reina, J. D. Cano-Moreno, F. V. Sánchez Martínez, and M. Islan Marcos, "Methodology for analysis and quantitative evaluation of teaching quality in university subjects," *Mathematics*, vol. 9, no. 8, p. 831, 2021.

[17] X. Sun, C. Cai, S. Pan, N. Bao, and N. Liu, "A university teachers' teaching performance evaluation method based on type-II fuzzy sets," *Mathematics*, vol. 9, no. 17, 2021.

[18] L. Ma and M. Ladisch, "Evaluation complacency or evaluation inertia? A study of evaluative metrics and research practices in Irish universities," *Research Evaluation*, vol. 28, no. 3, pp. 209–217, 2019.

[19] R. Lukman, D. Krajnc, and P. Glavič, "University ranking using research, educational and environmental indicators," *Journal of Cleaner Production*, vol. 18, no. 7, pp. 619–628, 2010.

[20] R. H. McGuckin and D. R. Winkler, "University resources in the production of education," *The Review of Economics and Statistics*, vol. 61, pp. 242–248, 1979.

[21] S. Ma, "The function of the establishment of digital archives in colleges and universities to the optimization of resources,"

IOP Conference Series: Materials Science and Engineering, vol. 750, no. 1, Article ID 012057, 2020.

[22] G. M. Nam, D. G. Kim, and S. O. Choi, "How resources of universities influence industry cooperation," *Journal of Open Innovation: Technology, Market, and Complexity*, vol. 5, no. 1, p. 9, 2019.

[23] J. K. Lind, "Resource environment and hierarchy in universities," *Science and Public Policy*, vol. 47, no. 2, pp. 184–193, 2020.

[24] B. Lepori, A. Geuna, and A. Mira, "Scientific output scales with resources. A comparison of US and European universities," *PLoS One*, vol. 14, no. 10, Article ID e0223415, 2019.

[25] L. Bergman, "Supporting academic literacies: university teachers in collaboration for change," *Teaching in Higher Education*, vol. 21, no. 5, pp. 516–531, 2016.

[26] X. Wang and H. Hu, "Sustainable evaluation of social science research in higher education institutions based on data envelopment analysis," *Sustainability*, vol. 9, no. 4, p. 644, 2017.

[27] M. Y. Romankina, N. V. Kuznetsova, and Y. A. Fedulova, "Scientific research as an important aspect of natural science education," *Journal of Physics: Conference Series*, vol. 1691, no. 1, Article ID 012011, 2020.

[28] A. Evarist, "Teacher-student ratio on classroom practices in universal secondary schools in Wakiso District-Uganda," *Journal of Emerging Trends in Educational Research and Policy Studies*, vol. 9, no. 4, pp. 167–177, 2018.

[29] H. Serin, "Student evaluations of teaching effectiveness: an instrument to increase teaching quality in higher education," *International Journal of Social Sciences & Educational Studies*, vol. 5, no. 4, pp. 168–173, 2019.

[30] S. Y. Sohn and Y. H. Ju, "Conjoint analysis for recruiting high quality students for college education," *Expert Systems with Applications*, vol. 37, no. 5, pp. 3777–3783, 2010.

[31] X. Xu and Y. He, "Blockchain application in modern logistics information sharing: a review and case study analysis," *Production Planning & Control*, pp. 1–15, 2022.

[32] H. Zhang, B. Xiao, J. Li, and M. Hou, "An improved genetic algorithm and neural network-based evaluation model of classroom teaching quality in colleges and universities," *Wireless Communications and Mobile Computing*, vol. 2021, Article ID 2602385, 7 pages, 2021.

[33] X. Xu, Z. Lin, and J. Zhu, "DVRP with limited supply and variable neighborhood region in refined oil distribution," *Annals of Operations Research*, vol. 309, no. 2, pp. 663–687, 2022.

[34] V. Sterzi, M. Pezzoni, and F. Lissoni, "Patent management by universities: evidence from Italian academic inventions," *Industrial and Corporate Change*, vol. 28, no. 2, pp. 309–330, 2019.

[35] P. Giuri, F. Munari, A. Scandura, and L. Toschi, "The strategic orientation of universities in knowledge transfer activities," *Technological Forecasting and Social Change*, vol. 138, pp. 261–278, 2019.

[36] M. F. Köse and M. Korkmaz, "Why are some universities better? An evaluation in terms of organizational culture and academic performance," *Higher Education Research and Development*, vol. 38, no. 6, pp. 1213–1226, 2019.

[37] D. D. Ridley, A. G. Matveev, and N. M. Cuevas, "Dynamics of internet visibility: mutual benefits for distinguished alumni and alma mater?" *International Journal of Educational Advancement*, vol. 5, no. 2, pp. 119–130, 2005.

[38] J. Xinrong and Z. Rong, "Model and empirical research on influencing factors of member selection of university think tank alliance," *Library and Information Service*, vol. 64, no. 23, p. 96, 2020.

[39] M. Muñoz-Suárez, N. Guadalajara, and J. M. Osca, "A comparative analysis between global university rankings and environmental sustainability of universities," *Sustainability*, vol. 12, no. 14, 2020.

[40] H. F. Moed, "A critical comparative analysis of five world university rankings," *Scientometrics*, vol. 110, no. 2, pp. 967–990, 2017.

[41] G. Buela-Casal, O. Gutiérrez-Martínez, M. P. Bermúdez-Sánchez, and O Vadillo-Munoz, "Comparative study of international academic rankings of universities," *Scientometrics*, vol. 71, no. 3, pp. 349–365, 2007.

[42] M. Mussard and A. P. James, "Engineering the global university rankings: gold standards, limitations and implications," *IEEE Access*, vol. 6, pp. 6765–6776, 2018.

[43] A. Schubert and G. Schubert, "Internationality at university level," *Scientometrics*, vol. 123, no. 3, pp. 1341–1364, 2020.

[44] T. L. Saaty, *Decision Making with Dependence and Feedback: the Analytic Network Process*, RWS Publications, Pennsylvania, PA, USA, 1996.

[45] A. Sala, T. M. Guerra, and R. Babuška, "Perspectives of fuzzy systems and control," *Fuzzy Sets and Systems*, vol. 156, no. 3, pp. 432–444, 2005.

[46] T. Chen, L. Wang, and J. Wang, "Transparent assessment of the supervision information in China's food safety: a Fuzzy-ANP comprehensive evaluation method," *Journal of Food Quality*, vol. 2017, Article ID 4340869, pp. 1–14, 2017.

Hindawi
Scientific Programming
Volume 2022, Article ID 4414093, 14 pages
https://doi.org/10.1155/2022/4414093

Research Article

Substation Equipment Temperature Prediction Method considering Local Spatiotemporal Relationship

Lijie Sun,[1] **Shuang Chen**,[2] **Junfei Zhu**,[3] **and Jianhua Li**[1]

[1]*School of Electronics and Information Engineering, Taizhou University, Taizhou 318000, Zhejiang, China*
[2]*School of Aeronautical Engineering, Taizhou University, Taizhou 318000, Zhejiang, China*
[3]*State Grid Taizhou Power Supply Company, Taizhou 317700, Zhejiang, China*

Correspondence should be addressed to Shuang Chen; shuangchen@tzc.edu.cn

Received 17 June 2022; Accepted 29 July 2022; Published 24 August 2022

Academic Editor: Hongguang Ma

Temperature prediction of substation equipment is one of the important means for intelligent inspection of substation equipment. However, there are still three challenges: (1) Limited extracted samples; (2) Typical nonlinearity, seasonality, and periodicity; (3) Changes in equipment and working conditions. To solve the problems above, a substation equipment temperature prediction method considering Spatio-temporal relationship (SETPM-CLSTR) is proposed. First, according to the time series of equipment temperature from two aspects of temporal and spatial, it is determined that the equipment temperature has seasonal, temporal, and spatial correlation; second, aiming at the problem that the spatial location correlation cannot be described quantitatively, grey relational analysis (GRA) is adopted to determine the spatial location monitoring points closely related to the prediction target; then, the daily maximum temperature and daily minimum temperature from the environment, the predicted target temperature from the past several times in time and the temperature from the spatial location monitoring point with close correlation in space are constructed as Spatio-temporal feature vectors; finally, CNN-BiLSTM double-layer depth network model is proposed to predict the equipment temperature. SETPM-CLSTR has applied to temperature prediction of phase A contact from primary equipment of a substation in Taizhou City, Zhejiang Province. Under the two prediction performance evaluation indexes of MASE and RMSE, compared with three correlation models of LSTM, BiLSTM, and CNN-LSTM from two aspects of different features and models, it is verified that SETPM-CLSTR in this study has better prediction performance.

1. Introduction

Since 2020, large-scale blackouts have occurred in Mumbai, India, Pakistan, and Texas, which has brought great impact and losses to local economic and social development [1]. The safe development of the power grid is related to national security development and is the "lifeline" to be firmly guarded. Power companies face a great test in ensuring the safe operation and reliable power supply of the power grid. At present, China's power system is developing towards ultra-high voltage (UHV) and large capacity. In addition, society puts forward higher requirements for power supply quality and reliability. Substation equipment is an important material basis for the intrinsic safety of the power grid and the first line of defense for the safety of a large power grid, and ensuring the safe operation of equipment is the top priority [2]. In the power system with large units, large capacity, and high voltage, how to ensure the safe and stable operation of power equipment and how to patrol and monitor the equipment have become important links in the equipment management and transformation of the power system. Monitoring the operation status of this equipment in real-time and giving a timely response mechanism can effectively prevent the occurrence of accidents caused by abnormal operation of equipment.

Online monitoring is mainly aimed at primary equipment, including circuit breaker, disconnector, grounding knife (knife switch), transformer, bus, switch cabinet, cable connector, etc., to conduct real-time monitoring on its key points. During the operation of this equipment, it is easy to

cause heating and increase the temperature of the equipment due to aging insulation, excessive voltage load during operation, a loose connection of joints, loose bolts at key points, oxidation and corrosion of conductor surface or excessive contact resistance of contact surface. If it is light, it will cause damage and burning of relevant electrical equipment, and then lead to substation operation failure; More importantly, it will lead to fire and safety accidents, resulting in huge economic losses and social impact on the substation. Therefore, it is very important to know the temperature of each piece of equipment in real-time.

In recent years, the thermal failure caused by excessive equipment temperature has caused the shutdown of electrical equipment in substations, large-scale power failure in nearby areas, and even serious fire accidents [3]. For example, a substation in Luoyang, Henan Province, caused over-voltage due to a cable grounding fault, resulting in the fire of the distribution cabinet, resulting in the shutdown of power generation equipment, the power failure of surrounding communities, and serious losses. Therefore, online real-time monitoring of the temperature condition of power equipment during operation and predicting the temperature in the future can prevent faults. At the same time, in order to promote the intellectualization, informatization, and digitization of substation management and meet the needs of social development, it is imperative to build and optimize the substation equipment temperature prediction and fault early warning management system.

However, there are still three challenges in substation equipment temperature prediction:

(1) Limited extracted samples. There are two main reasons for the small amount of historical equipment temperature data extracted. On the one hand, according to the common sense of substation operation, the equipment temperature will be affected by the load, but the load data and the substation are not in the same department, and the data submission process between them is complex, Therefore, it is difficult to obtain load data. On the other hand, the data storage space of an intelligent inspection system is insufficient. The substation intelligent inspection system is quite large, involving many aspects and a large amount of data. Therefore, the storage space will be cleared regularly, and the automatic clearing time is generally set to 1 year.

(2) Typical nonlinearity, seasonality, and periodicity. The temperature data of substation equipment is affected by the equipment itself and many other factors, such as climate, environment, load, and so on, and its internal variation is irregular, with typical nonlinearity and seasonality, which makes it challenging task to construct a reasonable prediction model.

(3) Changes in equipment and working conditions. Substations are usually built in remote rural areas far from the urban area. The existence of abnormal weather conditions such as climate, air and thunder, rain, and snow make it more difficult to accurately predict the equipment temperature.

At present, the research mainly focuses on the traditional statistical analysis methods and machine learning methods, using time series as input characteristics to predict the temperature of substation equipment. For example, the ARMA series model [4], random forest (RF) method [5], neural network [6], etc, in which used historical temperature data and ambient temperature to form time series as a feature vector. However, the traditional models of substation equipment temperature prediction ignore the spatial relationship information of equipment in the historical time, resulting in poor prediction accuracy. Therefore, it is particularly important to select what characteristics to characterize the temperature for prediction. Therefore, when solving the problem of substation equipment temperature prediction, inspired by references [7–11], from the perspective of environment and multi-objective thinking [12–14], this study carries out feature extraction for the construction of substation equipment temperature prediction model from the two aspects of time and space, from the three dimensions of the interaction between different monitoring points of equipment, the seasonal characteristics of the influence of ambient temperature on equipment temperature and the influence of equipment temperature at historical time on equipment future temperature [15].

However, the more features are not the better for the prediction results. To solve the problems above, this study applies the grey relational analysis (GRA) [16] to analyze the location correlation of substation equipment, calculates the correlation between temperatures at different test points, takes the ambient temperature as the seasonal influencing factor, and integrates the three types of data of equipment temperature in the past. In the prediction modeling stage, deep learning network is widely used in various research fields [17–19], and according to the existing research, Convolutional Neural Network (CNN) has strong advantages in depth feature extraction [20] and Long Short-term Memory Networks (LSTM) network is suitable for processing and predicting important events with relatively long interval and delay in time series [21]. However, there is still a problem when using LSTM for time series prediction modeling: it is unable to encode the information from back to front. Bidirectional LSTM (BiLSTM) is a sequence processing model, which is composed of two LSTMs: one receives input in the forward direction and the other receives input in the backward direction. BiLSTM effectively increases the amount of information available in the network [22, 23]. Based on the above research results, this study uses CNN-BiLSTM double-layer depth network and multivariate time series model to realize the temperature prediction of substation equipment. This is expected to achieve better prediction performance.

In short, for substation equipment temperature prediction, the study proposes a method considering a local Spatio-temporal relationship, which is organized as follows:

(i) Section 2 describes basic theories, including grey correlation degree, LSTM network, and CNN;

(ii) Section 3 introduces data sources, conducts a multivariate analysis of data;

(iii) Section 4 describes SETPM-CLSTR, including feature selection, spatial feature extraction based on GRA, double-layer depth network prediction model based on CNN-BiLSTM, comparative experiments, and analysis of prediction results;

(iv) The conclusions are covered in Section 5.

2. Methodology

2.1. GRA. Grey system theory [24] is the concept of grey correlation analysis of each subsystem proposed by Professor Deng Julong, which aims to find the numerical relationship between each subsystem (or factor) in the system through certain methods. Grey relational analysis (GRA) is a very active branch of grey system theory, which can provide quantitative measurement for the development and change trend of the system, and its basic idea is to toughen the original observation number of evaluation indexes, calculate the correlation coefficient and correlation degree, and sort the evaluation indexes according to the correlation degree. GRA is realized in the following five steps.

(1) Determine the formula of the reference sequence X_0 and comparison sequence X_i, which are described as the formula (1) and the formula (2):

$$X_0 = [X_0(1), X_0(2), \ldots, X_0(n)], \tag{1}$$

$$X_i = [X_i(1), X_i(2), \ldots, X_i(n)] \, (i = 0, 1, \ldots, m). \tag{2}$$

(2) The raw data is subject to unlimited tempering, and the average method is expressed as the formula:

$$X_i'(k) = \frac{X_i(k)}{X_i(l)} \, (k = 1, 2, \ldots, n; \; i = 0, 1, \ldots, n). \tag{3}$$

(3) Calculate the difference sequence, calculate the absolute value difference between the reference sequence and the comparison sequence, and find out the maximum value ΔMax and minimum value ΔMin:

$$\Delta_j(k) = \left| X_0'(k) - X_j'(k) \right| (k = 1, 2, \ldots, n; \; j = 0, 1, \ldots, m). \tag{4}$$

(4) Calculation the correlation coefficient by the formula:

$$\delta_j(k) = \frac{\Delta_{\min} + \rho \Delta_{\max}}{\Delta_j(k) + \rho \Delta_{\max}}, \tag{5}$$

where ρ is the resolution coefficient, between [0, 1]. Generally, the smaller ρ is, the stronger the resolution is. In this study, $\rho = 0.5$ is taken.

(5) Calculate the correlation degree by the formula (6) and sort from large to small:

$$R_j = \frac{1}{n} \sum_{k=1}^{N} \delta_j(k) \, (k = 1, 2, \ldots, n). \tag{6}$$

2.2. LSTM Deep Network

2.2.1. RNN. The traditional neural network has made many achievements in various fields, but it has one main disadvantage, that is, it can not do the temporal correlation of information [25]. Recurrent neural network (RNN) is a special neural network structure, which is based on the view that human cognition is based on past experience and memory and is different from DNN and CNN in that it not only considers the input of the previous moment but also endows the network with a "memory" function of the previous content. The cyclic structure of RNN is shown on the left side of the equal sign in Figure 1, in which the module A receives the input X_t and outputs the value h_t. The expanded view of the annular structure is shown on the right side of the equal sign in Figure 1, in which $X_0, X_1, X_2, \ldots, X_t$ are the input time series, and $h_0, h_1, h_2, \ldots, h_t$ are the output time series.

The cyclic structure allows information to be transferred from one network state to the next. and a recurrent neural network can be considered as multiple copies of the same network, and each network transmits a signal to its next network. Therefore, all recurrent neural networks are repetitive neural network chains. In the standard RNN, this repetitive sub-module has a very simple structure, such as a tanh layer, which is shown in Figure 2.

2.2.2. LSTM Network. Long short-term memory (LSTM) network is a special RNN, and it is an improved recurrent neural network, which can solve the problem that RNN cannot deal with long-distance dependence and is widely used in time series prediction [26]. Like RNN, the LSTM network also has a chain structure, but the repeated submodule structure is different from RNN. The chain structure of the LSTM network is shown in Figure 3. It not only has a single neural network layer but is cycled by a forgetting gate, input gate, and output gate in a special way. The forgetting gate refers to the retention degree of the previous moment of state information; the input gate determines the information update degree according to the input current information and the current unit state; the output gate is responsible for outputting specific memory from the unit structure at the current time.

Where the meanings are represented by all symbols in Figures 3 and 4, respectively.

In Figure 4, line symbols contain a vector from one node to another. The pink circle represents point-to-point operations, such as vector addition, and the yellow box represents the learned neural network. Intersecting lines represent information merging, and separated arrow lines represent information replication and splitting.

The hidden unit structure of the LSTM network can be expressed by formulas:

$$f_t = \sigma\left(W_f x_t + U_f h_{t-1} + b_f\right), \tag{7}$$

$$i_t = \sigma\left(W_i x_i + U_i h_{t-1} + b_i\right), \tag{8}$$

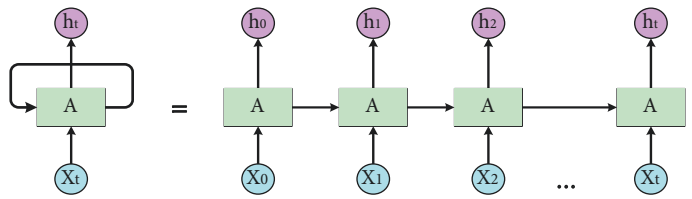


Figure 1: Cyclic structure and its expansion diagram of RNN.

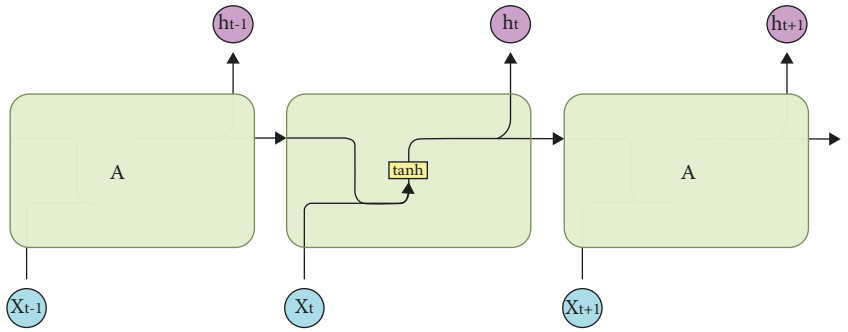


Figure 2: Sub module structure of RNN.

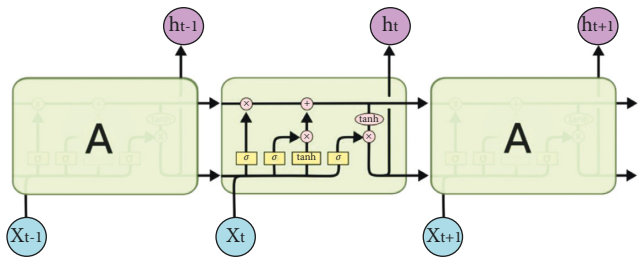


Figure 3: Sub module structure of LSTM.

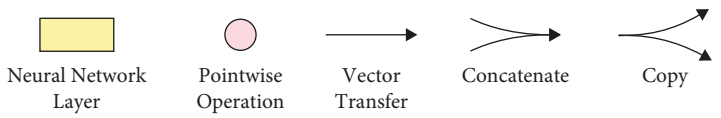


Figure 4: Schematic diagram of all symbols in Figure 3.

$$\tilde{c}_t = \tanh\left(W_c x_t + U_c h_{t-1} + b_c\right), \tag{9}$$

$$c_t = i_t \odot \tilde{c}_t + f_t \odot c_{t-1}, \tag{10}$$

$$o_t = \sigma\left(W_o x_t + U_o h_{t-1} + b_o\right), \tag{11}$$

$$h_t = o_t \odot \tanh\left(c_t\right), \tag{12}$$

where f_t and i_t respectively represent the forgetting gate and the input gate; $\tilde{c}_t$ and c_t respectively represent the candidate state and the unit state; o_t and h_t respectively represent the output gate and the final unit output; $W_f, U_f, b_f, w_i, U_i, b_i$, W_c, U_c, b_c, W_o, U_o and b_o are training parameter matrices; $\odot$ refers to Hadamard product.

2.3. CNN. Convolutional Neural Network (CNN) is one of the representative algorithms of deep learning. The basic structure of CNN is generally composed of the input layer, convolution layer, pooling layer, full connection layer, and output layer, in which the convolution layer and pooling layer are hidden layers [25]. In this study, CNN with three layers of convolution is used to characterize the temperature-depth of substation equipment, and the structure of CNN is shown in Figure 5.

In the convolution layer, each output feature map can combine and convolute the values of multiple feature maps, which are expressed as the formula:

$$x_j^l = f\left(u_j^l\right), \tag{13}$$

$$u_j^l = \sum_{i \in M_j} x_i^{l-1} * k_{ij}^l + b_j^l, \tag{14}$$

where u_j^l is the net activation of the j – th channel from the convolution layer l; M_j refers to the input characteristic graph subset of calculating net activation; k_{ij}^l and b_j^l represent the offset of convolution kernel matrix and

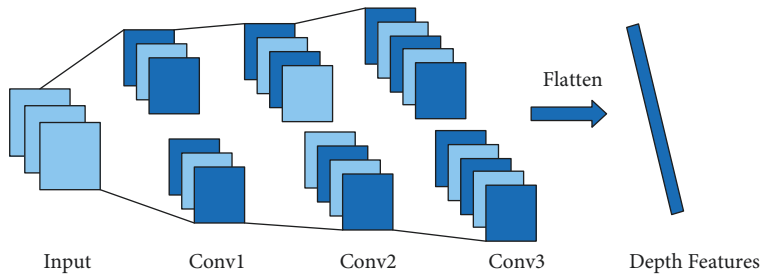


Figure 5: Structure diagram of CNN.

convoluted characteristic graph respectively; $*$ is the convolution symbol.

3. Problem Analysis

3.1. Description of Research Object. In January 2020, in order to promote the work deployment related to the ubiquitous power Internet of things of State Grid Corporation of China, promote the construction of intelligent deep supply, and jointly complete the intelligent inspection in the substation through a variety of high-tech means, so as to realize the intelligent automatic inspection in the substation instead of manual inspection, State Grid Zhejiang Electric Power Co., Ltd. specially formulated the pilot construction scheme of joint automatic inspection of high-definition video and intelligent inspection robot in the substation of Zhejiang Electric Power Co., Ltd. In response to this call, State Grid Companies in various cities in Zhejiang Province began to focus on establishing a joint automatic inspection system of substation high-definition video and intelligent inspection robot, so as to realize information connection, improve the joint intelligent inspection strategy, and expand the inspection functions of high-definition video and intelligent inspection robot.

The robot infrared temperature measurement subsystem is a functional branch of the video inspection system, which mainly aims at the infrared temperature measurement in the main transformer area, realizes the key temperature measurement monitoring of key equipment, sets the temperature measurement points of multi-point, multi-line, and multi-surface for the main transformer equipment, measures the temperature in real-time at the moving point of the mouse in the whole picture, and measures the temperature at multiple points in the whole picture at the same time. Moreover, the synchronous control function of the visible light camera makes the temperature measurement points more accurate. The outdoor robot inspection system can complete automatic inspection, remote inspection and special inspection tasks, and can replace personnel for daily inspection of equipment. Figure 6 shows that the wheeled robot in a substation in Taizhou is performing the task of infrared temperature measurement, the infrared imaging diagram of substation equipment is shown as the subgraph (a) in Figure 7, and the substation equipment diagram under visible light is shown as the subgraph (b) in Figure 7. The combined application of robot and physical identification (ID) adopts patrol correlation so that all abnormal data identified by the robot can be associated with the correct equipment. The goal of associating physical ID is to reverse write defects into the PMS system through physical ID information.

Figure 6: Process diagram of robot executing infrared temperature measurement task.

3.2. Data Acquisition. At present, the temperature early warning based on robot infrared temperature measurement is still in the trial operation stage, because the substation equipment is a large equipment, and there are many monitoring points and huge data. From the perspective of storage space and management, the initial inspection cycle of the robot is only set once a week, the shortest cycle is once a day, and there will be an interruption in the middle. Therefore, there is a serious practical problem that the early warning is not timely based on the robot's infrared temperature measurement to realize the equipment fault early warning task. Therefore, this study plays an important role in the equipment fault early warning based on the equipment temperature prediction.

The equipment of data acquisition is the primary equipment of the No. 2 main transformer from a 220 kV substation in Taizhou City, Zhejiang Province, and the temperature data is collected from October 1, 2019 to October 29, 2020 in this study, namely the data of 13 months. The primary equipment of the No. 2 main transformer consists of a 110 kV side and 220 kV side, and Table 1 shows the basic information of the equipment and the names of key points of equipment temperature inspection.

(a) (b)

Figure 7: Substation equipment diagram. (a) Infrared imaging diagram. (b) Substation equipment diagram of substation equipment under visible light.

Table 1: Summary of names for the equipment temperature monitoring key points.

Serial number	Equipment name
1	1–10 heat sinks at 220 side
2	19–23 heat sinks at 110 side
3	13–17 heat sinks at 110 side
4	110 side conservator
5	110 kV bushing phase C contact
6	110 kV bushing phase B contact
7	110 kV bushing phase A contact
8	220 kV bushing phase C contact
9	220 kV bushing phase B contact
10	220 kV bushing phase A contact
11	110 side equipment panorama
12	220 side equipment panorama

An intelligent inspection system is applied in the substation. The infrared equipment measures the temperature at 3 p.m. every day. The temperature data is exported in word form, that is, a multi-dimensional intelligent inspection report. The temperature data for more than a year is about 4G. The data in the database includes the name of key points of monitoring equipment, inspection time, inspection parts, inspection value (that is, the temperature value of key points of each equipment), temperature difference, infrared thermal imaging picture, alarm level manual review and description (fault description). In this study, the temperature monitored by 220 kV bushing phase A contact of No. 2 main transformer is selected for the experiment, including 370 days of data, the first 90% of the data set is used as the training set, and the remaining 10% of the data set is used as the test set.

3.3. Data Analysis

(1) Comparative analysis of the same equipment and the same monitoring point in different seasons.

According to the meteorological division method, in the meteorological department, usually, March to May of the Gregorian calendar is spring, June to August is summer, September to November is autumn, December to February of the next year is winter, and January, April, July, and October are often regarded as the representative months of winter, spring, summer, and autumn. Taking winter and summer as an example, this study analyzes the seasonal characteristics of substation equipment temperature, and Figure 8 shows the temperature change trend of phase A contact monitoring point of the bushing of No. 2 main transformer in Substation in winter and summer. Obviously, the average temperature of phase a contact is 30°C in winter and 50°C in summer. With the seasonal change, the equipment temperature changes significantly, which has an obvious positive correlation. Therefore, when predicting the equipment temperature, it is necessary to consider the ambient temperature factor.

(2) Comparative analysis of temperature at different monitoring points in the space-related position of the same equipment.

The primary equipment of the No. 2 main transformer consists of a 110 kV side and 220 kV side, and both sides of the equipment are independent of each other and have no intersection. Therefore, this study only analyzes the 220 kV side monitoring point where the phase A contact of the bushing is located for correlation analysis. The temperature trend diagram of 14 monitoring points at the 220 kV side is shown in Figure 9.

As can be seen from Figure 9, the temperatures at different monitoring points on the same side of the same equipment have a consistent trend and a typical linear correlation. There is a close relationship between the temperatures at most of the monitoring points. Therefore, it is necessary to consider spatial-related factors when predicting the temperature.

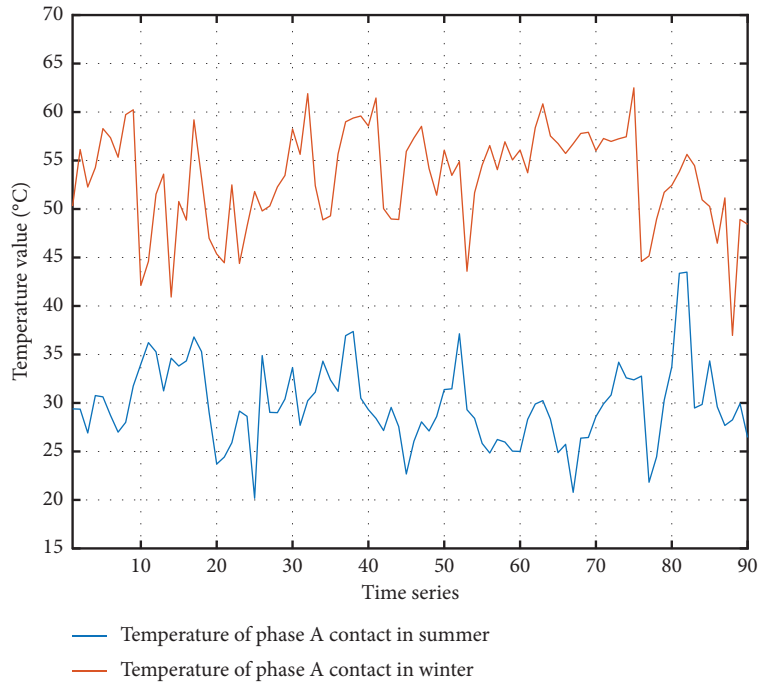


Figure 8: Seasonal trend chart of substation equipment temperature.

4. SETPM-CLSTR

According to the temperature data analysis for primary equipment of substation No. 2 main transformer in Section 3 of the previous article, the intelligent inspection of a substation based on a robot has the problems of long inspection cycle, short operation time and complex working environment, which lead that the substation equipment temperature prediction task has the characteristics of less characteristic parameters, small amount of data, instability, and seasonality. Based on the above problems, a substation equipment temperature prediction method considering local space-time relationship is proposed (recorded as SETPM-CLSTR), which is mainly realized through three links: feature selection, spatial feature extraction based on GRA, and double-layer depth network prediction model based on CNN-BiLSTM.

4.1. Feature Selection. For the task of substation equipment temperature prediction, the existing research only selects the characteristics of time, including the historical equipment temperature, and the daily maximum temperature and daily minimum temperature representing climate factors. This paper excavates the characteristics of space and establishes the feature vectors of space and time. In terms of time, the historical temperature data of several equipment-monitoring points and the daily maximum temperature and daily minimum temperature of the day are selected. In terms of space, the temperatures of all monitoring points on the same side of the equipment where the predicted target monitoring point is located are selected to form a Spatio-temporal feature set. Figure 10 shows the construction block diagram of the Spatio-temporal feature set.

4.2. Spatial Feature Extraction Based on GRA. The research object of this study is the primary equipment of the No. 2 main transformer in the substation. There are 14 monitoring points in 220 kV side space, one is the prediction target and 13 are the monitoring points related to spatial location. According to the spatial correlation analysis of the data in Chapter 3.3, it can be seen that different monitoring points have different correlation degrees with the prediction target, and the performance of the prediction model can not reach the best when the temperatures of all spatial monitoring points are used as spatial features. On the contrary, the existence of too many features with poor correlation will reduce the prediction performance of the model. Therefore, this study takes the temperature of phase A contact from bushing as the prediction target. In order to determine the monitoring points with high spatial correlation with phase A contact of bushing, this study uses GRA to calculate the grey correlation degree and quantitatively describe the contribution of the temperature of the monitoring point at the position of spatial correlation to the prediction target temperature, that is, calculate the correlation degree between the temperature of other 13 monitoring points and the temperature of phase A contact from the bushing. The greater the correlation degree, the higher the contribution to the prediction of phase A contact temperature of bushing, and the more it can characterize the temperature of phase A contact. Finally, the correlation degree is sorted in descending order. Grey correlation degree of temperature at local spatial correlation monitoring points at 220 kV side of the equipment is listed in Table 2.

According to the data in Table 2, this study selects the temperatures of the five spatial correlation monitoring points with the highest correlation degree as the spatial features, including phase B contact (recorded as B), phase C contact (recorded as C), No. 1 heat sink (recorded as 1#), No.

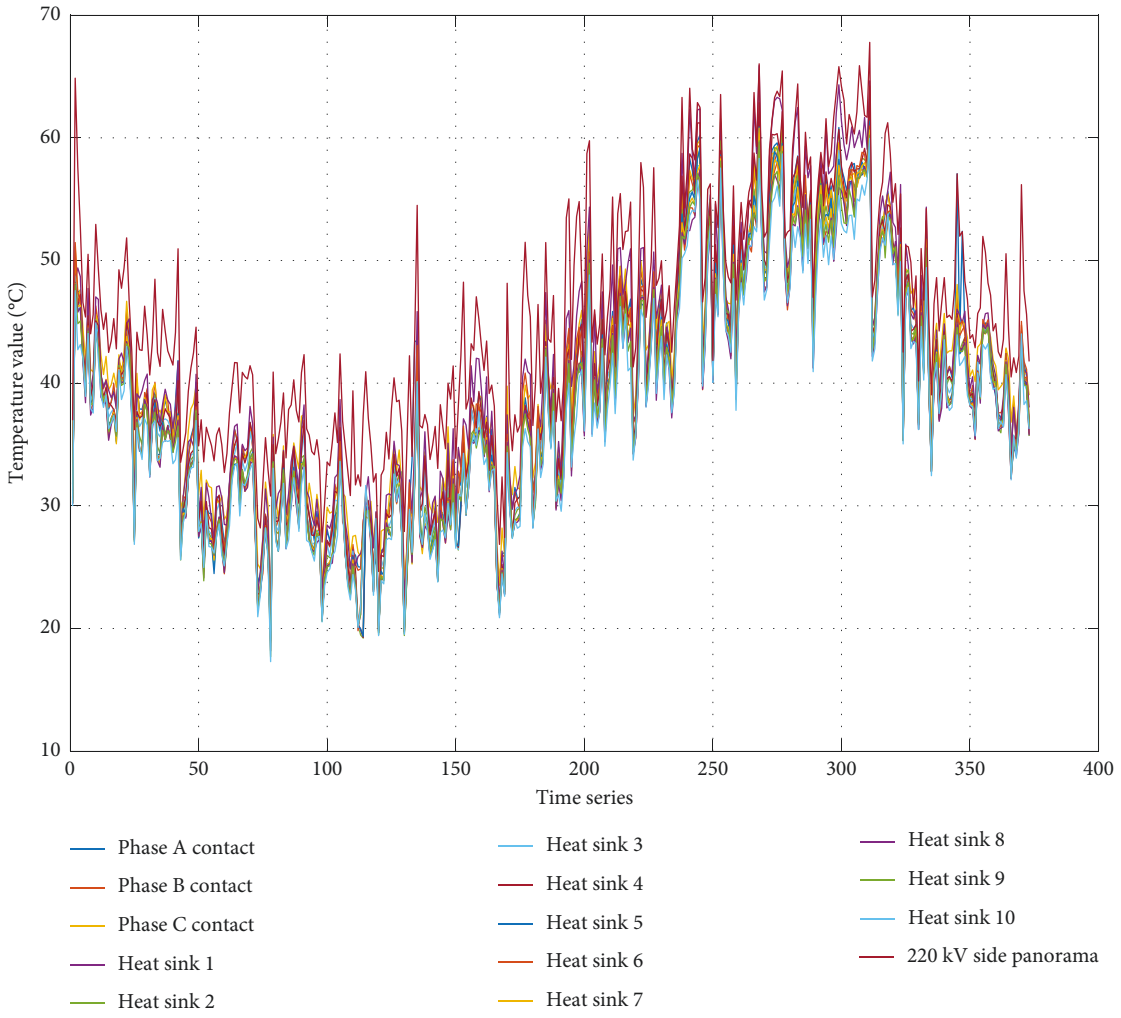


Figure 9: Temperature trend diagram of 14 monitoring points on 220 kV side.

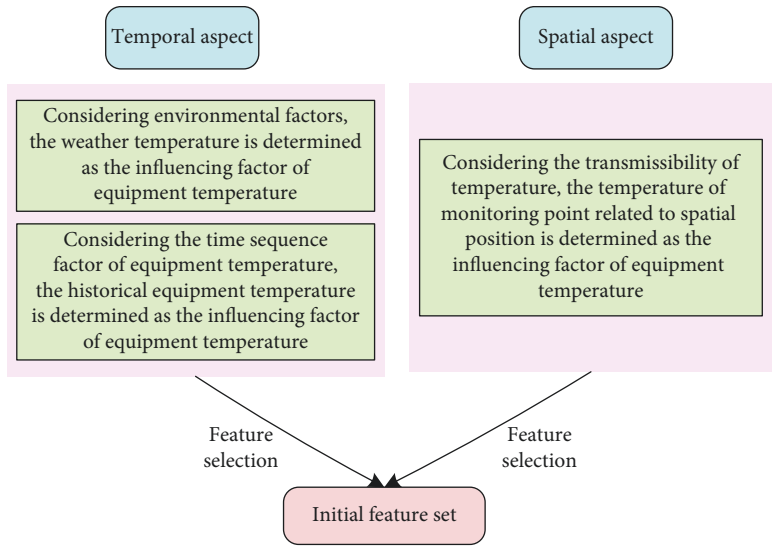


Figure 10: The construction block diagram of the Spatio-temporal feature set.

Table 2: Statistical table of grey correlation degree.

Local space location name	Correlation degree	Descending sort sequence number
220 kV bushing phase A contact	1	Prediction target (not sorted)
220 kV bushing phase B contact	0.9551	1
220 kV bushing phase C contact	0.9042	3
1# heat sink	0.8893	5
2# heat sink	0.8774	6
3# heat sink	0.8928	4
4# heat sink	0.9224	2
5# heat sink	0.8587	7
6# heat sink	0.8498	9
7# heat sink	0.8555	8
8# heat sink	0.8329	11
9# heat sink	0.8451	10
10# heat sink	0.8073	12
220 side equipment panorama	0.6657	13

Table 3: Partial samples of feature vector based on local Spatio-temporal relationship.

	Spatial features					Temporal features				
						Ambient characteristics		Historical temperature		
	B	C	1#	3#	4#	$D_{\max}$	$D_{\min}$	$T(t-1)$	$T(t-2)$	$T(t-3)$
1	47.55	47.21	47.62	48.69	46.67	52.70	31	22	47.41	48.90
2	42.64	42.68	44.51	42.82	42.23	46.58	30	20	47.55	47.41
3	39.95	40.37	41.76	39.92	39.13	45.40	25	18	42.64	47.55
...	...	...	...	...	...	...	...	...	...	...
370	51.79	50.72	51.37	54.35	51.63	54.22	21	18	46.06	47.97

3 heat sink (recorded as 1#) and No. 4 heat sink (recorded as 1#), selects the daily highest temperature (recorded as $D_{\max}$) and daily lowest temperature (recorded as $D_{\min}$) as the ambient characteristics, and selects the temperature of phase A contact in the first three days as the temporal features (recorded as $T(t-1)$, $T(t-2)$, $T(t-3)$ respectively), namely, the feature vector consists of 10 features based on local Spatio-temporal relationship, and some samples are shown in Table 3.

4.3. Normalization Processing. There is no comparability between the features of the feature vector based on the local Spatio-temporal relationship; therefore, normalization is needed before establishing the prediction model. Max − Min normalization method is used in this study, which is expressed as the formula:

$$X^* = \frac{X - \min}{\max - \min}, \tag{15}$$

where X is the temperature value; X^* is the value after Normalization processing; max is the maximum value of sample data, and min is the minimum value of sample data.

4.4. Double Layer Depth Network Prediction Model Based on CNN-BiLSTM. At this stage, it is necessary to establish a prediction model for substation equipment temperature prediction. This study proposes to use CNN and BiLSTM to build a two-layer depth network prediction model [26]. CNN is the depth feature extraction layer to mine the depth features of the feature vector based on the local space-time relationship after normalization, so as to obtain more information that can characterize the temperature of the predicted target monitoring point. BiLSTM is a bidirectional long and short memory depth network prediction layer to realize the regression prediction of substation equipment temperature.

BiLSTM is the abbreviation of Bi-directional Long Short-Term Memory, which is the combination of forwarding LSTM and backward LSTM, and both LSTM networks are connected to an output layer. This structure provides complete past and future time information for each point in the input sequence of the output layer. The BiLSTM network structure is shown in Figure 11, in which D_t represents the input of the network, refers to the temperature-depth characteristics of substation equipment extracted by CNN in this study; Y_t represents the output of the network, and refers the temperature prediction output of substation equipment in this paper.

4.5. The Implementation Process of SETPM-CLSTR. According to the above description, the specific implementation process of SETPM-CLSTR can be completed in the following five steps:

(1) According to the collected substation equipment temperature data and ambient temperature data, from the two aspects of time and space, the Spatio-

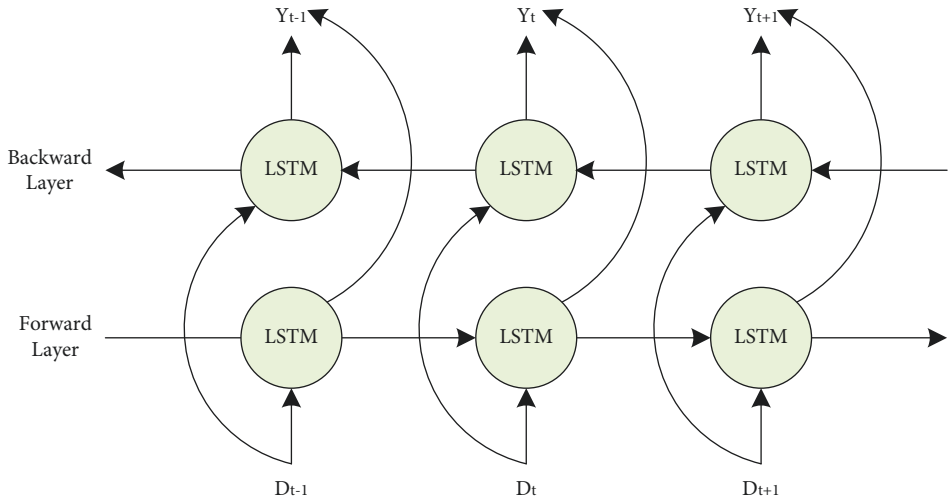


Figure 11: BiLSTM network structure.

temporal feature set is established under the three dimensions of equipment historical temperature, ambient temperature, and spatial local correlation monitoring point temperature.

(2) The grey correlation analysis method is used to analyze the spatial characteristics, determine the closely related spatial location related monitoring points, and combine n_1 historical past temperature of equipment, n_2 closely related temperature of spatial location monitoring points, the daily maximum and minimum temperature to form the feature vector based on local Spatio-temporal relationship;

(3) CNN is applied to extract the depth feature of the feature vector based on the local Spatio-temporal relationship to obtain the depth feature vector.

(4) Train BiLSTM depth network model with training set;

(5) For the test set, the trained BiLSTM model is applied to predict the temperature of substation equipment, and the prediction results are output.

The implementation process of SETPM-CLSTR is shown in Figure 12.

4.6. Temperature Prediction of Substation Equipment Based on CNN-BiLSTM

4.6.1. Specific Implementation Details of the Temperature Prediction. In this study, the CNN-BiLSTM network is used to predict the phase A contact of bushing, where, the CNN filter size is 10; the training cycle is 24 times per round, 60 rounds in total, and the total number of iterations is 1440; the learning rate is 0.005 and the error threshold is 0.001; The input of BiLSTM network is 128, the output is 32, and a full connection is added to get a temperature prediction value.

The prediction results for the test set based on CNN-BiLSTM are shown in Figure 13, and the testing relative error is shown in Figure 14.

It can be seen that the prediction effect of bushing phase A contact temperature based on the CNN-BiLSTM network is good, the main trend of the predicted value and the real value is basically consistent, and the relative error remains between $[-0.02, 0.04]$ in the test set from Figure 14, and this error range is acceptable and can meet the prediction accuracy requirements of substation equipment temperature early warning.

4.6.2. Prediction Performance Evaluation Indexes. In this study, mean absolute percentage error (MAPE) and root mean square error (RMSE) are used as the evaluation indexes of the prediction effect of the model [27, 28], and the calculation formula is shown in formula (16) and formula (17), respectively:

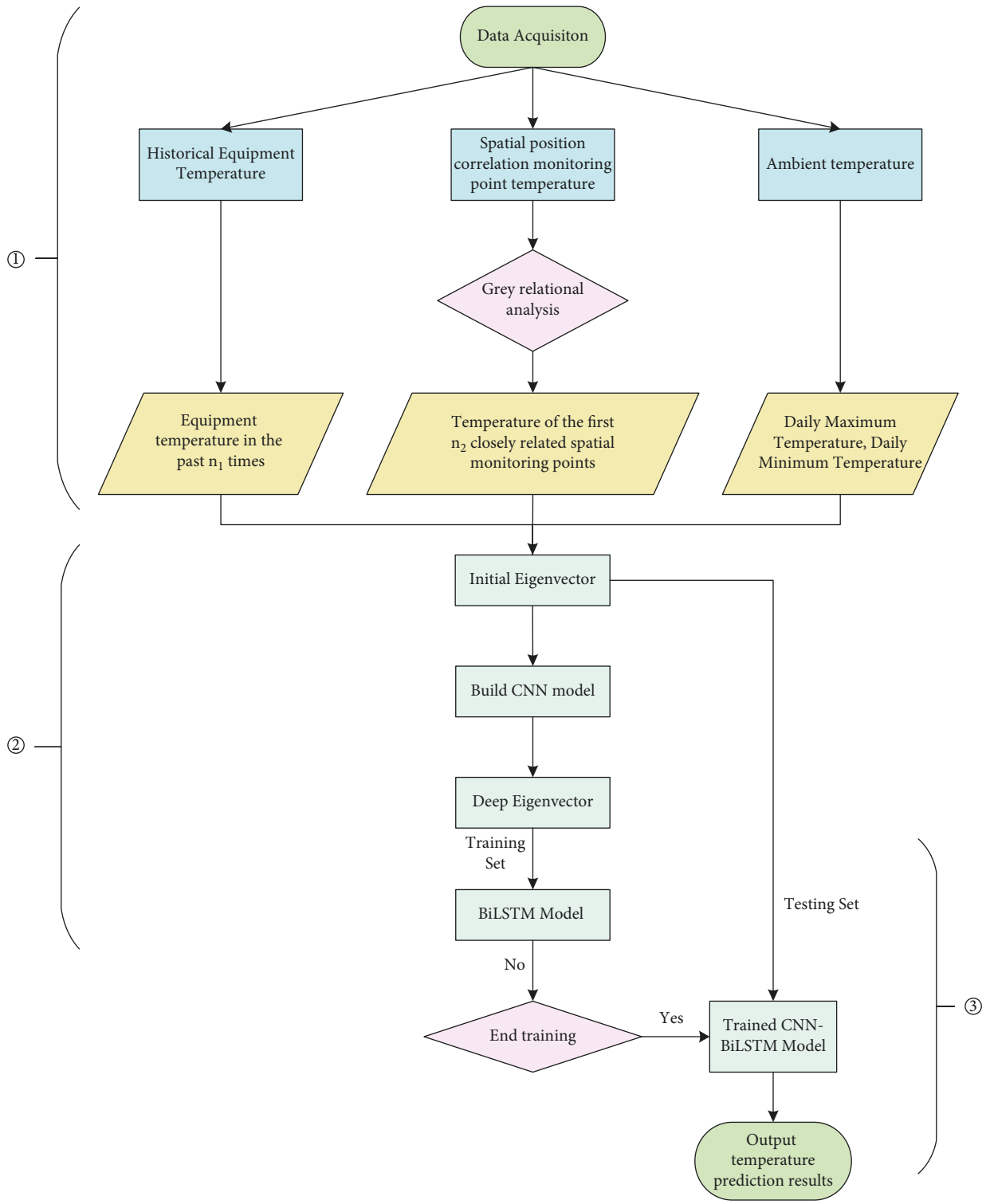
$$\delta_{\text{MAPE}} = \frac{1}{n}\sum_{r=1}^{n}\frac{\left|y_r - \widehat{y}_r\right|}{y_r} \times 100\%, \tag{16}$$

where n represents the number of prediction; y_r is the $r-$th real value of the temperature; $\widehat{y}_r$ is the $r-$th prediction of the temperature; the range of MAPE value is $[0, +\infty)$, the greater the error, the greater the value, therefore, the smaller the value, the better the performance.

$$\delta_{\text{RMSE}} = \sqrt{\frac{1}{n}\sum_{r=1}^{n}\left(y_r - \widehat{y}_r\right)^2}, \tag{17}$$

where n, y_r and $\widehat{y}_r$ have the same means with formula (16). The range of RMSE value is $[0, +\infty)$, the greater the error, the greater the value, and when the predicted value is completely consistent with the real value, RMSE = 0, namely, the perfect model.

4.6.3. Comparative Experiments. In order to verify the effectiveness of the method proposed in this study, two comparative experiments are carried out. On the one hand, to verify that the local Spatio-temporal relationship features are more effective than all features; On the other hand, it verifies the effectiveness of the CNN-BiLSTM double-layer depth network



① Constructing feature vector based on local spatio-temporal relationship, which includes n_1 historical temperature data of the equipment, the temperature of the first n_2 closely related spatial monitoring points obtained by using the grey correlation analysis method, dAily maximum temperature and daily minimum temperature;

② Train CNN-BiLSTM network Model: CNN is applied to extract the depth feature of the feature vector, and BiLSTM network is used to predict the temperature of the equipment;

③ Network moDel test: test the performance of the trained CNN-BiLSTM model and output the predicted temperature of substation equipmtnt.

Figure 12: The implementation process of SETPM-CLSTR.

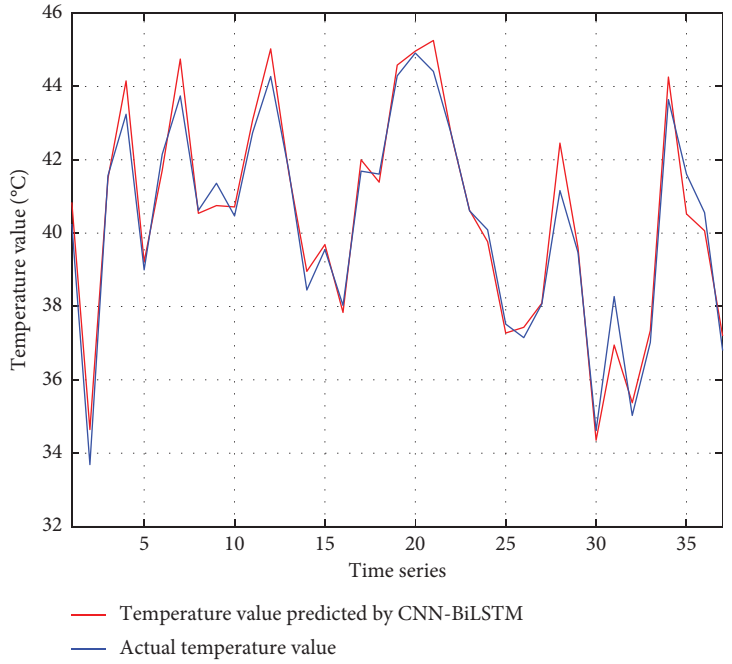


Figure 13: The prediction results for test set based on CNN-BiLSTM.

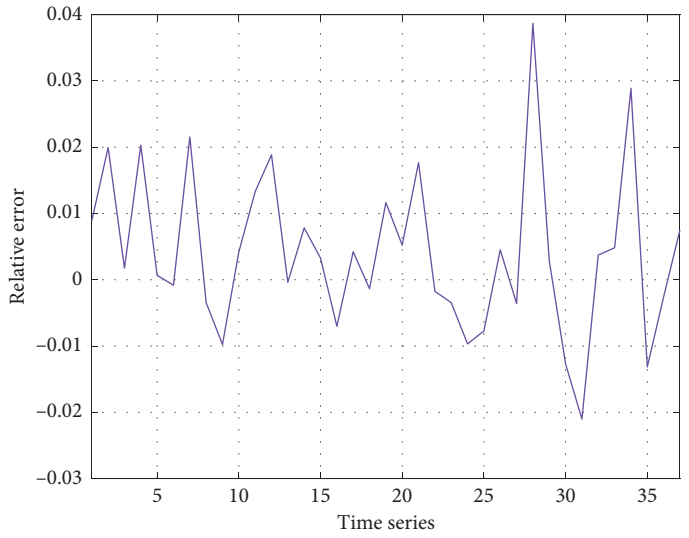


Figure 14: The testing relative error based on CNN-BiLSTM.

Table 4: Comparison results of prediction performance of different models.

Model	All features (18 dimensions)		CLSTR features (10 dimensions)	
	MAPE	RMSE	MAPE	RMSE
CNN-BiLSTM	1.03	60.09	0.94	52.86
CNN-LSTM	1.43	82.03	1.04	61.28
BiLSTM	5.77	634.74	5.47	552.96
LSTM	6.52	720.51	6.27	697.46

model in prediction performance. The comparison results of prediction performance are shown in Table 4. All features (18 dimensions) include two environmental characteristics of the daily maximum temperature and daily minimum temperature, the historical equipment temperature in the past three times and the temperature of 13 spatial location correlation monitoring points, a total of 18 characteristics, and CLSTR features refer to 10 features proposed in this study.

4.6.4. Analysis of Prediction Results. According to the above comparative experiments, this paper analyzes the prediction results according to the statistical results from Table 4:

(1) Under four models, CLSTR features proposed by this study have better performance than all features, which shows that when predicting the temperature of substation equipment, the more parameters, the better the prediction performance of the model;

(2) The model using CNN for depth feature extraction, including the CNN-BiLSTM network and the CNN-LSTM network, has significantly improved the prediction performance compared with the model without CNN(BiLSTM and LSTM), which shows that CNN plays a vital role in the temperature feature extraction of substation equipment.

(3) From the two aspects of feature extraction and combined modeling, the proposed SETPM-CLSTR has significantly improved the temperature prediction performance of substation equipment. Under the two evaluation indexes of MAPE and RMSE, this method has the best prediction performance compared with the other three models.

5. Conclusions

In order to ensure the stable operation of the power grid system and the safe operation of power equipment, an intelligent inspection system is gradually adopted in substations. Temperature early warning based on substation equipment is one of the main branches. Aiming at the difficulties of robot infrared temperature prediction in three aspects: less data, seasonality, and a complex working environment, SETPM-CLSTR is proposed in this study. Using the equipment temperature data of 220 kV side of No. 2 main transformer of a substation in Taizhou City, Zhejiang Province, comparative experiments are carried out in two aspects of different characteristics and different models. It is verified that the method proposed in this study has good prediction performance, and can provide a new idea for the temperature early warning system of intelligent patrol inspection of State Grid.

Although this research has achieved good prediction results, there is less research data in this study, and more research data should be collected in future research. The application of substation equipment temperature prediction for early equipment fault early warning is the content of future research [29, 30].

Data Availability

The equipment of data acquisition is the primary equipment of No. 2 main transformer from a 220 kV substation in Taizhou City, Zhejiang Province, and the temperature data is collected from October 1, 2019 to October 29, 2020 in this study, which can be obtained via e-mail to lijiesun@tzc.edu.cn.

Conflicts of Interest

The authors declare no conflicts of interest.

Authors' Contributions

Lijie Sun wrote the original draft, developed the methodology, helped with software, validated the study, and carried out the experiment. Shuang Chen visualized and investigated the study. Junfei Zhu wrote, reviewed, and edited the study, helped with software, and analysed the study. Jianhua Li supervised and editing the study and carried out typesetting.

Acknowledgments

This work was supported by Research Project of Education Department of Zhejiang Province (Y201840245).

References

[1] Y. Hu, S. Xue, H. Zhang et al., "Deep Cause Analysis and enlightenment of global blackout in recent 30 years," *China Electric Power*, vol. 4, no. 10, pp. 204–210, 2015.

[2] Y. Xin, *Design and Implementation of On-Line Temperature Monitoring and Early Warning System for Power Equipment*, Shandong University, Jinan, China, 2013.

[3] Y. J. Lin, "Substation equipment operation failure causes and treatment countermeasures," *China's new technologies and products*, vol. 13, pp. 60-61, 2017.

[4] M. Baptista, S. Sankararaman, I. P. Medeiros, C. Nascimento, H. Prendinger, and E. M. Henriques, "Forecasting fault events for predictive maintenance using data-driven techniques and ARMA modeling," *Computers & Industrial Engineering*, vol. 115, pp. 41–53, 2018.

[5] C. k. Lei, J. Deng, K. Cao, L. Ma, Y. Xiao, and L. Ren, "A random forest approach for predicting coal spontaneous combustion," *Fuel*, vol. 223, pp. 63–73, 2018.

[6] X. H. Kong and H. F. Zhang, "Substation Equipment Temperature prediction based on optimized generalized regression neural network," *China Electric Power*, vol. 49, no. 7, pp. 54–59, 2016.

[7] X. F. Xu, Z. R. Lin, and J. Zhu, "DVRP with limited supply and variable neighborhood region in refined oil distribution," *Annals of Operations Research*, vol. 309, no. 2, pp. 663–687, 2022.

[8] Z. Tian, C. Luo, H. Lu, S. Su, Y. Sun, and M. Zhang, "User and entity behavior analysis under urban big data," *ACM/IMS Transactions on Data Science*, vol. 1, no. 3, p. 1, 2020.

[9] X. F. Xu, J. Hao, and Y. Zheng, "Multi-objective artificial bee colony algorithm for multi-stage resource leveling problem in sharing logistics network," *Computers & Industrial Engineering*, vol. 142, no. 4, Article ID 106338, 2020.

[10] J. Hou, F. Chen, P. Li, and Z. Zhu, "Gray-box parsimonious subspace identification of Hammerstein-type systems," *IEEE Transactions on Industrial Electronics*, vol. 68, no. 10, pp. 9941–9951, 2021.

[11] T. Yu, Q. Gan, G. Feng, and G. Han, "A new fuzzy cognitive maps classifier based on capsule network," *Knowledge-Based Systems*, vol. 250, Article ID 108950, 2022.

[12] X. F. Xu, Z. R. Lin, X. Li, C. J. Shang, and Q. Shen, "Multi-objective robust optimisation model for MDVRPLS in refined

oil distribution," *International Journal of Production Research*, vol. 5, pp. 1–21, 2021.

[13] Y. Z. Hao and H. L. He, "Application analysis of infrared temperature measurement technology in substation operation and maintenance," *Electrical materials*, vol. 6, pp. 70–72, 2021.

[14] X. F. Xu, C. L. Wang, and P. Zhou, "GVRP considered oil-gas recovery in refined oil distribution: from an environmental perspective," *International Journal of Production Economics*, vol. 235, Article ID 108078, 2021.

[15] Z. Tian, W. Shi, Y. Wang et al., "Real time lateral movement detection based on evidence reasoning network for edge computing environment," *IEEE Transactions on Industrial Informatics*, vol. 15, no. 7, pp. 4285–4294, 2019.

[16] Y. H. Kim, S. H. Nam, S. B. Hong, and K. R. Park, "GRA-GAN: g," *Expert Systems with Applications*, vol. 198, Article ID 116792, 2022.

[17] Z. Tian, C. Luo, J. Qiu, X. Du, and M. Guizani, "A distributed deep learning system for web attack detection on edge devices," *IEEE Transactions on Industrial Informatics*, vol. 16, no. 3, pp. 1963–1971, 2020.

[18] X. Gao, X. Li, B. Zhao, W. Ji, X. Jing, and Y. He, "Short-term electricity load forecasting model based on EMD-GRU with feature selection," *Energies*, vol. 12, no. 6, p. 1140, 2019.

[19] X. F. Xu and Y. Y. He, "Blockchain Application in Modern Logistics Information Sharing: A Review and Case Study Analysis," *Production Planning & Control*, vol. 107, 2022.

[20] J. Yin, C. Ning, and T. Tang, "Data-Driven Models for Train Control Dynamics in High-Speed Railways: LAG-LSTM for Train Trajectory Prediction," *Information Sciences*, vol. 600, pp. 377–400, 2022.

[21] K. Feng and Z. Fan, "A novel bidirectional LSTM network based on scale factor for atrial fibrillation signals classification," *Biomedical Signal Processing and Control*, vol. 76, Article ID 103663, 2022.

[22] Y. Wang, X. Jiang, F. Yan, Y. Cai, and S. Liao, "The GRA-two algorithm for massive-scale feature selection problem in power system scenario classification and prediction," *Energy Reports*, vol. 7, pp. 293–303, 2021.

[23] F. L. Yin, X. M. Xue, C. Z. Zhang et al., "Multifidelity genetic transfer: an efficient framework for production optimization," *SPE Journal*, vol. 26, no. 4, pp. 1614–1635, 2021.

[24] L. Wu and L. Noels, "Recurrent Neural Networks (RNNs) with dimensionality reduction and break down in computational mechanics application to multi-scale localization step," *Computer Methods in Applied Mechanics and Engineering*, vol. 390, Article ID 114476, 2022.

[25] Y. Zhang, C. Li, Y. Jiang et al., "Accurate prediction of water quality in urban drainage network with integrated EMD-LSTM model," *Journal of Cleaner Production*, vol. 354, Article ID 131724, 2022.

[26] L. Yang, C. Ma, H. Hu, and D. Huang, "Study on Regional Medium and long-term temperature prediction based on improved LSTMs model," *Journal of Huizhou College*, vol. 41, no. 6, pp. 75–79, 2021.

[27] X. F. Xu, W. Liu, and L. A. Yu, "Trajectory prediction for heterogeneous traffic-agents using knowledge correction data-driven model," *Information Sciences*, vol. 608, pp. 375–391, 2022.

[28] Y. Wang, C. Zhu, Y. Wang, J. Sun, D. Ling, and L. Wang, "Survival risk prediction model for ESCC based on relief feature selection and CNN," *Computers in Biology and Medicine*, vol. 145, Article ID 105460, 2022.

[29] Z. Liang, G. Q. Sun, H. C. Li et al., "Short-term load forecasting based on deep belief network optimized by VMD and PSO," *Grid Technology*, vol. 42, no. 2, pp. 598–606, 2018.

[30] K. Hu, Y. Wang, W. Li, and L. Wang, "CNN-BiLSTM enabled prediction on molten pool width for thin-walled part fabrication using Laser Directed Energy Deposition," *Journal of Manufacturing Processes*, vol. 78, pp. 32–45, 2022.

Hindawi
Scientific Programming
Volume 2022, Article ID 6446903, 12 pages
https://doi.org/10.1155/2022/6446903

Research Article

An Orthogonal Matching Pursuit Variable Screening Algorithm for High-Dimensional Linear Regression Models

Yanxi Xie,[1] **Yuewen Li**,[1] **Victor Shi**,[2] **and Quan Lu**[3]

[1]*School of Management, Shanghai University of Engineering Science, Shanghai, China*
[2]*Lazaridis School of Business and Economics, Wilfrid Laurier University, Waterloo, Canada N2L 3C5*
[3]*School of Finance and Trade, Wenzhou Business College, Wenzhou, China*

Correspondence should be addressed to Quan Lu; luq@wzbc.edu.cn

Received 22 April 2022; Accepted 22 June 2022; Published 1 August 2022

Academic Editor: Hongguang Ma

Variable selection plays an important role in data mining. It is crucial to filter useful variables and extract useful information in a high-dimensional setup when the number of predictor variables d tends to be much larger than the sample size n. Statistical inferences can be more precise after irrelevant variables are moved out by the screening method. This article proposes an orthogonal matching pursuit algorithm for variable screening under the high-dimensional setup. The proposed orthogonal matching pursuit method demonstrates good performance in variable screening. In particular, if the dimension of the true model is finite, OMP might discover all relevant predictors within a finite number of steps. Throughout theoretical analysis and simulations, it is confirmed that the orthogonal matching pursuit algorithm can identify relevant predictors to ensure screening consistency in variable selection. Given the sure screening property, the BIC criterion can be used to practically select the best candidate from the models generated by the OMP algorithm. Compared with the traditional orthogonal matching pursuit method, the resulting model can improve prediction accuracy and reduce computational cost by screening out the relevant variables.

1. Introduction

Variable screening is an important technique in data mining. It captures informative variables by reducing the dimension in a high-dimensional setup when the number of predictor variables d tends to be much larger than the sample size n. However, statistical inference is difficult to compute in ultra-high dimensional linear models before variable screening due to the computational complexity. It is necessary to remove the irrelevant variables from the model before statistical inference. The core idea is to screening out the informative variables with the aim of building a relevant model for future prediction. By removing most irrelevant and redundant variables from the data, variable selection helps improve the performance of learning models in terms of obtaining higher estimation accuracy [1]. Then the AIC [2] or BIC [3] can be applied to further guarantee the accuracy of the relevant model.

The focus of this article is on ultra-high dimensional linear models, in which the number of predictor variables d tends to be much larger than the sample size n. In particular, the number of covariates may increase at an exponential rate. Such linear models have gained a lot of attention in practical areas, such as sentiment analysis and finance. Existing techniques in the past literature include forward selection [1], least absolute shrinkage and selection operator (Lasso) [4], smoothly clipped absolute deviation penalty (SCAD) [5], etc. These efforts have been devoted to the challenging ultra-high dimensionality problem, which is motivated by contemporary applications such as bio-informatics, genomics, finance, etc. In other words, it is becoming a major issue to investigate the existence of complex relationships and dependencies in data with the aim of building a relevant model for inference. A practically attractive approach is to first use a quick screening procedure to reduce the dimensionality of the covariates to a

reasonable scale, for example below the sample size, and then apply variable selection techniques such as LASSO and SCAD in the second stage.

Motivated by the current studies on variable screening approaches in ultra-high dimensional linear models, it is of our interests in showing the screening consistency property of the OMP under certain conditions, by restricting the technical conditions stated in Wang [6] and hence selecting a subset of predictors, which includes all relevant predictors, to ensure variable screening results.

The rest of this article is organized as follows: Section 2 provides the literature review on current variable screening methods. Section 3 demonstrates a variable screening algorithm based on the OMP. Furthermore, the asymptotic results of the estimators are studied. Section 4 examines via simulation that our proposed technique exhibits desired sample properties and can be useful in practical applications. Finally, Section 5 concludes the article and provides some future research direction. In particular, the proof of the asymptotic theories and lemmas can be found in Appendix.

2. Literature Review

In the content of variable selection, screening approaches have gained a lot of attention besides the penalty approaches such as Lasso [4] and SCAD [5]. When the predictor dimension is much larger than the sample size, the story changes drastically in the sense that the conditions for most of the Lasso-type algorithms cannot be satisfied. Therefore, to conduct model selection in the high-dimensional setup, variable screening is a reasonable solution.

Sure independence screening (SIS), which is proposed by Fan and Lv [7], has gained popularity under the condition when the number of predictor variables d tends to be much larger than the sample size n. Sure screening means a property that all the important variables are selected after applying a variable screening procedure with probability tending to 1. It is desired to have a dimensionality reduction method with the sure screening property. There are three facts why sure screening is of great importance and usage when dimension d is larger than sample size n, which is clearly stated in Fan and Lv [7]. First of all, the design matrix X is rectangular, having more columns than rows. In this case, the matrix $X^T X$ is giant in dimension and singular. The maximum spurious correlation between a covariate and a response can be large due to the dimensionality and the fact that an unimportant predictor can be highly correlated with the response variable owing to the presence of important predictors associated with the predictor. In addition, the population covariance matrix Σ may become ill conditioned as n grows, and it makes variable selection difficult. Third, the minimum nonzero absolute coefficient $|\beta_j|$ may decay with n and fall close to the noise level, say, the order $\{\log(d)/n\}^{(-1/2)}$. Hence, in general, it becomes challenging to estimate the sparse parameter vector β accurately when $d \gg n$.

To solve the abovementioned difficulties in variable selection, Fan and Lv [7] proposed a simple sure screening method using componentwise regression or equivalently correlation learning, to reduce dimensionality from high to moderate scale that is below sample size. Below is the description of the SIS method.

Let $\omega = (\omega_1, \ldots, \omega_d)^T$ be a d-vector that is obtained by componentwise regression, that is

$$\omega = X^T y, \tag{1}$$

where the $n \times d$ data matrix X is first standardized columnwise. For any given $\gamma \in (0, 1)$, we sort the d componentwise magnitudes of the vector ω in a descending order and define a submodel

$$M_\gamma = \left\{1 \le i \le d:\ |\omega_i| \text{ is among the first } [\gamma n] \text{ largest of all}\right\}, \tag{2}$$

where $[\gamma n]$ denotes the integer part of γn. It shrinks the full model $\{1, 2, \ldots, d\}$ down to a submodel M_γ with size $[\gamma n]$ smaller than the sample size n. This correlation learning ranks the importance of features according to their marginal correlations with the response variable. Moreover, it is called the independence screening because each feature is used independently as a predictor to decide the usefulness for predicting the response variable. The computational cost of SIS is of order $O(nd)$.

With dimension reduced accurately from high to below sample size, variable selection can be improved on both speed and accuracy, and can then be accomplished by a well-developed method such as SCAD, Lasso, or adaptive Lasso [8, 9], denoted by SIS-SCAD, SIS-Lasso, or SIS-AdapLasso, respectively. Moreover, sure screening property has been proven in Fan and Lv [7]. Intuitively, the core idea of SIS is to select the variables by two stages. In the first stage, an easy-to-implement method is used to remove the least important variables. In the second stage, a more sophisticated and accurate method is applied to reduce the variables further.

Though SIS enjoys sure screening property and is easy to be applied, it has several potential problems. First of all, if there is an important predictor jointly correlated but marginally uncorrelated with the response variable, it is not selected by SIS and thus cannot be included in the estimated model. Second, similar to Lasso, SIS cannot handle the collinearity problem between predictors in terms of variable selection. Third, when there are some unimportant predictors which are highly correlated with the important predictors, these unimportant predictors can have higher chance of being selected by SIS than other important predictors that are relatively weakly related to the response variable. In all, these three potential issues can be carefully treated when some extensions of SIS are proposed. In particular, iterative SIS (ISIS) is designed to overcome the weakness of SIS.

ISIS works in two steps. In the first step, a subset of k_1 variables $\mathbf{A}_1 = \left\{X_{i_1}, \ldots, X_{i_{k_1}}\right\}$ is selected by using an SIS-based model selection method such as SIS-SCAD or SIS-Lasso methods. There is an n-vector of residuals from regressing the response y over $X_{i_1}, \ldots, X_{i_{k_1}}$. In the second step, the residuals are treated as the new response variable and the previous step is repeated to the remaining $d - k_1$

variables. It returns a subset of k_2 variables $\mathbf{A}_2 = \left\{X_{j_1}, \ldots, X_{j_{k_2}}\right\}$. Fitting the residuals from the previous step on $\{X_1, \ldots, X_d\}\mathbf{A}_1$ can significantly weaken the prior selection of those unimportant variables that are highly correlated with the response through their relations with $X_{i_1}, \ldots, X_{i_{k_1}}$. In addition, the second step also makes those important variables which are missed out in the first step possible to be selected. Iteratively, the second step is iterated until l disjoint subsets $\mathbf{A}_1, \ldots, \mathbf{A}_l$ are obtained with the union $\mathbf{A} = \cup_{j=1}^{l}\mathbf{A}_j$ has a size $[\gamma n]$.

If SIS is used to select only one variable at each iteration, that is $|\mathbf{A_i}| = 1$, ISIS is equivalent to orthogonal matching pursuit (OMP) [10], which is a greedy algorithm for variable selection. This is discussed in the study by Barron and Cohen [11].

Kim [12] proposed a filter ranking method using the elastic net penalty with sure independence screening (SIS) on resampling technique to overcome the overfitting and high-performance computational issues. It is demonstrated via extensive simulation studies that SIS-LASSO, SIS-MCP, and SIS-SCAD with the proposed filtering method achieve superior performance of not only accuracy but also true positive detection compared to those with the marginal maximum likelihood ranking (MMLR) method.

Another very popular yet classical variable screening method is the forward regression (FR). As one type of important greedy algorithms, FR's theoretical properties have been investigated in the studies by Barron and Cohen [11], Donoho and Stodden [13], and Wang [6]. In particular, Wang [6] investigated FR's screening consistency property, under an ultra-high dimensional setup, by introducing the four technical conditions.

There are a few comments on those four technical conditions introduced in the study by Wang [6]. First of all, the normality assumption has been popularly used in the past literature for theory development. Second, the smallest and largest eigenvalues of the covariance matrix Σ need to be properly bounded. This bounded condition together with the normality assumption ensures the sparse Riesz condition (SRC) defined in the study by Zhang and Huang [14]. Third, the standard L_2 norm of the regression coefficients β is bounded above by some proper constant. It guarantees that the signal-to-noise ratio is convergent. Moreover, the minimum value of the nonzero β_js needs to be bounded below. This constraint on the minimal size of the nonzero regression coefficient ensures that relevant predictors can be correctly selected. Otherwise, if some of the nonzero coefficients converge too fast, they cannot be selected consistently. Last but not least, $\log(d)$ is bounded above in the order of n^{ξ} for some small constant ξ. This condition allows the predictor dimension d to diverge to infinity at an exponential fast speed, which implies that the predictor dimension can be substantially larger than the sample size n.

Under the assumption that the true model T exists, Wang [6] introduces the FR algorithm in the aim of discovering all relevant predictors consistently. The main step of FR algorithm is the iterative forward regression part. Consider the case where $k-1$ relevant predictors have been selected accordingly. Then the next step is to construct a candidate model that include one more predictor that belongs to the full set but excluding the selected $k-1$ predictors and calculate the residual sum of squares based on the constructed candidate model. This step is repeated for each predictor that belongs to the full set but excluding the selected $k-1$ predictors, and all the residual sum of squares are recorded accordingly. The minimum value of all the recorded residual sum of squares are found, and the kth relevant predictor is updated based on the index of the corresponding minimum residual sum of squares. A detailed algorithm can be found in the study by Wang [6].

Wang [6] showed the theoretical proof that FR can identify all relevant predictors consistently, even if the predictor dimension is considerably larger than the sample size. In particular, if the dimension of the true model is finite, FR might discover all relevant predictors within a finite number of steps. In other words, sure screening property can be guaranteed under the four technical conditions. Given the sure screening property, the recently proposed BIC of Chen and Chen [3] can be used to practically select the best candidate from the models generated by the FR algorithm. The resulting model is good in the sense that many existing variable selection methods, such as Adaptive Lasso and SCAD, can be applied directly to increase the estimation accuracy.

The extended Bayes information criterion (EBIC) proposed by Chen and Chen [3] is suitable for large model spaces. It has the following form:

$$\mathrm{BIC}(M) = \log\widehat{\sigma}^2_{(M)} + n^{-1}|M|(\log n + 2\log d), \tag{3}$$

where M is an arbitrary candidate model with $|M| \le n$, $\widehat{\sigma}^2_{(M)} = n^{-1}\mathrm{RSS}(M) = n^{-1}Y^T\left\{I_n - H_{(M)}\right\}Y$, and $H_{(M)} = X_{(M)}\left\{X^T_{(M)}X_{(M)}\right\}^{-1}X^T_{(M)}$. We then select the best model $\widehat{S} = S^{(\widehat{m})}$, where $\widehat{m} = \mathrm{argmin}_{1\le m\le n}\mathrm{BIC}(S^{(m)})$.

EBIC, which includes the original BIC as a special case, examines both the number of unknown parameters and the complexity of the model space. The model in Chen and Chen [3] is defined to be *identifiable* if no model of comparable size other than the true submodel can predict the response almost equally well. It has been shown that EBIC is selection consistent under some mild conditions. It also handles the heavy collinearity problem for the covariates. Furthermore, EBIC is easy to implement due to the fact the extended BIC family does not require a data adaptive tuning parameter procedure.

Other screening approaches include tournament screening (TS) [15], sequential Lasso [16], quantile-adaptive model-free variable screening [17], and conditional screening [18]. When $P \gg n$, the tournament screening possesses the sure screening property to reduce spurious correlation. Furthermore, the asymptotic properties of sequential Lasso for feature selection in linear regression models with ultra-high dimensional feature spaces are investigated. The advantage of sequential Lasso is that it is not restricted by the dimensionality of the feature space.

Quantile-adaptive model-free variable screening has two distinctive features, allowing the set of active variables to vary across quantiles and overcoming the difficulty in specifying the form of a statistical model in a high-dimensional space. Baranowski [19] proposed a workflow representation for scheduling, provenance, or visualization to resolve variable and method dependencies and evaluated the performance of screening properties. Samudrala [20] proposed a parallel algorithm by identifying key components for dimensionality reduction of large-scale data. It shows better performance for dimension reduction compared to the existing methods. Chen [21] proposed a model-free feature screening method when the censored response and error-prone covariates both exist. An iterative algorithm is developed in the presence of the censored response and error-prone covariates. In addition, we also develop the iteration method to improve the accuracy of selecting all important covariates. Choudalakis [22] proposed appropriate numerical methods for parameter estimation under the high-dimensional setup. A thorough comparison is considered among existing methods for both coefficient estimations and variable selection for supersaturated designs. Xu et al. [23–25] proposed several multi-objective robust optimization models for MDVRPLS in refined oil distribution. Ren [26] proposed an asymmetric learning to hash with variable bit encoding algorithm (AVBH) to solve the high-dimensional data problem, and a real data application is applied for the finite performance of the proposed AVBH algorithm. We proposed a parallel framework for dimensionality reduction of large-scale data. We also identified key components underlying the spectral dimensionality reduction techniques and proposed their efficient parallel implementation.

3. Main Results

Orthogonal matching pursuit (OMP) is an iterative greedy algorithm that selects at each step the column which is most correlated with the current residuals. The selected column is then added into the set of selected columns. Inspired by the idea of the FR algorithm in Wang (2009), it is shown that under some proper conditions, OMP can enjoy the sure screening property in the linear model setup.

3.1. Model Setup and Technical Conditions. Let (X_i, y_i) be the observation collected from the i^{th} subject $(1 \le i \le n)$, where $y_i \in R^1$ is the response and $X_i = (X_{i1}, \ldots, X_{id})^T \in R^d$ is the high-dimensional predictor with $d > n$ and $\mathrm{cov}(X_i) = \Sigma$. Moreover, $\beta = (\beta_1, \ldots, \beta_d)^T$ is the regression coefficient. In matrix representation, the design matrix is $X \in R^{n \times d}$ and the response vector is $y \in R^n$. Consider the linear regression model as

$$y = X\beta + \varepsilon. \tag{4}$$

Without loss of generality, it is assumed that the data are centered, that is the columns of X are orthonormal and y_is are conditionally independent given the design matrix X. Equivalently, $E(X_{ij}) = 0$ and $\mathrm{Var}(X_{ij}) = 1$. Moreover, the error term ϵ are independently and identically distributed with mean zero and finite variance σ^2. A model fitting procedure produces the vector of coefficients $\widehat{\beta} = (\widehat{\beta}_1, \ldots, \widehat{\beta}_d)^T$.

Before the main result for the screening property of OMP is presented, four technical conditions are needed as follows:

3.1.1. Assumption 1. Technical Conditions.

(C1) Normality assumption. Assume that ε follows the normal distribution.

(C2) Covariance matrix. $\lambda_{\min}(A)$ and $\lambda_{\max}(A)$ represent, respectively, the smallest and largest eigenvalues of an arbitrary positive definite matrix A. We assume that there exist a positive constant $1 < c < \infty$, such that $c^{-1} < \lambda_{\min}(\Sigma) \le \lambda_{\max}(\Sigma) < c$.

(C3) Regression coefficients. We assume that $\|\beta\| \le C_\beta$ for some constant $C_\beta > 0$, where $\|.\|$ denotes the standard L_2 norm. Also assume $\beta_{\min} \ge \nu_\beta n^{-\xi_{\min}}$ for some $\xi_{\min} > 0$, with $\beta_{\min} = \min_{j \in T} |\beta_j|$.

(C4) Divergence speed of d and d_0. We assume $\log(d) = O(n^\xi)$ and $d_0 = O(n^{\xi_0})$. In other words, there exists constants ξ, $\xi_{\min}$, and ν, such that $\log(d) \le \nu n^\xi$, $d_0 \le \nu n^{\xi_0}$, and $\xi + 6\xi_0 + 12\xi_{\min} < 1$.

3.2. OMP Algorithm. Under the assumption that the true set T exists, our main objective is to discover all relevant predictors consistently. To this end, we consider the following OMP algorithm (Algorithm1):

3.3. Theoretical Properties. To prove Theorem 1, the following lemmas are needed. For convenience, we define $\widehat{\Sigma} = n^{-1} X^T X$. Moreover, $\widehat{\Sigma}_{(M)}$ and $\Sigma_{(M)}$ are the submatrices of $\widehat{\Sigma}$ and Σ (corresponding to M), respectively.

Lemma 1. *If we have* $(1/c) < \lambda_{min}(\Sigma) < \lambda_{max}(\Sigma) < c$, *then eigenvalues of the submatrix of* Σ, *that is* Σ_{11}, *are also bounded. Suppose* Σ_{11} *contains* m *variables, that is* Σ_{11} *is a* $m \times m$ *matrix. Moreover, with probability tending to 1, we have*

$$(1/c) < \lambda_{\min}\left(\widehat{\Sigma}_{11}\right) < \lambda_{\max}\left(\widehat{\Sigma}_{11}\right) < c, \tag{5}$$

as long as $m = O(n^{((1-\xi)/3)})$.

The Proof of Lemma 1 can be found in Appendix.

Lemma 2. *Assume conditions (C1), (C2), (C4), and* $m = O(n^{((1-\xi)/3)})$. *Then, with probability tending to 1, we have*

$$\tau_{\min} \le \min_{|M| \le m} \lambda_{\min}\left\{\widehat{\Sigma}_{(M)}\right\} \le \max_{|M| \le m} \lambda_{\max}\left\{\widehat{\Sigma}_{(M)}\right\} \le \tau_{\max}. \tag{6}$$

Note that proof of Lemma 2 can be found in the study by Wang (2009), with only slight changes.

Before the theorems are established, we follow Wang (2009)'s idea on screening consistency of a solution path and define the solution path S to be screening consistent, if

Step 1 (Initialization). Set $S^{(0)} = \varnothing$. Set the residual $r_0 = y$.
Step 2 (Forward selection).
(i) (2.1) Evaluation. In the kth step $(k \geq 1)$, we are given $S^{(k-1)}$. Then, for every $j \in F \setminus S^{(k-1)}$, we compute $X_j^T r_{k-1}$,
where $r_{k-1} = (I_n - H_{(S^{(k-1)})})y$, $H_{(S^{(k-1)})} = X_{S^{(k-1)}}\left\{X_{(S^{(k-1)})}^T X_{(S^{(k-1)})}\right\}^{-1} X_{(S^{(k-1)})}^T$ is the projection onto the linear space spanned by the elements of $X_{(S^{(k-1)})}$ and $I_n \in R^{n\times n}$ is the identity matrix.
(ii) (2.2) Screening. We then find $a_k = \text{argmax}_{j\in F\setminus S^{(k-1)}} |X_j^T r_{k-1}|$,
and update $S^{(k)} = S^{(k-1)} \cup \{a_k\}$ accordingly.
Step 3 (Solution path). Iterating Step 2 for n times leads a total of n nested candidate models. We then collect those models by a solution path $S = \{S^{(k)}: 1 \leq k \leq n\}$ with $S^{(k)} = \{a_1, \ldots, a_k\}$.

Algorithm 1: The OMP algorithm.

$$P\left(T \subset S^{(k)} \in S \text{ for some } 1 \leq k \leq n\right) \longrightarrow 1. \tag{7}$$

Then OMP's screening consistency can be formally established by the following theorem.

Theorem 1. *Under model (4) and conditions (C1)–(C4), we have as* $n \longrightarrow \infty$

$$P\left(T \subset S^{\left(\left[Kn^{\xi_0+4\xi_{\min}}\right]\right)}\right) \longrightarrow 1, \tag{8}$$

where the constant $K = 2\tau_{max}\tau_{min^{-2}}C_\beta^2\nu_\beta^{-4}\nu$ *is independent of* n, *the constants* τ_{max}, τ_{min}, C_β, ν_β, *and* ν *are defined in conditions (C2)–(C4), and* $[t]$ *is the smallest integer no less than* t.

Theorem 1 proves that within $Kn^{\xi_0+4\xi_{\min}}$ steps, all relevant predictors will be identified by the OMP algorithm. This number of steps is much smaller than the sample size n under condition (C4). In particular, if the dimension of the true model is finite with $\xi_0 = \xi_{\min} = 0$, only a finite number of steps are needed to discover the entire relevant variable set.

Furthermore, Theorem 1 provides a theoretical basis for OMP, which enables us to empirically select the best model from S. On the other hand, the solution path S contains a total of n nested models. To further select relevant variables from the solution path S, the following BIC (Chen and Chen, 2008) is considered,

$$\text{BIC}(M) = \log\widehat{\sigma}_{(M)}^2 + n^{-1}|M|(\log n + 2\log d), \tag{9}$$

where M is an arbitrary candidate model with $|M| \leq n$, $\widehat{\sigma}_{(M)}^2 = n^{-1}\text{RSS}(M) = n^{-1}Y^T\left\{I_n - H_{(M)}\right\}Y$, and $H_{(M)} = X_{(M)}\left\{X_{(M)}^T X_{(M)}\right\}^{-1}X_{(M)}^T$. We then select the best model $\widehat{S} = S^{(\widehat{m})}$, where $\widehat{m} = \arg\min_{1\leq m\leq n} \text{BIC}(S^{(m)})$. We typically do not expect $\widehat{S}$ to be selection consistent (i.e., $P(\widehat{S} = T) \longrightarrow 1$). However, we are able to show that $\widehat{S}$ is indeed screening consistent.

Theorem 2. *Under model (4) and conditions (C1)–(C4), then as* $n \longrightarrow \infty$

$$P(T \subset \widehat{S}) \longrightarrow 1. \tag{10}$$

Define $k_{\min} = \min_{1\leq k\leq n}\{k: T \subset S^{(k)}\}$. By Theorem 2.1, we know that $k_{\min}$ satisfies $k_{\min} \leq Kn^{\xi_0+4\xi_{\min}}$ with probability tending to 1. Therefore, our aim is to prove that $P(\widehat{m} \leq k_{\min}) \longrightarrow 0$ as $n \longrightarrow \infty$. Then the theorem conclusion follows. Equivalently, it suffices to show that

$$P\left(\min_{1\leq k<k_{\min}}\left\{\text{BIC}\left(S^{(k)}\right) - \text{BIC}\left(S^{(k+1)}\right)\right\} > 0\right) \longrightarrow 1. \tag{11}$$

A detailed proof can be found in Appendix.

4. Numerical Analysis

4.1. Simulation Setup. For a reliable numerical comparison, the following three simulation examples on OMP algorithm are presented, to examine the finite performances of the screening consistent property of OMP. For each parameter setup, a total of $N = 100$ simulation replications are conducted.

Let $\widehat{S}^{(k)} = \left\{j: \widehat{\beta}_{j(k)} \neq 0\right\}$ be the model selected in the kth simulation replications and the corresponding average model size $= 100^{-1}\sum_k|\widehat{S}^{(k)}|$. Recall T represents the true model, and the coverage probability= $100^{-1}\sum_k I(\widehat{S}^{(k)^T})$, which measures how likely all relevant variables is discovered by the method, is evaluated. This defined coverage probability characterizes the screening property of a particular method.

To characterize the capability of a method in producing sparse solutions, we define

$$\text{Percentage of Correct Zeros}(\%) = \frac{1}{d-d_0}\left\{\frac{1}{100}\sum_{k=1}^{100}\sum_{j=1}^{d} I\left(\widehat{\beta}_{j(k)}=0\right)\times I\left(\beta_j=0\right)\right\}. \tag{12}$$

To characterize the method's underfitting effect, we further define

$$\text{Percentage of Incorrect Zeros}(\%) = \frac{1}{d_0}\left\{\frac{1}{100}\sum_{k=1}^{100}\sum_{j=1}^{d} I\left(\widehat{\beta}_{j(k)}=0\right)\times I\left(\beta_j\neq 0\right)\right\}. \tag{13}$$

If all sparse solutions are correctly identified for all irrelevant predictors and no sparse solution is mistakenly produced for all relevant variables, the true model is perfectly identified, that is $\widehat{S}^{(k)} = T$. To measure such a performance, we define the percentage of correctly fitted (%)= $100^{-1}\sum_k I(\widehat{S}^{(k)} = T)$, which characterizes the selection consistency property of a particular method.

As we need to know which variables are truly relevant or irrelevant, we create sparse regression vectors by setting $\beta_i = 0$ for all $i = 1, \ldots, d$, except for a chosen set T of coefficients, where β_i is defined in advance for every $1 \le i \le d_0$. Moreover, the noise vector $(\epsilon_1, \ldots, \epsilon_n)$ is chosen i.i.d. $N(0,1)$. Note that all the simulation runs are conducted in MATLAB.

Example 1. (independent predictors). This is an example borrowed from Fan and Lv [7]. X_i is generated independently according to a standard multivariate normal distribution. Thus, different predictors are mutually independent. $(n, d, d_0) = (100, 5000, 8)$ with $\beta_j = (-1)^{U_j}(4\log n\sqrt{n} + |Z_j|)$, where U_j is a binary random variable with $P(U_j) = 0.4$ and Z_j is a standard normal random variable.

Example 2. (autoregessive correlation). X_i is generated from a multivariate normal distribution with mean 0 and $\text{Cov}(X_{ij_1}, X_{ij_2}) = 0.5^{|j_1 - j_2|}$. This is called an autoregressive type correlation structure. Such type of correlation structure might be useful if a natural order exists among the predictors. As a consequence, the predictors with large distances in order are expected to be mutually independent approximately. This is an example from Tibshrani [4] with $(n, d, d_0) = (100, 5000, 3)$. In addition, the first, fourth, and seventh components of β are set to be 3, 1.5, and 2, respectively.

Example 3. (grouped variables). X_i is generated by the following rule. $X_{ij} = \sqrt{3/20}Z_1 + \sqrt{17/20}\epsilon_{x,j}$ for $j \in \{1, \ldots, d_0\}$, $X_{ij} = \sqrt{19/20}Z_2 + \sqrt{1/20}\epsilon_{x,j}$ for $j \in \{d_0 + 1, \ldots, d_0 + 5\}$, and $X_{ij} = \epsilon_{x,j}$ otherwise, where $Z_1 \sim N(0,1)$, $Z_2 \sim N(0,1)$, and $\epsilon_{x,j} \sim N(0,1)$ are independent. This creates within-group correlations of $\rho_{ij} = 0.15$ for $i, j \in \{1, \ldots, d_0\}$ and $\rho_{ij} = 0.95$ for $i, j \in \{d_0 + 1, \ldots, d_0 + 5\}$. This example presents an interesting scenario where a group of significant variables are mildly correlated and simultaneously a group of insignificant variables are strongly correlated. The settings are similar to those in Example 2. $(n, d, d_0) = (100, 5000, 3)$. In addition, the three nonzero components of β are set to be 3, 1.5, and 2, respectively.

4.2. Simulation Results for OMP Screening Consistent Property. Finite sample performance of OMP screening consistent property is investigated based on the above-mentioned three examples in Section 4.1. Simulation results are presented in Table 1.

First of all, simulation results for the independent predictor example are in good performance in terms of screening selection consistency for OMP. In other words, we have 100% coverage probability, which means all relevant variables can be discovered by OMP method. In addition, 94% of correctly fitted denotes that BIC selects the true set of variables correctly 94 times out of 100 simulation replications. This result is not surprising since Zhang [14] pointed out that OMP can select features or variables consistently under a certain irrepresentable condition. Furthermore, the percentage of correct zeros and the percentage of incorrect zeros are 99.9% and 1.6%, respectively. Last but not least, the average model size is 7.94, which is slightly below $d_0 = 8$.

Furthermore, simulation results for autoregressive correlation example are in very good performance in terms of screening selection consistency for OMP. Both of the coverage probability and the percentage of correctly fitted are 100%. Especially 100% of correctly fitted denotes BIC selects the true set of variables correctly 100 times out of 100 simulation replications. This is good news since the number of nonzero $\beta_j d_0$ is 3, which is a very sparse representation given $d = 5000$. On top of that, the percentage of correct zeros and the percentage of incorrect zeros are 100% and 0%, respectively. Last but not least, the average model size is 3. Therefore, it seems that our OMP algorithm works pretty well under this autoregressive correlation setup with a sparse representation of β.

Third, simulation results for grouped variables example are in worst performance among all the three examples in terms of screening selection consistency for OMP. However, the performance itself is not bad. Coverage probability is 96%, meaning that not all the relevant predictors can be discovered by OMP algorithm in some of the simulation replications. In addition, 34% of correctly fitted denotes that BIC selects the true set of variables correctly only 34 times out of 100 simulation replications. On top of that, the percentage of correct zeros and the percentage of incorrect zeros are 95.9% and 2%, respectively. Last but not least, the average model size is 3.84.

Besides a summary of simulation results of OMP algorithm, three plots are presented in Figure 1. For each of the three examples, one particular plot of number of variables included in the final model versus BIC is extracted for

Table 1: Simulation summary for OMP with $(n, d) = (100, 5000)$.

	d_0	Coverage probability (%)	Percentage of correct zeros (%)	Percentage of incorrect zeros (%)	Percentage of correctly fitted (%)	Average model size
IP	8	100	99.9	1.6	94	7.94
AC	3	100	100	0	100	3
GV	3	96	95.9	2	34	3.84

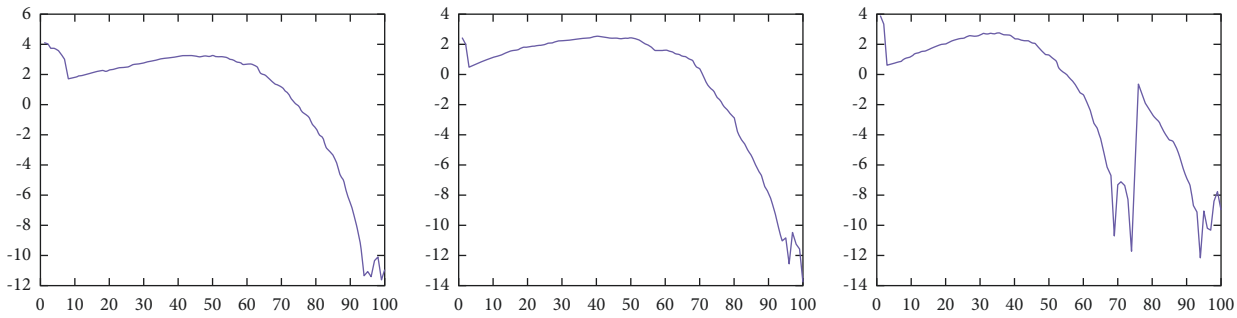


Figure 1: OMP simulation results of BIC Trends: X-axis for number of predictors included in the model and Y axis for BIC. (a): Example 1, (b): Example 2, and (c): Example 3.

reference. These graphs are not representable as a whole; however, they do provide trends of BIC casewise. Take BIC of Example 1 as an example. Please refer to Figure 1(a). BIC decreases as the number of variables included in the model increases. BIC reaches a local minimum when the number of variables included hits $d_0 = 8$. Thereafter, BIC increases as the number of variables included in the model increases until the number of variables gets near to the sample size. This is not surprising since BIC decreases as the model complexity increases. Similar trends can be observed for Example 2 and Example 3. Please refer to Figures 1(b) and 1(c).

One possible suggestion for OMP algorithm is that after the screening process with n candidate models, only $(n/2)$ BIC candidate models for minimum BIC values are compared. By doing so, computational time can be saved without loss of correctness of screening consistent property of OMP.

In conclusion, finite simulation performances in terms of screening selection consistency for OMP are good under all the three examples. Those performances support our theories proposed in Section 3.3.

5. Conclusion and Future Research

To conclude, this article shows the theoretical proof that OMP can identify all relevant predictors consistently, even if the predictor dimension is considerably larger than the sample size. In particular, if the dimension of the true model is finite, OMP might discover all relevant predictors within a finite number of steps. In other words, sure screening property can be guaranteed under the four technical conditions. Given the sure screening property, the recently proposed BIC of Chen and Chen (2008) can be used to practically select the best candidate from the models generated by the OMP algorithm. The resulting model is good in the sense that many existing variable selection methods, such as adaptive Lasso and SCAD, can be applied directly to increase the estimation accuracy. Compared with the traditional orthogonal matching pursuit method, the resulting model can improve prediction accuracy and reduce computational cost by screening out the relevant variables.

The abovementioned variable selection procedure only considers the fixed effect estimates in the linear models. However, in real life, a lot of existing data have both the fixed effects and random effects involved. For example, in the clinic trials, several observations are taken for a period of time for one particular patient. After collecting the data needed for all the patients, it is natural to consider random effects for each individual patient in the model setting since a common error term for all the observations is not sufficient to capture the individual randomness. Future research may include random effects in the model by imposing penalized hierarchical likelihood algorithm for accurate variable selection.

Appendix

Proof of lemmas and theorems

Proof of Lemma 1. Let $\mathbf{r} = (r_1, \ldots, r_2)^T \in R^d$ be an arbitrary d-dimensional vector and $\mathbf{r}_1$ be the subvector corresponding to Σ_{11}. By the condition, we know immediately,

$$\frac{1}{c} \leq \min_{\|\mathbf{r}_1\|=1} \inf \mathbf{r}_1^T \Sigma_{11} \mathbf{r}_1 \leq \max_{\|\mathbf{r}_1\|=1} \sup \mathbf{r}_1^T \Sigma_{11} \mathbf{r}_1 \leq c. \tag{A.1}$$

The argument is presented in the following. Suppose that the matrix Σ is positive definite and has the partition as given by $\begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$. Then the inverse of Σ has the following form:

$$\Sigma^{-1} = \begin{pmatrix} \Sigma_{11}^{-1} + \Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22,1}^{-1}\Sigma_{21}\Sigma_{11}^{-1} & -\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22,1}^{-1} \\ -\Sigma_{22,1}^{-1}\Sigma_{21}\Sigma_{11}^{-1} & \Sigma_{22,1}^{-1} \end{pmatrix}, \tag{A.2}$$

where $\Sigma_{22,1} = \Sigma_{22} - \Sigma_{21}\Sigma_{11}\Sigma_{12}$. In fact, the above formula can be derived from the following identity:

$$= \begin{pmatrix} \Sigma_{11} & 0 \\ 0 & \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12} \end{pmatrix}. \tag{A.3}$$

and the fact that

$$\begin{pmatrix} I & 0 \\ -\Sigma_{21}\Sigma_{11}^{-1} & I \end{pmatrix}^{-1} = \begin{pmatrix} I & 0 \\ \Sigma_{21}\Sigma_{11}^{-1} & I \end{pmatrix}. \tag{A.4}$$

Moreover, the largest singular value is referred to the operator norm of the linear operator (matrix) in a Hilbert space. If A is a $p \times n$ matrix of complex entries, then its singular values $S_1 \geq \ldots \geq S_q \geq 0, q = \min(p, n)$ are defined as the square roots of the q largest eigenvalues of the nonnegative definite Hermitian matrix AA^*. If A $(n \times n)$ is Hermitian, then let $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n$ denote its eigenvalues.

Stage 1. To prove $S_{\max}(\Sigma_{11}) < c$.

Let $T = (I \ 0)$, so $\Sigma_{11} = (I \ 0)\begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}\begin{pmatrix} I \\ 0 \end{pmatrix} = T\Sigma T'$. Then we have

$$\begin{aligned} & S_{\max}(\Sigma_{11}) \\ & \leq S_{\max}(T) S_{\max}(\Sigma) S_{\max}(T') \\ & = S_{\max}(\Sigma) \\ & < c, \end{aligned} \tag{A.5}$$

where $S_{\max}(T) = \sqrt{\lambda_{\max}(TT')}$. This part has been proven in page 334 of the book Spectral Analysis of Large Dimensional Random Matrices.

Stage 2. To prove $S_{\min}(\Sigma_{11}) > 1/c$.

Σ^{-1} is positive definite. So is the submatrix of Σ^{-1}, that is diagonal entries. Let $A = \Sigma_{11}^{-1}$, $B = \Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22,1}^{-1}\Sigma_{21}\Sigma_{11}^{-1}$, and $C = A + B$. We have that C is positive definite. We want to prove $A = \Sigma_{11}^{-1}$ is positive definite. We have B is also positive definite because for any vector x,

$$\begin{aligned} & x\prime Bx \\ & = \alpha' \Sigma_{22,1}^{-1} \alpha. \end{aligned} \tag{A.6}$$

Since $\Sigma_{22,1}^{-1}$ is positive definite. Therefore, for any vector β

$$\begin{aligned} & \beta' A\beta \\ & = S_{\max}(C) \leq S_{\max}\left(\Sigma^{-1}\right). \end{aligned} \tag{A.7}$$

So

$$S_{\max}\left(\Sigma_{11}^{-1}\right) \leq S_{\max}\left(\Sigma^{-1}\right) \Rightarrow S_{\min}(\Sigma_{11}) \geq S_{\min}(\Sigma) > \frac{1}{c}. \tag{A.8}$$

Now, the desired conclusion of Lemma 1 is implied by

$$P\left(\max_{\|\mathbf{r}_1\|=1} \sup \left| \mathbf{r}_1^T \left\{ \widehat{\Sigma}_{11} - \Sigma_{11} \right\} \mathbf{r}_1 \right| > \varepsilon \right) \longrightarrow 0, \tag{A.9}$$

where $\varepsilon > 0$ is an arbitrary positive number. The left-hand side of (A.9) is bounded by

$$\begin{aligned} & \left| \mathbf{r}_1^T \left\{ \widehat{\Sigma}_{11} - \Sigma_{11} \right\} \mathbf{r}_1 \right| \\ & \leq \max_{1 \leq j_1, j_2 \leq d} \left| \widehat{\sigma}_{j_1 j_2} - \sigma_{j_1 j_2} \right| \sum_{j_1, j_2 \in M} \left| r_{j_1} \right| \times \left| r_{j_2} \right| \\ & = \max_{1 \leq j_1, j_2 \leq d} \left| \widehat{\sigma}_{j_1 j_2} - \sigma_{j_1 j_2} \right| \left(\sum_{j \in M} \left| r_j \right| \right)^2 \\ & \leq |M| \max_{1 \leq j_1, j_2 \leq d} \left| \widehat{\sigma}_{j_1 j_2} - \sigma_{j_1 j_2} \right| \\ & = m \max_{1 \leq j_1, j_2 \leq d} \left| \widehat{\sigma}_{j_1 j_2} - \sigma_{j_1 j_2} \right|, \end{aligned} \tag{A.10}$$

where M is the set that contains m variables. Hence,

$$\begin{aligned} & \leq P\left(\max_{1 \leq j_1, j_2 \leq d} \left| \widehat{\sigma}_{j_1 j_2} - \sigma_{j_1 j_2} \right| > \frac{\varepsilon}{m} \right) \\ & \leq \sum_{1 \leq j_1, j_2 \leq d} P\left(\left| \widehat{\sigma}_{j_1 j_2} - \sigma_{j_1 j_2} \right| > \frac{\varepsilon}{m} \right) \\ & \leq d^2 C_1 \exp\left(-C_2 n \varepsilon^2 m^{-2} \right). \end{aligned} \tag{A.11}$$

By lemma A3 in Bickel and Elizaveta [28], there exists constants $C_1 > 0$ and $C_2 > 0$, such that $P(|\widehat{\sigma}_{j_1 j_2} - \sigma_{j_1 j_2}| > \varepsilon) \leq C_1 \exp(-C_2 n \varepsilon^2)$.

$$= C_1\left((m+3) \log d - C_2 n \varepsilon^2 m^{-2} \right). \tag{A.12}$$

By C4, we have $d \leq \nu n^{\xi}$. Therefore, $\leq C_1 \exp((m+3)\nu n^{\xi} - C_2 \varepsilon^2 n m^{-2})$

Since $m = O(n^{((1-\xi)/3)})$, $m + 3 \leq \nu_m n^{((1-\xi)/3)}$ for some constant $\nu_m \geq 0$.

$$= C_1 \exp\left(\nu \nu_m n^{(((1-\xi)/3)+\xi)} \left(1 - C_2 \varepsilon^2 \nu^{-1} \nu_n^{-1} n^{(1-3(((1-\xi)/3)-\xi))} \right) \right). \tag{A.13}$$

Proof of Theorem 1

Proof. For every $k \le [Kn^{\xi_0+4\xi_{\min}}]$, we have

$$\begin{aligned}
&\Omega(k)\\
&= \left\| H^{(k)}_{a_{k+1}} \left\{ I_n - H_{(S^{(k)})} \right\} Y \right\|^2\\
&= \frac{Y^T \left\{ I_n - H_{(S^{(k)})} \right\} X^{(k)}_{a_{k+1}} X^{(k)T}_{a_{k+1}} \left\{ I_n - H_{(S^{(k)})} \right\} Y}{\left\| X^{(k)}_{a_{k+1}} \right\|^2}\\
&= \frac{\left| X^{(k)T}_{a_{k+1}} \left\{ I_n - H_{(S^{(k)})} \right\} Y \right|^2}{\left\| X^{(k)}_{a_{k+1}} \right\|^2},
\end{aligned} \tag{A.14}$$

where $H^{(k)}_j = X^{(k)}_j X^{(k)T}_j \| X^{(k)}_j \|^{-2}$ and $X^{(k)}_j = \{I_n - H_{(S^{(k)})}\} X_j$. Note that a_{k+1} is selected to solve the maximization problem:

$$\max_{j \in F \setminus S^{(k)}} X_j^T \mathrm{RSS}\left(S^{(k)}\right) X_j = \max_{j \in F \setminus S^{(k)}} \left| X_j^T \left\{ I_n - H_{(S^{(k)})} \right\} Y \right|^2. \tag{A.15}$$

In addition,

$$\left\| X^{(k)}_{a_{k+1}} \right\|^2 \le \left\| X_{a_{k+1}} \right\|^2 \le \max_j \left\| X_j \right\|^2, \tag{A.16}$$

By Lemma 1, $\max_j \| X_j \|^2 / n \le \tau_{\max}$, with probability tending to 1. Therefore,

$$\left\| X^{(k)}_{a_{k+1}} \right\|^2 \le n\tau_{\max}, \tag{A.17}$$

with probability tending to 1.
Suppose that $T \subset S^{[Kn^{\xi_0+4\xi_{\min}}]}$.

$$\begin{aligned}
&\Omega(k)\\
&\ge \frac{\max_{j \in S^*_k} \left| X^{(k)T}_j \left\{ I_n - H_{(S^{(k)})} \right\} X_{(T)} \beta_{(T)} \right|^2 - \max_{j \in S^*_k} \left| X^{(k)T}_j \left\{ I_n - H_{(S^{(k)})} \right\} \varepsilon \right|^2}{n\tau_{\max}}\\
&\ge \frac{\max_{j \in S^*_k} \left| X^{(k)T}_j \left\{ I_n - H_{(S^{(k)})} \right\} X_{(T)} \beta_{(T)} \right|^2 - \max_{j \in T} \left| X^{(k)T}_j \left\{ I_n - H_{(S^{(k)})} \right\} \varepsilon \right|^2}{n\tau_{\max}},
\end{aligned} \tag{A.18}$$

where $S^*_k = T \setminus S^{(k)} \ne \varnothing$, $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_n)^T \in R^n$. In what follows the two terms involved in (A.18) will be carefully evaluated separately. □

Step 1. (The first term in (A.18)). For convenience, we define $Q_{(S^{(k)})} = I_n - H_{(S^{(k)})}$. Then,

$$\begin{aligned}
&= \max_{j \in S^*_k} \left| X^{(k)T}_j Q_{(S^{(k)})} X_{(T)} \beta_{(T)} \right|^2\\
&= \max_{j \in S^*_k} \left| X^T_j Q^T_{(S^{(k)})} Q_{(S^{(k)})} X_{(T)} \beta_{(T)} \right|^2\\
&= \max_{j \in S^*_k} \left| X^T_j Q_{(S^{(k)})} X_{(T)} \beta_{(T)} \right|^2\\
&= \max_{j \in S^*_k} \left| X^T_j Q_{(S^{(k)})} X_{(S^*_k)} \beta_{(S^*_k)} \right|^2.
\end{aligned} \tag{A.19}$$

Note that

$$\begin{aligned}
&= \beta^T_{(S^*_k)} X^T_{(S^*_k)} Q_{(S^{(k)})} X_{(S^*_k)} \beta_{(S^*_k)}\\
&= \sum_{j \in S^*_k} \beta_j X^T_j Q_{(S^{(k)})} X_{(S^*_k)} \beta_{(S^*_k)}\\
&\le \left(\sum_{j \in S^*_k} \beta_j^2 \right)^{(1/2)} \left(\sum_{j \in S^*_k} \left[X^T_j Q_{(S^{(k)})} X_{(S^*_k)} \beta_{(S^*_k)} \right]^2 \right)^{1/2}\\
&\le C_\beta \times \sqrt{d_0} \times \max_{j \in S^*_k} \left| X^T_j Q_{(S^{(k)})} X_{(S^*_k)} \beta_{(S^*_k)} \right|.
\end{aligned} \tag{A.20}$$

Applying (A.20) back to (A.19), we have

$$\max_{j \in S^*_k} \left| X^{(k)T}_j \left\{ I_n - H_{(S^{(k)})} \right\} X_{(T)} \beta_{(T)} \right|^2 \ge \frac{\left\| Q_{(S^{(k)})} X_{(S^*_k)} \beta_{(S^*_k)} \right\|^4}{d_0 C^2_\beta}. \tag{A.21}$$

Recall that $S^*_k = T \setminus S^{(k)} \ne \varnothing$. Then by conditions (C1)–(C3), and Lemma 2.1, we have

$$\begin{aligned}
&\left\|Q_{(S^{(k)})}X_{(S_k^*)}\beta_{(S_k^*)}\right\|^2 \\
&= \sum_{i,j\in S_k^*}\beta_i\beta_j X_i^T Q_{(S^{(k)})}X_j \\
&\geq \beta_{\min^2}\left\|Q_{(S^{(k)})}X_{(S_k^*)}\right\|^2 \\
&= \beta_{\min^2}\left\|X^{(k)}_{(S_k^*)}\right\|^2 \\
&= \beta_{\min^2}\left\|X_{(S_k^*)}\right\|^2 \\
&\geq n\tau_{\min}\beta_{\min^2}.
\end{aligned} \tag{A.22}$$

with probability tending to 1. Applying this result back to (A.21) and also noting the technical conditions (C3) and (C4), we have

$$\begin{aligned}
&\max_{j\in S_k^*}\left|X_j^{(k)T}\left\{I_n - H_{(S^{(k)})}\right\}X_{(T)}\beta_{(T)}\right|^2 \\
&\geq \frac{\nu_\beta^4 n^{-4\xi_{\min}} n^2\tau_{\min^2}}{\nu n^{\xi_0}C_\beta^2} \\
&= \nu_\beta^4\tau_{\min^2}\nu^{-1}C_\beta^{-2}n^{2-\xi_0-4\xi_{\min}}.
\end{aligned} \tag{A.23}$$

Step 2. (The second term in (A.18)).

$$\begin{aligned}
&\left|X_j^{(k)T}\left\{I_n - H_{(S^{(k)})}\right\}\varepsilon\right|^2 \\
&= \left|X_j^T\left\{I_n - H_{(S^{(k)})}\right\}\varepsilon\right|^2 \\
&= \left(X_j^T Q_{(S^{(k)})}\varepsilon\right)^2 \\
&\leq \max_{j\in T}\max_{|M|\leq m^*}\left(X_j^T Q_{(M)}\varepsilon\right)^2,
\end{aligned} \tag{A.24}$$

where $m^* = Kn^{\xi_0+4\xi_{\min}}$. Note that $X_j^T Q_{(M)}\varepsilon$ is a normal random variable with mean 0 and variance given by $\|Q_{(M)}X_j\|^2 \leq \|X_j\|^2$. Thus, the right-hand side of (A.24) can be bounded further by

$$\max_{j\in T}\left\|X_j\right\|^2 \max_{j\in T}\max_{|M|\leq m^*}\chi_1^2, \tag{A.25}$$

where χ_1^2 stands for a chi-square random variable with one degree of freedom. By conditions (C1) and (C2) and Lemma 1, we know that $n^{-1}\max_{j\in T}\|X_j\|^2 \leq \tau_{\max}$ with probability tending to 1. On the other hand, the total number of combinations for $j \in T$ and $|M| \leq m^*$ is no more than d^{m^*+2}. Then we can proceed by using Bonferroni's inequality to obtain

$$\begin{aligned}
&\leq \sum_{j\in T}\sum_{|M|\leq m^*} P\left(\chi_1^2 \geq 2(m^*+2)\log(d)\right) \\
&= d^{m^*+2}P\left(\chi_1^2 \geq 2(m^*+2)\log(d)\right) \\
&= (2\pi)^{(-1/2)}d^{m^*+2}\int_{2(m^*+2)\log(d)}^{\infty} x^{(-1/2)}e^{(-x/2)}\mathrm{d}x \\
&\leq \frac{(2\pi)^{(-1/2)}d^{m^*+2}}{\sqrt{2(m^*+2)\log(d)}}\int_{2(m^*+2)\log(d)}^{\infty} e^{(-x/2)}\mathrm{d}x \\
&= \frac{(2\pi)^{(-1/2)}d^{m^*+2}}{\sqrt{2(m^*+2)\log(d)}}2d^{-(m^*+2)} \\
&= \frac{1}{\sqrt{\pi(m^*+2)\log(d)}} \\
&\longrightarrow 0,
\end{aligned} \tag{A.26}$$

which implies that $\max_{j\in T}\max_{|M|\leq m^*}\chi_1^2 \leq 2(m^*+2)\log(d)$ with probability tending to 1 as $d \longrightarrow \infty$. Then in conjunction with (C4), we have

$$\begin{aligned}
&\max_{j\in T}\left\|X_j\right\|^2 \max_{j\in T}\max_{|M|\leq m^*}\chi_1^2 \\
&\leq n\tau_{\max}\times 3Kn^{\xi_0+4\xi_{\min}}\times\nu n^{\xi} \\
&= n\tau_{\max}\times 3K\nu n^{\xi+\xi_0+4\xi_{\min}},
\end{aligned} \tag{A.27}$$

with probability tending to 1. Combining (A.23) and (A.27) and putting them back to (A.18), we can show

$$\begin{aligned}
&n^{-1}\Omega(k) \\
&= \tau_{\max^{-1}}\tau_{\min^2}C_\beta^{-2}\nu_\beta^4\nu^{-1}n^{-\xi_0-4\xi_{\min}} \\
&\quad\times\left\{1 - 3K\nu^2\tau_{\max}\tau_{\min^{-2}}C_\beta^2\nu_\beta^{-4}n^{\xi+2\xi_0+8\xi_{\min}-2}\right\} \\
&= \frac{2}{Kn^{\xi_0+4\xi_{\min}}} \\
&\quad\times\left\{1 - 3K\nu^2\tau_{\max}\tau_{\min^{-2}}C_\beta^2\nu_\beta^{-4}n^{\xi+2\xi_0+8\xi_{\min}-2}\right\},
\end{aligned} \tag{A.28}$$

uniformly for every $k \leq Kn^{\xi_0+4\xi_{\min}}$. Recall that $K = 2\tau_{\max}\tau_{\min^{-2}}C_\beta^2\nu_\beta^{-4}\nu$. Under condition (C4), we have

$$\begin{aligned}
&n^{-1}\|Y\|^2 \\
&\geq 2\times\left\{1 - 3K\nu^2\tau_{\max}\tau_{\min^{-2}}C_\beta^2\nu_\beta^{-4}n^{\xi+2\xi_0+8\xi_{\min}-2}\right\} \\
&\longrightarrow 2.
\end{aligned} \tag{A.29}$$

In contrast, under the assumption $\mathrm{Var}(Y_i) = 1$, we have $n^{-1}\|Y\|^2 \longrightarrow_p 1$. This contradicts with the result of (A.29). Hence, it implies that it is impossible to have $S^{(k)} \cap T = \varnothing$ for every $1 \le k \le Kn^{\xi_0+4\xi_{\min}}$. Consequently, with probability tending to 1, all relevant predictors should be recovered within a total of $Kn^{\xi_0+4\xi_{\min}}$ steps. This completes the proof.

Proof of Theorem 2

Proof. It suffices to show that

$$P\left(\min_{1\le k<k_{\min}}\left\{\mathrm{BIC}\left(S^{(k)}\right)-\mathrm{BIC}\left(S^{(k+1)}\right)\right\}>0\right)\longrightarrow 1. \tag{A.30}$$

Note that

$$\begin{aligned}&\mathrm{BIC}\left(S^{(k)}\right)-\mathrm{BIC}\left(S^{(k+1)}\right)\\&\quad-n^{-1}(\log n+2\log d)\\&\ge\log\left(\frac{\widehat{\sigma}^2_{\left(S^{(k)}\right)}}{\widehat{\sigma}^2_{\left(S^{(k+1)}\right)}}\right)-3n^{-1}\log d\\&=\log\left(1+\frac{\widehat{\sigma}^2_{\left(S^{(k)}\right)}-\widehat{\sigma}^2_{\left(S^{(k+1)}\right)}}{\widehat{\sigma}^2_{\left(S^{(k+1)}\right)}}\right)\\&\quad-3n^{-1}\log d,\end{aligned} \tag{A.31}$$

under the assumption $d > n$. By definition, we have $\widehat{\sigma}^2_{\left(S^{(k+1)}\right)} \le n^{-1}\|Y\|^2 \longrightarrow_p 1$. Then, with probability tending to 1, the right-hand side of (A.14) is

$$\begin{aligned}&=\log\left[1+\frac{\mathrm{RSS}\left(S^{(k)}\right)-\mathrm{RSS}\left(S^{(k+1)}\right)}{2n}\right]-3n^{-1}\log d\\&=\log\left[1+\frac{\Omega(k)}{2n}\right]-3n^{-1}\log d,\end{aligned} \tag{A.32}$$

by the definition of $\Omega(k) = \mathrm{RSS}(S^{(k)}) - \mathrm{RSS}(S^{(k+1)})$. In addition, we use the fact that $\log(1+x) \ge \min\{\log 2, 0.5x\}$ for any $x > 0$. Then the right-hand side of (A.32) is

$$\ge\min\left\{\log 2, 4^{-1}K^{-1}n^{-\xi_0-4\xi_{\min}}\right\}-3n^{-1}\log d, \tag{A.33}$$

according to (A.29). Moreover, the right-hand side of (A.33) is independent of k, hence, it is a uniform lower bound for $\mathrm{BIC}(S^{(k)}) - \mathrm{BIC}(S^{(k+1)})$ with $1 \le k < k_{\min}$. Thus, it suffices to show that the right-hand side of (A.33) is positive with probability tending to 1. To this end, we first note that

$$\log 2-3n^{-1}\log d\longrightarrow 0, \tag{A.34}$$

under condition (C4). Therefore, we can look at

$$\begin{aligned}&4^{-1}K^{-1}n^{-\xi_0-4\xi_{\min}}-3n^{-1}\log d\\&=4^{-1}K^{-1}n^{-\xi_0-4\xi_{\min}}\\&\quad\cdot\left(1-12\nu\mathrm{Kn}^{\xi+\xi_0+4\xi_{\min}-1}\right)\\&\ge 0,\end{aligned} \tag{A.35}$$

with probability tending to 1 under condition (C4). This completes the proof. □

Data Availability

The data used to support the findings of the study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

A preprint has previously been published as a part of the Ph.D. thesis of the first author [27]: https://scholarbank.nus.edu.sg/bitstream/10635/43427/1/thesis.pdf. This work was funded by the Shanghai Planning Project of Philosophy and Social Science (Grant no. 2021EGL004). The authors gratefully acknowledge the Shanghai Planning Office of Philosophy and Social Science for the technical and financial support.

References

[1] Y. Xie, Y. Li, Z. Xia, and R. Yan, "An improved forward regression variable selection algorithm for high-dimensional linear regression models," *IEEE Access*, vol. 8, Article ID 129032, 2020.

[2] H. Akaike, "Information theory and an extension of the maximum likelihood principle," in *Proceedings of the Second International Symposium of Information Theory*, pp. 199–213, Springer, Budapest, Hungary, 1973.

[3] J. Chen and Z. Chen, "Extended Bayesian information criteria for model selection with large model spaces," *Biometrika*, vol. 95, no. 3, pp. 759–771, 2008.

[4] R. Tibshirani, "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society: Series B*, vol. 58, no. 1, pp. 267–288, 1996.

[5] J. Fan and R. Li, "Variable selection via nonconcave penalized likelihood and its oracle properties," *Journal of the American Statistical Association*, vol. 96, no. 456, pp. 1348–1360, 2001.

[6] H. Wang, "Forward regression for ultra-high dimensional variable screening," *Journal of the American Statistical Association*, vol. 104, no. 488, pp. 1512–1524, 2009.

[7] J. Fan and J. Lv, "Sure independence screening for ultra-high dimensional feature space," *Journal of the Royal Statistical Society: Series B*, vol. 70, no. 5, pp. 849–911, 2008.

[8] H. Zou, "The adaptive lasso and its oracle properties," *Journal of the American Statistical Association*, vol. 101, no. 476, pp. 1418–1429, 2006.

[9] H. Zou and R. Li, "One-step sparse estimates in nonconcave penalized likelihood models," *Annals of Statistics*, vol. 36, no. 4, pp. 1509–1533, 2008.

[10] T. T. Cai and L. Wang, "Orthogonal matching pursuit for sparse signal recovery with noise," *IEEE Transactions on Information Theory*, vol. 57, no. 7, pp. 4680–4688, 2011.

[11] A. R. Barron, A. Cohen, W. Dahmen, and R. DeVore, "Approximation and learning by greedy algorithms," *Annals of Statistics*, vol. 36, no. 1, pp. 64–94, 2008.

[12] S. Kim and J. M. Kim, "Two-stage classification with sis using a new filter ranking method in high throughput data," *Mathematics*, vol. 7, no. 6, p. 493, 2019.

[13] D. Donoho and V. Stodden, "Breakdown point of model selection when the number of variables exceeds the number of observations, the 2006," in *Proceedings of the IEEE International Joint Conference on Neural Network Proceedings IEEE*, Vancouver, BC, Canada, July 2006.

[14] C. H. Zhang and J. Huang, "The sparsity and bias of the lasso selection in high-dimensional linear regression," *Annals of Statistics*, vol. 36, no. 4, pp. 1567–1594, 2008.

[15] Z. Chen and J. Chen, "Tournament screening cum ebic for feature selection with high-dimensional feature spaces," *Science in China(Series A:Mathematics)*, vol. 52, no. 6, pp. 1327–1341, 2009.

[16] S. Luo and Z. Chen, "Sequential lasso cum EBIC for feature selection with ultra-high dimensional feature space," *Journal of the American Statistical Association*, vol. 109, no. 507, pp. 1229–1240, 2014.

[17] X. He, L. Wang, and H. G. Hong, "Quantile-adaptive model-free variable screening for high-dimensional heterogeneous data," *Annals of Statistics*, vol. 41, no. 1, pp. 342–369, 2013.

[18] H. G. Hong, L. Wang, and X. He, "A data-driven approach to conditional screening of high-dimensional variables," *Stat*, vol. 5, no. 1, pp. 200–212, 2016.

[19] M. Baranowski, A. Belloum, M. Bubak, and M. Malawski, "Constructing workflows from script applications," *Scientific Programming*, vol. 20, no. 4, pp. 359–377, 2012.

[20] S. K. Samudrala, J. Zola, S. Aluru, and B. Ganapathysubramanian, "Parallel framework for dimensionality reduction of large-scale datasets," *Scientific Programming*, vol. 2015, Article ID 180214, 12 pages, 2015.

[21] L. P. Chen, "Feature screening based on distance correlation for ultrahigh-dimensional censored data with covariate measurement error," *Computational Statistics*, vol. 36, no. 2, pp. 857–884, 2021.

[22] S. Choudalakis, M. Mitrouli, A. Polychronou, and P. Roupa, "Solving high-dimensional problems in statistical modelling: a comparative study," *Mathematics*, vol. 9, no. 15, pp. 1806–1816, 2021.

[23] X. Xu, Z. Lin, X. Li, C. Shang, and Q. Shen, "Multi-objective robust optimisation model for mdvrpls in refined oil distribution," *International Journal of Production Research*, pp. 1–21, 2021.

[24] X. Xu, J. Hao, and Y. Zheng, "Multi-objective artificial bee colony algorithm for multi-stage resource leveling problem in sharing logistics network," *Computers & Industrial Engineering*, vol. 142, no. 2, Article ID 106338, 2020.

[25] X. Xu, Z. Lin, and J. Zhu, "DVRP with limited supply and variable neighborhood region in refined oil distribution," *Annals of Operations Research*, vol. 309, no. 2, pp. 663–687, 2022.

[26] Y. Ren, J. Qian, Y. Dong, Y. Xin, and H. Chen, "Avbh: asymmetric learning to hash with variable bit encoding," *Scientific Programming*, vol. 2020, Article ID 2424381, 11 pages, 2020.

[27] Y. Xie, *Variable Selection Procedures in Linear Regression Models*, Ph.D. dissertation, Stats Dept., National University of Singapore, Singapore, 2013.

[28] P. J. Bickel and E. Levina, "Covariance regularization by thresholding," *Annals of Statistics*, vol. 36, no. 6, pp. 2577–2604, 2008.

Hindawi
Scientific Programming
Volume 2022, Article ID 1549899, 13 pages
https://doi.org/10.1155/2022/1549899

Research Article

User-Characteristic-Oriented Bilateral Matching between Online Learning Service Demanders and Providers

Qirui Wu,[1] **Wenbo Zhang**,[2] **Hong Yong**,[1] **and Xi Chen**[2]

[1]*School of Foreign Languages, Xidian University, Xi'an 710126, China*
[2]*School of Economics & Management, Xidian University, Xi'an 710126, China*

Correspondence should be addressed to Qirui Wu; qrwu@xidian.edu.cn

Received 2 June 2022; Accepted 1 July 2022; Published 20 July 2022

Academic Editor: Xiaofeng Xu

Under the background of information technology and the Internet era, the matching problem in different application scenarios is becoming increasingly prominent. With respect to the matching problem in knowledge services, enabling users to choose suitable knowledge out of massive information has become an urgent demand to be satisfied. Initiating from interdisciplinary perspective, this paper proposes a matching method for online learning services according to user characteristics, which focuses on the matching of decision making for knowledge service with user relevance and characteristic under social network environment. Firstly, the complex network among users is constructed, and the user group is subcategorized into subgroups, thereby aggregating the subgroup information effectively. Secondly, the weight of the indices that evaluate the matching subject is determined by conducting the best-worst method. Thirdly, considering the difference between the expectation and actual levels of the matching subject, the cumulative prospect theory is adopted to calculate the satisfaction degree of both sides. Aiming at maximizing the satisfaction degree of the subjects, a multi-objective optimization model is established to obtain the optimal matching pairs. Finally, the validity and rationality of the proposed method are verified, offering interdisciplinary perspective and theoretical foundation for knowledge service matching and the education reform of humanities.

1. Introduction

As the Internet and information industry continue to prosper, the exponentially rapid growth in online educational sources and services undoubtedly results in mixed qualities, to which the first and foremost thing for people to finish before self-teaching is to carefully discriminate between the expected online services and the confusing lousy commercials [1, 2]. The former can further be personalized for end-users according to their different requirements, whereas the latter may bring them nothing but a waste of money and energy, not to mention the underlying unpleasant and disappointing experiences induced. As for users who crave online educational services, they are forced to bear the bulk of the deleterious junk information reluctantly, which severely disturbs their initial pursuit of educational knowledge and learning skills [3–6]. Therefore, according to users' practical needs, it is of critical necessity to conduct in-depth analysis on their intrinsic characteristics such as learning interests and preferences in advance before the optimal tailor-made online educational services can be matched and recommended for them [7].

In this study, we therefore perform in-depth investigation on end-users' preferences from the perspective of social network analysis (SNA) and decision-making analysis. Considering the expectations of users and providers of online educational services, we propose a bilateral matching method for these two sides. Firstly, we establish and analyze the framework of our proposed matching method. Secondly, considering the relevance among users, we subcategorize users into different subgroups according to the results of SNA, upon which we adopt cumulative prospect theory to calculate the satisfaction degree between users and online educational services. Thirdly, we establish a matching model of knowledge services, generating the optimal matching pairs for users and services.

The key contributions of this study can be formulated as follows:

(1) From the perspective of SNA, we first mine the relevance between users, and then, we conduct GN algorithm to detect the community of user groups and subcategorize them into multiple user subgroups to realize the optimal matching between user groups and online educational services, thereby improving the matching efficiency between two sides.

(2) We introduce decision analysis method into the scenario of supply-demand matching for online educational services. Considering the differences in their expectations and actual levels, which exist between users and online services, we introduce cumulative prospect theory to characterize the satisfaction degrees that exist between users and online educational services, establishing a tailor-made bilateral matching model between these two sides. Our proposal not only helps users determine the strategy of how to make the best choice among diverse online educational services but also matches the most appropriate users for online service providers.

(3) Initiating from interdisciplinary perspective, we establish a set of online educational service matching methods according to users' preferences and apply them to facilitating the optimal selection of online English learning services in colleges and universities, which not only solves the problem of supply-demand matching in educational knowledge services but also offers innovative approaches and theoretical foundation for China's education reforms of humanistic disciplines.

The remainder of this paper is organized as follows. Section 2 outlines the related work of knowledge matching and decision-making problems. Section 3 highlights the corresponding problem description of our study. Section 4 formulates the methodologies adopted in this study, involving social network analysis, decision matrix aggregation, and multi-objective model construction. Section 5 elaborates a real case study that is engaged to validate the rationality of our proposed matching model. Section 6 discusses our future research focus and summarizes conclusions of this paper.

2. Related Work

During the past several decades, the matching problems that occur in the field of knowledge service have attracted much attention to foreign and domestic, which can be basically categorized into two aspects: the matching among online teaching/learning services and that among diverse knowledge services. One of the most important educational and knowledge services is to determine the appropriate knowledge matching pair (MP) between end-users and online services, which can be regarded as a representative and typical bilateral matching problem. The bilateral matching was firstly proposed by Gale and Shapley as a classic decision-making method [8] in response to conducting marriage matching and solving problems concerning higher education recruitment, and was further applied in various applications and fields [9–16].

Concerning online learning services, Lee discussed the quality of online learning services in South Korea and the United States in 2010. Through conducting empirical study adopting factor analysis, structural equation, and other models, the results suggested that online learning services are useful to students in different regions and countries worldwide, and the awareness of the quality of online learning services is an important factor affecting students' online acceptance and satisfaction degree [17]. In 2019, He et al. argued that online learning services are closely correlated with students' performances. After discussing and analyzing the influence of online learning services on students' satisfaction degree and learning results, He et al. reported that the use and frequency of online learning services depend heavily on students' participation paradigm and involvement [18]. Later, Zhao et al. provided a new set of teaching evaluation methods for online learning services and used the Internet big data to obtain learning process data from online learning service platforms, thereby deeply mining learners' characteristics and preferences. The study not only established learners' evaluation system and methods but also further provided students with intelligent and personalized services [19]. In 2021, Jung et al. explored online distance learning under the influence of COVID-19 and adopted AMOS21.0 model to conduct empirical study. The results revealed that online learning services and the quality of corresponding information and systems have positive impact on students as well as on their satisfaction degrees [20].

Regarding knowledge matching problems, Chen et al. proposed a knowledge supply and demand matching model in 2016, which aimed to maximize the satisfaction degree between demanders and suppliers [21]. Compared with our proposed method that characterizes with diverse user attributes and parameters, this approach also conducted the matching between knowledge demander and supplier; however, it adopted fuzzy linguistic information from the principle of axiomatic design.

In 2017, Li et al. proposed a decision-making method for matching experts and demand groups under the background of fuzzy language environment [22]. This approach targeted on optimizing the matching between expert groups and demand groups to elevate the efficiency of tacit knowledge sharing. Moreover, the approach further expanded the research work of tacit knowledge sharing methodologies and was widely applied in other fields where matching decision-making problems predominate. In contrast to our proposed method, this approach focused on tacit knowledge and emphasized the significance of experts, during which process the fuzzy linguistic environment was considered.

Considering network effect, Chang et al. developed a matching method for digital platforms [23] in 2019, which adopted fuzzy multi-attribute decision-making model to calculate the value of the network with respect to the platform, which solved the matching problem between

knowledge providers and demanders on digital platforms. This approach conducted knowledge matching with respect to three major dimensions, which focused on the operating rules of digital platforms rather than the platform service model. By contrast, our proposed method does not depend on platforms and/or their operating rules, by which the robustness and the flexibility of our proposal are therefore highlighted.

Later, in view of the matching for online learning in service platforms, in 2020, Pan et al. proposed an algorithm for online situational learning and resource allocation based on perishable resources [24]. Instead of making decisions between supply and demand, this approach was motivated by one healthcare application and can be easily extended to other service applications. Specifically, Pan et al. incorporated contextual learning with online healthcare service platforms, in which customers' arrival satisfies nonstationary Poisson process and the corresponding rewards can be predicted for subsequence resource allocation. Compared with our proposed method, the rationale of this approach can be understood as "random" and "iterative" rather than "stationary," while our proposal is more "dynamic" and "comprehensive" without requiring any iteration epochs.

In 2021, Zhang et al. considered the self-confidence fuzzy preference relevance of decision makers, proposing a bilateral matching method [25] based on FPRs-SC. This method was further applied to studies of supply-demand matching of knowledge services. In his research, Zhang et al. firstly proposed an extended logarithmic least squares method (LLSM) to derive a priority weight vector, upon which the bilateral matching decision-making approach was proven to be effective for the matching between knowledge suppliers and demanders. It is worth noting that this approach adopted an unbiased weight attribute to conduct the optimal matching decision making, during which the improved algorithms were validated by practical examples, providing enlightening and constructive theoretical foundation for our study. Based on this approach, we further apply more specific matching decision making with respect to other sort of demanders and suppliers from an interdisciplinary perspective.

In addition, Gao et al. proposed an entropy beta method in 2021, which was based on the dynamic matching process of user knowledge to achieve the optimal matching between users and knowledge services [26]. Differentiating from our proposal that selects various user preferences as multi-attributes, Gao et al. introduced the expert knowledge recommendation system (EKRS), which were further improved by the entropy beta algorithm in terms of accuracy, efficiency, dynamic regulation, etc. The principle of this approach can be translated as to filter out and obtain the optimal matching probability through the established algorithm and matching model, thereby proving the effectiveness and validity of both the algorithm and the model dynamically. The above two cases have great thought-provoking significance on our works.

3. Problem Description

In order to solve the problem of knowledge service matching in the context of the Internet, this study introduces the bilateral matching theory into the effective matching between users and online learning services, in which the users are regarded as demanders, while the online services are providers. Considering the preferences of users, we select appropriate online services for users and vice versa, thus realizing the optimal matching between supply and demand.

Online learning service matching is a matching problem involving multiple subjects under the Internet background. When users choose suitable online learning services on their own discretion, they usually seek for advices from people around them and make choices according to their personalized needs. Online learning services offer knowledge services to users who need them, throughout the whole process the service providers can also choose users. Therefore, concerning the problem of knowledge service matching between users (demanders) and online learning services (providers), users' preferences must not be neglected, and the requirements of online learning services should be further considered. Therefore, in this section, we investigate how to generate an optimal matching for users and online learning services according to their expectations and needs for each other.

The following lists out corresponding descriptions and definitions of the sets and variables involved in the decision-making problem of online educational service matching.

$D = \{d_i | i = 1, 2, \ldots, m\}$: a nonempty set with m users.

$S = \{s_j | j = 1, 2, \ldots, n\}$: a nonempty set with n language learning services.

$C = \{c_p | p = 1, 2, \ldots, l\}$: a nonempty set with l properties evaluating online learning services.

$W = \{w_1, w_2, \ldots, w_l\}$: a set of attributes weight of online learning services properties.

$A = \{a_q | q = 1, 2, \ldots, k\}$: a nonempty set with k properties of evaluating users.

$V = \{v_1, v_2, \ldots, v_k\}$: a set of attributes weight of user properties.

$B = [b_{ii'}]_{m \times m}$: the associative adjacency matrices among m users, where $b_{ii'} = 1$ denotes the association existing between the i-th and i'-th user, and $b_{ii'} = 0$ denotes no association between them.

$E = [e_{ip}]_{m \times l}$: it denotes the decision matrix of users' expectation level of online learning service, where $e_{ip} \in U$.

$R = [r_{jq}]_{n \times k}$: it denotes the decision matrix of online language learning service regarding users' expectation level, where $r_{jq} \in U$.

$E' = [e'_{jp}]_{n \times l}$: it denotes the decision matrix of the actual level of online language learning services, where $e'_{jp} \in U$.

$R' = [r_{iq}']_{m\times k}$: it denotes the decision matrix of user's actual level, where $r_{iq}' \in U$.

In this study, the research objective can be elaborated as the problem of matching decision making that exists between online learning services and users, in which users' learning interest and the relevance among them are considered. Specifically, we first subdivide the users into subgroups based on the relevance among users, after which the expectations of users within the subgroups are aggregated. Then, through combining the decision matrices of the expectation levels of users and of online learning services with those of the actual levels, we determine the optimal matching pairs for users and online learning services.

4. Methodology

In this study, we propose a matching method of decision making for online learning services, which considers users' learning preferences and the relevance among them. Firstly, according to the relevance among users who choose online learning services, we adopt social network analysis [27] (SNA) to subcategorize them into different subgroups. Secondly, based on multiple attributes of the users, the decision matrices of their expectation level and those of the actual level of the user subgroups are aggregated. Furthermore, considering the difference between the expectation and the actual levels of the user subgroups and the online learning services, we calculate the satisfaction degree matrices of user subgroups and those of online learning services adopting cumulative prospect theory. Finally, we establish a multi-objective optimization model to obtain the optimal matching pairs between users and online services, to which the specific details of our proposal are formulated as follows.

4.1. Community Detection-Based Algorithm for Division and Aggregation of User Subgroup. Considering the actual online learning services, we must accurately acquire diverse users' learning demands, preferences, and interests. Moreover, we need precisely obtain on whom the major and key groups of different online learning services target before achieving the optimal pairwise matching. Unfortunately, in real world, the enormous number of users and different user demands are posing disadvantageous obstacles and great challenges to the matching. Since users are intercorrelated with each other rather than isolated individuals in real world, they usually share the same preferences for choosing the same one or several services. Therefore, by referring to the principle of complex network theory, in this study, we adopt the notion of undirected weighted graph in network to demonstrate the incidence matrix among users, after which we subcategorize the user group into different subgroups and aggregate the preference information of subgroups, thereby realizing the optimal bilateral matching between users and services. The specific steps of user subgroup division and aggregation through using community detection algorithm are presented as follows.

Step 1. Divide user subgroups.

Firstly, in this study, we adopt topological graph $G = (D, H)$ in complex network to represent the association network among users, where G denotes undirected network connection graph, and $D = \{d_1, d_2, \ldots, d_m\}$ denotes a collection of nodes within the network. The users within the network are regarded as the nodes of the network, and $H = \{h_1, h_2, \ldots, h_o\}$ denotes the edges of the network, where o is the number of edges being expressed as $o = m(m-1)/2$. The weight of the edges can be expressed by the following adjacency matrix B.

$$B = \begin{bmatrix} b_{11} & b_{12} & \ldots & b_{1m} \\ b_{21} & b_{22} & \ldots & b_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ b_{m1} & b_{m2} & \ldots & b_{mm} \end{bmatrix}, \tag{1}$$

where $b_{ii'}$ is the edge weight between i-th node and i'-th node (in this study, $b_{ii'}$ refers to the correlation degree between i-th user and i'-th user). In order to perform effective subgroup division with respect to users group, based on the weighted adjacency matrix among user groups, this study adopts the GN algorithm [28] used in community detection algorithm to divide user groups into subgroups, and the divided subgroups can be expressed by

$$\mathfrak{R} = \{\mathfrak{R}_1, \mathfrak{R}_2, \ldots, \mathfrak{R}_Y\}. \tag{2}$$

Step 2. Calculate user weights.

After subcategorizing the user group into user subgroups, we determine the weight of each individual user included in the subgroups, thereby aggregating the expectation levels of online learning services of users in the same user subgroup. Within the same subgroup, the user weight can be determined by

$$\overline{\xi}(d_i) = \sum_{i'}^{\#\mathfrak{R}_y} b_{ii'}, \tag{3}$$

where $d_i, d_{i'} \in \mathfrak{R}_y$, $\mathfrak{R}_y$ denotes the y-th user subgroup and $y = 1, 2, \cdots, o$, $\#\mathfrak{R}_y$ denotes the number of users included in the y-th user subgroup. The weight of each individual user within the subgroup can be obtained by normalizing $\overline{\xi}(d_i)$, which is expressed as

$$\xi(d_i) = \frac{\overline{\xi}(d_i)}{\sum_{d_i \in \mathfrak{R}_y} \overline{\xi}(d_i)}, \tag{4}$$

where $\xi(d_i)$ denotes the weight of user d_i in the y-th subgroup.

Step 3. Aggregate the decision matrices of user subgroups.

In order to aggregate users' expectation levels concerning online learning services within the same user subgroup, this study uses network density operator to aggregate

the user's decision matrix, to which the specific process is expressed by

$$NDWA\left(d_i \in \Re_y\right) = \sum_{d_i \in \Re_y} \xi(d_i) \Lambda D_y. \tag{5}$$

In the above equation (5), $NDWA$ represents the operator of network density-weighted averaging (NDWA) [29], Λ represents the information aggregation operator, And D_y is the clustering of the y-th user subgroup.

4.2. Calculation of Attribute Weights Based on Best-Worst Method. According to the aforementioned user subgroup division and decision matrix aggregation, this study adopts best-worst method (BWM) [30, 31] and cumulative prospect theory to calculate the satisfaction degree between user subgroups and online learning services, thereby measuring the matching degree between the expectation level and actual level of users and online learning services. Firstly, the BWM is performed to calculate the weights of corresponding indicators, to which the specific procedures are presented as follows.

Step 4. Determine the set of evaluation indicators.

Taking as an example the attribute set of online learning services, the evaluation indicator set to be accessed can be expressed as

$$C = \{c_1, c_2, \ldots, c_l\}. \tag{6}$$

Step 5. Determine the most and least important indicators in the evaluation indicator set.

The most important indicator and the least important indicator within the evaluation indicator set are represented by c_B and c_W, respectively.

Step 6. Construct relative preference vector.

We use the scale ranging from 1 to 9 to construct the vector of the most important indicator U_B relative to other indicators (also known as the best-to-others vector), as well as the vector of other indicators U_W relative to the least important indicator (also known as the others-to-worst vector), which are expressed as

$$U_B = (u_{B1}, u_{B2}, \ldots, u_{Bl}). \tag{7}$$

$$U_W = (u_{1W}, u_{2W}, \ldots, u_{lW}). \tag{8}$$

In the above equation (7), each element in U_B denotes the relative preference information of the most important indicator relative to other indicators, where $u_{BB} = 1$. In the above equation (8), each element in U_W denotes the relative preference information of other indicators relative to the least important ones, where $u_{WW} = 1$.

Step 7. Solve the optimal weight.

The optimal weight of the indicator can be solved by using the following linear programming model.

$$\begin{aligned} &\min \max_p \left\{ \left| \frac{w_B}{w_p} - u_{Bp} \right|, \left| \frac{w_p}{w_W} - u_{pW} \right| \right\}, \\ &s.t. \sum_{p=1}^{l} w_p = 1, \\ &w_p \geq 0, p = 1, 2, \ldots, l. \end{aligned} \tag{9}$$

For ease of calculation, the above model is transformed into the following linear model for solution

$$\begin{aligned} &\min \varsigma, \\ &s.t. \left| w_B - u_{Bp} w_p \right| \leq \varsigma, \\ &\left| w_p - u_{pW} w_W \right| \leq \varsigma, \\ &\sum_{p=1}^{l} w_p = 1, \\ &w_p \geq 0, p = 1, 2, \ldots, l. \end{aligned} \tag{10}$$

By using lingo software to solve the above model, we obtain the weight W and the optimal target value ς^* of each attribute l of online learning services.

Step 8. Verify the consistency.

In order to further verify the consistency and effectiveness of the evaluation results, the consistency ratio can be calculated using the following equation:

$$CR = \frac{\varsigma^*}{CI}, \tag{11}$$

where CR is the consistency ratio, and CI is the consistency index whose specific values are listed out in the following Table 1. When $CR < 0.1$, it can be considered that the evaluation results satisfy the consistency requirements.

Performing the same method, the weight V of attributes of users can be obtained.

4.3. Prospect Theory-Based Calculation of Satisfaction Degree. During the actual process of matching decision making, the individual preferences of both sides to be matched may impose certain impact on decision-making results. For instance, the difference between the user's expectation level of online learning services and the actual level of the services themselves may affect the users' choice. Therefore, from the perspective of cumulative prospect theory proposed by Kahneman and Tversky [32], this study fully considers the expectation level and actual level of users and online learning services, thereby matching the satisfaction degree of both sides. The detailed steps of calculating satisfaction degree based on cumulative prospect theory are presented as follows.

Table 1: Consistency indices.

a_{BW}	1	2	3	4	5	6	7	8	9
CI	0.00	0.44	1.00	1.63	2.30	3.00	3.73	4.47	5.23

Step 9. Determine the set of evaluation indicators.

Taking as an example the assessment of the online learning services with a set of l attributes, the set of evaluation indices to be assessed is represented by

$$C = \{c_1, c_2, \ldots, c_l\}. \tag{12}$$

Step 10. Determine the reference point.

Setting as the reference point the expectation level of online learning service of the user subgroups, the reference point regarding user subgroup $\mathfrak{R}_y$ can be expressed as

$$E_y = \begin{bmatrix} e_{y1} & e_{y2} & \cdots & e_{yl} \end{bmatrix}. \tag{13}$$

Step 11. Determine the benefit.

Concerning different types of indicators, the gains between the actual level of online learning services and the expectation level of users are calculated by

$$G_{yj}^{p} = \begin{cases} e_{jp}' - e_{yp}c_p \text{ is the cost attribute,} \\ e_{yp} - e_{jp}'c_p \text{ is the benefit attribute,} \end{cases} \tag{14}$$

where G_{yj}^{p} denotes the gain value of the p-th index of the j-th online learning services of the user subgroup $\mathfrak{R}_y$, $G_{yj}^{p} > 0$ denotes the value of gains, and $G_{yj}^{p} < 0$ denotes the value of losses.

Step 12. Determine the value function.

The value function is commonly expressed as

$$v\left(G_{yj}^{p}\right) = \begin{cases} \left(G_{yj}^{p}\right)^{\alpha} G_{yj}^{p} \geq 0 \\ -\lambda\left(-G_{yj}^{p}\right)^{\beta} G_{yj}^{p} < 0 \end{cases}, \tag{15}$$

where $v(G_{yj}^{p})$ denotes the value of the p-th index of the j-th online learning services of the user subgroup $\mathfrak{R}_y$; α and β denote the coefficients of incomes and losses, respectively; and λ denotes the risk aversion coefficient. According to the study results of Kahneman and Tversky [32], in this study, we set $\alpha = \beta = 0.88$, and $\lambda = 2.25$.

Step 13. Determine the prospect value.

According to the value function, we calculate the prospect value of user subgroups about online learning services, which can be express by

$$f_{yj} = \sum_{p=1}^{l} \pi_p^{+} v\left(G_{yj}^{p}\right) + \sum_{p=1}^{l} \pi_p^{-} v\left(G_{yj}^{p}\right), \tag{16}$$

where f_{yj} denotes the prospect value of the user subgroup $\mathfrak{R}_y$ regarding the j-th online learning service provider, and π_p^{+} and π_p^{-} denote the probability weight function that is used to evaluate the p-th attribute gains and losses of online learning, respectively. The probability weight function can be calculated by

$$\pi_p^{+} = \frac{\left(w_p\right)^{\gamma}}{\left(\left(w_p\right)^{\gamma} + \left(1 - w_p\right)^{\gamma}\right)^{1/\gamma}}, \qquad \pi_p^{-} = \frac{\left(w_p\right)^{\sigma}}{\left(\left(w_p\right)^{\sigma} + \left(1 - w_p\right)^{\sigma}\right)^{1/\sigma}}, \tag{17}$$

where γ and σ denote the probability weight coefficients of gains and losses, respectively. According to the study results of Kahneman and Tversky [32], the values of γ and σ are 0.61 and 0.69, respectively.

Step 14. Determine the satisfaction degree.

According to the prospect value f_{yj} of the user subgroup $\mathfrak{R}_y$ regarding the j-th online learning service provider, we obtain the prospect value matrix F of the user subgroup about the online learning services, which can be expressed as

$$F = \left[f_{yj}\right]_{o\times n} = \begin{bmatrix} f_{11} & f_{12} & \cdots & f_{1l} \\ f_{21} & f_{22} & \cdots & f_{2l} \\ \vdots & \vdots & \ddots & \vdots \\ f_{o1} & f_{o2} & \cdots & f_{ol} \end{bmatrix}_{o\times n}. \tag{18}$$

By normalizing F, we obtain the satisfaction degree f_{yj}' of the user subgroup $\mathfrak{R}_y$ with the online learning service provider s_j, which is calculated by

$$f_{yj}' = \frac{f_{yj} - \min\left\{f_{yj} | y = 1, 2, \ldots, o\right\}}{\max\left\{f_{yj} | y = 1, 2, \ldots, o\right\} - \min\left\{f_{yj} | y = 1, 2, \ldots, o\right\}}. \tag{19}$$

Through conducting the same method, we obtain the satisfaction degree t_{yj}' of the online learning service provider s_j on the user subgroup $\mathfrak{R}_y$.

4.4. Online Learning Matching Method considering User's Characteristics. Based on the above analysis, we obtain the satisfaction degree f_{yj} of the user subgroup $\mathfrak{R}_y$ regarding the online learning service provider s_j, as well as the satisfaction degree t_{yj}' of the service provider s_j regarding the user subgroup $\mathfrak{R}_y$, through which a multi-objective optimization model can be established to maximize the satisfaction degree of users and online learning service providers, thereby achieving the best result of matching the satisfaction degree of both sides.

Specifically, this study introduces the variable x_{yj} that ranges from 0 to 1 to describe the matching between users and online learning services, to which $x_{yj} = 1$ suggests that a matching exists between the user subgroup $\mathfrak{R}_y$ and the online learning service provider s_j, whereas $x_{yj} = 0$ indicates that a mismatching exists between these two sides. In order to solve this bilateral matching problem between users and online learning services, we establish a multi-objective optimization model as follows:

$$\max Z_1 = \sum_{y=1}^{Y}\sum_{j=1}^{n} x_{yj} f'_{yj},$$

$$\max Z_2 = \sum_{y=1}^{Y}\sum_{j=1}^{n} x_{yj} t'_{yj},$$

$$s.t. \sum_{j=1}^{n} x_{yj} \le 1,\ y = 1, 2, \ldots, Y,$$

$$\sum_{y=1}^{Y} x_{yj} \le 1,\ j = 1, 2, \ldots, n,$$

$$x_{yj} = 0\ or\ 1,\ y = 1, 2, \ldots, Y;\ j = 1, 2, \ldots, n. \tag{20}$$

The above multi-objective optimization model determines the most appropriate matching pairs between the two sides by maximizing the satisfaction degree of users and online learning service providers. There are various methods to solve the multi-objective optimization model, among which the commonly used approach is to transform two objectives into a single-objective underlying the rationale of linear weighting method. Following the principle of linear weighting, we therefore transform the above multi-objective optimization model into a single-objective optimization model, which can be expressed by

$$\max Z = \tau_1 \times \sum_{y=1}^{Y}\sum_{j=1}^{n} x_{yj} f'_{yj} + \tau_2 \times \sum_{y=1}^{Y}\sum_{j=1}^{n} x_{yj} t'_{yj},$$

$$s.t. \sum_{j=1}^{n} x_{yj} \le 1,\ y = 1, 2, \ldots, Y,$$

$$\sum_{y=1}^{Y} x_{yj} \le 1,\ j = 1, 2, \ldots, n,$$

$$x_{yj} = 0\ or\ 1,\ y = 1, 2, \ldots, Y;\ j = 1, 2, \ldots, n,$$

$$\tau_1 + \tau_2 = 1,$$

$$0 \le \tau_1, \tau_2 \le 1, \tag{21}$$

where τ_1 and τ_2 are weighting coefficients. This single-objective optimization model can be solved by using MATLAB software, thereby obtaining the optimal matching pairs between users and online learning services.

5. Case Study

In this section, we introduce a real case to verify the effectiveness and rationality of our proposed online learning service matching model, in which the users' learning interests and relevance are considered.

In spite of science and engineering or humanities, China's higher education highly emphasizes English-related courses, which accounts for a high proportion throughout the teaching syllabus of almost all courses of nonlanguage majors. As the Internet and information technology boom, China's higher education is accelerating the implementation of sound and comprehensive informatization in colleges and universities. Characterizing with accessibility, convenience, and efficiency under this social background, online teaching/learning and online educational resources have becoming a popular pedagogy for both teachers and learners.

In recent years, a large number of universities in China have been launching a series of online English teaching and learning services, including online courses and online interactive platforms, which help teachers and students to jointly make their own choices. Nevertheless, many students often feel confused when choosing among diverse online English learning services, which may inhibit them from making the most appropriate preferences. It has become an increasingly prominent problem for those students who deserve the right choices that match their own inherent characteristics involving learning preferences, interests, and advantages. On the one hand, students may choose certain services they like according to their own interests. On the other hand, students may also choose some services following the advices from their classmates and/or friends.

For online service providers such as colleges and universities, offering diversified optional services for students is to enable students to improve their English proficiency according to their own characteristics. Therefore, they may consider many factors that originate from students, such as students' examination results and fields of expertise, thus recommending appropriate courses for students. Therefore, in order to realize the optimal matching between students and online English learning services, how to comprehensively consider students' learning interests and association as well as the expectations of online English learning service providers is a crucial and urgent problem to be solved for the education and teaching reform of China's colleges and universities.

In this study, we take *X* University of China as an example to investigate the influences of four online English teaching/learning services offered by the University for its students. These online services for students to choose are special training of English writing (s_1), special training of English reading (s_2), special training of English listening comprehension (s_3), special training of oral English (s_4), special training of translation and interpretation (s_5), and special training of English grammar (s_6). When launching the above special training courses, the University mainly considers the following six aspects: student's willingness to study abroad (c_1), student's learning interest (c_2), student's grade (c_3), student's major (c_4), student's score (c_5), and student's personal preference (c_6). The students choose online English learning services mainly based on the following points: course credits (a_1), course hours (a_2), course difficulty (a_3), course practicality (a_4), course emphasis (a_5), and course type (a_6) that can only be selected for certain majors.

We randomly selected thirty undergraduate students in the *X* University. They were asked to choose their own online English teaching services according to their own learning interests and the association among them. Specifically, the incidence matrix among students is presented in Table 2, the expectation level of students about online services is shown

Table 2: The incidence matrix and adjacency matrix among students.

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**	**21**	**22**	**23**	**24**	**25**	**26**	**27**	**28**	**29**	**30**
1	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	1	0	0	0	0
3	0	0	0	1	1	0	0	0	0	0	0	0	1	0	0	1	1	0	0	0	0	1	0	0	0	0	0	0	0	0
4	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
5	0	0	1	1	0	0	0	0	0	0	0	0	1	0	0	1	1	1	0	0	1	0	0	0	0	0	0	0	0	0
6	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0	1	0	0	0	1	0	0	0	0
7	0	0	0	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
8	0	0	0	0	0	0	1	0	1	0	1	1	0	1	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0
9	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1	0	1	0	0	0	1	0	0	1	0	0	1	0	0	0
10	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
11	0	0	0	0	0	0	0	1	0	0	0	1	0	0	0	1	1	0	0	1	0	0	0	0	0	0	1	0	0	0
12	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0	1	0	1	0	0	0	0	1	0	0	0	0	1	0	0
13	0	0	1	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1	1	0	1	1	0	0	0	0	0
14	0	1	0	0	0	0	1	1	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0	1
15	0	0	0	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	1	1	0	0	0
16	0	0	1	0	1	0	0	0	0	0	1	1	0	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
17	0	0	1	0	1	0	0	0	1	0	1	0	0	0	0	0	0	1	0	0	1	0	0	0	0	1	0	0	0	0
18	0	0	0	0	1	1	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	1	0	0	0	1	0	0	0	0
19	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0	0
20	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	1	0	1	0	0	0	0	1	0	0	0	0
21	0	0	0	1	1	0	0	0	1	0	0	0	1	0	1	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0
22	1	0	1	0	0	1	1	0	0	1	0	0	1	0	0	0	0	1	0	0	0	0	1	0	0	0	1	0	0	0
23	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0
24	0	0	0	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
25	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0
26	0	1	0	0	0	1	0	0	0	0	0	0	0	1	1	0	1	1	0	1	0	0	0	0	1	0	0	0	0	0
27	0	0	0	0	0	0	0	0	1	0	1	0	0	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
28	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1	1
29	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
30	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0

The significance for bold value represents different students.

in Table 3, the expectation level of the X University about students is listed out in Table 4, the actual level of students themselves is demonstrated in Table 3, and the actual level of online English teaching services is illustrated in Table 4.

Firstly, based on the incidence adjacency matrix among thirty students, these students are subdivided into six subgroups by performing GN algorithm, as shown in Table 5.

Then, the weight of each student within the four subgroups is calculated, to which the results are listed out in Table6.

Through using the weight of students in the subgroups, the decision matrix of students' actual level and that of their expectation level about online learning services are aggregated, thereby obtaining the actual level of students' subgroups and their expectation level about online English learning services, as shown in Table 7.

Further, the weight of each attribute is determined by performing BWM, and five experts are invited to assess the indicators of evaluating students and online services, to which the decision matrix of these experts is shown in Tables 8 and 9.

Taking as an example the evaluation of the attributes of online English learning services, we establish the following linear objective programming model based on BWM.

$$\begin{aligned}
&\min \varsigma,\\
&s.t.\left|w_4 - 5w_1\right| \le \varsigma, \left|w_4 - 6w_1\right| \le \varsigma, \left|w_4 - 4w_2\right| \le \varsigma, \left|w_4 - 3w_2\right| \le \varsigma,\\
&\left|w_4 - 3w_3\right| \le \varsigma, \left|w_4 - 2w_3\right| \le \varsigma, \left|w_4 - 3w_5\right| \le \varsigma, \left|w_4 - 2w_5\right| \le \varsigma,\\
&\left|w_4 - 6w_6\right| \le \varsigma, \left|w_4 - 5w_6\right| \le \varsigma, \left|w_1 - 2w_6\right| \le \varsigma, \left|w_1 - 3w_6\right| \le \varsigma,\\
&\left|w_2 - 3w_6\right| \le \varsigma, \left|w_2 - 4w_6\right| \le \varsigma, \left|w_3 - 5w_6\right| \le \varsigma, \left|w_3 - 4w_6\right| \le \varsigma,\\
&\left|w_3 - 5w_6\right| \le \varsigma, \left|w_4 - 6w_6\right| \le \varsigma, \left|w_4 - 3w_6\right| \le \varsigma, \left|w_4 - 2w_6\right| \le \varsigma,\\
&\sum_{p=1}^{l} w_p = 1,\\
&w_p \ge 0, p = 1, 2, \ldots, l.
\end{aligned} \tag{22}$$

By using MATLAB software to solve the above model, we obtain the weight of evaluating the attributes of online English learning services, which can be expressed as

$$W = \left[\, 0.0845 \;\; 0.1268 \;\; 0.1690 \;\; 0.3944 \;\; 0.1690 \;\; 0.0563 \,\right], \tag{23}$$

where the consistency ration CR is expressed by

$$CR = \frac{\varsigma^*}{CI} = \frac{0.1127}{3.00} = 0.0376 < 0.1. \tag{24}$$

Table 3: Students' actual level and their expectation level about online English teaching services.

	c_1	c_2	c_3	c_4	c_5	c_6	a_1	a_2	a_3	a_4	a_5	a_6
d_1	2	4	3	2	3	3	3	4	3	2	3	3
d_2	1	3	4	3	2	4	4	3	4	3	3	2
d_3	1	3	4	3	2	3	4	2	4	3	3	4
d_4	1	3	3	4	2	3	3	2	3	2	3	4
d_5	1	2	4	3	3	3	4	3	4	3	4	3
d_6	2	4	3	3	3	2	3	4	3	2	2	3
d_7	1	3	3	4	3	4	4	3	4	4	3	3
d_8	2	3	4	4	3	3	4	2	4	4	3	2
d_9	3	2	2	4	3	4	2	3	2	4	2	3
d_{10}	2	4	3	2	2	2	3	4	2	2	3	2
d_{11}	4	3	3	3	2	2	2	1	3	2	3	4
d_{12}	4	4	3	3	3	2	2	1	3	2	3	3
d_{13}	3	3	2	3	2	3	2	3	2	3	2	4
d_{14}	1	3	3	3	3	4	3	2	3	3	3	2
d_{15}	3	4	3	3	3	3	3	3	2	3	2	3
d_{16}	4	3	4	3	3	2	3	2	4	3	2	3
d_{17}	1	3	4	4	2	3	4	2	4	3	2	4
d_{18}	1	3	3	4	2	3	4	3	4	3	2	3
d_{19}	2	4	3	2	3	3	3	3	3	2	3	3
d_{20}	2	3	4	4	2	3	3	3	4	3	3	3
d_{21}	3	3	4	3	4	4	3	2	2	3	3	4
d_{22}	2	3	2	3	3	2	4	3	3	2	3	2
d_{23}	4	3	3	4	3	2	4	3	4	3	2	3
d_{24}	3	3	2	3	4	3	2	3	3	4	3	3
d_{25}	1	3	4	3	3	4	3	3	4	3	3	3
d_{26}	2	3	3	4	2	3	4	3	4	3	3	2
d_{27}	2	3	3	2	3	3	3	2	3	3	4	4
d_{28}	1	4	1	2	2	1	3	4	3	2	4	3
d_{29}	2	4	1	3	3	2	3	3	2	3	3	3
d_{30}	1	3	1	2	2	2	4	3	3	2	3	4

Table 4: Online English learning services' actual level and their expectation level about students.

	a_1	a_2	a_3	a_4	a_5	a_6	c_1	c_2	c_3	c_4	c_5	c_6
s_1	2	2	2	2	3	2	2	3	2	2	2	3
s_2	2	2	1	2	2	3	2	3	2	1	2	4
s_3	3	3	3	2	4	2	1	3	4	2	3	3
s_4	4	3	4	3	3	3	3	3	3	3	2	2
s_5	3	4	3	2	2	4	4	3	2	2	3	3
s_6	2	3	3	4	3	4	2	2	3	4	4	2

Table 5: Subgroup division result of 30 students.

Subgroup	Students
$\mathfrak{R}_1$	d_1, d_6, d_{10}, d_{19}, d_{22}
$\mathfrak{R}_2$	d_3, d_4, d_5, d_{17}, d_{18}
$\mathfrak{R}_3$	d_2, d_7, d_8, d_{14}, d_{20}, d_{25}, d_{26}
$\mathfrak{R}_4$	d_9, d_{13}, d_{15}, d_{21}, d_{24}, d_{27}
$\mathfrak{R}_5$	d_{11}, d_{12}, d_{16}, d_{23}
$\mathfrak{R}_6$	d_{28}, d_{29}, d_{30}

Through verifying the consistency, the rationality of the above weight is therefore validated.

By performing the same method, we obtain the weight of evaluating students' attributes, which is written as

$$V = \left[\, 0.1662 \;\; 0.3878 \;\; 0.0554 \;\; 0.0997 \;\; 0.1662 \;\; 0.1247 \,\right]. \tag{25}$$

Using the obtained attribute weights, we can calculate the satisfaction degree matrix between the student subgroups and the online English learning services, as listed out in the following Tables 10 and 11.

Subsequently, according to the satisfaction degree matrices of the two sides, we establish a bilateral matching model between the student subgroups and the online English learning services, in which the weighting coefficient $\tau_1 = \tau_2 = 0.5$ is taken. By using MATLAB software, we therefore obtain the matching results between the student subgroups and the online English learning services, to which the schematic diagram is presented in Figure 1.

Table 6: Students' weight in subgroups.

$\mathfrak{R}_1$	Student	d_1	d_6	d_{10}	d_{19}	d_{22}		
	Weight	0.1600	0.2000	0.1200	0.1600	0.3600		
$\mathfrak{R}_2$	Student	d_3	d_4	d_5	d_{17}	d_{18}		
	Weight	0.2069	0.1034	0.2414	0.2414	0.2069		
$\mathfrak{R}_3$	Student	d_2	d_7	d_8	d_{14}	d_{20}	d_{25}	d_{26}
	Weight	0.0811	0.0811	0.1892	0.1892	0.1351	0.1081	0.2162
$\mathfrak{R}_4$	Student	d_9	d_{13}	d_{15}	d_{21}	d_{24}	d_{27}	
	Weight	0.1875	0.2188	0.1563	0.2188	0.0938	0.1250	
$\mathfrak{R}_5$	Student	d_{11}	d_{12}	d_{16}	d_{23}			
	Weight	0.2857	0.2857	0.2857	0.1429			
$\mathfrak{R}_6$	Student	d_{28}	d_{29}	d_{30}				
	Weight	0.5000	0.1250	0.3750				

Table 7: Student subgroups' actual level and their expectation level about online English learning services.

	c_1	c_2	c_3	c_4	c_5	c_6	a_1	a_2	a_3	a_4	a_5	a_6
$\mathfrak{R}_1$	2.0000	3.6400	2.6400	2.5600	2.8800	2.3200	3.3600	3.4800	2.8800	2.0000	2.8000	2.5200
$\mathfrak{R}_2$	1.0000	2.7586	3.6897	3.5517	2.2414	3.0000	3.8966	2.4483	3.8966	2.8966	2.7931	3.5517
$\mathfrak{R}_3$	1.5405	3.0000	3.5135	3.6216	2.5676	3.4595	3.5676	2.6216	3.8108	3.2703	3.0000	2.3243
$\mathfrak{R}_4$	2.8756	2.9694	2.7193	3.0631	3.0944	3.4069	2.5005	2.6568	2.2192	3.2819	2.5630	3.5632
$\mathfrak{R}_5$	4.0000	3.2857	3.2857	3.1429	2.7143	2.0000	2.5715	1.5715	3.4286	2.4286	2.5714	3.2857
$\mathfrak{R}_6$	1.1250	3.6250	1.0000	2.1250	2.1250	1.5000	3.3750	3.5000	2.8750	2.1250	3.5000	3.3750

Table 8: The best and worst attributes and corresponding vectors evaluating online English learning services.

	Best—all vectors of attributes							Worst— all vectors of attributes						
	Best	a_1	a_2	a_3	a_4	a_5	a_6	Worst	a_1	a_2	a_3	a_4	a_5	a_6
Expert 1	a_4	5	4	2	1	3	6	a_6	2	3	4	5	3	1
Expert 2	a_4	6	4	3	1	3	5	a_6	3	4	5	5	2	1
Expert 3	a_4	5	3	2	1	3	5	a_6	2	3	4	6	3	1
Expert 4	a_4	6	3	2	1	2	6	a_6	2	3	5	5	3	1
Expert 5	a_4	5	4	2	1	3	6	a_6	3	4	5	6	2	1

Table 9: The best and worst attributes and corresponding vectors evaluating student subgroups.

	Best—all vectors of attributes							Worst—all vectors of attributes						
	Best	c_1	c_2	c_3	c_4	c_5	c_6	Worst	c_1	c_2	c_3	c_4	c_5	c_6
Expert 1	c_2	2	1	5	4	2	3	c_3	4	5	1	2	3	2
Expert 2	c_2	2	1	6	5	3	4	c_3	5	6	1	2	4	3
Expert 3	c_2	2	1	5	4	3	3	c_3	5	5	1	3	3	3
Expert 4	c_2	3	1	6	5	3	4	c_3	4	6	1	2	4	3
Expert 5	c_2	2	1	5	4	3	3	c_3	5	6	1	2	3	3

Table 10: Satisfaction degree of student subgroups on online English learning services.

	s_1	s_2	s_3	s_4	s_5	s_6
$\mathfrak{R}_1$	0.8212	0.8083	0.9737	0.6443	1	0.6714
$\mathfrak{R}_2$	0.1557	0.1579	0.1398	0.2908	0.1257	0.0485
$\mathfrak{R}_3$	0	0	0	0	0	0
$\mathfrak{R}_4$	0.6728	0.6332	0.6764	0.5184	0.6246	0.9470
$\mathfrak{R}_5$	1	1	1	1	0.9093	1
$\mathfrak{R}_6$	0.7208	0.7071	0.8593	0.5883	0.8972	0.6171

Table 11: Satisfaction degree of online English learning services about student subgroups.

	$\mathfrak{R}_1$	$\mathfrak{R}_2$	$\mathfrak{R}_3$	$\mathfrak{R}_4$	$\mathfrak{R}_5$	$\mathfrak{R}_6$
s_1	0.8549	0.8048	0.8483	0.8312	0.7660	0.8453
s_2	1	0.9558	1	1	1	1
s_3	0.7148	1	0.9220	0.6265	0.5395	0.9190
s_4	0	0.1096	0.1214	0.1174	0	0
s_5	0.0516	0	0	0	0.0748	0.1371
s_6	0.4666	0.8001	0.6954	0.5293	0.3542	0.4097

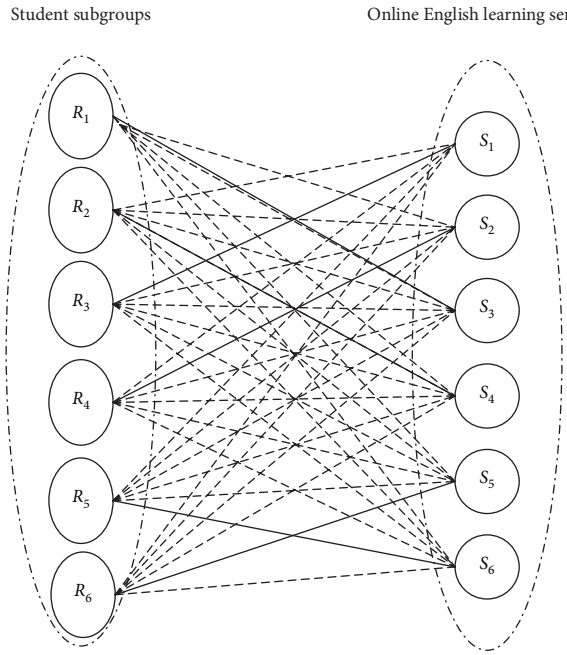


Figure 1: Matching results between student subgroups and online English learning services.

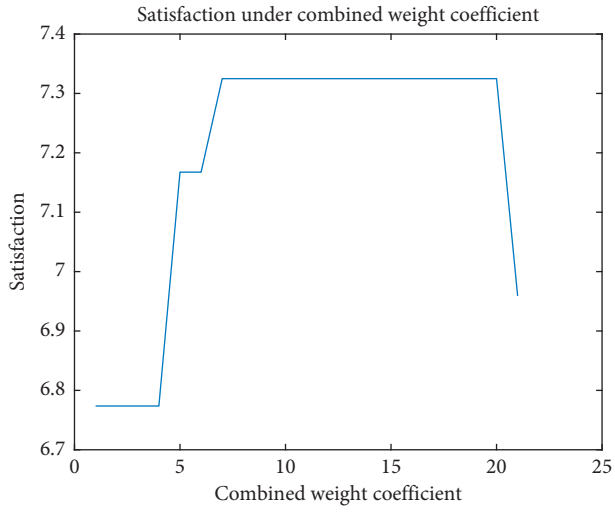


Figure 2: Sensitivity analysis.

It can be observed from Figure 1 that the best/optimal matching pairs between the subgroups and the services are $\mathfrak{R}_1 \leftrightarrow s_3$, $\mathfrak{R}_2 \leftrightarrow s_4$, $\mathfrak{R}_3 \leftrightarrow s_1$, $\mathfrak{R}_4 \leftrightarrow s_2$, $\mathfrak{R}_5 \leftrightarrow s_6$, and $\mathfrak{R}_6 \leftrightarrow s_5$.

Finally, we conduct sensitivity analysis with respect to the weighting coefficients τ_1 and τ_2 in our established model. In response to τ_1 and τ_2, their values adopted in this study are within their value ranges, to which we assign values to τ_1 and τ_2 with a step size of 0.05, satisfying $\tau_1 + \tau_2 = 1$. We performed a total of 21 tests on τ_1 and τ_2, and the result is presented as follows.

It can be observed from Figure 2 that when $\tau_1 \in [0.3, 0.95]$ and the corresponding $\tau_2 \in [0.05, 0.7]$, the matching result between the student subgroups and the online English learning services remains unchanged, to which the maximum satisfaction degree is 7.3249. As τ_1 gradually increases and τ_2 decreases in range $[0, 0.3]$, the above matching result changes, to which the corresponding maximum satisfaction degree also decreases. To summarize, with appropriate generality, in this study, we set $\tau_1 = \tau_2 = 0.5$ in the above case study.

6. Conclusions

Under the current harsh and severe situation of COVID-19 epidemic, the rapid development of the Internet and informatization are ushering in a new wave of opportunities for online learning services. How to choose suitable matching pairs for users and online learning services has become a critical research trend in the study of knowledge service. From the perspective of bilateral matching, we propose an online learning service matching method, which jointly considers user preferences by combining social network analysis (SNA), cumulative prospect theory and best-worst method. In this study, we matched the undergraduate students of X University in China with online English learning services provided by the University and verified the practicability and effectiveness of our proposed method. Our proposal is capable of extending existing studies, making innovative contributions in practice.

Firstly, we extend knowledge service matching method to the matching problem between users and online learning services, thereby proposing a clear logical framework while forming an ingenious perspective. Our proposed decision-making method can be extended to practical matching problems in other domains such as the matching between students and colleges or that between students and job vacancies.

Secondly, when considering the preferences and user demands, we explore the relationship between users from social network analysis perspective, thereby dividing the selected users into subgroups. Compared with conventional bilateral matching methods, the subdividing method we adopted effectively improves the efficiency and accuracy of matching.

Thirdly, performing as one of the most effective approaches to cope with the gap between the actual and the expectation levels, we obtain the satisfaction degree between users and online learning services by adopting cumulative prospect theory, which comprehensively and precisely describes the satisfaction degree of the two sides to be matched.

Despite the above research findings, certain shortcoming may exist in our proposed method, which may potentially

impose detrimental encumbrance or disadvantageous impact on matching decision-making problems. Our future research interests will be focusing on further enriching the relevant evaluation information, enabling more robust generalization ability of the proposed method for solving other matching problems in complicated scenarios [33, 34].

Data Availability

The structured, semistructured, and unstructured data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This paper was supported by the Fundamental Research Funds of China's Central Universities under Grant no. 20101215636.

References

[1] N. Haefner, J. Wincenta, V. Paridac, and O. Gassmanna, "Artificial intelligence and innovation management: a review, framework, and research agenda," *Technological Forecasting and Social Change*, vol. 162, pp. 1–10, 2021.

[2] R. Truzoli, V. Conte, and S. Conte, "The impact of risk and protective factors on online teaching experience in high school Italian teachers during the COVID-19 pandemic," *Journal of Computer Assisted Learning*, vol. 37, no. 4, pp. 940–952, Aug, 2021.

[3] H. Aziz, T. James, D. Remulla et al., "Effect of COVID-19 on surgical training across the United States: a national survey of general surgery residents," *Journal of Surgical Education*, vol. 78, no. 2, pp. 431–439, Mar, 2021.

[4] R. Gopal, V. Aggarwal, and A. Aggarwal, "Impact of online classes on the satisfaction and performance of students during the pandemic period of COVID 19," *Education and Information Technologies*, vol. 26, no. 6, pp. 6923–6947, 2021.

[5] K. Chaturvedi, D. K. Singh, and N. Singh, "COVID-19 and its impact on education, social life and mental health of students: a survey," *Children and Youth Services Review*, vol. 121, Article ID 105866, 2021.

[6] R. Scherer, S. K. Howard, J. Siddiq, and F. Siddiq, "Profiling teachers' readiness for online teaching and learning in higher education: who's ready?" *Computers in Human Behavior*, vol. 118, Article ID 106675, 2021.

[7] X. F. Xu, C. L. Zhou, and P. Zhou, "GVRP considered oil-gas recovery in refined oil distribution: from an environmental perspective," *International Journal of Production Economics*, vol. 235, Article ID 108078, 2021.

[8] T. Fleiner, R. W. Manlove, and D. F. Manlove, "Efficient algorithms for generalized stable marriage and roommates problems," *Theoretical Computer Science*, vol. 381, no. 1-3, pp. 162–176, 2007.

[9] Z. Zhang, J. Gao, Y. Yu, and W. Yu, "Two-sided matching decision making with multi-granular hesitant fuzzy linguistic term sets and incomplete criteria weight information," *Expert Systems with Applications*, vol. 168, Article ID 114311, 2021.

[10] Y. Q. Ma, Q. Y. Li, Z. Z. Lin, Z. Ge, Y. Zhao, and X. L. Zhao, "Premium power value-added service product decision-making method based on multi-index two-sided matching," *IEEE Access*, vol. 9, pp. 62166–62181, 2021.

[11] S. Q. Chen, Y. M. Wang, H. L. Zhang, and X. X. Zhang, "A decision-making method for uncertain matching between volunteer teams and rescue tasks," *International Journal of Disaster Risk Reduction*, vol. 58, Article ID 102138, 2021.

[12] Q. Yue, L. Zhang, Y. S. Peng, and Q. YuHongXiao, "Decision method for two-sided matching with interval-valued intuitionistic fuzzy sets considering matching aspirations," *Journal of Intelligent and Fuzzy Systems*, vol. 31, no. 6, pp. 2903–2910, 2016.

[13] X. Ma and H. Ma, "A two-sided matching decision model based on uncertain preference sequences," *Mathematical Problems in Engineering*, vol. 2015, 1-10 pages, Article ID 241379, 2015.

[14] Z. P. Fan, M. Y. Li, and X. Zhang, "Satisfied two-sided matching: a method considering elation and disappointment of agents," *Soft Computing*, vol. 22, no. 21, pp. 7227–7241, 2018.

[15] I. Korkmaz, H. Gökçen, and T. Çetinyokuş, "An analytic hierarchy process and two-sided matching based decision support system for military personnel assignment," *Information Sciences*, vol. 178, no. 14, pp. 2915–2927, 2008.

[16] J. Han, B. Li, H. M. Liang, and K. K. Lai, "A novel two-sided matching decision method for technological knowledge supplier and demander considering the network collaboration effect," *Soft Computing*, vol. 22, no. 16, pp. 5439–5451, 2018.

[17] J.-W. Lee, "Online support service quality, online learning acceptance, and student satisfaction," *The Internet and Higher Education*, Dec, vol. 13, no. 4, pp. 277–283, 2010.

[18] H. He, Q. H. Zheng, D. H. Di, and B. Dong, "How learner support services affect student engagement in online learning environments," *IEEE Access*, vol. 7, pp. 49961–49973, 2019.

[19] Y. Y. Shan and S. N. Shan, "Online learning support service system architecture based on location service architecture," *Mobile Information Systems*, vol. 2021, Article ID 6663934, 1-11 pages, 2021.

[20] J.-H. Jung and J.-I. Shin, "Assessment of university students on online remote learning during COVID-19 pandemic in Korea: an empirical study," *Sustainability*, Oct, vol. 13, no. 19, pp. 1–13, 2021.

[21] X. Chen, Z. W. Li, Z. P. Fan, X. Zhou, and X. Zhang, "Matching demanders and suppliers in knowledge service: A method based on fuzzy axiomatic design," *Information Sciences*, vol. 346-347, pp. 130–145, 2016.

[22] M. Li, J. Wang, and Y. Xu, "An approach to the match between panels of experts and groups of demanders in fuzzy linguistic environments," *Kybernetes*, Dec, vol. 46, no. 5, pp. 854–875, 2017.

[23] J. Chang, H. Sun, and B. Z. Sun, "Matching knowledge suppliers and demanders on a digital platform: a novel method," *IEEE Access*, vol. 7, pp. 21331–21342, 2019.

[24] X. Pan, J. Song, J. T. Truong, and V. A. Truong, "Online contextual learning with perishable resources allocation," *Iise Transactions*, vol. 52, no. 12, pp. 1343–1357, 2020.

[25] Z. Zhang, X. Y. Kou, W. Y. Gao, and Y. Gao, "Consistency improvement for fuzzy preference relations with self-confidence: an application in two-sided matching decision making," *Journal of the Operational Research Society*, vol. 72, no. 8, pp. 1914–1927, 2021.

[26] L. Gao, Y. Gan, Z. Zhang, and X. L. Zhang, "A user-knowledge dynamic pattern matching process and optimization strategy

based on the expert knowledge recommendation system," *Applied Intelligence*, vol. 51, no. 11, pp. 7601–7613, 2021.

[27] E. Rousseau and R. Rousseau, "Social network analysis: a powerful strategy, also for the information sciences," *Journal of Information Science*, Jan, vol. 28, no. 6, pp. 441–453, 2002.

[28] M. Arasteh and S. Alizadeh, "A fast divisive community detection algorithm based on edge degree betweenness centrality," *Applied Intelligence*, vol. 49, no. 2, pp. 689–702, Feb, 2019.

[29] B. Mozaffari and M. Mozaffari, "A distributed averaging-based evidential Expectation-Maximization algorithm for density estimation in unreliable sensor networks," *Measurement*, vol. 165, Article ID 108162, 2020.

[30] J. Rezaei, "Best-worst multi-criteria decision-making method," *Omega*, vol. 53, pp. 49–57, 2015.

[31] S. Guo and H. R. Zhao, "Fuzzy best-worst multi-criteria decision-making method and its applications," *Knowledge-Based Systems*, vol. 34, no. 8, pp. 1953–1967, 2019.

[32] C. R. Tversky and A. Tversky, "A belief-based account of decision under uncertainty," *Management Science*, vol. 44, no. 7, pp. 879–895, Jul, 1998.

[33] X. F. Xu, Z. R. Lin, X. Li, C. J. Shen, and Q. Shen, "Multi-objective robust optimisation model for MDVRPLS in refined oil distribution," *International Journal of Production Research*, vol. 5, pp. 1–21, 2021.

[34] X. F. Xu, J. Zheng, and Y. Zheng, "Multi-objective artificial bee colony algorithm for multi-stage resource leveling problem in sharing logistics network," *Computers & Industrial Engineering*, vol. 142, no. 4, Article ID 106338, 2020.

Hindawi
Scientific Programming
Volume 2022, Article ID 3946744, 18 pages
https://doi.org/10.1155/2022/3946744

Research Article

Optimization and Benefit Analysis of Intelligent Networked Vehicle Supply Chain Based on Stackelberg Algorithms

Cheng Che, Xin Geng, Huixian Zheng, Yi Chen, and Xiaoguang Zhang

School of Economics and Management, China University of Petroleum, Qingdao 266580, Shandong, China

Correspondence should be addressed to Xin Geng; 1371178916@qq.com

Received 14 April 2022; Revised 13 May 2022; Accepted 9 June 2022; Published 5 July 2022

Academic Editor: Hongguang Ma

With the rapid development of intelligent technology, the construction of smart cities with the goal of creating a harmonious human life has become a new hot spot in the world. A new round of information technology revolution, represented by the Internet, big data, cloud computing, artificial intelligence (AI), and fifth-generation mobile communications (5G), is driving profound changes in the automotive industry. Smart vehicles (also known as intelligent connected vehicles) that integrate many high-tech technologies to provide safer, more convenient, and low-carbon comprehensive travel solutions have become an inevitable form of future vehicles. This paper constructs a three-level supply chain consisting of high-performance chip suppliers, smart car manufacturers, and retailers. On the basis of considering the level of product innovation and sales effort, Stackelberg game method is used to study the influence of each player in the supply chain on each parameter and the profit of each player under three scenarios: centralized decision making, nonsharing of innovation cost, and sharing of innovation cost. The results show that: when the level of sales effort has nothing to do with the level of innovation, the level of product innovation and the total profit of the supply chain increase with the improvement of the retailer's sales ability; when the level of sales effort is related to the level of innovation, the level of product innovation under the cost-sharing decision model is greater than the case of no cost sharing, but the total profit of the supply chain is less than the case of no cost sharing.

1. Introduction

From smart buildings to smart residential quarters, how to realize the sustainable development of cities has always been the focus of human attention and exploration. At present, with the rapid development of intelligent technology, the construction of smart cities with the goal of creating a harmonious human life has become a new hot spot in the world. In the construction of smart cities, new energy vehicles have become an important part of construction and promotion in recent years due to the energy saving and environmental protection. In 2020, the new energy vehicle industry has made breakthrough progress, transforming from the primary stage of development to the middle and advanced stage, from focusing on solving the "three electricity problems" to promoting the integration of electrification and renewable energy, intelligent transportation, and smart city. The coordinated development of intelligent connected vehicles and smart cities constitutes a vehicle-city-integrated development system. The development of smart cities requires cars as the entry point and driving force, and the development of smart cars requires the supporting support and guarantee of urban energy and digital infrastructure more than ever, so as to further optimize urban infrastructure, improve travel services, and enhance urban operating efficiency. In the future, intelligent networked vehicles will become the "neurons" moving in smart cities, connecting with intelligent traffic control systems to form an integrated smart travel service system. Intelligent transportation refers to an efficient integrated transportation system that integrates information, communication, sensing, control, computer technology, and transportation management system [1]. Smart city (SC) refers to the use of Internet of Things, cloud computing, and other technologies to change the interaction mode of various entities urban subjects, quickly respond to various social needs, and build a new type of livable city that operates efficiently [2]. In a smart city, transportation is a necessary part of residents' life,

so the intelligent development of transportation is an important foundation for a smart city [3].

At present, the world's energy demand is highly dependent on fossil fuels, namely coal, oil, and natural gas [4]. Due to the continuous growth of energy demand, the use of fossil fuels has led to a rapid increase in the total amount of global carbon emissions, which has caused serious damage to the Earth's ecological environment [5]. In response to global climate change, many countries have taken joint actions and successively announced the timetable for banning the sale of gasoline-powered vehicles. The Netherlands will completely ban the sale of fuel vehicles in 2025, India, England, and France will also fully implement the ban in 2030–2040, and China will soon formulate a timetable for banning the sale of fuel vehicles. Accelerating the development of the intelligent connected vehicle industry, including intelligent connected vehicles, has become one of the important strategies to solve global environmental problems [6, 7]. ICV has not only involved the automobile industry, but also the communication, electronic parts, and other industries. It has become a competitive highland for the development of automobile technology in various countries. It has received the attention of the Chinese government and has risen to the height of national strategy [8]. Intelligent connected vehicle is the integration of the Internet and the Internet of Things, which are two high and new technologies. It deeply integrates the innovation achievements of the Internet and the Internet of Things technology into the field of intelligent vehicles. Relying on the development of modern communication information technology and the Internet, the multidirectional docking between vehicles and *X* elements (vehicles, people, roads, network platforms, etc.) is realized, and the intelligent automatic driving function of vehicles is comprehensively improved, so as to improve traffic congestion and provide intelligent, efficient, safe, comfortable, energy-saving, and environment friendly user experience [9]. With the promotion of China's intelligent and connected automobile industry to the national strategic level, the industry positioning is shifting from automatic driving and Internet of vehicles to intelligent manufacturing and intelligent network integration. In recent years, relying on 5G, AI industry, and O_2O service industry, China has made good progress in the field of intelligent and connected market operations [10].

The ICV industry is the product of the deep integration of the automotive industry and the new generation of information technology. It integrates technologies in many fields such as automobiles, transportation, communications, the Internet, and big data, forming an industrial chain that integrates multiple industries [9, 11]. The industrial chain of intelligent networked vehicles is divided into three parts: upper, middle, and lower. Upstream is mainly the technology layer, including perception, communication, decision-making, and execution system, involving information fusion, cloud platform and big data, vehicle-integrated control, and other technologies; On the basis of technology integration, midstream provides vehicle, intelligent cockpit, and autonomous driving solutions, involving a large number of system integration technology, interaction technology, vehicle involvement, and manufacturing technology. Downstream is the application level, including car sales and other related application services [12]. Intelligent connected vehicles will be an inevitable form in the future automotive industry market. In the era of intelligent connected vehicles, it is very important to do a good job in the marketing and service of automobiles. As a new type of automotive product, intelligent connected vehicles are quite different from traditional fuel vehicles in terms of technology and handling performance. These differences will affect whether consumers are willing to accept such products. Therefore, consumers should increase their awareness of intelligent connected vehicles. The cognition of the automobile market and the understanding of the policy will promote the selection of this type of automobile [13]. In the actual marketing environment, consumers' perceived value of cars can be used as the result of consumers' evaluation of the product and then affect their purchase intentions [14, 15]. In the automotive market, because intelligent connected vehicles are new technology-oriented products, in the early stage of development, technology-related attributes have become the focus of customers' attention. Empirical studies by some scholars have found that the improvement of car performance and the reduction of price will increase consumers' willingness to buy, thereby further expanding the consumer market [16]. Egbue et al. also confirmed that car performance has a positive impact on customer acceptance [17]. Intelligent connected vehicles are an opportunity recognized by the global automotive industry as an opportunity with huge social benefits. In the future, it is worth exploring how the automotive industry can upgrade its technology to improve car performance and introduce it into consumers' lives through good marketing methods [18].

Based on the above background, this paper studies the supply chain of intelligent networked vehicles. Different from previous literature, this paper considers the impact of retailers' sales efforts and product innovation levels on the supply chain. By constructing a three-level supply chain model consisting of high-performance chip suppliers, car manufacturers, and retailers in the automotive market, this paper analyzes the maximum profit level of each subject in the supply chain under three different situations. First of all, compared with the secondary supply chain in most previous studies, we study the tertiary supply chain, which is more in line with the industrial chain structure of intelligent connected vehicles. Second, unlike previous studies, we also considered retailers' sales efforts and product innovation levels. Because we understand that consumers' knowledge of intelligent connected vehicles and the performance and quality of the chips installed in the car will affect the market demand for cars. Finally, through the establishment of the Stackelberg game model to analyze different situations, it further broadens the research ideas of ICV supply chain and enriches related research theories.

2. Literature Review

This paper focuses on the impact on the profit of each entity in the three-tier supply chain of intelligent connected

vehicles when the factors of product innovation level and sales effort are considered. There are three categories of domestic and foreign literature related to this study: intelligent connected vehicles and their industrial chain composition, automotive secondary supply chain, and research on collaborative innovation decision-making in automotive supply chain. In this section, we will review relevant literature and summarize these three parts.

2.1. Intelligent Connected Vehicle and Its Industrial Chain Composition. Smart city and intelligent transportation are the themes of urban development in the new century. In particular, the use of the Internet and big data to improve the operational efficiency of cities and improve traffic efficiency is the goal pursued by city managers. In this context, automobile manufacturers reposition their products and turn to provide more information-based and intelligent products and services, so that enterprises can adapt to the new living environment in the future [19]. As a strategic emerging industry involving multiindustry intersection, the research and development of its key technologies has a very significant impact on industrial development [10]. In the future, intelligent connected vehicles will be able to effectively ensure traffic safety, and the intelligent cockpit system in the car will also bring extraordinary experience and fun to drivers and passengers. Moreover, due to the integration of multiple technologies, intelligent connected vehicles have become the strategic commanding heights of technological transformation in the automotive industry [20, 21]. In recent years, the field of smart cars has seen tremendous growth in the lot market. The number of intelligent connected vehicles will continue to increase in the next few years. This is due to the continuous upgrading of Internet of Things technology, which provides more innovation opportunities for applications in the automotive field. In the construction of smart cities in the future, the role of intelligent connected vehicles will become more and more prominent [22, 23].

At the stage of rapid development of science and technology in China, the application of various new technologies in the automotive field has promoted the development of the automotive industry, and intelligent connected vehicles will be the competitive point of future automotive development. Modern intelligent connected cars are more in line with people's expectations. At present, intelligent connected vehicles have formed a complete industrial chain. According to the research on its industrial chain, we can think that the industrial chain of ICV is from the supply of chips to the distribution and manufacture of vehicle parts, and then to the sales link in the downstream market [24]. Wang et al. mentioned in their research that the ICV industry chain involves many fields, but according to the relationship between the upstream and downstream of the industry chain, it specifically covers upstream key systems such as decision-making and execution, and midstream system integration such as intelligent cockpit and intelligent connected vehicle. And downstream application services such as sales [25]. At present, the key upstream area of the ICV industry chain is automotive chips, but there are still technical barriers, and the development is constrained by foreign first-tier suppliers [26]. Zeng believes that chips, as the key components of automobiles, are the foundation of automobiles and the foundation of a strong country. If the chips of intelligent connected vehicles are not strong, then the whole vehicle is not strong, and the car is not strong [27]. The R&D and production of key components represented by connected and intelligent technologies will be a key link in the future ICV industry chain, and chips will become the core components of the car to realize the interconnection between people, vehicles, and homes [28].

By reviewing the related literature on intelligent connected vehicles and their industrial chains, we can find that with the continuous advancement of technology in the future, intelligent interconnection technology will support a new automobile civilization, and its complete industrial chain structure will further promote automobiles. The transformation and progress of the industry have brought new development opportunities to the existing traditional automobile industry. Therefore, intelligent connected vehicles are regarded by major auto-industries as a major opportunity for future development and become the focus of industrial layout. From the above analysis of the industrial chain structure of intelligent connected vehicles, we can think that its supply chain is a three-level supply chain system composed of chip suppliers that dominate the upstream, vehicle manufacturers in the midstream, and sellers in the downstream.

2.2. Secondary Supply Chain of Automobiles. Although intelligent connected vehicles have good market prospects, they are faced with many practical problems such as imperfect market development, low consumer maturity, and not yet fully opened demand. In order to increase the market demand for intelligent connected vehicles, it is very important to design a reasonable supply chain system. In recent years, there has been a lot of literature on automotive secondary supply chains. In the related research on the automobile supply chain under the subsidy policy, Shen et al. studied two supply chain models of direct government subsidies to automobile manufacturers and direct subsidies to distributors. Through comparison, they found that manufacturers, as the dominant player in the market, have more bargaining power. Moreover, with the payment transfer of subsidies, the profits of one enterprise will also drive the profits of the other [29, 30]. Ju et al. incorporated policy features such as industrial quantity targets and subsidy budget constraints into the model for research [31]. Sheuet al. studied the impact of government subsidies and green taxes on competition in green supply chains [32]. In the research on this aspect, some scholars discuss from the perspective of government subsidy policy and efficiency, and some scholars study the problem of subsidy strategy from the perspective of automobile supply chain. In order to promote the transformation and upgrading of the automobile industry, many countries have implemented active automobile subsidy policies for their development in order to reduce the price of automobiles. And many governments

directly provide subsidies to car buyers or sellers. In the related research on the automotive secondary supply chain model under the "double points" policy, Lu et al. considered new energy vehicles under the two-level supply chain model composed of two automobile manufacturers and one autodealer. They considered the impact of emission reduction and endurance of new energy vehicles on the supply chain [33, 34]. Yu et al. studied the importance of the quality game of electric vehicles under the policy of retreat and double points, and deeply explored the production decision optimization problem of the secondary supply chain [35, 36]. Tang et al. introduced the double-point system implemented by the government and consumer preferences into the supply chain model. Considering the impact of the two on the optimal decision-making of the supply chain, the market size of the development of the automotive industry and the benefits obtained by the supply chain are closely both related to the fluctuation of the point price [37, 38]. Kang and Zhang studied the game change process between fuel vehicle and electric vehicle manufacturers under the two models of centralized decision-making and decentralized decision-making under the double-point trading policy, and further introduced consumers' concern about low carbon as an important factor affecting purchasing behaviour [39]. Wang constructed a production R&D game model for traditional automakers and the new energy vehicle market entered in two ways: self-production and outsourcing [40]. As the government's low-carbon subsidy policy "recedes," the "double credit" policy came into being. In the future, the double credit policy will play an indispensable role in the transformation of the automobile industry from fuel vehicles to new energy vehicles and intelligent connected vehicles. In the related research on the automotive supply chain from the perspective of "closed-loop supply chain," Ma et al. constructed a two-channel battery recycling game model between automobile manufacturers and retailers to implement cost-sharing contract and liability sharing contract, and studied the impact of different conditions on automobile retail price, market demand, and profit of each member of the supply chain [41]. Qiu et al. used evolutionary game method to analyze the choice of battery recycling strategy of new energy vehicles in the two-level supply chain composed of original equipment manufactures and 4S stores, and found that the key factor affecting the subsidy strategy is the revenue increase rate of both sides after the subsidy [42]. Covindan et al. studied the patterns of coordination among agents in a two-level closed-loop supply chain, classified them, and compared their advantages and disadvantages [43, 44]. Xie et al. focused on the revenue sharing contract of forward and reverse supply chains. Considering the closed-loop supply chain system composed of a single manufacturer and a single retailer in the market without consumption preference, they studied the influence of different sharing ratios on the profit maximization of manufacturers and retailers [45, 46]. From the research of the above scholars, it can be found that the closed-loop supply chain has advantages in environmental protection and resource conservation, and can improve the efficiency of resource recovery, thereby bringing huge social and economic benefits. However, most of the current researches on closed-loop supply chain and coordination mechanism are relatively scattered, and there are many studies on different dominant modes and recycling modes, but they fail to summarize the characteristics of automobile sales and recycling and build a corresponding supply chain system.

Through the summary and review of the above literature, we can find that scholars have conducted a large number of researches on the secondary supply chain of automobiles and achieved a lot of research results. It can be seen that most of the relevant researches on automobile supply chain are conducted from the perspective of government subsidy policy, "double integral" policy, and closed-loop supply chain. With the development of science and technology in the future, the traditional automobile industry will accelerate industrial transformation and upgrading, and intelligent connected vehicles will become the leader of the automobile industry in the future. For intelligent connected vehicles, the three-level supply chain is more in line with its industrial chain structure, but the relevant research on the three-level supply chain of intelligent connected vehicles is still in its infancy. Based on the existing research, we try to construct a three-level supply chain model composed of high-performance chip suppliers, intelligent vehicle manufacturers, and retailers in the automobile market, and we also discuss the influence on the profits of each body in the supply chain under different circumstances.

2.3. Research on Collaborative Innovation Decision-Making in Automotive Supply Chain. The collaborative innovation of the supply chain has become an important strategy for companies to deal with market competition. The collaborative innovation between the main bodies of the supply chain can make the operation of the supply chain more efficient and then benefit each member. Therefore, scholars have conducted a lot of research on collaborative innovation decision-making in the automotive supply chain. In the literature that studies the impact of collaborative innovation on the automotive supply chain under different circumstances, Jing et al. discussed the optimal allocation and promotion effort level of the automotive supply chain under decentralized decision-making, centralized decision-making, revenue sharing, and cost-sharing contracts. The study found that system coordination cannot be achieved under decentralized decision-making, while the combined contract of cost sharing and benefit sharing can achieve coordination and achieve Pareto improvement [47]. Zhang et al. studied the choice of supply chain entities in R&D investment and technological innovation mode under two innovation modes. The study showed that when the market demand for new products was very high, the profits would be the largest when the subjects cooperated with each other, and a win-win situation could be achieved [48]. Gong and Xiong pointed out that knowledge sharing promotes collaborative innovation in supply chains, and in the case of centralized decision-making in supply chains, supply chains can be coordinated by promising wholesale price contracts when knowledge costs are high [49]. Zhu and Sun studied the issue

of how to make product quality decisions when automakers make decentralized and centralized decision-making, and analyzed three distribution strategies of online direct sales, traditional retail, and mixed channels [50]. By establishing models in which auto-suppliers independently carry out product innovation under dominant and nondominant conditions and obtain financing from retailers for product innovation, Yu and Li obtained the optimal decision of each subject in the supply chain under different product innovation strategies and analyzed the impact of innovation ability on profits [35]. From what has been discussed above, we can find that by suppliers, manufacturers and sellers in the car on a full range of collaborative innovation, product lifecycle each member of the supply chain can improve the innovation ability of new products, new technologies and innovation efficiency, and market marketing ability, so as to improve enterprise profit and consumer satisfaction. At the same time, each enterprise in the automotive supply chain can also share the cost of technological innovation through collaborative innovation, thereby reducing the innovation risk of a single enterprise and achieving a win-win situation for multiple parties. In consideration of the automotive supply chain sales efforts and green low-carbon aspects, Jiang studied the influence of retailer's sales effort behaviour on the decision-making of each subject in the supply chain, and found that retailer's sales effort cost and recycler's recovery difficulty were positively correlated with wholesale price, but negatively correlated with the profit of each subject in the supply chain [51]. Liu et al. studied the secondary supply chain considering consumers' green preferences and retailers' sales efforts [52]. Shang and Teng discussed product greenness, sales effort, price decision, and the profit of each member of the supply chain in three cases [53]. Jia et al. introduced two cost contracts and combined contracts, respectively, to study the decision-making and coordination problem of automobile supply chain based on bilateral sales efforts [54].

By summarizing relevant literature, we found that these studies have considered the impact of different situations and factors on the supply chain. Most of the existing scholars' studies have explored the automotive supply chain from the perspectives of consumers' green and low-carbon preference, retailers' sales efforts, and price decisions. There were few literature that included both product innovation and sales effort into the market demand function and discussed the relationship between the two factors and their impact on the supply chain. Therefore, on the basis of constructing the three-level supply chain of intelligent connected vehicles, Stackelberg game is used to study the influence of each body of the supply chain on various parameters and the profit of each body of the supply chain under centralized decision-making, nonsharing of innovation cost, and sharing of innovation cost, considering the innovation level and sales effort of products.

3. Problem Description and Model Hypotheses

3.1. Problem Description. Based on the summary of the upper, middle, and lower reaches of the ICV industry chain [15, 24, 25], this paper will discuss the three-level supply chain system of ICV. Compared with traditional fuel vehicles, the degree of electronization and informatization of intelligent connected vehicles is greatly improved, which increases the demand for automotive chips in the automotive industry. However, due to the influence of supply chain and long chip production cycle, the chips should be produced according to the demand [55]. At present, the level of innovation and research and development of chips is not enough, and there is still a gap with the international leading level. In the future, chip suppliers must also continuously improve the quality and performance of automotive chips in order to meet the future development needs of ICV [56]. At the same time, the realization of the value of smart cars also requires publicity efforts in the sales link to make consumers perceive and generate purchase intentions, which will further motivate suppliers and manufacturers to increase investment in innovative products and form a virtuous circle [57].

To sum up, the research object of this paper is a three-level supply chain system consisting of high-performance chip suppliers, car manufacturers, and retailers in the auto-trading market. Chip suppliers in the supply chain upstream intelligent chip made car production, and sales at a wholesale price to auto-vehicle manufacturers in the middle reaches of the intelligent snatched, again by the smart car manufacturers to produce sales at wholesale price for the car market retailers, finally by the retailers sold at retail prices to consumers.

For the intelligent and connected automobile industry, the quantity of chips produced by high-performance chip suppliers is mainly based on the actual demand of vehicle manufacturers, that is, production on demand. Due to the need of future development, chip suppliers are also constantly innovating to improve the quality and performance of automotive chips. As a result, the increasing cost of innovation is also inhibiting the development of advanced chips. In the main body of the supply chain, the retailer in the auto-market is the main body closest to the consumer. In the future popularization and promotion of intelligent connected vehicles, the sales effort of retailers will also affect the sales of cars. Therefore, this paper uses the Stackelberg game model to analyze the game models of the ICV supply chain under the three modes of centralized decision-making, nonsharing of innovation costs and sharing of innovation costs, taking into account the level of product innovation and sales effort. And obtain the maximum profit function and maximum profit value of each subject, and finally summarize some conclusions through analysis and comparison.

3.2. Model Hypotheses

Assumption 1. This paper adopts the Stackelberg game model. According to [49, 58, 59] mentioned in the literature, suppliers and vehicle manufacturers participate in the production mode of intelligent connected vehicles. This prioritization enables followers to infer information and

make corresponding choices based on the behaviour of the predecessors [60]. As the core component of the car in the intelligent connected car [61], the importance of the chip in the automobile manufacturing industry is becoming more and more obvious [62], and it is an indispensable part for the vehicle manufacturer to produce the car. Therefore, we set the high-performance chip supplier as the leader in the game model, and the manufacturers and retailers of ICV as the followers in this game model.

Assumption 2. According to literature [51,63], we set the market demand function as: $q = a - bp + \alpha\theta + \lambda e$, where a represents the potential demand in the intelligent connected vehicle market, b represents the sensitivity coefficient of consumers to product prices, and [44, 45], p represents the retail price of intelligent connected vehicles, α represents the innovation effect of products, and θ represents the innovation level of products. Referring to the research of Zerang et al. [64, 65], it is assumed that market demand is a linear function of price and sales effort, and then e of retailers' sales effort is introduced, and the marginal impact of sales effort on market demand is λ. Sales efforts refer to sales policies or measures taken to expand market demand. Intelligent vehicle manufacturers maximize profits by determining the optimal wholesale price, while retailers sell intelligent connected vehicles to consumers by determining the optimal retail price p and sales effort level e, thus maximizing their profits [66].

Assumption 3. The cost of innovation input of chip supplier is $\theta^2/2k_1$ [67, 68], θ represents the level of product innovation, which can be measured by the level of product technical quality. k_1 represents the innovation capability of the chip supplier. From this function model, it can be seen that the higher the innovation level of the product, the higher the innovation cost of the chip supplier, which is unfavourable to the chip supplier, because they will bear more capital expenditure [69]. Similarly, we can assume that the retailer's sales effort cost is $e^2/2\omega$ [51].

Assumption 4. When an intelligent connected vehicle manufacturer and a chip supplier adopt a cooperative innovation decision, the two innovative technical standards are the same, that is, the degree of innovation of the two is the same [70, 71].

Assumption 5. In this paper, the profit function of chip supplier is expressed by parameter π_s. Similarly, the profit function of intelligent connected vehicle manufacturers is expressed by parameter π_m. The profit function of automobile retailers is expressed by parameter π_r, and the profit function of the whole supply chain is expressed by π_c. This article discusses the profit situation of the supply chain in three decision-making situations. If the mark c above the parameter corresponding to the profit function of the main body of the supply chain represents the centralized decision model, the mark d above the parameter represents the cost nonsharing decision model, the cost-sharing model includes two models, one is that the sales effort is independent of the product innovation level, which is represented by the parameter cs. The other is that the sales effort is related to the level of product innovation, expressed by the parameter cd. For example, π_s^c represents the profit function of the chip supplier in the case of centralized decision-making, and π_r^{cs} represents the profit function of the automobile retailer when the degree of sales effort in the cost-sharing decision model is independent of the level of product innovation [72].

3.3. Model Parameters. A summary of all the parameters and their definitions involved in this article is shown in Table 1.

4. Decision Model of Intelligent Connected Vehicle

4.1. Centralized Decision-Making in the Supply Chain of Intelligent Connected Vehicle (Case C). Under the centralized decision-making model, the main bodies of the supply chain cooperate closely to make joint decisions. Taking the maximum profit of the entire supply chain as the ultimate goal of decision-making, jointly determine the market retail price p of intelligent connected vehicle, the product innovation level θ, and the sales effort level e. From the above basic assumptions, we can see that the retailer's sales effort cost is $e^2/2\omega$.

The profit maximization function of the whole intelligent connected vehicle supply chain is

$$\max_{p,\theta,e} \pi_c^c = (p-c)(a - bp + \alpha\theta + \lambda e) - \frac{e^2}{2\omega} - \frac{\theta^2}{2k_1} - \frac{\theta^2}{2k_2}. \quad (1)$$

(1) can be obtained as follows:

$$\begin{aligned}
\frac{\partial \pi_c^c}{\partial \theta} &= (p-c)\alpha - \frac{\theta}{k_1} - \frac{\theta}{k_2} \\
&= p\alpha - c\alpha - \frac{\theta}{k_1} - \frac{\theta}{k_2}, \\
\frac{\partial \pi_c^c}{\partial p} &= a - bp + \alpha\theta - pb + cb, \\
\frac{\partial^2 \pi_c^c}{\partial e} &= \lambda p - \lambda c - \frac{e}{\omega}, \\
\frac{\partial^2 \pi_c^c}{\partial p^2} &= -b - b \\
&= -2b, \\
\frac{\partial^2 \pi_c^c}{\partial \theta^2} &= -\frac{1}{k_1} - \frac{1}{k_2} \\
&= -\frac{k_2 + k_1}{k_1 k_2}, \\
\frac{\partial^2 \pi_c^c}{\partial e^2} &= -\frac{1}{\omega}.
\end{aligned} \quad (2)$$

Table 1: Parameter meaning.

Parameters	Definition
p	Intelligent connected vehicle market retail prices
q	Intelligent connected vehicle market demand
θ	Product innovation level
w_1	The wholesale price of high-performance chips
w_2	The wholesale price of intelligent connected vehicle
k_1	Chip supplier innovation capability
k_2	Innovation capacity of intelligent connected vehicle manufacturers
α	Product innovation effect
c	The unit cost of the chip produced by the chip supplier
a	Potential market demand
b	The price sensitivity of a consumer to a product
β	The proportion of innovation costs borne by carmakers
π_s	Profits of chip suppliers
π_m	Profits of intelligent connected vehicle manufacturers
π_r	Retailer's profit
π_c	Total profit in the supply chain
e	Retailer's sales effort
ω	Retailer's sales ability
λ	The marginal influence of retailer's sales effort on market demand
φ	The influence coefficient of innovation level on retailers' sales effort
Annotation	$k_1 \in [0,1]$, $k_2 \in [0,1]$, $b \in [0,1]$, $\beta \in [0,1]$, $a > 0$

The result is: $\partial^2\pi_c^c/\partial p\,\partial e = \lambda$, $\partial^2\pi_c^c/\partial p\,\partial\theta = \alpha$, $\partial^2\pi_c^c/\partial e\,\partial p = \lambda$, $\partial^2\pi_c^c/\partial e\,\partial\theta = 0$, $\partial^2\pi_c^c/\partial\theta\,\partial p = \alpha$, $\partial^2\pi_c^c/\partial\theta\,\partial e = 0$.

In conclusion, the Hessian matrix of π_c^c is

$$H = \begin{pmatrix} \frac{\partial^2\pi_c^c}{\partial p^2} & \frac{\partial^2\pi_c^c}{\partial p\,\partial e} & \frac{\partial^2\pi_c^c}{\partial p\,\partial\theta} \\ \frac{\partial^2\pi_c^c}{\partial e\,\partial p} & \frac{\partial^2\pi_c^c}{\partial e^2} & \frac{\partial^2\pi_c^c}{\partial e\,\partial\theta} \\ \frac{\partial^2\pi_c^c}{\partial\theta\,\partial p} & \frac{\partial^2\pi_c^c}{\partial\theta\,\partial e} & \frac{\partial^2\pi_c^c}{\partial\theta^2} \end{pmatrix} \tag{3}$$

$$= \begin{pmatrix} -2b & \lambda & \alpha \\ \lambda & -\frac{1}{\omega} & 0 \\ \alpha & 0 & -\frac{k_1+k_2}{k_1k_2} \end{pmatrix}.$$

When $2b - \lambda^2\omega > 0$, $\alpha^2k_1k_2 < (2b-\lambda^2\omega)(k_1+k_2)$, the determinant of the first-order subform is $H_1 = -2b < 0$, the determinant of the second-order subform is $H_2 = 2b - \lambda^2\omega/\omega > 0$, the determinant of the third-order subform is $H_3 = \alpha^2k_1k_2 - (2b-\lambda^2\omega)(k_1+k_2)/k_1k_2\omega < 0$, because the Hessian matrix is negative definite, the function has a maximum value. It can be seen that π_c^c is a concave function of p, e, θ, so π_c^c has a maximum value and there is a maximum profit.

Combining the first-order partial derivatives $\partial\pi_c^c/\partial p = 0, \partial\pi_c^c/\partial e = 0, \partial\pi_c^c/\partial\theta = 0$ of p, e, θ obtained above and solving the equation, the optimal retail price of intelligent connected vehicles in the market can be obtained as

$$p^{c*} = \frac{(a - \mathrm{bc})(k_1+k_2)}{(2b-\lambda^2\omega)(k_1+k_2) - \alpha^2k_1k_2} + c. \tag{4}$$

The retailer's optimal sales effort is

$$e^{c*} = \frac{(a-bc)(k_1+k_2)\lambda\omega}{(2b-\lambda^2\omega)(k_1+k_2) - \alpha^2k_1k_2}. \tag{5}$$

The optimal product innovation level is

$$\theta^{c*} = \frac{\alpha k_1k_2(a-bc)}{(2b-\lambda^2\omega)(k_1+k_2) - \alpha^2k_1k_2}. \tag{6}$$

Substituting $p^{c*}, e^{c*}, \theta^{c*}$ into formula (1), the maximum total profit of the intelligent connected vehicle supply chain can be obtained as

$$\pi_c^{c*} = \frac{(a-bc)^2(k_1+k_2)}{2\left[(2b-\lambda^2\omega)(k_1+k_2) - \alpha^2k_1k_2\right]}. \tag{7}$$

4.2. Decision-Making Model for Nonsharing of Innovation Costs in the Supply Chain of Intelligent Connected Vehicles (Case D). In the nonsharing decision-making model of supply chain innovation costs, intelligent connected vehicle manufacturers, chip supplier, and auto-retailers all only consider the maximization of their own interests, and all three use their own objective function of maximizing profit to make the final decision. In the Stackelberg model, chip supplier is the leader in the game.

(1) As the leader in the Stackelberg game, chip suppliers consider the principle of maximizing their own interests and determine the product innovation level θ and the wholesale price of batteries w_1. At this time, the objective function of the chip supplier's maximum profit is

$$\max_{w_1,\theta} \pi_s^d = (w_1 - c)(a - bp + \alpha\theta + \lambda e) - \frac{\theta^2}{2k_1}. \tag{8}$$

(2) As chip suppliers determine their own product innovation level, intelligent connected vehicle manufacturers also need to determine their own level of innovation based on the chip innovation level θ. The intelligent connected vehicle manufacturer here considers maximizing their own interests and decides their own wholesale price w_2. The objective function of the maximum profit of an intelligent connected vehicle manufacturer is

$$\max_{w_2} \pi_m^d = (w_2 - w_1)(a - bp + \alpha\theta + \lambda e) - \frac{\theta^2}{2k_2}. \quad (9)$$

(3) Intelligent connected vehicle retailers wholesale vehicles to intelligent connected vehicle manufacturers. The retailer will consider the degree of sales effort based on the level of product innovation, and determine its own retail price p in consideration of maximizing its own interests. The objective function of the retailer's maximum profit is

$$\max_{p,e} \pi_r^d = (p - w_2)(a - bp + \alpha\theta + \lambda e) - \frac{e^2}{2\omega}. \quad (10)$$

The entire supply chain function is

$$\pi_c^d = \pi_s^d + \pi_m^d + \pi_r^d. \quad (11)$$

According to (10),

$$\frac{\partial \pi_r^d}{\partial p} = a - 2bp + \alpha\theta + \lambda e + bw_2, \quad (12)$$

$$\frac{\partial \pi_r^d}{\partial e} = \lambda p - \lambda w_2 - \frac{e}{\omega}. \quad (13)$$

Further derivation: $\partial^2 \pi_r^d / \partial p^2 = -2b$, $\partial^2 \pi_r^d / \partial e^2 = -(1/\omega)$, $\partial^2 \pi_r^d / \partial p\ \partial e = \lambda$, $\partial^2 \pi_r^d / \partial e\ \partial p = \lambda$

The Hessian matrix of π_r^d is: $H = \begin{pmatrix} -2b & \lambda \\ \lambda & -(1/\omega) \end{pmatrix}$

When $2b - \lambda^2\omega > 0$, $|H| = 2b - \lambda^2\omega/\omega > 0$ and the determinant of the first-order subform of a matrix is $H_1 = -2b < 0$, it can be seen that π_r^d is a concave function of p and e, so π_r^d has a maximum value, and the maximum profit of the function can be obtained at the maximum value point.

Let (12) and (13) be equal to zero, that is, make the function equal to zero with respect to p and e, and obtain the two formulas together:

$$p = \frac{a + \alpha\theta - bw_2}{2b - \lambda^2\omega} + w_2, \quad (14)$$

$$e = \frac{(a + \alpha\theta - bw_2)\lambda\omega}{2b - \lambda^2\omega}. \quad (15)$$

Substituting (14) and (15) into (9), find the first-order partial derivative of w_2:

$$\frac{\partial \pi_m^d}{\partial w_2} = \frac{b(a + \alpha\theta + bw_1 - 2bw_2)}{2b - \lambda^2\omega}. \quad (16)$$

Due to $\partial^2 \pi_m^d / \partial w_2^2 < 0$, So, (9) has a maximum value. Let (16) be equal to zero, we get:

$$w_2 = \frac{a + \alpha\theta + bw_1}{2b}. \quad (17)$$

Substituting (14) (15) (17) into (8), we get:

$$\max_{w_1,\theta} \pi_s^d = \frac{b(w_1 - c)(a + \alpha\theta - bw_2)}{2(2b - \lambda^2\omega)} - \frac{\theta^2}{2k_1}. \quad (18)$$

From (18), we can see that the Hessian matrix of π_s^d is

$$H = \begin{pmatrix} -\dfrac{b^2}{2b - \lambda^2\omega} & \dfrac{\alpha b}{2(2b - \lambda^2\omega)} \\ \dfrac{\alpha b}{2(2b - \lambda^2\omega)} & -\dfrac{1}{k_1} \end{pmatrix}. \quad (19)$$

When $4(2b - \lambda^2\omega) - \alpha^2 k_1 > 0$, $|H| = b^2[4(2b - \lambda^2\omega) - \alpha^2 k_1]/2k_1(2b - \lambda^2\omega)^2 > 0$, and the determinant of the first-order subform of the matrix $H_1 = -(b^2/2b - \lambda^2\omega) < 0$, it can be seen that π_s^d is a concave function of w_1 and θ, and π_s^d has a maximum value, and the value obtained at the maximum value point is the maximum profit of this function.

Calculating the first-order partial derivatives of w_1 and θ in (19) gives:

$$\frac{\partial \pi_s^d}{\partial w_1} = \frac{b(a + \alpha\theta + bc - 2bw_1)}{2(2b - \lambda^2\omega)}, \quad (20)$$

$$\frac{\partial \pi_s^d}{\partial \theta} = \frac{\alpha b(w_1 - c)}{2(2b - \lambda^2\omega)} - \frac{\theta}{k_1}. \quad (21)$$

Let (20) and (21) be equal to 0, and the simultaneous equations can be solved:

$$w_1^{d*} = \frac{2(a - bc)(2b - \lambda^2\omega)}{b[4(2b - \lambda^2\omega) - \alpha^2 k_1]} + c, \quad (22)$$

$$\theta^{d*} = \frac{\alpha k_1 (a - bc)}{4(2b - \lambda^2\omega) - \alpha^2 k_1}. \quad (23)$$

Substituting (22) and (23) into (17), we get:

$$w_2^{d*} = \frac{3(a - bc)(2b - \lambda^2\omega)}{b[4(2b - \lambda^2\omega) - \alpha^2 k_1]} + c. \quad (24)$$

Substituting (23) and (24) into (14) and (15) can be obtained:

$$p^{d*} = \frac{b(7a + bc) - (3a + bc)\lambda^2\omega - \alpha^2 k_1 bc}{b[4(2b - \lambda^2\omega) - \alpha^2 k_1]}, \quad (25)$$

$$e^{d*} = \frac{(a - bc)\lambda\omega}{4(2b - \lambda^2\omega) - \alpha^2 k_1}. \quad (26)$$

Substituting $p^{d*}, e^{d*}, w_1^{d*}, w_2^{d*}, \theta^{d*}$ into (8), (9), (10), and (11), respectively, we get:

$$\pi_s^{d*} = \frac{(a-bc)^2}{2\left[4\left(2b-\lambda^2\omega\right)-\alpha^2k_1\right]},$$
$$\pi_m^{d*} = \frac{(a-bc)^2\left[2k_2\left(2b-\lambda^2\omega\right)-\alpha^2k_1^2\right]}{2k_2\left[4\left(2b-\lambda^2\omega\right)-\alpha^2k_1\right]^2}, \tag{27}$$

$$\pi_r^{d*} = \frac{(a-bc)^2\left(2b-\lambda^2\omega\right)}{2\left[4\left(2b-\lambda^2\omega\right)-\alpha^2k_1\right]^2}, \tag{28}$$

$$\pi_c^{d*} = \frac{(a-bc)^2\left[7k_2\left(2b-\lambda^2\omega\right)-\alpha^2k_1\left(k_1+k_2\right)\right]}{2k_2\left[4\left(2b-\lambda^2\omega\right)-\alpha^2k_1\right]^2}. \tag{29}$$

Nature 1: Under the supply chain cost-sharing decision model, the retail price of intelligent connected vehicles p, the product innovation level θ, the wholesale price of chip w_1, the wholesale price of intelligent connected vehicles w_2, and the retailer's sales effort e have no relationship between the manufacturer's innovative ability k_2.

Prove: In the supply chain cost nonsharing decision model, the above (22), (23), (24), (25), and (26) can be seen that there are no parameters in the function expression. And the first-order partial derivatives of $\partial\theta^{d*}/\partial k_2 = 0$, $\partial w_1^{d*}/\partial k_2 = 0$, $\partial w_2^{d*}/\partial k_2 = 0$, $\partial p^{d*}/\partial k_2 = 0$, $\partial e^{d*}/\partial k_2 = 0$ are all zero, so nature 1 is correct. Regardless of the innovative capabilities of intelligent connected vehicle manufacturers, it has no impact on supply chain innovation. In the intelligent connected vehicle supply chain, the innovation capabilities of chip suppliers will only have an impact on this supply chain.

4.3. Decision Model for Cost Sharing of Intelligent Connected Vehicle Supply Chain Innovation

4.3.1. The Degree of Sales Effort has Nothing to do with the Level of Product Innovation (Case CS). The chip suppliers and intelligent connected vehicle manufacturers in the main body of the supply chain cooperate to innovate together, and at the same time share the cost of innovation together. In the Stackelberg game model, the chip suppliers are still the leader of the game. Intelligent connected vehicle manufacturers take the initiative to propose innovation cooperation, assuming that the proportion of cooperative innovation costs borne by intelligent connected vehicle manufacturers is β and $\beta \in [0, 1]$, the innovation cost is $\beta(k_1+k_2)\theta^2/2k_1k_2$. After the chip suppliers and the car manufacturer reach a cooperation agreement, the proportion of the innovation cost shared by the chip suppliers is $(1-\beta)$, and the innovation input cost is $(1-\beta)(k_1+k_2)\theta^2/2k_1k_2$. Chip suppliers will determine the product innovation level θ according to the maximum benefits they can obtain, and at the same time determine the wholesale price w_1 of their batteries to intelligent connected vehicle manufacturers.

In the case of shared cost sharing between the two entities, the innovation level of the intelligent connected vehicle manufacturer in the production of intelligent connected vehicles should be determined according to the innovation level θ of the batteries. In this case, the vehicle manufacturer determines the intelligent connected vehicle's innovation level. Wholesale price is w_2. For the intelligent connected vehicle retailer in the main body of the supply chain, its decision variable is the degree of sales effort e, and the cost of the sales effort is $e^2/2\omega$. Determine your final market retail price p and sales effort level e based on the wholesale price given by the manufacturer and on the premise of guaranteeing the maximum profit. It can be concluded that the objective function of the maximum profit of the chip suppliers is

$$\max_{w_1,\theta} \pi_s^{cs} = (w_1-c)(a-bp+\alpha\theta+\lambda e) - (1-\beta)\frac{(k_1+k_2)}{2k_1k_2}\theta^2. \tag{30}$$

The objective function of the maximum profit of an intelligent connected vehicle manufacturer is

$$\max_{w_2} \pi_m^{cs} = (w_2-w_1)(a-bp+\alpha\theta+\lambda e) - \beta\frac{(k_1+k_2)}{2k_1k_2}\theta^2. \tag{31}$$

The objective function of the retailer's maximum profit in the intelligent connected vehicle market is

$$\max_{p,e} \pi_r^{cs} = (p-w_2)(a-bp+\alpha\theta+\lambda e) - \frac{e^2}{2\omega}. \tag{32}$$

The profit function of the total supply chain is

$$\pi_c{}^{cs} = \pi_s{}^{cs} + \pi_m{}^{cs} + \pi_r{}^{cs}. \tag{33}$$

Derivation of (32) can be obtained:

$$\frac{\partial\pi_r^{cs}}{\partial p} = a+\alpha\theta+\lambda e+bw_2-2bp, \tag{34}$$

$$\frac{\partial\pi_r^{cs}}{\partial e} = \lambda(p-w_2)-\frac{e}{\omega}. \tag{35}$$

And then we get: $\partial^2\pi_r^{cs}/\partial p^2 = -2b$, $\partial^2\pi_r^{cs}/\partial e^2 = -(1/\omega)$, $\partial^2\pi_r^{cs}/\partial p\,\partial e = \lambda$, and $\partial^2\pi_r^{cs}/\partial e\,\partial p = \lambda$.

The Hessian matrix of π_r^{cs} is

$$H = \begin{pmatrix} -2b & \lambda \\ \lambda & -\dfrac{1}{\omega} \end{pmatrix}. \tag{36}$$

When $2b-\lambda^2\omega > 0$, $|H| = 2b-\lambda^2\omega/\omega > 0$. The first-order subdeterminant of a matrix $H_1 = -2b < 0$, π_r^{cs} is the concave function of p and e. At this time, π_r^{cs} has a maximum value, and the function obtains the maximum profit at the maximum value point. Combine the above (34) and (35) together, set $\partial\pi_r^{cs}/\partial p = 0$ and $\partial\pi_r^{cs}/\partial e = 0$, and obtain:

$$p = \frac{a+\alpha\theta-bw_2}{2b-\lambda^2\omega} + w_2, \tag{37}$$

$$e = \frac{(a+\alpha\theta-bw_2)\omega\lambda}{2b-\lambda^2\omega}. \tag{38}$$

Substituting (37) and (38) into (31), and calculating the first-order partial derivative of w_2, we get:

$$\frac{\partial \pi_m^{cs}}{\partial w_2} = \frac{b\left(a + \alpha\theta + bw_1 - 2bw_2\right)}{2b - \lambda^2\omega}. \tag{39}$$

Because $\partial^2 \pi_m^{cs}/\partial w_2^2 < 0$. Therefore, (31) has a maximum value. Set (39) equal to 0 to solve:

$$w_2 = \frac{a + \alpha\theta + bw_1}{2b}. \tag{40}$$

In summary, substituting (37), (38), and (40) into (30) can be obtained:

$$\max_{w_1,\theta} \pi_s^{cs} = \frac{b\left(w_1 - c\right)\left(a + \alpha\theta - bw_1\right)}{2\left(2b - \lambda^2\omega\right)} - (1-\beta)\frac{k_1 + k_2}{2k_1k_2}\theta^2. \tag{41}$$

Calculating the first-order partial derivative of w_1 and θ from (41) can be obtained:

$$\frac{\partial \pi_s^{cs}}{\partial w_1} = \frac{b\left(a + \alpha\theta + bc - 2bw_1\right)}{2\left(2b - \lambda^2\omega\right)}, \tag{42}$$

$$\frac{\partial \pi_s^{cs}}{\partial \theta} = \frac{\alpha b\left(w_1 - c\right)}{2\left(2b - \lambda^2\omega\right)} - (1-\beta)\frac{k_1 + k_2}{k_1k_2}\theta. \tag{43}$$

And then we get: $\partial^2 \pi_s^{cs}/\partial w_1^2 = -(b^2/2b - \lambda^2\omega)$, $(\partial^2 \pi_s^{cs}/\partial \theta^2) = -(1-\beta)(k_1 + k_2/k_1k_2)\partial^2 \pi_s^{cs}/\partial w_1 \partial\theta = \alpha b/2(2b - \lambda^2\omega)$, and $\partial^2 \pi_s^{cs}/\partial\theta\ \partial w_1 = \alpha b/2(2b - \lambda^2\omega)$

From the above derivation, we can see that the Hessian matrix of π_s^{cs} is

$$H = \begin{pmatrix} -\dfrac{b^2}{2b - \lambda^2\omega} & \dfrac{\alpha b}{2\left(2b - \lambda^2\omega\right)} \\ \dfrac{\alpha b}{2\left(2b - \lambda^2\omega\right)} & -(1-\beta)\dfrac{k_1 + k_2}{k_1k_2} \end{pmatrix}. \tag{44}$$

When $4(1-\beta)(k_1 + k_2)(2b - \lambda^2\omega) - \alpha^2 k_1 k_2 > 0$, the determinant of the first-order subform of the matrix is $H_1 = -(b^2/2b - \lambda^2\omega) < 0$, therefore, it can be seen that π_s^{cs} is a concave function of w_1 and θ. At this time, there is a maximum value, and the maximum profit can be obtained when the function reaches the maximum value. By finding the first-order partial derivatives of w_1 and θ, and then setting (42) and (43) equal to 0, the simultaneous equations can be obtained:

$$w_1^{cs*} = \frac{2(a - bc)(1-\beta)(k_1 + k_2)\left(2b - \lambda^2\omega\right)}{b\left[4(1-\beta)(k_1 + k_2)\left(2b - \lambda^2\omega\right) - \alpha^2 k_1 k_2\right]} + c, \tag{45}$$

$$\theta^{cs*} = \frac{\alpha^2 k_1 k_2 (a - bc)}{4(1-\beta)(k_1 + k_2)\left(2b - \lambda^2\omega\right) - \alpha^2 k_1 k_2}. \tag{46}$$

Substituting (45) and (46) into (40), we can get:

$$w_2^{cs*} = \frac{3(a - bc)(1-\beta)(k_1 + k_2)\left(2b - \lambda^2\omega\right)}{b\left[4(1-\beta)(k_1 + k_2)\left(2b - \lambda^2\omega\right) - \alpha^2 k_1 k_2\right]} + c. \tag{47}$$

Substituting (47) and (46) into (37) and (38), we get:

$$p^{cs*} = \frac{b((7a + bc) - (3a + bc)\lambda^2\omega - \alpha^2 k_1 k_2 bc}{b\left[4(1-\beta)(k_1 + k_2)\left(2b - \lambda^2\omega\right) - \alpha^2 k_1 k_2\right]}, \tag{48}$$

$$e^{cs*} = \frac{(a - bc)(1-\beta)(k_1 + k_2)\lambda\omega}{4(1-\beta)(k_1 + k_2)\left(2b - \lambda^2\omega\right) - \alpha^2 k_1 k_2}. \tag{49}$$

Substituting (48), (45), (47), (49), and (46) into (30), (31), (32), and (33), the maximum profits of different entities can be obtained:

$$\pi_s^{cs*} = \frac{(a - bc)^2(1-\beta)(k_1 + k_2)}{2\left[4(1-\beta)(k_1 + k_2)\left(2b - \lambda^2\omega\right) - \alpha^2 k_1 k_2\right]}, \tag{50}$$

$$\pi_m^{cs*} = \frac{(a - bc)^2\left[2(1-\beta)^2(k_1 + k_2)^2\left(2b - \lambda^2\omega\right) - \beta\alpha^2 k_1 k_2 (k_1 + k_2)\right]}{2\left[4(1-\beta)(k_1 + k_2)\left(2b - \lambda^2\omega\right) - \alpha^2 k_1 k_2\right]^2}, \tag{51}$$

$$\pi_r^{cs*} = \frac{\left(2b - \lambda^2\omega\right)(a - bc)^2(1-\beta)^2(k_1 + k_2)^2}{2\left[4(1-\beta)(k_1 + k_2)\left(2b - \lambda^2\omega\right) - \alpha^2 k_1 k_2\right]^2}, \tag{52}$$

$$\pi_c^{cs*} = \frac{(a - bc)^2(k_1 + k_2)\left[7\left(2b - \lambda^2\omega\right)(1-\beta)^2(k_1 + k_2) - \alpha^2 k_1 k_2\right]}{2\left[4(1-\beta)(k_1 + k_2)\left(2b - \lambda^2\omega\right) - \alpha^2 k_1 k_2\right]^2}. \tag{53}$$

Nature 2: In the above three models, when the retailer's sales effort has nothing to do with the product innovation level, the retailer's sales effort e, product innovation level θ, market retail prices p, and total supply chain profits π_c all increase with the increase of retailer's sales ability.

Prove: $\partial\theta^{c*}/\partial\omega = \alpha k_1 k_2 (a - bc)(k_1 + k_2)\lambda^2/[(2b - \lambda^2\omega)(k_1 + k_2) - \alpha^2 k_1 k_2]^2 > 0$, $\partial\theta^{d*}/\partial\omega = 4\alpha k_1 (a - bc)\lambda^2/[4(2b - \lambda^2\omega) - \alpha^2 k_1]^2 > 0$, $\partial\theta^{cs*}/\partial\omega = 4\alpha k_1 k_2 (a - bc)(1 - \beta)(k_1 + k_2)\lambda^2/[4(1 - \beta)(k_1 + k_2)(2b - \lambda^2\omega) - \alpha^2 k_1 k_2]^2 > 0$. Because the first-order partial derivatives are all greater than 0, the function is monotonically increasing. Similarly, it can be proved that e, p, π_c increase with the increase in sales ability ω. When the degree of sales effort has nothing to do with the level of product innovation, no matter what, as long as the retailer improves its own sales ability, it will increase the overall profit for the supply chain. Increased retailer sales capacity will reduce the retailer's sales input costs, but strong sales capacity will increase market demand. Increased demand will increase the profits of the main supply chain entities, and chip suppliers will also improve product innovation. The improvement of the level, product performance and quality has increased the wholesale price of batteries. Eventually, retail businesses will increase the retail price of their cars to protect their profits.

4.3.2. The Degree of Sales Effort Is Related to the Level of Product Innovation (Case CD). In this case, it is assumed that the retailer's sales effort level in the automobile market increases with the improvement of the product innovation level. At this time, the retailer's sales effort level is $e = \varphi\theta$. The profit maximization function of chip suppliers, intelligent connected vehicle manufacturers, retailers, and the entire supply chain is as follows:

$$\max_{w_1,\theta} \pi_s^{cd} = (w_1 - c)(a - bp + \alpha\theta + \lambda\varphi\theta) - (1 - \beta)\frac{(k_1 + k_2)}{2k_1k_2}\theta^2, \tag{54}$$

$$\max_{w_2} \pi_m^{cd} = (w_2 - w_1)(a - bp + \alpha\theta + \lambda\varphi\theta) - \beta\frac{(k_1 + k_2)}{2k_1k_2}\theta^2, \tag{55}$$

$$\max_{p} \pi_r^{cd} = (p - w_2)(a - bp + \alpha\theta + \lambda\varphi\theta) - \frac{(\varphi\theta)^2}{2\omega}, \tag{56}$$

$$\pi_c^{cd} = (p - w_2)(a - bp + \alpha\theta + \lambda\varphi\theta) - \frac{(\varphi\theta)^2}{2\omega}. \tag{57}$$

To find the first-order partial derivative with respect to (56), we can get:

$$\frac{\partial\pi_r^{cd}}{\partial p} = a + \alpha\theta + \lambda\varphi\theta + bw_2 - 2bp. \tag{58}$$

Because $\partial^2\pi_r^{cd}/\partial p^2 < 0$, therefore (56) has a maximum value, let (58) equal to 0, and the solution is: $p = a + \alpha\theta + \lambda\varphi\theta + bw_2/2b$, substituting (55) and finding the partial derivative of, we can get:

$$\frac{\partial\pi_m^{cd}}{\partial w_2} = \frac{1}{2}(a + \alpha\theta + \lambda\varphi\theta + bw_1 - 2bw_2). \tag{59}$$

Because $\partial^2\pi_m^{cd}/\partial w_2^2 < 0$, therefore, (55) has a maximum value, let (59) equal to 0, and the solution is: $w_2 = a + \alpha\theta + \lambda\varphi\theta + bw_1/2b$, substituting it into (54), we can get:

$$\max_{w_1,\theta} \pi_s^{cd} = \frac{1}{4}(w_1 - c)(a + \alpha\theta + \lambda\varphi\theta) - (1 - \beta)\frac{(k_1 + k_2)}{2k_1k_2}\theta^2. \tag{60}$$

From (60), the Hessian matrix of π_s^{cd} is: $H = \begin{pmatrix} -(b/2) & \alpha + \lambda\varphi/4 \\ \alpha + \lambda\varphi/4 & -(1 - \beta)((k_1 + k_2)/k_1k_2) \end{pmatrix}$, $|H| = 8b(1 - \beta)(k_1 + k_2) - (\alpha + \lambda\varphi)^2 k_1k_2/16k_1k_2 > 0$, the determinant of the first-order principal and subform is $H_1 = -(b/2) < 0$. At this time, π_s^{cd} is the concave function of w_1 and θ, so π_s^{cd} has a maximum value, and the maximum profit is obtained at the maximum value point. Find the first-order partial derivative of w_1 and θ:

$$\frac{\partial\pi_s^{cd}}{\partial w_1} = \frac{1}{4}(a + \alpha\theta + \lambda\varphi\theta + bc - 2bw_1), \tag{61}$$

$$\frac{\partial\pi_s^{cd}}{\partial\theta} = \frac{(\alpha + \lambda\varphi)(w_1 - c)}{4} - (1 - \beta)\frac{(k_1 + k_2)}{k_1k_2}\theta. \tag{62}$$

Set (61) and (62) equal to 0, and solve the equations in parallel, we can solve:

$$w_1^{cd*} = \frac{4(a - bc)(1 - \beta)(k_1 + k_2)}{8b(1 - \beta)(k_1 + k_2) - (\alpha + \lambda\varphi)^2 k_1k_2} + c, \tag{63}$$

$$\theta^{cd*} = \frac{(\alpha + \lambda\varphi)k_1k_2(a - bc)}{8b(1 - \beta)(k_1 + k_2) - (\alpha + \lambda\varphi)^2 k_1k_2}, \tag{64}$$

$$e^{cd*} = \varphi\theta^{cd*} = \frac{\varphi(\alpha + \lambda\varphi)k_1k_2(a - bc)}{8b(1 - \beta)(k_1 + k_2) - (\alpha + \lambda\varphi)^2 k_1k_2}. \tag{65}$$

Substituting (63) and (64) into w_2, we get:

$$w_2^{cd*} = \frac{6(a - bc)(1 - \beta)(k_1 + k_2)}{8b(1 - \beta)(k_1 + k_2) - (\alpha + \lambda\varphi)^2 k_1k_2} + c. \tag{66}$$

Substituting (63) and (64) into p, we get:

$$p^{cd*} = \frac{7(a-bc)(1-\beta)(k_1+k_2)}{8b(1-\beta)(k_1+k_2)-(\alpha+\lambda\varphi)^2 k_1 k_2} + c. \tag{67}$$

Substituting (67), (63), (66), and (64) into (54), (55) (56), (57), respectively, we can get:

$$\begin{aligned}
\pi_s^{cd*} &= \frac{(a-bc)^2(1-\beta)(k_1+k_2)}{2\left[8b(1-\beta)(k_1+k_2)-(\alpha+\lambda\varphi)^2 k_1 k_2\right]}, \\
\pi_m^{cd*} &= \frac{(a-bc)^2\left[4b(1-\beta)^2(k_1+k_2)^2-\beta(\alpha+\lambda\varphi)^2 k_1 k_2(k_1+k_2)\right]}{2\left[8b(1-\beta)(k_1+k_2)-(\alpha+\lambda\varphi)^2 k_1 k_2\right]^2},
\end{aligned} \tag{68}$$

$$\pi_r^{cd*} = \frac{(a-bc)^2\left[2b\omega(1-\beta)^2(k_1+k_2)^2-\varphi^2(\alpha+\lambda\varphi)^2 k_1^2 k_2^2\right]}{2\omega\left[8b(1-\beta)(k_1+k_2)-(\alpha+\lambda\varphi)^2 k_1 k_2\right]^2}, \tag{69}$$

$$\pi_c^{cd*} = \frac{(a-bc)^2\left[14b\omega(1-\beta)^2(k_1+k_2)^2-\varphi^2(\alpha+\lambda\varphi)^2 k_1^2 k_2^2-\omega(\alpha+\lambda\varphi)^2(k_1+k_2)k_1 k_2\right]}{2\omega\left[8b(1-\beta)(k_1+k_2)-(\alpha+\lambda\varphi)^2 k_1 k_2\right]^2}. \tag{70}$$

Nature 3: In the cost-sharing decision-making model of the main body of the supply chain, when the degree of sales effort is related to the innovation level of the product, the degree of sales effort e, the level of product innovation θ, and the retail price of intelligent connected vehicles in the market p have nothing to do with the retailer's sales ability ω, the total profit of the supply chain increases as the retailer's sales capacity increases.

Prove: Find the first-order partial derivative with respect to ω for $\theta^{cd*}, e^{cd*}, p^{cd*}$, we can get: $\partial\theta^{cd*}/\partial\omega = 0$, $\partial e^{cd*}/\partial\omega = 0$, $\partial p^{cd*}/\partial\omega = 0$, this shows that the level of product innovation, the degree of sales effort, and the retail price of the intelligent connected vehicle market have nothing to do with the retailer's sales ability.

$\partial\pi_c^{cd*}/\partial\omega = (a-bc)^2\varphi^2(\alpha+\lambda\varphi)^2 k_1^2 k_2^2\omega^2/2[8b(1-\beta)(k_1+k_2)-(\alpha+\lambda\varphi)^2 k_1 k_2]^2 > 0$, it can be seen that the first-order partial derivative of the function is greater than 0, and the function is monotonically increasing, so the total profit of the supply chain increases with the increase of the retailer's sales ability. In the supply chain cost-sharing decision-making model, when the retailer's sales effort is related to the product innovation level, the retailer's sales ability ω can only affect the retailer's sales input cost $e^2/2\omega$, and has no effect on other parameters. Because in the game model, the chip supplier is still the leader of the game and determines the level of product innovation. For retailers, the stronger the sales ability ω, the lower the sales input cost, the weaker the sales ability, and the higher the sales input cost. When the sales input cost is low, the maximum profit that the retailer can obtain will increase, and the overall profit of the supply chain will also increase.

5. Numerical Simulation

In order to better verify the change relationship between the various parameters, this section assigns values to each parameter, and compares and analysis various indicators under different models, to more clearly express and verify the properties described above. Refer to the previously related literature and set the values of fixed parameters. These values are as follows: $a = 50$, $b = 1$, $c = 5$, $k_1 = 0.6$, $k_2 = 0.4$, $\alpha = 0.8$, $\lambda = 0.4$, $\omega = 0.4$, $\varphi = 1$, $\beta = 0.7$ [72–74]. In the drawings in this section, the centralized decision-making model is represented by model C, and the cost-sharing model is represented by model D. There are two situations in the cost-sharing model. The first is the case where the degree of sales effort has nothing to do with the level of innovation, using model CS Representative, the second is the situation where the degree of sales effort is related to the level of innovation is represented by the model CD.

5.1. The Impact of Auto-Retailers' Sales Capabilities on the Supply Chain. As can be seen from Figure 1, model C, model D, and model CS, as the retailer's sales ability improves, the degree of sales effort has also become an upward trend. In the case of model CD, the retailer's sales effort has nothing to do with sales ability, but it is much greater than the retailer's sales effort when the sales effort is not related to the innovation level. Regardless of sales ability, if auto-retailers want to achieve a higher degree of sales effort, they should sell based on the level of product innovation.

It can be seen from Figure 2 that in the case of Model C, Model D, and Model CS, the innovation level of products increases with the increase of retailer's sales ability. In the model CD, the product innovation level has nothing to do with the retailer's sales ability, and it is a horizontal straight line. However, the product innovation level is higher than the other three cases at this time, because the retailer's decision variable is not the sales ability but the chip supplier's product innovation level, and the chip suppliers does not consider the retailer's investment when determining the product's innovation level how much sales cost and profit. If

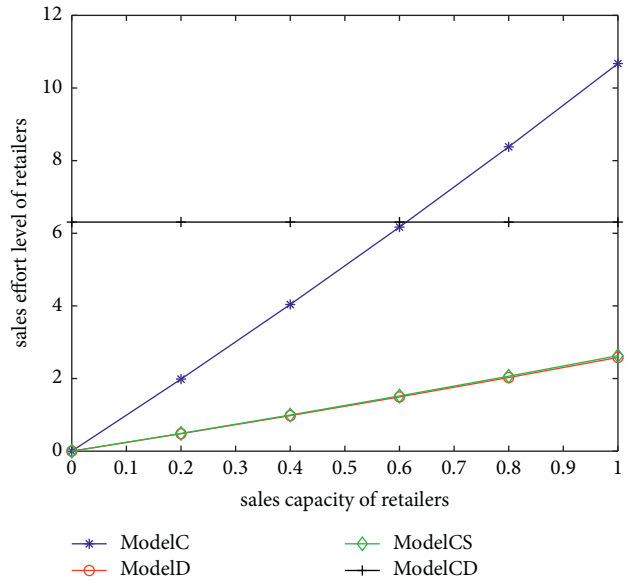


Figure 1: Sales effort level of retailers with the changing trend of sales capacity.

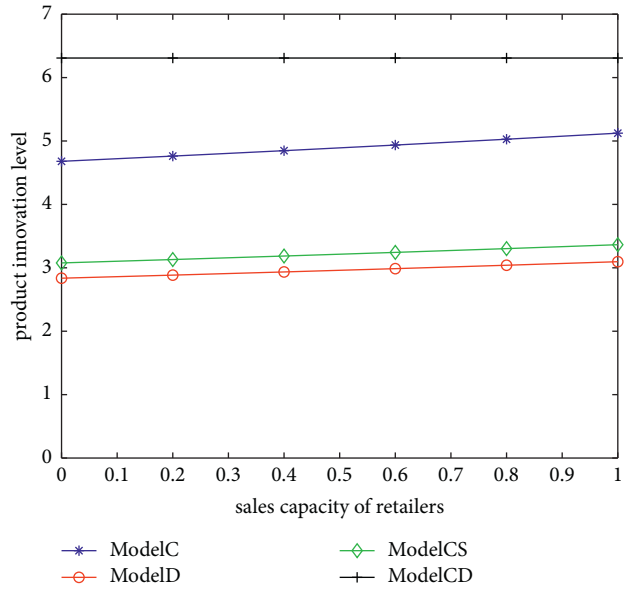


Figure 2: Product innovation level with the changing trend of sales capacity.

the overall profit of the supply chain of intelligent connected vehicles is improved, automobile retailers with stronger sales capacity should be selected.

From (28), (52), and (69), the relationship between the profit of the automobile retailer and the retailer's sales ability is obtained. In the three models in Figure 3, we can see that the retailer's profit level increases with the improvement of sales capabilities. By comparing Model D and Model CS, when the degree of sales effort has nothing to do with the level of product innovation, the cost sharing of each main body of the supply chain is more profitable than when the cost is not shared. This is because when chip suppliers in the

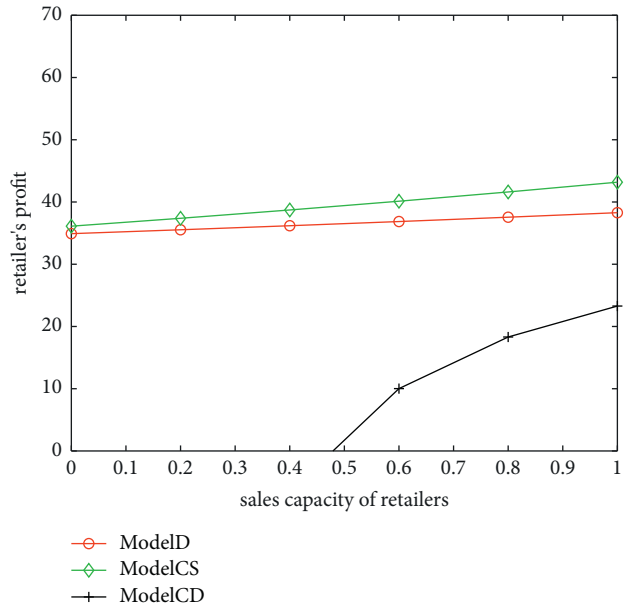


Figure 3: Retailer's profit with the changing trend of sales capacity.

supply chain cooperate with intelligent connected vehicle manufacturers to share the cost of innovation, the innovation level of batteries will increase, and the performance improvement of batteries will promote the increase in market demand. At this time, retailer's higher profits can also be obtained, so the profit of the retailer is higher when the cost is shared than when the cost is not shared. It can be seen from the model CD that when the auto-retailer is weak in its own sales ability, the retailer's profit will still show a negative value. With the improvement of the sales ability, the retailer's profit will increase, but the magnitude is small. Because the degree of sales effort is related to product innovation at this time, chip suppliers will vigorously promote the improvement of product innovation under the cost-sharing model, but will not consider the relevant investment and profit level of the retailer. The retailer is costly due to a large amount of sales effort. As a result, the profit level is not high.

5.2. The Impact of the Proportion of Innovation Costs Borne by Automakers in the Supply Chain Innovation Cost Sharing Model. From (8), (23), (46), and (64), the relationship between the level of product innovation and the proportion of innovation costs borne by automakers is obtained. As shown in Figure 4, the model CS and model CD can be seen that regardless of whether the degree of sales effort is related to the innovation level of the product, the innovation level of the product increases with the increase in the proportion of the innovation cost borne by the automaker, and the two are directly proportional, the model CS is always higher than the model CD. It shows that auto-retailers determine their own marketing efforts based on the level of product innovation, which is more conducive to the improvement of product innovation level. When automakers bear a greater proportion of innovation costs, the cost-sharing model CS and

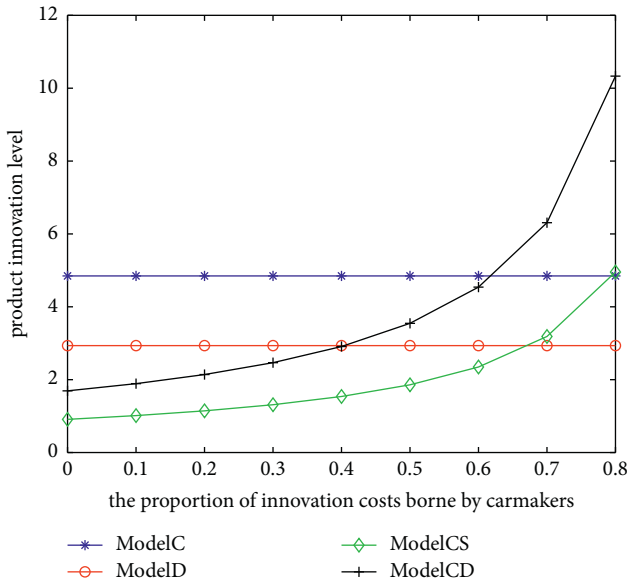


Figure 4: Innovation level changes with the proportion of innovation cost borne by manufacturers.

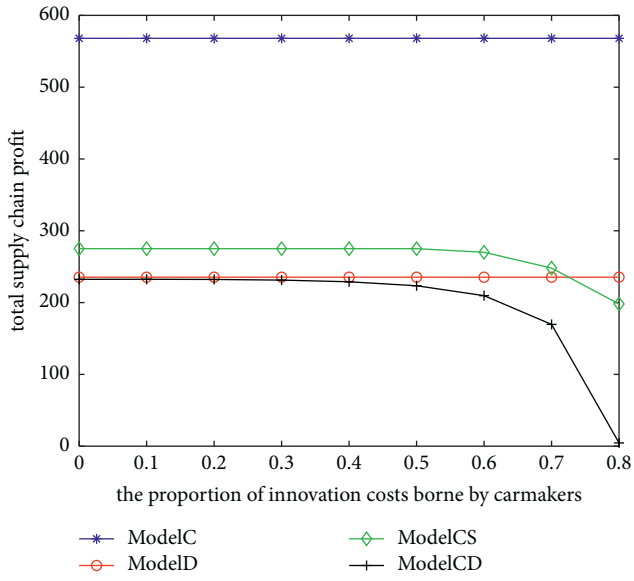


Figure 5: Total supply chain profit changes with the proportion of innovation cost borne.

model CD even surpass the centralized decision-making model C and the non-cost-sharing model D.

From (7), (29), (53), and (70), the relationship between the total profit of the supply chain and the proportion of innovation costs borne by automakers is obtained. In the case of the main body of the supply chain sharing costs, namely the model CS and the model CD, as the proportion of automakers' innovation costs increases, (Figure 5 the total profit of the supply chain gradually decreases. When the retailer's sales effort is not related to the level of product innovation, the total profit of the supply chain is greater than the situation when the cost is not shared, indicating that the total profit of the supply chain has also increased after the introduction of retail sales efforts. When the degree of sales effort is related to the level of product innovation, it can be seen that the total profit of the supply chain is lower than the total profit of the supply chain of the non-cost-sharing model.

Because of the excessively high cost of sales and product innovation costs, the profit of each main body of the supply chain has decreased, which has led to a decrease in the total profit of the supply chain.

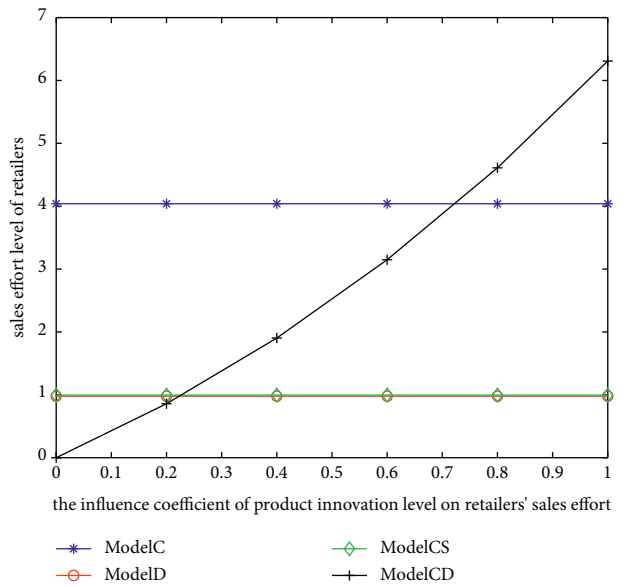


Figure 6: Sales effort level of retailers changes with the influence coefficient of sales effort.

5.3. The Influence of the Influence Coefficient of Retailer's Sales Effort on Supply Chain Decision-Making. From (5), (26), (49), and (65), the relationship between the in fluence factor φ of product innovation level on retailer's marketing effort and the degree of sales effort is obtained. When the retailer's sales effort is correlated with the product innovation level, it can be seen in Figure 6 that the sales ability level increases with the increase. When the value of φ is very small, since the level of product innovation has little impact on the retailer's marketing efforts, the auto-retailer will not invest too much marketing cost at this time, so the degree of sales effort is less than the other three cases. When the value of φ increases, the product innovation level at this time has a great impact on the retailer's marketing efforts. In model CD, the chip supplier and the car manufacturer share the cost of innovation, and the product innovation level is high, and the retailer. The degree of marketing effort will also increase and be greater than the case of model D. When the value increases to a certain value, the degree of marketing effort exceeds that of model C.

From (6), (23), (46), and (64), the relationship between the product innovation level and the influence factor φ of the product innovation level on the retailer's marketing efforts can be obtained as shown in Figure 7. When the retailer's sales effort is related to the product innovation level, the

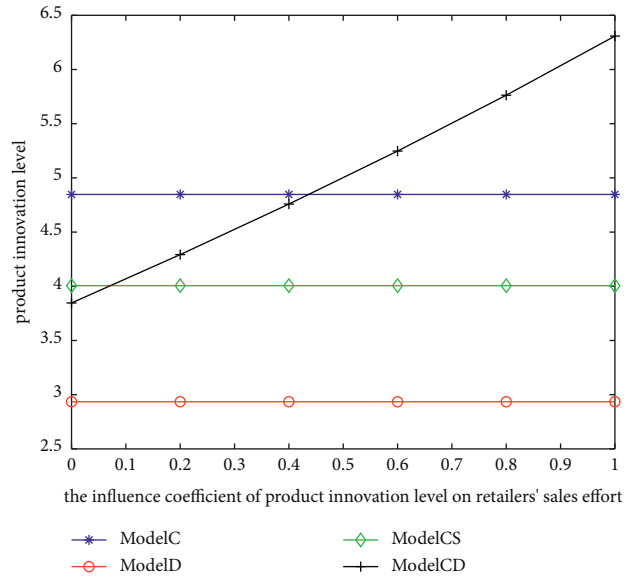


Figure 7: Product innovation level changes with the influence coefficient of sales effort.

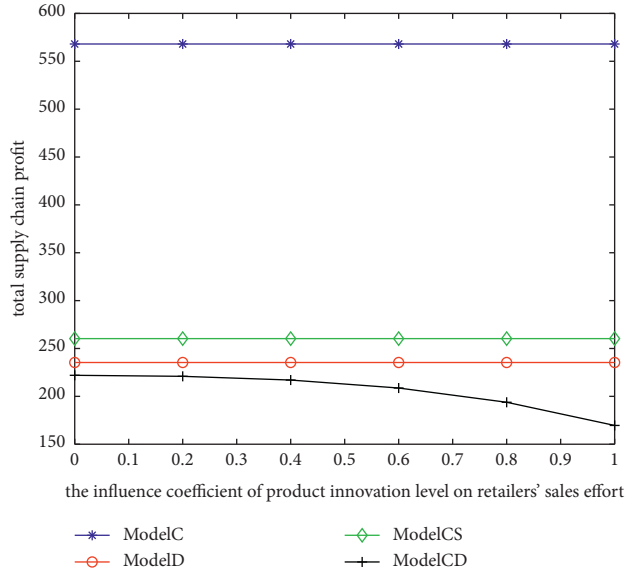


Figure 8: Total supply chain profit changes with the influence coefficient of sales effort.

product innovation level increases with the increase of φ, that is, the product innovation level and the product innovation level have a positive relationship with the retailer's influence factor φ on the retailer's marketing effort. When all entities in the supply chain share the cost of innovation, the product innovation level of the two cases is greater than the case of no cost sharing.

It can be seen from Figure 8 that when the retailer's sales effort is related to the product innovation level, there is an inverse relationship between the total profit of the supply chain and the product innovation level's impact factor φ on the retailer's marketing effort. As φ increases, the supply chain total profit gradually decreases and is less than the total profit of the supply chain in the other three situations. Because the retailer's sales effort is related to the chip supplier's product innovation level, the greater the impact factor φ of the innovation level on the marketing effort, the greater the sales effort will lead to an increase in demand in the automotive market. To obtain more profits, chip suppliers will further increase their investment in product innovation costs. The larger the impact factor φ, the more cost input of each main body of the supply chain. When the growth rate of cost input is greater than the rate of profit increase, the overall profit of the supply chain will become a downward trend. The analysis shows that from the perspective of the total profit of the supply chain, the retailer's sales effort should not be determined by the level of product innovation. But from the perspective of product innovation, if you want to improve the performance and functions of batteries, retailers can choose this situation to sell.

6. Conclusion

In this article, the Stackelberg game model is used to form a three-tier supply chain for chip suppliers, intelligent connected vehicle manufacturers, and retailers in the automotive market. They are in the centralized decision model, the supply chain cost nonsharing model, and the supply chain cost-sharing model. Under the two marketing methods of the model, the influence of retailer's sales effort and chip product innovation level on the supply chain is especially considered, and the model is simulated and analyzed to study the relationship between various parameters. As well as the profit situation of each main body of the supply chain, the following conclusions are obtained.

(1) When automakers and chip suppliers share the cost of innovation, and the retailer's sales effort has nothing to do with the product innovation level, the product innovation level and the total profit of the supply chain are greater than the innovation in the non-cost-sharing model It is beneficial for the supply chain to share the cost of innovation by all parties in the supply chain. In this case, the improvement of auto-retailer's sales ability can further improve the product innovation level, the retailer's sales effort, and the retailer's profit.

(2) When the automaker and the chip supplier share the cost of innovation, and the retailer's sales effort is related to the product innovation level, the product innovation level is greater than the case where the cost is not shared, but the total profit of the supply chain is less than the cost without sharing the situation, and chip suppliers will not choose to cooperate with car manufacturers. In this case, the level of product innovation and marketing effort has nothing to do with the retailer's sales ability. The total profit of the intelligent connected vehicle supply chain increases with the increase in the capacity of retailers and sellers.

(3) When the degree of sales effort is related to the level of product innovation, the retailer's sales effort and

product innovation level increase with the increase of the product innovation level's influence factor φ on the retailer's marketing effort, forming a positive correlation. The total profit of the supply chain decreases as the innovation level increases the retailer's marketing effort factor φ, which is negatively correlated.

Based on the above research conclusions, the following management enlightenment is put forward.

For chip suppliers and automakers, according to their respective development goals, retailers should be guided to choose to improve the level of product innovation to obtain innovation benefits, establish corporate image, or pursue higher profits. If chip suppliers and automobile manufacturers want a higher level of product innovation, they should choose retailers who are willing to make marketing efforts based on the level of product innovation; if the profit of the supply chain is considered, the retailer with stronger marketing ability should be selected.

For retailers in the automotive market, marketing activities have allowed consumers to have more knowledge about the use of intelligent connected vehicles and the performance and advantages of innovative products. Different levels of sales effort have different effects on the automotive market demand. When determining the level of sales effort, retailers should comprehensively consider the impact on the level of product innovation and supply chain profits. If you want to improve your own sales efforts and quickly improve product reputation in a short time, regardless of the retailer's sales capabilities, a series of sales activities should be carried out based on the innovation level of intelligent connected vehicles products. But if you consider the profitability of the supply chain when retailers choose to determine their own sales efforts, they need to improve their sales capabilities. For example, they need to promote more information about intelligent connected vehicles, increase consumers' environmental awareness, and guide consumers to buy green, low-carbon and environmentally friendly travel tools, and also increase the sales of green products in the entire automotive market.

There are several limitations in this research. Firstly, this article does not consider the impact of government subsidies and tax incentives on the supply chain of intelligent connected vehicles, future research can explore how to achieve a win-win cooperation between all parties in the intelligent connected vehicles supply chain under the effect of government subsidies and tax preferential policies and other external incentives, and further improve the level of product innovation. Secondly, the supply chain studied in this article only considers the situation that is dominated by chip suppliers. In reality, there are also situations in which automobile manufacturers dominate, and this aspect can be continued in the future [75].

Data Availability

The original contributions presented in the study are included within the article and can be obtained from the corresponding author.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

The authors are grateful to all the foundations that support us. This research was funded by Shandong Provincial Key Research and Development Program (Soft Science) (2020RKE28013) and Qingdao Social Science Planning Research Project (QDSKL2101039).

References

[1] Y. P. Wang and X. P. Yan, *Probability of Intelligent Transportation Technology"*, Shanghai People's Publishing House, Shanghai, China, 2017.

[2] Y. Y. Wang, "Analysis of smart city concept and future urban development," *Urban Housing*, vol. 27, no. 11, pp. 128-129, 2020.

[3] Z. W. Liu, H. K. Song, and F. Q. Zhao, "Research on the innovative development strategy of a new generation of intelligent vehicles based on 4S integration," *Chinese Engineering Science*, vol. 23, no. 03, pp. 153–162, 2021.

[4] A. Pareek, R. Dom, and J. Gupta, "Insights into renewable hydrogen energy: recent advances and prospects," *Materials Science for Energy Technologies*, vol. 3, pp. 319–327, 2020.

[5] Y. Z. Chen and J. D. Liu, "Industrial development and support policies of hydrogen energy for vehicles in Japan," *Modern Japanese Economy*, vol. 02, pp. 80–94, 2021.

[6] C. L. Feng, "The development dilemma and coping strategies of my country's intelligent networked automobile industry," *Contemporary Economic Management*, vol. 40, no. 05, pp. 55–58, 2018.

[7] X. F. He and Y. Y. He, "Blockchain application in modern logistics information sharing: a review and case study analysis," *Production Planning & Control*, vol. 33, pp. 1–15, 2022.

[8] Q. Lu and X. Z. Zhang, "Talking about the development of intelligent networked vehicles in my country," *Chinese Educational Technology Equipment*, vol. 11, pp. 22–24+27, 2020.

[9] M. Y. Li and K. Q. Li, "Strategic analysis on establishing an automobile power in China based on intelligent connected vehicles," *Chinese Journal of Engineering Science*, vol. 20, no. 1, pp. 52–58, 2018.

[10] Y. Q. Liu, R. Xi, and F. Zhou, "Research on the optimization path of technological innovation in the intelligent networked vehicle industry: an empirical analysis based on DEMATEL and system dynamics model," *Soft Science*, vol. 35, no. 09, pp. 37–46, 2021.

[11] X. Xu, "Value creation for intelligent connected vehicles: an industry value-chain perspective," *Digital Business Models*, vol. 51, no. 02, pp. 25–29, 2019.

[12] Y. Q. Liu, F. Zhou, and R. Xi, "Technology research and cooperation network of China's intelligent connected vehicle industry in post-epidemic period: international patent perspective," *China Science and Technology Forum*, vol. 05, pp. 32–45, 2021.

[13] R. Yu, J. L. Wu, and S. T. Shi, "Research on the influencing factors of consumers' willingness to buy new energy vehicles under the third wave of car manufacturing peak," *Modernization of Shopping Malls*, vol. 06, pp. 18–20, 2022.

[14] Y. F. Zhang, L. Xu, and A. Q. Zhao, "Can consumer brand knowledge improve the purchase intention of new energy vehicles—a dual-path adjustment-oriented research," *Modern Marketing*, vol. 1, pp. 89–93, 2022.

[15] J. Z. Li, Y. P. Zhou, and D. K. Yu, "Consumers' purchase intention of new energy vehicles: do product-life-cycle policy portfolios matter?" *Sustainability*, vol. 12, no. 05, p. 66, 2022.

[16] G. Cecere, N. Guerzoni, and M. Guerzoni, "Price or performance? A probabilistic choice analysis of the intention to buy electric vehicles in European countries," *Energy Policy*, vol. 118, pp. 19–32, 2018.

[17] O. Egbue, S. Samaranayake, and V. A. Samaranayake, "Mass deployment of sustainable transportation: evaluation of factors that influence electric vehicle adoption," *Clean Technologies and Environmental Policy*, vol. 19, no. 7, pp. 1927–1939, 2017.

[18] F. Q. Zhao, H. Tan, and Z. W. Liu, "Analysis of the business models of the intelligent and connected vehicle industry," *MATEC Web of Conferences*, vol. 28, no. 03, pp. 325–329, 2020.

[19] J. Li, S. B. Qiu, and H. J. Li, "Smart cars in smart cities," *Science China Information Sciences*, vol. 46, no. 05, pp. 551–559, 2016.

[20] W. Miao, "Six key tasks for the development of intelligent networked vehicles," *Auto Vertical and Horizontal*, vol. 7, no. 07, pp. 21–23, 2017.

[21] M. Y. Bian and K. Q. Li, "Top-level design of automobile power strategy based on intelligent networked vehicles," *China Engineering Science*, vol. 20, no. 01, pp. 52–58, 2018.

[22] D. M. Li, L. B. Cai, and Z. M. Cai, "Intelligent vehicle network system and smart city management based on genetic algorithms and image perception," *Mechanical Systems and Signal Processing*, vol. 141, Article ID 106623, 2020.

[23] T. J. Kim, "Automated autonomous vehicles: prospects and impacts on society," *Journal of Transportation Technologies*, vol. 08, no. 03, pp. 137–150, 2018.

[24] M. J. Chen, "The development status and trend of intelligent connected vehicle (ICV) technology," *Auto Interiors*, vol. 22, pp. 8-9, 2021.

[25] J. M. Wang, C. Wei Chen, and Y. L. Xu, "Analysis of the competitive situation of global intelligent networked vehicle technology from the perspective of patents," *Automotive Technology*, vol. 08, pp. 20–29, 2021.

[26] X. Hu, H. Y. Xie, and P. C. Zhao, "Research on the development situation of intelligent connected vehicle industry," *Auto Interiors*, vol. 18, pp. 135–137, 2019.

[27] Q. H. Zeng, "Suggestions on accelerating the development of intelligent networked new energy vehicles," *China Science and Technology Industry*, vol. 02, pp. 10-11, 2022.

[28] Y. H. Jin, "Research on the upgrading of my country's new energy vehicle industry chain in the period of new development opportunities," *Economics*, vol. 01, pp. 38–90, 2022.

[29] C. R. Shen and L. Z. Li, "Research on automotive supply chain decision-making considering government subsidies under the background of new energy," *Practice and Understanding of Mathematics*, vol. 49, no. 08, pp. 79–88, 2019.

[30] X. F. Xu, J. Hao, and Y. Zheng, "Multi-objective artificial bee colony algorithm for multi-stage resource leveling problem in sharing logistics network," *Computers & Industrial Engineering*, vol. 142, no. 4, Article ID 106338, 2020.

[31] Q. J. Ju, P. Ju, and W. Q. Dai, "Research on the impact of new energy vehicle subsidy policy and ownership: unit subsidies, sales incentives and product differentiation," *Journal of Management Science*, vol. 24, no. 06, pp. 101–116, 2021.

[32] J. B. Sheu and Y. Chen, "Impact of government financial intervention on competition among green supply chains," *International Journal of Production Economics*, vol. 138, no. 1, pp. 201–213, 2012.

[33] C. Che, X. L. Qi, and W. Q. Ma, "An empirical study on the influencing factors of mobile social network marketing effectiveness," *Chinese Journal of Management Science*, vol. 25, no. 05, pp. 145–149, 2017.

[34] C. Lu, Q. Q. Wang, and Q. Chen, "Automobile supply chain coordination considering price, emission reduction and endurance under the "double points" policy," *Systems Engineering Theory and Practice*, vol. 41, no. 10, pp. 2595–2608, 2021.

[35] H. Yu and Y. X. Li, "Supplier's product innovation strategy choice from the perspective of supply chain: financing VS No financing," *Journal of Management in Engineering*, vol. 35, no. 03, pp. 172–180, 2021.

[36] C. Che, W. Q. Ma, and S. F. Cao, "Research on time distance, social distance and the effect of online shopping decision framework," *Commercial Research*, vol. 58, no. 09, pp. 13–136, 2015.

[37] J. H. Tang, F. Yang, and J. W. Xu, "Research on the production and pricing of the secondary automobile supply chain considering consumer preferences under the dual-point system," *Industrial Engineering & Management*, vol. 26, no. 01, pp. 121–129, 2021.

[38] X. F. Xu, C. L. Wang, and P. Zhou, "GVRP considered oil-gas recovery in refined oil distribution: from an environmental perspective," *International Journal of Production Economics*, vol. 235, Article ID 108078, 2021.

[39] K. Kang and S. Y. Zhang, "Research on the game of electric vehicle manufacturers based on the influence of double-point policy," *Research on Technology Economics and Management*, vol. 04, pp. 59–66, 2022.

[40] Y. Wang, "Research on R&D strategy of automobile supply chain considering market entry method under double points policy," *Soft Science*, vol. 35, no. 09, pp. 28–36, 2021.

[41] L. Ma, Y. J. Liu, and H. Zhu, "Design of dual-channel recycling contract for new energy vehicle batteries from the perspective of closed-loop supply chain," *Technology Management Research*, vol. 41, no. 20, pp. 184–193, 2018.

[42] Z. G. Qiu, Y. Zheng, and Y. Q. Xu, "New energy vehicle power battery closed-loop supply chain recovery subsidy strategy——based on evolutionary game analysis," *Business Research*, vol. 08, pp. 28–36, 2020.

[43] K. Covindan, M. N. Popiuc, and A. Diabat, "Overview of coordination contracts within forward and reverse supply chains," *Journal of Cleaner Production*, vol. 47, pp. 319–334, 2013.

[44] C. Che, X. G. Zhang, and Y. Chen, "A model of waste price in a symbiotic supply chain based on Stackelberg algorithm," *Sustainability*, vol. 13, no. 4, pp. 17–40, 2021.

[45] J. P. Xie, L. Liang, and Y. Y. Li, "Research on revenue sharing contract mechanism strategy in closed-loop supply chain," *Journal of Management in Engineering*, vol. 31, no. 02, pp. 185–193, 2017.

[46] C. Che, Y. Chen, and X. G. Zhang, "Study on emission reduction strategies of dual-channel supply chain considering green finance," *Frontiers in Environmental Science*, vol. 09, pp. 197–203, 2021.

[47] Y. G. Jing, P. Q. Guo, and K. D. Qin, "New energy vehicle leasing system coordination contract whose demand rate is affected by the level of promotion effort," *Chinese Management Science*, vol. 26, no. 03, pp. 94–100, 2018.

[48] J. Zhang, Z. Y. Wang, and F. F. Yu, "New product technology innovation mode selection in vertical supply chain," *Journal of Management*, vol. 17, no. 11, pp. 1697–1705, 2020.
[49] Q. G. Gong and Q. Xiong, "Research on collaborative innovation management of automotive supply chain based on knowledge sharing," *Science, Technology and Development*, vol. 15, no. 05, pp. 464–473, 2019.
[50] L. L. Zhu and S. H. Sun, "Research on product quality decision-making of supply chain distribution channels under innovation investment," *Management Review*, vol. 31, no. 5, p. 231, 2019.
[51] M. T. Jiang, "The impact of retailers' sales efforts on closed-loop supply chain decision-making: taking the supply chain of new energy vehicles as an example," *Times Economic and Trade*, vol. 19, no. 01, pp. 17–21, 2022.
[52] G. D. Liu, T. J. Yang, and X. M. Zhang, "Green supply chain pricing and product green degree decisions considering green preference and sales efforts," *Journal of Beijing Jiaotong University*, vol. 18, no. 01, pp. 115–126, 2019.
[53] W. F. Shang and L. L. Teng, "Retailer-led green supply chain game strategy considering government subsidies and sales efforts," *Systems Engineering*, vol. 38, no. 02, pp. 40–50, 2020.
[54] C. J. Jia, J. J. Zhou, and M. Shang, "Decision-making and coordination of automobile supply chain considering bilateral sales efforts under O2O mode," *Logistics Technology*, vol. 40, no. 10, pp. 76–83, 2021.
[55] C. Bai, "Chip chip! Hurry up," *Auto Technician*, vol. 19, no. 08, pp. 22–24+29, 2021.
[56] H. Feng and Z. Q. Xiong, "Opportunities and development strategies of new energy vehicle industry under the background of dual carbon," *Decision and Information*, vol. 05, pp. 68–76, 2022.
[57] J. H. Liu, Y. Huang, and F. Wang, "Choice of three-level low-carbon supply chain cost-sharing contract based on joint emission reduction and sales efforts," *Business Research*, vol. 01, pp. 123–132, 2022.
[58] Y. Y. Yi and J. Chen, "Research on the optimal pricing of trade-in in automobile supply chain," *Systems Engineering Theory and Practice*, vol. 42, no. 4, pp. 1072–1085, 2022.
[59] W. B. Wang, J. Lv, and M. Y. Zhang, "Research on resource investment strategy of new energy vehicle supply chain under cooperation mode," *Industrial Technology and Economics*, vol. 40, no. 10, pp. 33–41, 2021.
[60] J. Z. Li, C. L. Liu, and G. N. Cai, "Analysis of cooperative decision-making of inter-chain dynamic game in cluster supply chain," *Chinese Journal of Management Engineering*, vol. 04, pp. 20–24, 2006.
[61] Z. B. Hui, "The coming era of intelligent networked vehicles," *Scientific Chinese*, vol. 17, pp. 63–65, 2019.
[62] Y. Tian, "How far is the realization of "autonomous and controllable" automotive chips?" *Intelligent Connected Vehicles*, vol. 01, pp. 42–45, 2021.
[63] Y. Yin and J. C. Liu, "Research on new energy vehicle supply chain pricing strategy based on game theory and particle swarm optimization," *Scientific Journal of Intelligent Systems Research*, vol. 3, no. 4, p. 46, 2021.
[64] E. Sane Zerang, A. A. Taleizadeh, J. Razmi, S. Zerang, A. Taleizadeh, and J. Razmi, "Analytical comparisons in a three-echelon closed-loop supply chain with price and marketing effort-dependent demand: game theory approaches," *Environment, Development and Sustainability*, vol. 20, no. 1, pp. 451–478, 2018.
[65] A. A. Taleizadeh, E. Sane-Zerang, and T. Choi, "The effect of marketing effort on dual-channel closed-loop supply chain systems," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 48, no. 2, pp. 265–276, 2016.
[66] X. L. Huang and M. X. Hong, "Fuzzy supply chain game dominated by retailers—considering the situation of sales efforts," *Operations Management*, vol. 29, no. 01, pp. 57–68, 2020.
[67] Z. Y. Li, X. D. Qi, and Y. Liu, "An empirical study on the effect of government R&D subsidies in the new energy vehicle industry," *Journal of Northeastern University*, vol. 19, no. 04, pp. 356–363, 2017.
[68] B. L. Shao and L. J. Wang, "Research on supplier innovation investment considering fairness preference under different contracts," *Industrial Engineering*, vol. 24, no. 02, pp. 85–91, 2021.
[69] J. H. Gao, H. S. Han, and L. T. Hou, "Research on retailer-led closed-loop supply chain decision-making considering product greenness and sales effort," *Management Review*, vol. 27, no. 04, pp. 187–196, 2015.
[70] T. Luomaranta and M. Martinsuo, "Supply chain innovations for additive manufacturing," *International Journal of Physical Distribution & Logistics Management*, vol. 50, no. 1, pp. 54–79, 2016.
[71] J. G. Wang, J. H. Chang, and Y. C. Wu, "The optimal production decision of competing supply chains when considering green degree: a game-theoretic approach," *Sustainability*, vol. 12, no. 18, p. 7413, 2020.
[72] Z. W. Kong, J. P. Xie, and W. S. Zhang, "New energy vehicle battery R&D cost sharing contract optimization based on quality sensitivity," *Shanghai Management Science*, vol. 40, no. 02, pp. 108–114, 2018.
[73] F. L. Yin, X. M. Xue, and C. Z. Zhang, "Multifidelity genetic transfer: an efficient framework for production optimization," *SPE Journal*, vol. 26, no. 04, pp. 1614–1635, 2021.
[74] J. P. Xie, J. Li, and F. F. Yang, "Multi-level contract decision optimization for closed-loop supply chain of new energy vehicles," *Journal of Management in Engineering*, vol. 34, no. 02, pp. 180–193, 2020.
[75] K. Zhang, J. D. Ma, and X. P. Zhang, "History matching of naturally fractured reservoirs using a deep sparse autoencoder," *SPE Journal*, vol. 26, no. 04, pp. 1700–1721, 2021.

Hindawi
Scientific Programming
Volume 2022, Article ID 2725113, 17 pages
https://doi.org/10.1155/2022/2725113

Research Article

Revealing the Pattern of Causality in Processes of Urbanization and Economic Growth: An Evidence from China

Huaxia Lv,[1,2] **Xiaojing Zheng**,[2] **and Shuang Chen**[3]

[1]*Research Institute of Rural Revitalization and Regional Economic Development in the New Era, Weifang University of Science and Technology, Shouguang 262700, China*
[2]*School of Economic and Management, Weifang University of Science and Technology, Shouguang 262700, China*
[3]*School of Aeronautical Engineering, Taizhou University, Taizhou, Zhejiang 318000, China*

Correspondence should be addressed to Xiaojing Zheng; zhengxj@wfust.edu.cn

Received 19 April 2022; Accepted 24 May 2022; Published 30 June 2022

Academic Editor: Xiaofeng Xu

This paper introduces a method of combining phase space reconstruction and symbolic dynamics to study the causality between urbanization and economic growth at different regional levels in Shandong Province and finds that there is a strong positive relationship between urbanization and economic growth from China, indicating that the development of urbanization can drive the economic growth. Then, according to the results of correlation analysis between respective subvariables belonging to urbanization and economic growth and the principle of "deleting strong and reserving weak," the paper selects the independent variable and dependent variable to explore the hidden causal mechanisms that drive the developing of urbanization and economic growth from China. The results show that (1) the pattern causality between the independent variable and the dependent variable is dominated by dark causality at the provincial level; (2) the pattern causality between the independent variable and the dependent variable is dominated by dark causality at the Jinan economic circles and the Lunan economic circles, but the positive causality is dominated at the Jiaodong economic circles; (3) the types of causality between the same evaluation index and PU in different regions are different, and furthermore the degrees of positive, negative, and dark causality are different at two levels and three regions. The conclusion shows that although there is an obvious positive interaction between urbanization and economic growth, the influences of many factors are neither positive nor negative causality, but dark causality, which plays an important role in developing urbanization and economic growth. This work is useful for studying the law of causality between urbanization and economic growth, and this interesting result can be extended to other economic events.

1. Introduction

The study of social systems is hard-wired with the understanding of time series interdependencies of indicators. The research on the relationship between urbanization and economic growth and its driving factors has always been the focus of academic circles. Knowing the evolution of this relationship is vital to planning effective urbanization and economic growth strategies. It is even more important to identify hidden causality in the processes of urbanization and economic growth if the shift in driving forces is considered.

A straightforward tool to study the interaction between urbanization and economic growth is correlation, but it is known to be misleading in nonlinear relationships of social systems [1, 2], when causality is more appropriate. Causality in social system has been approached via methods such as Granger causality [3–7], cointegration [8–10], transfer entropy [11], Mann–Kendall [12], regression model [13–15], differential equation [16], structural equation [17], spatial Durbin model [18–20], and convergent cross mapping [21]. Nevertheless, none of the aforementioned methods provide insight regarding more complex interactions, namely, dark causality [22], which cannot be classified as purely positive or negative. Such causality also exists in the processes of urbanization development and economic growth.

Positive causality (and positive correlation) is anticipated in the relationship between urbanization and

economic growth. Urbanization can promote population agglomeration [16], agglomeration of production factors [13], and scale of urban infrastructure and institutional settings [23] and is also conducive to promoting innovation and social division of labor [12], thereby promoting industrial structure upgrading and economic growth. At the same time, economic growth will bring more employment opportunities and attract more labor force to gather in cities and towns, so as to promote the further improvement of urbanization level [24]. On the other hand, negative causality is proved to be between urbanization and economic growth in recent literature. When urbanization reaches a certain degree, it will impede economic growth [7, 25], which is consistent with the inverted U-shaped relationship between the old and the new urbanization and economic growth obtained by Chinese scholars [26]. Such shift in the above relationship between urbanization and economic growth suggests that a weak influence in one period may become a dominant influence in the next period, which reflects the effect from environment to social system [27]. The hidden nature of interdependencies between urbanization and economic growth may be the missing link leading to a deeper understanding of relationship between urbanization and economic growth. These hidden causalities have neither been focused on by scientists, nor been discovered. Therefore, it is important to uncover dark causality between urbanization and economic growth, which is an important innovation.

To analyze the types of causal interactions, both transparent (positive, negative) and opaque (dark) in the processes of urbanization development and economic growth, we introduce a method of combining phase space reconstruction and symbolic dynamics proposed by Stavros et al. [28] to explore the hidden causality in the processes of urbanization and economic growth at different regional levels in Shandong Province from China. The rest of the paper is arranged as follows. Section 2 provides the evaluation indicator system, data, and methodology used in this paper. In Section 3, we use Pearson correlation analysis of the indicators of urbanization and economic growth to recognize the weak factors defined as independent variables and dependent variables. In Section 4, we introduce the discrete algorithm of entropy for the continuous data, analyze the empirical results, and discover the complex causalities between subvariables belonging to urbanization and economic growth, which consist of positive causality, negative causality, and hidden causality. Finally, the discussion of conclusion and the future research possibility are provided in the last section.

2. Data and Methodology

2.1. Construction of Indicator System. According to [5, 29], urbanization and economic growth go hand in hand, and urbanization both reflects and contributes to economic growth and economic development patterns. Therefore, in this paper, the percent of urban population in total population (PU) reflecting urbanization level is used as dependent variable, the factors of which are used as independent variables. Based on the literature [30–37], we establish an indicator system consisting of 17 evaluation indexes calculated from 26 statistical indicators, as shown in Table 1. These indicators reflect the development level of regional politics (LGBR, LGBE), economy (PGDP, RTS, RRUHCE, IN), society (RRUE, RRUEP), culture (PLCPT), education (EE, RST), healthcare (BPT, MEPPT), and infrastructure (communication and transportation) (PT, VFT, BVPTS). Further, in order to better reflect the urban-rural gap, we simply calculate some statistical indicators to obtain the evaluation indexes such as RRUHCE [38, 39], RRUE, and RRUEP [40, 41].

2.2. Data Resource. The statistic indicator data are derived from the 2001–2018 Shandong Statistical Year book. Due to the consistency of the research unit and given the availability of data, this study selected 17 cities in Shandong Province from 2000 to 2017 as samples to measure the hidden causality among the processes of urbanization and economic growth.

2.3. Method. We simplify the method proposed in [28] to explore the types of causality that exist in the process of urbanization and economic growth. In this method, arbitrary variable can be projected at multidimension space, which increases the observability of the data and could simply the complexity of the data, and then, by analyzing the corresponding relationship between variables, the causality could be clearer. Details are as follows.

2.3.1. Reconstructing the Shadow Attractors $M_{\mathbf{X}}$ and $M_{\mathbf{y}}$. Firstly, we reconstruct the shadow attractor $M_{\mathbf{X}}$ of the independent variable $\mathbf{x}_t$, and the formula is as follows:

$$M_x = \begin{pmatrix} X_1 = < x_{1+(E-1)\tau}, \ldots, x_{1+\tau}, x_1 > \\ X_2 = < x_{2+(E-1)\tau}, \ldots, x_{2+\tau}, x_2 > \\ \vdots \\ X_{L-(E-1)\tau} = < x_{L-(E-1)\tau}, \ldots, x_{L+\tau}, x_L > \end{pmatrix}. \quad (1)$$

Similarly, the reconstruction formula of the shadow attractor $M_{\mathbf{y}}$ of the dependent variable y_t is as follows:

$$M_y = \begin{pmatrix} Y_1 = < y_{1+(E-1)\tau}, \ldots, y_{1+\tau}, y_1 > \\ Y_2 = < y_{2+(E-1)\tau}, \ldots, y_{2+\tau}, y_2 > \\ \vdots \\ Y_{L-(E-1)\tau} = < y_{L-(E-1)\tau}, \ldots, y_{L+\tau}, y_L > \end{pmatrix}. \quad (2)$$

In order to reduce the difficulty of causal mode judgment, refer to the Takens theorem to select the appropriate embedding dimension E and time lag τ to construct the shadow attractors and phase space. In this paper, we use $E = 3$ and $\tau = 1$. Then, the formulas of $M_{\mathbf{X}}$ and $M_{\mathbf{y}}$, respectively, become

Table 1: Indicator system.

Number	Evaluation indexes	Name and calculation	Statistic indicators
1	PGDP (yuan)	Per capita GDP	Per capita GDP
2	PU (%)	Population urbanization rate (urban population/total registered population) × 100%	Total registered population; urban population
3	RTS	Ratio of tertiary-secondary industry (tertiary industry/secondary industry)	Secondary industry and tertiary industry in GDP
4	RRUHCE	Ratio of rural-urban household consumption expenditure (value of rural household consumption expenditure/value of urban household consumption expenditure)	Value of urban household consumption expenditure; value of rural household consumption expenditure
5	RRUE	Ratio of rural/urban employment (number of employed persons in rural area/number of employed persons in urban area)	Number of employed persons in urban area; number of employed persons in rural area
6	RRUEP	Ratio of the proportion of rural employees to agricultural population divided by the proportion of urban employees to nonagricultural population (number of employed persons in rural area/agricultural population/number of employed persons in urban area/nonagricultural population)	Number of employed persons in urban area; number of employed persons in rural area; agricultural population; nonagricultural population
7	LGBR (10000 yuan)	Local government budgetary revenue	Local government budgetary revenue
8	LGBE (10000 yuan)	Local government budgetary expenditure	Local government budgetary expenditure
9	EE (10000 yuan)	Education expenditure	Education in local government budgetary expenditure
10	PLCPT (volumes per thousand)	Total collections per thousand	Total collections in public libraries

Table 1: Continued.

Number	Evaluation indexes	Name and calculation	Statistic indicators
11	BPT (beds per thousand)	Beds per thousand	Number of beds in medical institutions
12	MEPPT (person per thousand)	Number of medical personnel per thousand	Number of medical personnel
13	IN (10000 yuan)	Investment in fixed assets	Investment in fixed assets (excluding investment in fixed assets of rural households)
14	RST	Ratio of student-teacher (total enrollment in primary, senior, and junior secondary schools/full − time teachers in primary, senior, and junior secondary schools)	Total enrollment and full-time teachers in senior and junior secondary schools, total enrollment and full-time teachers in primary schools
15	PT (10000 persons)	Passenger traffic	Passenger traffic
16	VFT (10000 tons)	Volume of freight traffic	Volume of freight traffic
17	BVPTS (100 million yuan)	Business volume of post and telecommunication services	Business volume of post and telecommunication services

$$M_x = (X_T) = (X_T, X_{T+1}, X_{T+2}), \tag{3}$$

$$M_y = (Y_T) = (Y_T, Y_{T+1}, Y_{T+2}). \tag{4}$$

2.3.2. Calculating the Distance Matrices $D_{\mathbf{x}}$ and $D_{\mathbf{y}}$. The first-order difference between variables at two adjacent time points is used as the element to calculate distance matrices $D_{\mathbf{x}}$ and $D_{\mathbf{y}}$ of the shadow attractors $M_{\mathbf{X}}$ and $M_{\mathbf{y}}$; then, the distance matrix $D_{\mathbf{x}}$ of $M_{\mathbf{X}}$ can be expressed as

$$D_X = \begin{pmatrix} d(X_1, X_1), d(X_1, X_2), \ldots d\left(X_1, X_{L-(E-1)\tau}\right) \\ d(X_2, X_1), d(X_2, X_2), \ldots, d\left(X_2, X_{L-(E-1)\tau}\right) \\ \vdots \\ d\left(X_{L-(E-1)\tau}, X_1\right), d\left(X_{L-(E-1)\tau}, X_2\right), \ldots, d\left(X_{L-(E-1)\tau}, X_{L-(E-1)\tau}\right) \end{pmatrix}, \tag{5}$$

where $d(X_m, X_n) = \Delta X_{mn} = X_n - X_m$. Similarly, the distance matrix $D_{\mathbf{y}}$ of $M_{\mathbf{y}}$ can be expressed as

$$D_y = \begin{pmatrix} d(Y_1, Y_1), d(Y_1, Y_2), \ldots, d\left(Y_1, Y_{L-(E-1)\tau}\right) \\ d(Y_2, Y_1), d(Y_2, Y_2), \ldots, d\left(Y_2, Y_{L-(E-1)\tau}\right) \\ \vdots \\ d\left(Y_{L-(E-1)\tau}, Y_1\right), d\left(Y_{L-(E-1)\tau}, Y_2\right), \ldots, d\left(Y_{L-(E-1)\tau}, Y_{L-(E-1)\tau}\right) \end{pmatrix}, \tag{6}$$

where $d(Y_m, Y_n) = \Delta Y_{mn} = Y_n - Y_m$.

2.3.3. Identifying the Change Pattern of x_t and y_t and Signature Calculation. In order to eliminate the influence of the value level and measurement unit of the variable on the judgment of the causality pattern, we define the change pattern of x_t and y_t as $s_n^{\mathbf{x}}$ and $s_n^{\mathbf{y}}$, and the formulas of $s_n^{\mathbf{x}}$ and $s_n^{\mathbf{y}}$ are as follows:

$$s_n^x = \left(\frac{x_n - x_m}{x_m}\right), \tag{7}$$

$$s_n^y = \left(\frac{y_n - y_m}{y_m}\right). \tag{8}$$

Since $E = 3$ is used as an embedding parameter, the change of variables and corresponding pattern signatures follow the symbolic rules, as follows:

(i) $\nearrow\nearrow$: $X_T < X_{T+1} < X_{T+2}$.
(ii) $\longrightarrow\nearrow$: $X_T = X_{T+1} < X_{T+2}$.
(iii) $\searrow\nearrow$: $X_T > X_{T+1} < X_{T+2}$.
(iv) $\nearrow\longrightarrow$: $X_T < X_{T+1} = X_{T+2}$.
(v) $\longrightarrow\longrightarrow$: $X_T = X_{T+1} = X_{T+2}$.
(vi) $\searrow\longrightarrow$: $X_T > X_{T+1} = X_{T+2}$.
(vii) $\nearrow\searrow$: $X_T < X_{T+1} > X_{T+2}$.
(viii) $\longrightarrow\searrow$: $X_T = X_{T+1} > X_{T+2}$.
(ix) $\searrow\searrow$: $X_T > X_{T+1} > X_{T+2}$.

Furthermore, according to the pattern signatures, we calculate the pattern probability $P_{\mathbf{X}_T}$ and $P_{\mathbf{Y}_T}$, and the formulas of $P_{\mathbf{X}_T}$ and $P_{\mathbf{Y}_T}$ are as follows:

$$P_{X_T} = \text{Signature}\left(s_n^x\right), \quad P_{Y_T} = \text{Signature}\left(s_m^y\right). \tag{9}$$

2.3.4. Pattern Causality Matrix and Separating PC Pattern. We fill $P_{\mathbf{X}_T}$ and $P_{\mathbf{Y}_T}$ into pattern matrix for $E = 3$ and then divide the pattern matrix to calculate the cumulative probability (denoted as PC) of the positive, negative, and dark causality patterns, and the functions are as follows [22]:

$$\mathrm{PC}(\mathrm{Positive}) = \frac{1}{L}\sum \mathrm{diag}_{\mathrm{main}}(\mathrm{PC}),\ \mathrm{PC}(\mathrm{Positive}) \in [0,1],$$

$$\mathrm{PC}(\mathrm{Negative}) = \frac{1}{L}\sum \mathrm{diag}_{\mathrm{counter}}(\mathrm{PC}),\ \mathrm{PC}(\mathrm{Negative}) \in [0,1], \tag{10}$$

$$\mathrm{PC}(\mathrm{Dark}) = \frac{1}{L}\sum\left(PC \notin \left(\mathrm{diag}_{\mathrm{main}}(\mathrm{PC})\right) \cup \left(\mathrm{diag}_{\mathrm{counter}}(\mathrm{PC})\right)\right),\ \mathrm{PC}(\mathrm{Dark}) \in [0,1],$$

where PC(positive) + PC(negative) + PC(dark) = 1.

2.3.5. Theorems and Proofs. According to [22], the reason that this method enables identification of the nature of influence is that it satisfies two important definitions.

Definition 1

(1) If X causes same-direction changes to Y, then we say that X has a positive influence on Y.

(2) If X causes opposite-direction changes to Y, then we say that X has a negative influence on Y.

(3) If X causes changes to Y which are of neither the same nor the opposite direction, then we say that X has a "dark" influence on Y.

(4) $\rightarrowtail\!\!\!\!\!\rightarrow$ symbolizes diffeomorphism.

In terms of Definition 1, four lemmas are introduced to analyze all kinds of causalities in the processes of urbanization and economic growth.

Lemma 1. *Let M be an m-dimensional compact manifold and $X: M \longrightarrow \mathbb{R}$, $Y: M \longrightarrow \mathbb{R}$ be smooth observation functions. Let $\varphi: MM$ be a smooth diffeomorphism. If there exists $\psi: M_X M_Y$ such that ψ is bijective, then M_X causes M_Y.*

Lemma 2. *Let M be an m-dimensional compact manifold and $X: M \longrightarrow \mathbb{R}$, $Y: M \longrightarrow \mathbb{R}$ be smooth observation functions. If M_X causes M_Y (through a bijective map ψ) and there exist $h: M_X X$ and $g M_Y$ Y, with h, g being bijective, then X causes Y as well.*

Lemma 3. *Let M be an m-dimensional compact manifold and $X: M \longrightarrow \mathbb{R}$, $Y: M \longrightarrow \mathbb{R}$ be smooth observation functions. Let $\varphi: MM$ be a smooth diffeomorphism. If there exists $\psi: M_X \rightarrowtail M_Y$ such that ψ is injective, then M_X weakly causes M_Y.*

Lemma 4. *Let M be an m-dimensional compact manifold and $X: M \longrightarrow \mathbb{R}$, $Y: M \longrightarrow \mathbb{R}$ be smooth observation functions. If M_X weakly causes M_Y (through an injective map ψ) and there exist $h: M_X \rightarrowtail X$ and $g: M_Y \rightarrowtail Y$ with h, g being injective, then X weakly causes Y as well.*

Definition 2

(1) P: $M_X \longrightarrow^{+} M_Y$, $\longrightarrow^{+} : \{\underline{x} \longrightarrow \underline{y} \,|\, p\,\underline{x} \overset{+}{\leftrightarrow} +p\,\underline{x}\}$P corresponds to the same patterns (i.e., positive mapping).

(2) N: $M_X \overset{-}{\rightarrow} M_Y$, $\overset{-}{\rightarrow} : \{\underline{x} \longrightarrow \underline{y} \,|\, p\,\underline{x} \overset{-}{\leftrightarrow} -p\,\underline{x}\}$N corresponds to opposite patterns (i.e., negative mapping).

(3) D: $M_X \longrightarrow * M_Y$, $\longrightarrow * : \{\underline{x} \longrightarrow \underline{y} \,|\, p\,\underline{x} \overset{*}{\leftrightarrow} p\,\underline{x}\}$. D corresponds to patterns that are neither the same nor opposite (i.e., dark mapping).

Definition 2 is used extensively in the formulation and proof of the following three important theorems.

Theorem 1. *Let M be an m-dimensional compact manifold and $X: M \longrightarrow \mathbb{R}$, $Y: M \longrightarrow \mathbb{R}$ be smooth observation functions. Let $\varphi: MM$ be a smooth diffeomorphism. Let M_X and M_Y be the shadow attractors of X and Y, respectively. If* P: $M_X \longrightarrow^{+} M_Y$ *such that* P *is bijective (or injective) and there exist $h: M_X \longrightarrow X$ and $g: M_Y \longrightarrow Y$, with h, g being bijective (or injective), then X exerts a positive influence on Y.*

Theorem 2. *Let M be an m-dimensional compact manifold and $X: M \longrightarrow \mathbb{R}$, $Y: M \longrightarrow \mathbb{R}$ be smooth observation functions. Let $\varphi: MM$ be a smooth diffeomorphism. Let M_X and M_Y be the shadow attractors of X and Y, respectively. If* N: $M_X \overset{-}{\rightarrow} M_Y$ *is bijective (or injective) and there exist $h: M_X \longrightarrow X$ and $g: M_Y \longrightarrow Y$ that are bijective (or injective), then X exerts a negative influence on Y.*

Theorem 3. *Let M be an m-dimensional compact manifold and $X: M \longrightarrow \mathbb{R}$, $Y: M \longrightarrow \mathbb{R}$ be smooth observation functions. Let $\varphi: MM$ be a smooth diffeomorphism. Let M_X and M_Y be the shadow attractors of X and Y, respectively. If* D: $M_X \longrightarrow * M_Y$ *such that* D *is bijective (or injective) and there exist $h: M_X \longrightarrow X$ and $g: M_Y \longrightarrow Y$ that are bijective (or injective), then X exerts a dark influence on Y.*

Proofs of the above theorems are given in the supplementary information in [22].

3. Correlation Analysis

3.1. Population Urbanization Rate and Per Capita GDP of Shandong Province. As shown in Figure 1, since 1952, the processes of economic growth and urbanization in Shandong Province have experienced two periods of slow and rapid development: the relatively slow period from 1952 to 1987 and the rapid development period from 1987 to 2019. From 1952 to 1987, per capita GDP increased from y91 to y1,131, an increase of 11.4 times, and the urbanization rate rose from 5.99% to 13.25%. From 1987 to 2019, per capita GDP increased from y1,131 to y70,653, an increase of 61 times, and the urbanization rate increased from 13.25% to 49.94% (the highest value in 2018 was 50.94%).

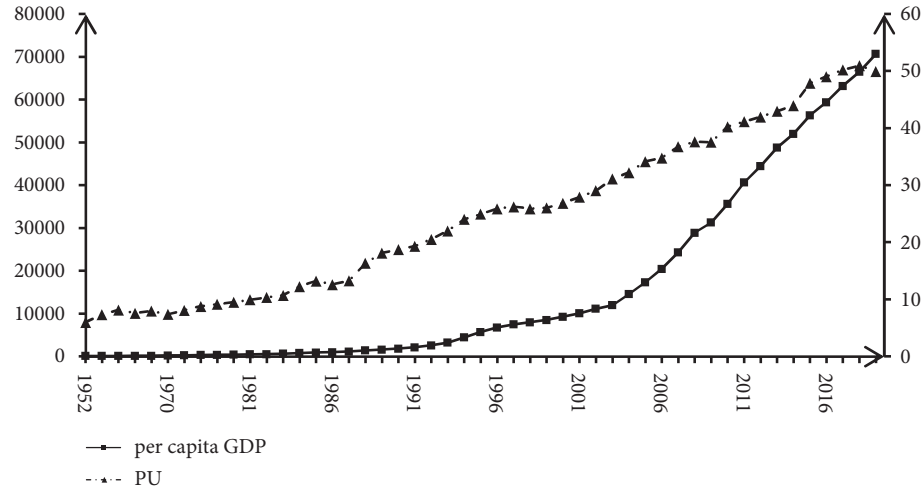


Figure 1: The relationship between per capital GDP and PU from 1952 to 2019.

Figure 1 shows that from 1952 to 2019, per capital GDP increased in the way of logistic, and PU increased linearly. The linear growth of PU promoted the growth of per capital GDP in logistic pattern, and the causality between per capital GDP and PU was significant.

3.2. Correlation Analysis of Per Capita GDP and Population Urbanization Rate. The correlation between per capita GDP and PU of 17 cities from 2000 to 2017 is analyzed year by year, and the results are shown in Figure 2. It can be observed that the correlation between per capita GDP and PU shows three stages: (1) from 2000 to 2004, there was a strong positive correlation between per capita GDP and PU, indicating a strong interaction between economic growth and urbanization development; (2) from 2004 to 2012, the positive correlation between per capita GDP and PU gradually decreased, and the correlation coefficient decreased from 0.861 to 0.503, indicating that the interaction between the two is weakening; (3) from 2012 to 2013, the correlation between per capita GDP and PU returned to the highly positive correlation, which remained consistent until 2017, indicating that the relation between economic growth and urbanization became strongly interactive. The main reason for the above changes is the shift of urbanization strategy, which was proposed at the Central Economic Work Conference in December 2012: from "traditional urbanization" to "new urbanization." New urbanization has become an important engine of economic growth, which is proved by the high positive correlation between per capita GDP and PU from 2013 to 2017.

3.3. Correlation Analysis of Evaluation Indexes. The correlation analysis of the sequences of 17 evaluation indicators in 17 cities from 2000 to 2017 is carried out, and the results are shown in Figure 3. The correlation coefficients of arbitrary two indexes of LGBR, LGBE, EE, IN, and the correlation coefficient between BPT and MEPPT, are all larger than 0.9. The correlation coefficients between PU, BPT, MEPPT, and PGDP and between RRUE, BPT, MEPPT, and PU were greater than 0.8 and less than 0.9. The correlation coefficients between PGDP and RRUE, LGBR, and IN, between RTS and LGBR and LGBE, between PU and LGBR, LGBE, EE, and IN, between RRUE and RRUEP, BPT, and MEPPT, and between LGBR/LGBE/EE and BPT, MEPPT, VFT, and BVPTS are greater than 0.6 and less than 0.8. It shows that there is strong interaction among these evaluation indexes. However, except the correlation coefficients between RRUEP and RRUE, PT and VFT, the correlation coefficients of arbitrary two indexes of RRUHCE, RRUEP, PLCPT, RST, PT, and the correlation coefficients between the above five indexes and the other indexes are all less than 0.5. The correlation coefficients of arbitrary two indexes of LGBR, LGBE, EE, IN, and the correlation coefficient between BPT and MEPPT, are all larger than 0.9, indicating weak interaction among them, which is exactly the problem to be studied in this paper.

Figure 3 shows the following. (1) In the matrix, the diagonal represents the probability density distribution function of variables, from which it can be seen that the vast majority of variables follow normal distribution or biased normal distribution; the distribution of some variables, such as PLCPT, LGBR, LGBE, EE and IN, is far from the standard normal distribution. (2) The data in the upper triangle of the matrix represent the correlation coefficient of two variables, which expresses the strength of correlation between two variables, where $*$, $**$, and $***$ denote significance at 10, 5, and 1% levels, respectively. (3) The lower triangle of the matrix represents the scatter diagram between two variables, from which we can intuitively observe the relationships, such as positive correlation, negative correlation, or no correlation. (4) In summary, our research object meets certain causality conditions, and among the above correlations, the correlations of two variables including PU, RST, RRUHCE, RRUEP, PLCPT, RST, and PT are not obvious. We suspect that there is some dark causality in the progress of urbanization and economic growth, which will be further analyzed in the following sections.

3.4. Selection of Independent Variable and Dependent Variable. Our purpose is mainly to judge the causality

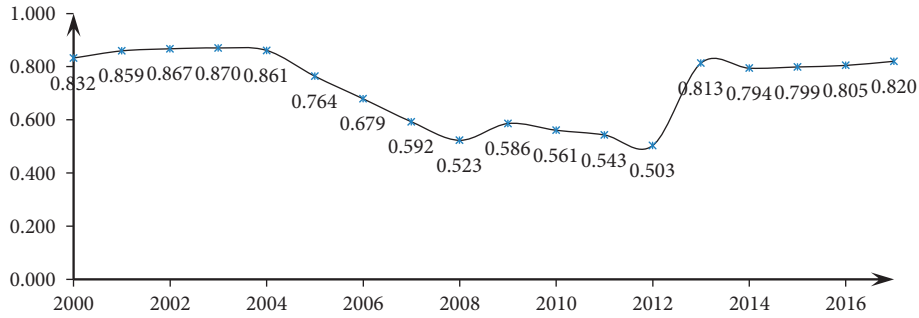


Figure 2: The correlation coefficients between per capital GDP and PU of 17 cities from 2000 to 2017.

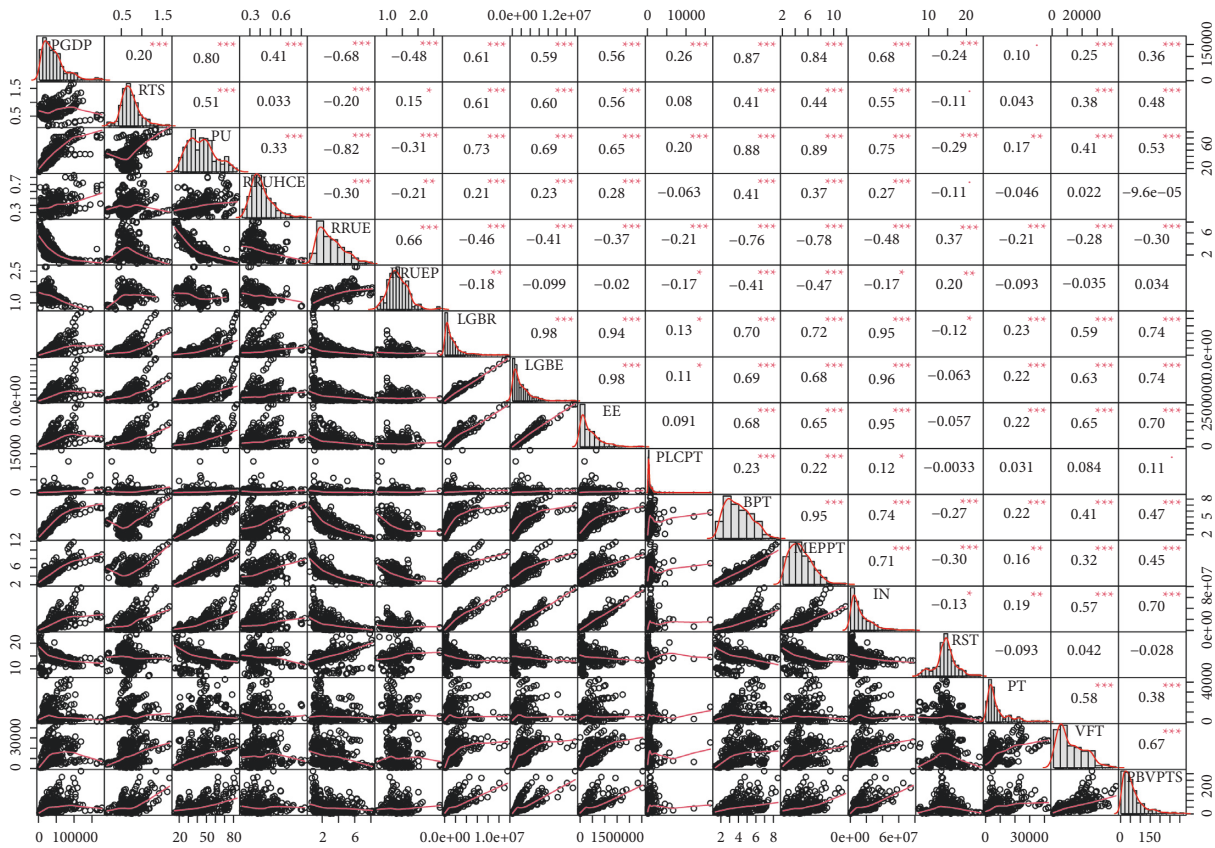


Figure 3: Pairwise correlation coefficient between 17 evaluation indexes.

between weakly correlated variables. Therefore, we omit the evaluation indexes with large correlation coefficients and retain those with small correlation coefficients, including PU, RTS, RRUHCE, RRUEP, PLCPT, RST, and PT, as shown in Figure 4, to study the hidden causality by multiobjective artificial bee colony algorithm [32] among the processes of urbanization development and economic growth. Among them, PU, which has the strongest correlation with per capita GDP, is used as the dependent variable to reflect the level of urbanization and economic development, denoted as $\mathbf{y}_t = (\mathrm{PU}_{\mathrm{it}})$. Let $y = \mathrm{PU}$; then, $\mathbf{y}_t = (y_{\mathrm{it}})$. The other six evaluation indexes are used as independent variables, denoted as $\mathbf{x}_t = (\mathrm{RST}_{\mathrm{it}}, \mathrm{RRUHCE}_{\mathrm{it}}, \mathrm{RRUEP}_{\mathrm{it}}, \mathrm{PLCPT}_{\mathrm{it}}, \mathrm{RST}_{\mathrm{it}}, \mathrm{PT}_{\mathrm{it}})$. Let $\mathrm{RST} = x^1, \mathrm{RRUHCE} = x^2, \mathrm{RRUEP} = x^3, \mathrm{PLCPT} = x^4, \mathrm{RST} = x^5, \mathrm{PT} = x^6$; then, $x_t = (x_{it}^1, x_{it}^2, x_{it}^3, x_{it}^4, x_{it}^5, x_{it}^6)$, where $i = 1, 2, \ldots, 17$, respectively, indicate 17 cities, and $t = 1, 2, \ldots, 18$ indicate the year from 2000 to 2017.

4. Empirical Results of Pattern Causality Matrix in Shandong Province

4.1. Discrete Algorithm of Entropy for the Continuous Data. We adopted the discrete algorithm of entropy [42] to treat the time series data of continuous variables, and the details are as follows.

X is a continuous variable, its time series value is $(x_1, x_2, \cdots, x_n)$, $P(x_i)$ represents the probability of occurrence of x_i, and $\sum P(x_i) = 1$. Then, the information entropy of X is

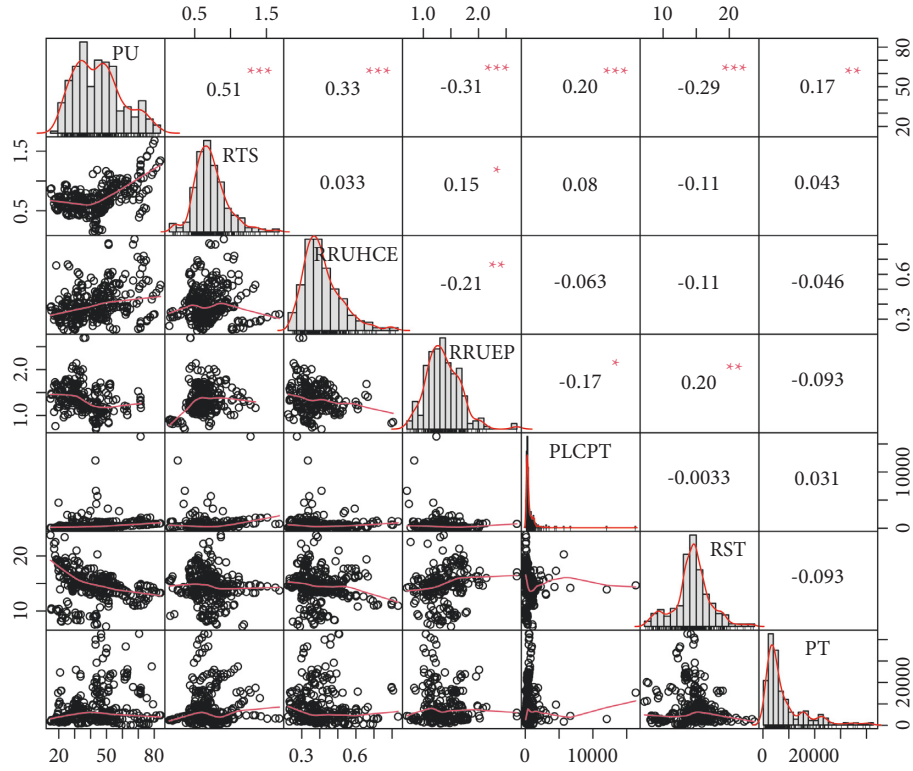


Figure 4: Pairwise correlation coefficient between 7 evaluation indexes.

$$H(X) = E[-\log P(x_i)] = -\sum_{i=1}^{N} P(x_i)\log P(x_i). \quad (11)$$

In 1993, Fayyad and Irani proposed the minimum information entropy discretization method [43]. The discrete threshold boundary value is selected by the class information entropy of the candidate interval. The information entropy of the subsets S_1 and S_2 is $\mathrm{Ent}(S_1)$ and $\mathrm{Ent}(S_2)$; then, the class information entropy of the breakpoint T on attribute A can be expressed as

$$E(A,T;S) = \left(\frac{|S_1|}{|S|}\right)\cdot \mathrm{Ent}(S_1) + \left(\frac{|S_2|}{|S|}\right)\cdot \mathrm{Ent}(S_2), \quad (12)$$

where $|S|$ represents the number of elements in the set S.

For a given attribute A, the point with the smallest value of $E(A,T;S)$ is taken as the partition point of discretization, denoted as T min. The partition point T min divides the interval into two parts, and the sample set is also divided into two sets S_1 and S_2. In the calculation of all partition points of S_1 and S_2, it is assumed that T_1 and T_2 are the best partition points of S_1 and S_2, respectively, and their corresponding class information entropy is $E(A,T_1,S_1)$ and $E(A,T_2,S_2)$, respectively. If $E(A,T_1,S_1) > E(A,T_2,S_2)$, continue to divide S_1; otherwise, divide S_2. In other words, whichever is bigger is divided. Then, they are recursively invoked until the recursive stop condition is met. The stopping condition is to use the minimum discriminant length criterion to judge whether the recursive discretization stops, that is, the partition interval of recursion in the set S stops when the following conditions are met:

$$\mathrm{Gain}(A,T;S) < \left(\frac{\log_2(N-1)}{N}\right) + \left(\frac{(\Delta A,T;S)}{N}\right), \quad (13)$$

where N is the number of samples in the set $S\mathrm{Gain}(A,T;S) = \mathrm{Ent}(s) - \mathrm{Ent}(A,T,S)$.

$$\Delta(A,T;S) = \log_2\left(3^k - 2\right) - \left[k\cdot \mathrm{Ent}(S) - k_1\cdot \mathrm{Ent}(S_2) - k_2\cdot \mathrm{Ent}(S_2)\right], \quad (14)$$

where k_1 and k_2 represent the number of categories in the sets S_1 and S_2, namely, $k = |S|$, $k_1 = |S_1|$, $k_2 = |S_2|$.

According to this process, all continuous variables involved in this paper are discretized.

After the discretization of variable data, we invoke the corresponding algorithms and theorems mentioned in Section 2.3, i.e., equations (1)–(10), to analyze the hidden causalities in the progress of urbanization and economic growth in Shandong Province.

4.2. Empirical Results at the Provincial Level. PC (from x to y) pattern to pattern matrix at the provincial level is shown in Table 2. Cumulative probability of the positive, negative, and dark PC from RST to PU is 0.393382, 0.209559, and 0.397059, respectively, calculated according to Table 2(a). Cumulative probability of the positive, negative, and dark PC from RRUHCE to PU is 0.316176, 0.183824, and 0.5, respectively, calculated according to Table 2(b). Cumulative probability of the positive, negative, and dark PC from RRUEP to PU is 0.26738, 0.31016, and 0.42246, respectively,

Table 2: PC (from x to y) pattern to pattern matrix at the provincial level.

(a) PC (from RTS to PU) pattern to pattern matrix

	y:↗↗	y:→↗	y:↘↗	y:↗→	y:→→	y:↘→	↗↘	y:→↘	y:↘↘
x1:↗↗	0.371324	0	0.102941	0	0	0	□.095588	0	0.022□59
x1:→↗	0.007353	0	0	0	0	0	0	0	0
x1:↘↗	0.069853	0	0.018382	0	0	0	0.007353	0	0
x1:↗→	0.007353	0	0	0	0	0	0	0	0
x1:→→	0	0	0	0	0	0	0	0	0
x1:↘→	0	0	0	0	0	0	0	0	0
x1:↗↘	0.091912	0	0.007353	0	0	0	0.003676	0	0.003676
x1:→↘	0	0	0	0	0	0	0	0	
x1:↘↘	0.172794	0	0.011029	0	0	0	0.007353	0	0

(b) PC (from RRUHCE to PU) pattern to pattern matrix

	y:↗↗	y:→↗	y:↘↗	y:↗→	y:→→	y:↘→	y:↗↘	y:→↘	y:↘↘
x2:↗↗	0.279412	0	0.073529	0	0	0	0.066176	0	0.018382
x2:→↗	0	0	0	0	0	0	0	0	0
x2:↘↗	0.158088	0	0.029412	0		0	0.011029	0	0.0□3676
x2:↗→	0	0	0	0	0	0	0	0	0
x2:→→	0	0	0	0	0	0	0	0	0
x2:↘→	0	0	0	0	0	0	0	0	0
x2:↗↘	0.161765	0	0.022059	0	0	0	0.007353	0	0
x2:→↘	0	0	0	0	0	0	0	0	0
x2:↘↘	0.132353	0	0.029412	0	0	0	0.007353	0	0

(c) PC (from RRUEP to PU) pattern to pattern matrix

	y:↗↗	y:→↗	y:↘↗	y:↗→	y:→→	y:↘→	y:↗↘	y:→↘	y:↘↘
x3:↗↗	0.240642	0	0.02139	0	0	0	0.005348	0	0.005348
x3:→↗	0	0	0	0	0	0	0	0	0
x3:↘↗	0.15508	0	0.016043	0	0	0	0.0106□5	0	0
x3:↗→	0	0	0	0	0	0	0	0	0
x3:→→	0	0	0	0	0	0	0	0	0
x3:↘→	0	0	0	0	0	0	0	0	0
x3:↗↘	0.176471	0	0.053476	0	0	0	0.010695	0	0
x3:→↘	0	0	0	0	0	0	0	0	0
x3:↘↘	0.240642	0	0.037433	0	0	0	0.010695	0	0.016043

(d) PC (from PLCPT to PU) pattern to pattern matrix

	y:↗↗	y:→↗	y:↘↗	y:↗→	y:→→	y:↘→	y:↗↘	y:→↘	y:↘↘
x4:↗↗	0.397059	0	0.066176	0	0		0.077206	0	0.0330□8
x4:→↗	0	0	0	0	0	0	0	0	0
x4:↘↗	0.180147	0	0.007353	0	0	0	0.007353	0	0
x4:↗→	0	0	0	0	0	0	0	0	0
x4:→→	0	0	0	0	0	0	0	0	0
x4:↘→	0	0	0	0	0	0	0	0	0
x4:↗↘	0.1□6029	0	0.040441	0	0	0	0.007353	0	0
x4:→↘	0	0	0	0	0	0	0	0	0
x4:↘↘	0.040441	0	0.003676	0	0	0	0.0□3676	0	0

(e) PC (from RST to PU) pattern to pattern matrix

	y:↗↗	y:→↗	y:↘↗	y:↗→	y:→→	y:↘→	y:↗↘	y:→↘	y:↘↘
x5:↗↗	0.080882	0	0.014706	0	0	0	0.007353	0	0
x5:→↗	0	0	0	0	0	0	0	0	0
x5:↘↗	0.161765	0	0.033088	0	0	0	0.018382	0	0
x5:↗→	0	0	0	0	0	0	0	0	0
x5:→→	0	0	0	0	0	0	0	0	0
x5:↘→	0	0	0	0	0	0	0	0	0
x5:↗↘	0.172794	0	0.014706	0	0	0	0.00735	0	0.003676
x5:→↘	0	0	0	0	0	0	0	0	0
x5:↘↘	0.319853	0	0.0625	0	0	0	0.0698	0	0.033088

(f) PC (from PT to PU) pattern to pattern matrix

	y:↗↗	y:→↗	y:↘↗	y:↗→	y:→→	y:↘→	y:↗↘	y:→↘	y:↘↘
x6:↗↗	0.352941	0	0.069853	0	0	0	0.077206	0	0.033088
x6:→↗	0	0	0	0	0	0		0	0
x6:↘↗	0.158088	0	0	0	0	0		0	0
x6:↗→	0.003676	0	0.007353	0	0	0	0.0□3□76	0	0.003676
x6:→→	0	0	0	0	0	0	0	0	0
x6:↘→	0	0	0	0	0	0	0	0	0
x6:↗↘	0.15□412	0	0.033088	0	0	0	0.011029	0	0.003676
x6:→↘	0.003676	0	0	0	0	0	0	0	0
x6:↘↘	0.077206	0		0	0	0	0.0□7353	0	0

Note. (1) Each cell takes values from 0 to 1. Blue cells denote positive influence. Red cells denote negative. Purple cells denote dark influence. (2) The first column represents the change pattern of cause (x), the first line represents the change pattern of effect (y), the numbers represent the probability of effect (y) caused by cause (x), "0" represents impossible event, and "1" represents certain event. (3) From the table, we can find that the dominant causalities between RTS, RRUHCE, RRUEP, PLCPT, RST, PT, and PU are dark.

calculated according to Table 2(c). Cumulative probability of the positive, negative, and dark PC from PLCPT to PU is 0.411765, 0.121324, and 0.466912, respectively, calculated according to Table 2(d). Cumulative probability of the positive, negative, and dark PC from RST to PU is 0.121324, 0.352941, and 0.525735, respectively, calculated according to Table 2(e). Cumulative probability of the positive, negative, and dark PC from PT to PU is 0.363971, 0.143382, and 0.492647, respectively, calculated according to Table 2(f). To sum up, the dark PC dominates in PC (from **x** to **y**) pattern at the provincial level, as shown in Table 3.

4.3. Empirical Results at the Regional Level. According to the current administrative division and development strategy of Shandong Province, it is generally divided into the Jinan economic circle, the Jiaodong economic circle, and the Lunan economic circle (as shown in Figure 5). From the spatial structure, the Jinan economic circle includes Jinan (including Laiwu), Zibo, Tai'an, Liaocheng, Dezhou, Binzhou, and Dongying; the Jiaodong economic circle includes Qingdao, Yantai, Weihai, Weifang, and Rizhao; the Lunan economic circle includes Linyi, Zaozhuang, Jining, and Heze.

In Figure 5, green denotes the range of Jinan economic circle, blue denotes the range of Jiaodong economic circle, and pink denotes the range of Lunan economic circle.

4.3.1. Analysis of Pattern Causality Matrix in the Jinan Economic Circle. PC (from x to y) pattern to pattern matrix in the Jinan economic circle is shown in Table 4. Cumulative probability of the positive, negative, and dark PC from RST to PU is 0.515625, 0.21875, and 0.265625, respectively, calculated according to Table 4(a). Cumulative probability of the positive, negative, and dark PC from RRUHCE to PU is 0.3671875, 0.171875, and 0.4609375, respectively, calculated according to Table 4(b). Cumulative probability of the positive, negative, and dark PC from RRUEP to PU is 0.3984375, 0.1953125, and 0.40625, respectively, calculated according to Table 4(c). Cumulative probability of the positive, negative, and dark PC from PLCPT to PU is 0.4453125, 0.0703125, and 0.484375, respectively, calculated according to Table 4(d). Cumulative probability of the positive, negative, and dark PC from RST to PU is 0.1015625, 0.421875, and 0.4765625, respectively, calculated according to Table 4(e). Cumulative probability of the positive, negative, and dark PC from PT to PU is 0.421875, 0.1171875, and 0.4609375, respectively, calculated according to Table 4(f). To sum up, in addition to that the positive influence dominates in PC (from RTS to PU) pattern, and the dark PC dominates in PC (from the other independent variable in **x** to **y**) pattern in the Jinan economic circle, as shown in Table 3.

4.3.2. Analysis of Pattern Causality Matrix in the Jiaodong Economic Circle. PC (from **x** to **y**) pattern to pattern matrix in the Jiaodong economic circle is shown in Table 5. Cumulative probability of the positive, negative, and dark PC from RST to PU is 0.5, 0.2, and 0.3, respectively, calculated according to Table 5(a). Cumulative probability of the positive, negative, and dark PC from RRUHCE to PU is 0.4875, 0.15, and 0.3625, respectively, calculated according to Table 5(b). Cumulative probability of the positive,

Table 3: Comparison of PC pattern at different regional levels in Shandong Province.

Regional level	PC pattern with PU					
	RTS	RRUHCE	RRUEP	PLCPT	RST	PT
Provincial level	Dark > positive > negative (0.397059,0.393382,0.209559)	Dark > positive > negative (0.5,0.316176,0.183824)	Dark > negative > positive (0.42246,0.310160.26738)	Dark > positive > negative (0.466912,0.411765,0.121324)	Dark > negative > positive (0.525735,0.352941,0.121324)	Dark > positive > negative (0.492647,0.363971,0.143382)
The Jinan economic circle	Positive > dark > negative (0.515625,0.265625,0.21875)	Dark > positive > negative (0.4609375,0.3671875,0.171875)	Dark > positive > negative (0.40625,0.3984375,0.1953125)	Dark > positive > negative (0.484375,0.4453125,0.0703125)	Dark > negative > positive (0.4765625,0.421875,0.1015625)	Dark > positive > negative (0.4609375,0.421875,0.1171875)
The Jiaodong economic circle	Positive > dark > negative (0.5,0.3,0.2)	Positive > dark > negative (0.4875,0.3625,0.15)	Positive > dark > negative (0.45,0.3375,0.2125)	Positive > dark > negative (0.7125,0.2875,0)	Negative > dark > Positive (0.475,0.45,0.075)	Positive > dark > negative (0.55,0.375,0.075)
The Lunan economic circle	Dark > positive > negative (0.421875,0.390625,0.1875)	Dark > positive > negative (0.5,0.296875,0.203125)	Dark > positive > negative (0.515625,0.28125,0.203125)	Dark > positive > negative (0.546875,0.3125,0.140625)	Dark > negative > positive (0.5625,0.296875,0.140625)	Dark > positive > negative (0.515625,0.359375,0.125)

Note. Positive, negative, and dark causalities are ranked according to their importance, from which it is easy to find the dominant causality.

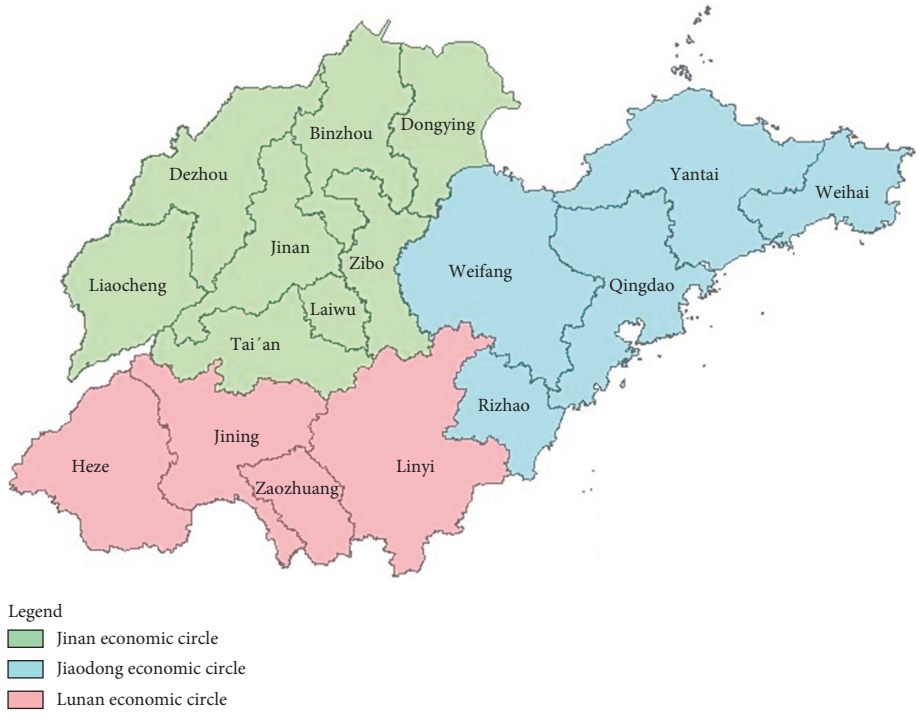


Figure 5: Three economic circles' range in Shandong Province.

Table 4: PC (from **x** to **y**) matrix in the Jinan economic circle.

(a) PC (from RTS to PU) pattern to pattern matrix

	y:↗↗	y:→↗	y:↘↗	y:↗→	y:→→	y:↘→	y:↗↘	y:→↘	y:↘↘
x1:↗↗	0.5	0	0.0625	0	0	0	0.054688	0	0.023438
x1:→↗	0	0	0	0	0	0	0	0	0
x1:↘↗	0.0625	0	0	0	0	0	0.015625	0	0
x1:↗→	0	0	0	0	0	0	0	0	0
x1:→→	0	0	0	0	0	0	0	0	0
x1:↘→	0	0	0	0	0		0	0	0
x1:↗↘	0.070313	0	0.007813		0	0	0.015625	0	0
x1:→↘	0	0	0	0	0	0	0	0	0
x1:↘↘	0.171875	0	0	0	0	0	0	0	0.015625

(b) PC (from RRUHCE to PU) pattern to pattern matrix

	y:↗↗	y:→↗	y:↘↗	y:↗→	y:→→	y:↘→	y:↗↘	y:→↘	y:↘↘
x2:↗↗	0.320313	0	0.054688	0	0	0	0.05488	0	0.007813
x2:→↗	0	0	0	0	0	0		0	0
x2:↘↗	0.148438	0	0.03125	0	0	0	0.007813	0	0.007813
x2:↗→	0	0	0	0	0	0	0	0	0
x2:→→	0	0	0	0	0	0	0	0	0
x2:↘→	0	0	0	0	0	0	0	0	0
x2:↗↘	0.164063	0	0.015625	0	0	0	0.015625	0	0
x2:→↘	0	0	0	0	0	0	0	0	0
x2:↘↘	0.140625	0	.015625	0	0	0	0.015625	0	0

(c) PC (from RRUEP to PU) pattern to pattern matrix

	y:↗↗	y:→↗	y:↘↗	y:↗→	y:→→	y:↘→	y:↗↘	y:→↘	y:↘↘
x3:↗↗	0.351563	0	0.046875	0	0	0	0.0625	0	0
x3:→↗	0	0	0	0	0	0	0	0	0
x3:↘↗	0.132813	0	0.03125	0	0	0	0.015625	0	0
x3:↗→	0	0	0	0	0	0	0	0	0
x3:→→	0	0	0	0	0	0	0	0	
x3:↘→	0	0	0	0	0	0	0	0	0
x3:↗↘	0.132813	0	0.03125	0	0	0	0.015625	0	0
x3:→↘	0	0	0	0	0	0	0	0	0
x3:↘↘	0.148438	0	0.015625	0	0	0	0.007813	0	0.007813

(d) PC (from PLCPT to PU) pattern to pattern matrix

	y:↗↗	y:→↗	y:↘↗	y:↗→	y:→→	y:↘→	y:↗↘	y:→↘	y:↘↘
x4:↗↗	029688	0	0.070313	0	0	0	0.0625	0	0.015625
x4:→↗	0	0	0	0	0	0	0	0	0
x4:↘↗	0.171875	0	0	0	0	0	0.05625	0	0
x4:↗→	0	0	0	0	0	0	0	0	0
x4:→→	0	0	0	0	0	0	0	0	0
x4:↘→	0	0	0	0	0	0	0	0	0
x4:↗↘	0.148438	0	0.007813	0	0	0	0.01562	0	0
x4:→↘	0	0	0	0	0	0	0	0	0
x4:↘↘	0.03125	0	0.015625	0	0	0	0.015625	0	0

(e) PC (from RST to PU) pattern to pattern matrix

	y:↗↗	y:→↗	y:↘↗	y:↗→	y:→→	y:↘→	y:↗↘	y:→↘	y:↘↘
x5:↗↗	0.078125	0	0.007813	0	0	0	0.015625	0	0
x5:→↗	0	0	0	0	0	0	0	0	0
x5:↘↗	0.1565	0	0.015625	0	0	0	0.015625	0	0
x5:↗→	0	0	0	0	0	0	0	0	0
x5:→→	0	0	0	0	0	0	0	0	0
x5:↘→	0	0	0	0	0	0	0	0	0
x5:↗↘	0.164063	0	0.00783	0	0	0	0.007813	0	0.007813
x5:→↘	0	0	0	0	0	0	0	0	0
x5:↘↘	0.398438	0	0.0625	0	0	0	0.046875	0	0.015625

(f) PC (from PT to PU) pattern to pattern matrix

	y:↗↗	y:→↗	y:↘↗	y:↗→	y:→→	y:↘→	y:↗↘	y:→↘	y:↘↘
x6:↗↗	0.40625	0	0.046875	0	0	0	0.046875	0	0.015625
x6:→↗	0	0	0	0	0	0	0	0	0
x6:↘↗	0.164063	0	0	0	0	0	0.007813	0	0.007813
x6:↗→	0.007813	0	0	0	0	0	0	0	0
x6:→→	0	0	0	0	0	0	0	0	0
x6:↘→	0	0	0	0	0	0	0	0	0
x6:↗↘	0.125	0	0.03125	0	0	0	0.015625	0	0
x6:→↘	0.007813	0	0	0	0	0	0	0	0
x6:↘↘	0.0625	0	0.007813	0	0	0	.015625	0	0

Note. The meaning is the same as Table 2. From the table, we can find that the dominant causalities between RRUHCE, RRUEP, PLCPT, RST, PT, and PU are dark, and the dominant causality between RTS and PU is positive.

negative, and dark PC from RRUEP to PU is 0.45, 0.2125, and 0.3375, respectively, calculated according to Table 5(c). Cumulative probability of the positive, negative, and dark PC from PLCPT to PU is 0.7125, 0, and 0.2875, respectively, calculated according to Table 5(d). Cumulative probability of the positive, negative, and dark PC from RST to PU is 0.075, 0.475, and 0.45, respectively, calculated according to Table 5(e). Cumulative probability of the positive, negative, and dark PC from PT to PU is 0.55, 0.075, and 0.375, respectively, calculated according to Table 5(f). To sum up, in addition to that the negative influence dominates in PC (from RST to PU) pattern, and the positive PC dominates in PC (from the other independent variable in **x** to **y**) pattern in the Jiaodong economic circle, as shown in Table 3.

Table 5: PC (from x to y) pattern to pattern matrix in the Jiaodong economic circle.

(a) PC (from RTS to PU) pattern to pattern matrix

	y:↗↗	y:→↗	y:↘↗	y:↗→	y:→→	y:↘→	y:↗↘	y:→↘	y:↘↘
x1:↗↗	0.4875	0	0.0125	0	0	0	0.0125	0	0
x1:→↗	0.025	0	0	0	0	0	0	0	0
x1:↘↗	0.1	0	0.0125	0	0		0	0	0
x1:↗→	0.025	0	0	0	0	0	0	0	0
x1:→→	0	0	0	0		0	0	0	0
x1:↘→	0	0	0	0	0	0	0	0	0
x1:↗↘	0.125	0	0	0	0	0	0	0	0
x1:→↘	0	0	0	0	0	0	0	0	0
x1:↘↘	0.2	0	0	0	0	0	0	0	0

(b) PC (from RRUHCE to PU) pattern to pattern matrix

	y:↗↗	y:→↗	y:↘↗	y:↗→	y:→→	y:↘→	y:↗↘	y:→↘	y:↘↘
x2:↗↗	0.475	0	0	0	0	0	0.012	0	0
x2:→↗	0	0	0	0	0	0	0	0	0
x2:↘↗	0.1875	0	0.0125	0	0	0	0	0	0
x2:↗→	0	0	0	0	0	0	0	0	0
x2:→→	0	0	0	0	0		0	0	0
x2:↘→	0	0	0	0	0	0	0	0	0
x2:↗↘	0.1625	0	0.0125	0	0	0	0	0	0
x2:→↘	0	0	0	0	0	0	0	0	0
x2:↘↘	0.1375	0		0	0	0		0	0

(c) PC(from RRUEP to PU) pattern to pattern matrix

	y:↗↗	y:→↗	y:↘↗	y:↗→	y:→→	y:↘→	y:↗↘	y:→↘	y:↘↘
x3:↗↗	.45	0	0.0125	0	0	0	0.0125	0	0
x3:→↗	0	0	0	0	0	0	0	0	0
x3:↘↗	0.1625	0	0	0	0	0	0	0	0
x3:↗→	0	0	0	0	0	0	0	0	0
x3:→→	0	0	0	0	0	0	0	0	0
x3:↘→	0	0	0	0	0	0	0	0	0
x3:↗↘	0.1375	0	0	0	0	0	0	0	0
x3:→↘	0	0	0	0	0	0	0	0	0
x3:↘↘	0.2125	0	0.0125	0	0	0	0	0	0

(d) PC (from PLCPT to PU) pattern to pattern matrix

	y:↗↗	y:→↗	y:↘↗	y:↗→	y:→→	y:↘→	y:↗↘	y:→↘	y:↘↘
x4:↗↗	0.7125	0	0.0125	0	0	0	0.0125	0	0
x4:→↗	0	0	0	0	0	0	0	0	0
x4:↘↗	0.1375	0	0	0	0	0	0	0	0
x4:↗→	0	0	0	0	0	0	0	0	0
x4:→→	0	0	0	0	0	0	0	0	0
x4:↘→	0	0	0	0	0	0	0	0	0
x4:↗↘	0.125	0	0	0	0	0	0	0	0
x4:→↘	0	0	0	0	0	0	0	0	0
x4:↘↘	0	0	0	0	0	0	0	0	0

(e) PC (from RST to PU) pattern to pattern matrix

	y:↗↗	y:→↗	y:↘↗	y:↗→	y:→→	y:↘→	y:↗↘	y:→↘	y:↘↘
x5:↗↗	.075	0	0	0	0	0	0	0	0
x5:→↗	0	0	0	0	0	0	0	0	0
x5:↘↗	0.225	0	0	0	0		0	0	0
x5:↗→	0	0	0	0	0	0	0	0	0
x5:→→	0	0	0	0	0	0	0	0	0
x5:↘→	0	0	0	0	0	0	0	0	0
x5:↗↘	0.2	0	0	0	0	0	0	0	0
x5:→↘	0	0	0	0	0	0	0	0	0
x5:↘↘	0. 75	0	0.0125	0	0	0	0.0125	0	0

(f) PC (from PT to PU) pattern to pattern matrix

	y:↗↗	y:→↗	y:↘↗	y:↗→	y:→→	y:↘→	y:↗↘	y:→↘	y:↘↘
x6:↗↗	0.55	0	0.0125	0	0	0	0.0125	0	0
x6:→↗	0	0	0	0	0	0	0	0	0
x6:↘↗	0.1625	0	0	0	0	0	0	0	0
x6:↗→	0	0	0	0	0	0	0	0	0
x6:→→		0	0	0		0	0	0	0
x6:↘→	0	0	0	0	0	0	0	0	0
x6:↗↘	0 1875	0	0	0	0	0	0	0	0
x6:→↘	0	0	0	0	0	0	0	0	0
x6:↘↘	0.075	0	0	0	0		0	0	0

Note. The meaning is the same as Table 2. From the table, we can find that the dominant causalities between RTS,RRUHCE, RRUEP, PLCPT, PT, and PU are positive, and the dominant causality between RST and PU is negative.

Table 6: PC (from **x** to **y**) pattern to pattern matrix in the Lunan economic circle.

(a) PC (from RTS to PU) pattern to pattern matrix

	y:↗↗	y:→↗	y:↘↗	y:↗→	y:→→	y:↘→	y:↗↘	y:→↘	y:↘↘
x1:↗↗	0.□75		0.09375	0	0	0	0.109375	0	0.015625
x1:→↗	0	0	0	0	0	0	0	0	0
x1:↘↗	0.09375	0	0	0	0	0	0.015625	0	0
x1:↗→	0	0	0	0	0	0	0	0	0
x1:→→	0	0	0	0	0	0	0	0	0
x1:↘→	0	0	0	0	0	0	0	0	0
x1:↗↘	0.0625	0	0.03125	0	0	0	0.015625	0	0
x1:→↘	0	0	0	0	0	0	0	0	0
x1:↘↘	0.1□5	0	0.03125	0	0	0	0.03125	0	0

(b) PC (from RRUHCE to PU) pattern to pattern matrix

	y:↗↗	y:→↗	y:↘↗	y:↗→	y:→→	y:↘→	y:↗↘	y:→↘	y:↘↘
x2:↗↗	0.21875	0	0.0625	0	0	0	0.078125	0	0.015625
x2:→↗	0	0	0	0	0	0	0	0	0
x2:↘↗	0.140625	0	0.046875	0	0	0	0.015625	0	0.0156□5
x2:↗→	0	0	0	0	0	0	0	0	0
x2:→→	0	0	0	0		0	0	0	0
x2:↘→		0	0	0	0	0	0	0	0
x2:↗↘	0.140625	0	0.03125	0	0	0	0.03125	0	0
x2:→↘	0	0	0	0	0	0	0	0	0
x2:↘↘	0.140625	0	0.03125	0	0	0	0.03125	0	0

(c) PC (from RRUEP to PU) pattern to pattern matrix

	y:↗↗	y:→↗	y:↘↗	y:↗→	y:→→	y:↘→	y:↗↘	y:→↘	y:↘↘
x3:↗↗	0.2	0	0□046875	0	0	0	0.09375	0	0
x3:→↗	0	0	0	0	0	0	0	0	0
x3:↘↗	0.140625	0	0.015625	0	0	0	0.03125	0	0
x3:↗→	0	0	0	0	0	0	0	0	0
x3:→→	0	0	0	0	0	0	0	0	0
x3:↘→	0	0	0	0	0	0	0	0	0
x3:↗↘	0.109375	0	0.03125	0	0	0	0.01□625		0.015625
x3:→↘	0	0	0	0	0	0	0	0	0
x3:↘↘	0□140625	0	0□078125	0	0	0	0.03125	0	0

(d) PC (from PLCPT to PU) pattern to pattern matrix

	y:↗↗	y:→↗	y:↘↗	y:↗→	y:→→	y:↘→	y:↗↘	y:→↘	y:↘↘
x4:↗↗	0.21875	0	0.0625	0	0	0	0.0625	0	0.015625
x4:→↗	0	0	0	0	0	0	0	0	0
x4:↘↗	0.1875	0	0.0625	0	0	0	0.015625	0	0.01□625
x4:↗→	0	0		0	0	0	0	0	0
x4:→→	0	0	0	0	0	0	0	0	0
x4:↘→	0	0	0	0	0	0	0	0	0
x4:↗↘	0.203125	0	0.046875	0	0	0	0.03125	0	0
x4:→↘	0	0	0	0	0	0	0	0	0
x4:↘↘	0.0625	0	0	0	0	0	0.015625	0	0

(e) PC (from RST to PU) pattern to pattern matrix

	y:↗↗	y:→↗	y:↘↗	y:↗→	y:→→	y:↘→	y:↗↘	y:→↘	y:↘↘
x5:↗↗	0.078125	0	0.03125	0	0	0	0.015625	0	0.015625
x5:→↗	0	0	0	0	0	0	0	0	0
x5:↘↗	0.1875	0	0.03125	0	0	0	0.015625	0	0.015625
x5:↗→	0	0	0	0	0	0	0	0	0
x5:→→	0	0	0	0	0	0	0	0	0
x5:↘→	0	0	0	0	0	0	0	0	0
x5:↗↘	0.15625	0	0.03125	0	0	0	0.□3125	0	0
x5:→↘	0	0	0	0	0	0	0	0	0
x5:↘↘	0.234375	0	0.0625	0	0	0	0.078125	0	0.015625

(f) PC (from PT to PU) pattern to pattern matrix

	y:↗↗	y:→↗	y:↘↗	y:↗→	y:→→	y:↘→	y:↗↘	y:→↘	y:↘↘
x6:↗↗	0.3125	0	0.109375	0	0	0	0□09375	0	0
x6:→↗	0	0	0	0	0	0	0	0	0
x6:↘↗	0□125	0	0.0□5625	0	0	0	0.03125	0	0
x6:↗→	0	0	0	0	0	0	0	0	0
x6:→→	0	0	0	0	0	0	0	0	0
x6:↘→	0	0	0	0	0	0	0	0	0
x6:↗↘	0.15625	0	0.03125	0	0	0	0.03125	0	0
x6:→↘	0	0	0	0	0	0	0	0	0
x6:↘↘	0.0625	0	0.015625	0	0		0.015□25	0	0

Note. The meaning is the same as Table 2. From the table, we can find that the dominant causalities between RST, RRUHCE, RRUEP, PLCPT, RST, PT, and PU are dark.

4.3.3. Analysis of Pattern Causality Matrix in the Lunan Economic Circle. PC (from **x** to **y**) pattern to pattern matrix in the Lunan economic circle is shown in Table 6. Cumulative probability of the positive, negative, and dark PC from RST to PU is 0.390625, 0.1875, and 0.421875, respectively, calculated according to Table 6(a). Cumulative probability of the positive, negative, and dark PC from RRUHCE to PU is 0.296875, 0.203125, and 0.5, respectively, calculated according to Table 6(b). Cumulative probability of the positive, negative, and dark PC from RRUEP to PU is 0.28125, 0.203125, and 0.515625, respectively, calculated according to Table 6(c). Cumulative probability of the positive, negative, and dark PC from PLCPT to PU is 0.3125, 0.140625, and 0.546875, respectively, calculated according to Table 6(d). Cumulative probability of the positive, negative, and dark PC from RST to PU is 0.140625, 0.296875, and 0.5625, respectively, calculated according to Table 6(e). Cumulative probability of the positive, negative, and dark PC from PT to PU is 0.359375, 0.125, and 0.515625, respectively, calculated according to Table 6(f). To sum up, the

dark PC dominates in PC (from **x** to **y**) pattern in the Lunan economic circle, as shown in Table 3.

4.4. Comparisons at Different Regional Levels. Table 3 summarizes the above-mentioned PC patterns at different regional levels. In general, except for the Jiaodong economic circle, the dark influence dominates in the PC patterns from **x** to **y**, which proves that there is a nonpositive/negative causality in the processes of urbanization and economic growth. However, in the Jiaodong economic circle, the dominant causality between **x** and **y** is still a positive influence, which indicates that the increase in the relevant economic indicators can obviously drive the development of urbanization and economic growth. In addition to the dominant causality, the relationship between RTS and PU is dominated by positive causality in the Jinan economic circle, indicating that promoting the developing of tertiary industry can obviously drive the development of urbanization and economic growth. The cases that the negative influence is stronger than the positive influence are observed in the causality between RST and PU and in the causality between RRUEP and PU at the provincial level. The former indicates that the decline in RST can promote urbanization and economic growth, which means that it needs to increase number of full-time teachers and optimize educational resources at the provincial level. The emergence of the latter is mainly due to the large differences between dark and positive causality in different regions, indicating that there is a large spatial difference in employment participation rates between urban and rural population, and its policy implication is to promote the employment of rural population in regions dominated by dark causality.

There is a clear interaction between urbanization, economic growth, and industrial structure in the region [44], and such interaction shows temporal and spatial differences. The processes of urbanization and economic growth start with early industrialization when people migrate from rural to urban areas [45]. However, as time goes by, industrialization has limited attractiveness to the rural population and brings about the problems such as waste of resources, increase of labor cost, and environmental damage which hinder the process of urbanization and economic growth [46]. At this period, upgrading the industrial structure and promoting the development of tertiary industry, which is strongly attractive to rural labor force, become the driving forces for urbanization and economic growth [47]. In the Jinan and Jiaodong economic circles, the development of tertiary industry plays a stronger role in promoting urbanization and economic growth, resulting in the fact that the positive causality dominates in the relationship between RST and PU. However, the industrial structure level of the Lunan economic circle is lower than that of the above two regions, which means that the secondary industry development attracts less rural population, and the tertiary industry development is relatively weak, leading to dark causality as the dominant relationship between RST and PU. The above results lead to the fact that the dark and positive causalities dominate at the provincial level.

According to the traditional view, urbanization led to widening urban-rural gap and prompting population migration from rural to urban areas [48], so as to result in the problems, such as the loss of rural labor force and hollow villages. The above problems are not conducive to agricultural and rural development, which forces the proposal of new urbanization. The research proves that there is a significant correlation between new urbanization and narrowing of the urban-rural gap [49]. In this paper, the causality between RRUHCE (gap between urban and rural residents' consumption), RRUEP (employment gap between urban and rural residents), and PU reflects the above correlation. (1) In the Jinan and Lunan economic circles, the gaps of urban-rural residents' consumption and employment are large, and the increase of the rural residents' consumption and employment positively causes the urbanization development and economic growth, but the positive causality is not continuous in time, finally resulting in the fact that dark causality is dominant among RRUHCE, RRUEP, and PU, followed by positive causality. (2) However, in Jiaodong economic circle, the consumption gap and employment gap between urban and rural residents are narrowing, and there is an obvious positive causality between the increases of rural residents' consumption and employment and urbanization and economic growth. As a result, the relationships between RRUHCE, RRUEP, and PU are dominated by positive causality. (1) and (2) result in the fact that the dark causality becomes the dominant relationship between RRUHCE, RRUEP, and PU at the provincial level.

With the deepening of urbanization and economic growth, the role of "people" becomes increasingly prominent, so education plays an increasingly important role in urbanization and economic growth [50]. But in different stages of economic development, the role of education is not the same. Generally speaking, when the level of economic development is low, the role of basic education is greater, while the role of higher education is relatively small, but when the level of economic development is high, the role of basic education is relatively small, while the role of higher education is greater [51]. In our paper, this relationship is proved by the causality between RST and PU. (1) In Jiaodong economic circle with a relatively high level of economic development, basic education is developed well and the allocation of basic education resources is relatively reasonable. As a result, the negative causality is dominant relationship between RST and PU. However, due to the differences between cities within the region, there is also an obvious dark causality between RST and PU. (2) In Jinan economic circle, the economic development levels of cities are quite different. There is a big gap between the basic education development of Dongying, Jinan, Zibo, and Tai 'an, which are relatively developed, and that of Laiwu, Binzhou, Liaocheng, and Dezhou, which are relatively backward. As a result, dark causality and negative causality are dominant relationships between RST and PU. (3) The overall economic level of Lunan economic circle is relatively low, the development of basic education is relatively backward, and there are large gaps between cities. As a result,

in Lunan economic circle, the dark causality is a dominant relationship between RST and PU, and negative causality is obviously weaker than that of Jinan economic circle. (1), (2), and (3) result in the fact that dark causality dominates in the relationship between RST and PU, followed by negative causality at the provincial level.

PLCPT, as an auxiliary indicator for improving population quality, is often used as a basic indicator in the index system to measure the level of urbanization together with education indicators in research [31]. The impact of PLCPT on urbanization and economic growth is similar to that of higher education, which makes the conclusions of the causality between PLCPT and PU similar to the causality between RST and PU in this paper, as follows:

(1) In Jiaodong economic circle with high level of economic development, there is a strong positive causality between PLCPT and PU.

(2) In Jinan economic circle, due to the large gap in the economic development level of internal cities, dark causality and positive causality are dominant.

(3) In Lunan economic circle with low level and small gap of economic development, the dark causality is the dominant relationship between PLCPT and PU. (1), (2), and (3) result in the fact that at the provincial level, the relationship between PLCPT and PU is dominated by dark causality, followed by positive causality.

As shown in existing works, population mobility and urbanization have made important contributions to the rapid economic growth [52, 53]. Generally speaking, the population usually flows from the areas with low economic development level to the areas with high economic development level [54], which makes the population flow in the areas with high economic development level larger and that in the areas with low economic development level relatively small. As a result, there is an obvious positive causality between population mobility and urbanization and economic growth. In this paper, such relationship is shown as follows:

(1) In the Jiaodong economic circle with high economic development level, the relationship between PT and PU is dominated by positive causality based on PT: ↗↗⟶PU:↗↗causal pattern.

(2) In the two economic circles of Jinan and Lunan, the positive causality between PT and PU was weakened due to the large differences among cities within them, showing that the dark causality is dominant. (1) and (2) result in the fact that the dark causality becomes the dominant relationship between PT and PU at the provincial level.

5. Conclusion

This paper introduces a method of nearest neighbor phase space reconstruction based on symbolic dynamics to study the causality between urbanization and economic growth at different regional levels in Shandong Province, and the conclusions are as follows:

(1) The analysis of the correlation between per capita GDP and population urbanization rate shows that there is a strong positive relationship between urbanization and economic growth both on the whole (Figure 1) and at the municipal level (Figure 2), indicating that the urbanization process in Shandong Province plays a significant role in driving economic growth both. In other words, it is feasible to drive economic growth by promoting urbanization development.

(2) The correlation analysis of evaluation indexes (Figures 3 and 4) shows that urbanization has a strong positive correlation with public finance, medical and health, education expenditure, and fixed asset investment and has a strong negative relationship with RRUE but has no obvious correlation with RTS, RRUHCE, RRUEP, PLCPT, RST, and PT, indicating that increasing the positive relationship variables and reducing the negative relationship variables can promote the development of urbanization and economic growth. The specific policy suggestions are as follows:

(1) The government should appropriately expand budgetary expenditures (including the expenditures in education, medical, and health) and optimize the structure of fiscal expenditures to strengthen government functions.

(2) Increasing investment in fixed assets to promote the upgrading of industrial structure.

(3) Employment opportunities or preferential policies for rural residents are provided to increase their income and consumption level, so as to narrow the urban-rural gap and promote high-quality development of urbanization.

(3) The analysis of pattern causality matrix indicates that there is a dark causality in the processes of urbanization and economic growth, and except for Jiaodong economic circle, the dark influence dominates in the PC patterns from **x** to **y**. The positive influence dominates in the Jiaodong economic circle. In addition, the types of causality between the same evaluation index and PU in different regions are different. The relationship between RTS and PU is dominated by positive causality in the Jinan economic circle and the Jiaodong economic circle, while in the other two regions, dominant influence is dark causality. The relationship between RST and PU is dominated by negative causality in the Jiaodong economic circle, while in the other regions, dominant influence is dark causality. Based on the findings, policy implications about the driving factors of urbanization developing and economic growth have been proposed not only at the provincial scale but also on the three economic circles.

Although we discover the complex causalities in the processes of urbanization and economic growth and put forward some policy recommendations based on the results, other problems, such as "how these policies are implemented" and "which one is the most scientific and feasible," have not been estimated yet, and these problems would be studied in forthcoming papers.

Data Availability

The data used to support the findings of this study are included within the article.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This research was supported by the Natural Science Foundation of Shandong Province (ZR2020MG004) and Soft Science Planning Program of Shandong Province (2020RKA07033).

References

[1] G. U. Yule, "Why do we Sometimes get Nonsense-Correlations between Time-Series?--A Study in Sampling and the Nature of Time-Series," *Journal of the Royal Statistical Society*, vol. 89, no. 1, pp. 1–63, 1926.

[2] K. Pearson, "Notes on regression and inheritance in the case of two parents," *Proceedings of the Royal Society of London*, vol. 58, pp. 240–242, 1895.

[3] C. W. J. Granger, "Investigating causal relations by econometric models and cross-spectral methods," *Econometrica*, vol. 37, no. 3, pp. 424–438, 1969.

[4] X. Xu, Z. Lin, X. Li, C. Shang, and Q. Shen, "Multi-objective robust optimisation model for MDVRPLS in refined oil distribution," *International Journal of Production Research*, vol. 202121 pages, 2021.

[5] Pragati, "Re-examining the relationship between urbanization and economic growth: an analysis," *Journal of Resources, Energy and Development*, vol. 14, no. 2, pp. 73–82, 2018.

[6] Y. Hong, "Study on the relationship between population urbanization and economic growth, industrial structure: taking bijie city as an example," in *Proceedings of the 1st International Seminar on Education, Innovation and Economic Management (SEIEM, 2016)*, pp. 126–129, Chongqing, China, December 2016.

[7] H. M. Nguyen and L. D. Nguyen, "The relationship between urbanization and economic growth," *International Journal of Social Economics*, vol. 45, no. 2, pp. 316–339, 2018.

[8] R. F. Engle and C. W. J. Granger, "Co-integration and error correction: representation, estimation, and testing," *Econometrica*, vol. 55, no. 2, pp. 251–276, 1987.

[9] D. M. Ibrahiem, "Road energy consumption, economic growth, population and urbanization in Egypt: cointegration and causality analysis," *Environment, Development and Sustainability*, vol. 20, no. 3, pp. 1053–1066, 2018.

[10] Z. Ahmed, M. M. Asghar, M. N. Malik, and K. Nawaz, "Moving towards a sustainable environment: the dynamic linkage between natural resources, human capital, urbanization, economic growth, and ecological footprint in China," *Resources Policy*, vol. 67, Article ID 101677, 2020.

[11] T. Schreiber, "Measuring information transfer," *Physical Review Letters*, vol. 85, no. 2, pp. 461–464, 2000.

[12] W. Guan, Y. Yao, X. Peng, Y. Wei, and H. Zhang, "The relationship of urbanization and economic growth in China based on the provincial panel data in 1978-2014," *Scientia Geographica Sinica*, vol. 36, no. 6, pp. 813–819, 2016.

[13] H. Qi, X. Xi, and Q. Gao, "Measurement on the development level of China's urbanization and the time—varying characteristics of the effects of economic growth," *Economist*, vol. 11, no. 4, pp. 26–34, 2015.

[14] S. Tripathi and K. Mahey, "Urbanization and economic growth in Punjab (India): an empirical analysis," *Urban Research & Practice*, vol. 10, no. 4, pp. 379–402, 2017.

[15] Z. L. Ma and H. Y. Ma, "Dynamic impact analysis of urbanization progress and industrial structure change on VAR model," *Applied Mechanics and Materials*, vol. 631-632, pp. 23–26, 2014.

[16] J. R. Faria and A. Mollick, "Urbanization, economic growth, and welfare," *Economics Letters*, vol. 52, no. 1, pp. 109–115, 1996.

[17] J. Pearl, "Causality," *Uspekhi Khimii: Cambridge University Press*, vol. 19, p. 2, 2009.

[18] X. Xu, Z. Lin, and J. Zhu, "DVRP with limited supply and variable neighborhood region in refined oil distribution," *Annals of Operations Research*, vol. 309, no. 2, pp. 663–687, 2022.

[19] Y. Liu, M. Liu, G. Wang, L. Zhao, and P. An, "Effect of environmental regulation on high-quality economic development in China-an empirical analysis based on dynamic spatial Durbin model," *Environmental Science and Pollution Research*, vol. 28, no. 39, 54678 pages, Article ID 54661, 2021.

[20] X. Zhang, Z. Gang, and X. Dong, "Effects of government healthcare expenditure on economic growth based on spatial Durbin model: evidence from China," *Iranian Journal of Public Health*, vol. 49, no. 2, pp. 283–293, 2020.

[21] G. Sugihara, R. May, H. Ye et al., "Detecting causality in complex ecosystems," *Science*, vol. 338, no. 6106, pp. 496–500, 2012.

[22] S. K. Stavroglou, A. A. Pantelous, H. E. Stanley, and K. M. Zuev, "Unveiling causal interactions in complex systems," *Proceedings of the National Academy of Sciences*, vol. 117, no. 14, pp. 7599–7605, 2020.

[23] I. Turok and G. McGranahan, "Urbanization and economic growth: the arguments and evidence for Africa and Asia," *Environment and Urbanization*, vol. 25, no. 2, pp. 465–482, 2013.

[24] Z. Yang, X. Zhang, J. Li, J. Lei, and Z. Duan, "The spatial-temporal pattern of the relationship between urbanization and economic development at prefecture-level units in China: a quantitative analysis based on 2000 and 2010 census data," *Geographical Research*, vol. 39, no. 1, pp. 25–40, 2020.

[25] C.-W. Su, T.-Y. Liu, H.-L. Chang, and X.-Z. Jiang, "Is urbanization narrowing the urban-rural income gap? A cross-regional study of China," *Habitat International*, vol. 48, pp. 79–86, 2015.

[26] S. Hao, "Analysis of the inverted U-shaped relationship between China's new urbanization and economic growth -based on Comparison with traditional urbanization," *International Business and Management*, vol. 16, no. 2, pp. 27–33, 2018.

[27] X. Xu, C. Wang, and P. Zhou, "GVRP considered oil-gas recovery in refined oil distribution: from an environmental

perspective," *International Journal of Production Economics*, vol. 235, Article ID 108078, 2021.

[28] S. K. Stavroglou, A. A. Pantelous, H. E. Stanley, and K. M. Zuev, "Hidden interactions in financial markets," *Proceedings of the National Academy of Sciences*, vol. 116, no. 22, pp. 10646–10651, 2019.

[29] A. F. Ades and E. L. Glaeser, "Trade and circuses: explaining urban giants," *The Quarterly Journal of Economics*, vol. 110, no. 1, pp. 195–227, 1995.

[30] G. Ma and Y. Wang, "Research on the influencing factors of new urbanization quality based on the Comparison between the eastern and western regions," *Forecasting*, vol. 40, no. 6, pp. 61–67, 2021.

[31] H. Dong, Q. Ru, Q. Fu et al., "Research on new urbanization in Shaanxi Province under the Rural Revitalization Strategy : Take Hanzhong city as a case," *E3S Web of Conferences*, vol. 185, Article ID 02037, 2020.

[32] X. Xu, J. Hao, and Y. Zheng, "Multi-objective artificial bee Colony algorithm for multi-stage resource leveling problem in sharing logistics network," *Computers & Industrial Engineering*, vol. 142, no. 4, Article ID 106338, 2020.

[33] J. Tang, Q. He, and Y. Liu, "Spatiotemporal dislocation and influencing factors of urbanization quality and scale in Yangtze River Delta urban agglomeration," 2021, https://kns.cnki.net/kcms/detail/42.1178.N.20211207.0205.004.html.

[34] X. Xu and Y. Wang, "Analysis on the coupling and coordinated development and influencing factors of new urbanization, rural revitalization and economic growth quality in Chinese provinces," *Inquiry Into Economic Issues*, vol. 10, pp. 13–26, 2021.

[35] F. Zhang, "Research on the dynamic factors of new urbanization in hebei province based on factor regression," *Statistics and Applications*, vol. 10, no. 5, pp. 761–769, 2021.

[36] J. Liu and H. Shi, "Analysis on the dynamic changes of the income gap between urban and rural residents in qinghai province," *International Journal of Educational Management*, vol. 7, no. 1, pp. 183–185, 2022.

[37] H. Wei and H. Yao, "Environmental regulation, roundabout production, and industrial structure transformation and upgrading: evidence from China," *Sustainability*, vol. 14, no. 7, p. 3810, 2022.

[38] J. Wang and K. Zhao, "Study on the impact of urbanization and aging on the consumption gap between urban and rural residents: theoretical model and empirical analysis," *Journal of Yunnan University of Finance and Economics*, vol. 36, no. 8, pp. 38–54, 2020.

[39] M. Qiao and S. Qin, "Dynamic evolution and threshold effects of population structure and urban-rural consumption inequality," *Journal of Technical Economics & Management*, vol. 5, pp. 98–102, 2018.

[40] W. Liu and Y. Han, "Urban-rural non-agricultural employment structure, population transfer mode and urbanization level: based on China's data," *Statistics & Information Forum*, vol. 29, no. 8, pp. 85–92, 2014.

[41] W. Zhang, W. Zeng, and X. Xu, "The dynamic relationship between employment structure optimization and urban-rural residential income distribution—empirical analysis of PVAR model based on Chinese data," *Reformation & Strategy*, vol. 32, no. 1, pp. 132–137, 2016.

[42] M. Zhu, "The study and implementation of discretization algorithm based on information entropy," *Software Engineering and Applications*, vol. 8, no. 6, pp. 358–363, 2019.

[43] U. M. Fayyad and K. B. Irani, "Multi-interval discretization of continuous-valued attributes for classification learning," in *Proceedings of the 13th international joint Conference on artificial intelligence (IJCAI-93)*, pp. 1022–1027, Chambèry, France, August 1993.

[44] Y. Song and Z. Zhou, "Study on the coordination relationship between new-type urbanization and industrial structure in south anhui," *World Scientific Research Journal*, vol. 6, no. 6, pp. 86–94, 2020.

[45] H. M. Siddaram and H. H. Bharadi, "Understanding the processes of urbanization and spatial concentration in India: a study of Karnataka," *International Journal of Research in Social Sciences*, vol. 7, no. 11, pp. 143–151, 2017.

[46] Y. Hao, S. Zheng, M. Zhao, H. Wu, Y. Guo, and Y. Li, "Reexamining the relationships among urbanization, industrial structure, and environmental pollution in China-New evidence using the dynamic threshold panel model," *Energy Reports*, vol. 6, pp. 28–39, 2020.

[47] X. Han, P. L. Wu, and W. L. Dong, "An analysis on interaction mechanism of urbanization and industrial structure evolution in Shandong, China," *Procedia Environmental Sciences*, vol. 13, pp. 1291–1300, 2012.

[48] D. Hu, "Trade, rural-urban migration, and regional income disparity in developing countries: a spatial general equilibrium model inspired by the case of China," *Regional Science and Urban Economics*, vol. 32, no. 3, pp. 311–338, 2002.

[49] J. Li, T. S. Cheong, J. Shen, and D. Fu, "Urbanization and Rural-urban Consumption Disparity: Evidence from China," *The Singapore Economic Review*, vol. 64, no. 4, pp. 983–996, 2019.

[50] T. Xie and W. Yu, "Whether the development of education will lead to a balanced urbanization of inter-provincial population - empirical analysis based on conditional convergence model of inter-provincial panel data," *Macroeconomics*, vol. 8, pp. 107–114, 2021.

[51] P. E. Petrakis and D. Stamatakis, "Growth and educational levels: a comparative analysis," *Economics of Education Review*, vol. 21, no. 5, pp. 513–521, 2002.

[52] F. Cai, "Urbanization and the contribution of migrant workers - a reflection on the potential of china's economic growth in the post-crisis period," *Chinese Journal of Population Science*, vol. 111, no. 1, pp. 2–10, 2010.

[53] G. Shi and Z. Li, "Research on the mechanism of population mobility promoting regional economic growth:based on panel data of the yangtze river delta urban agglomeration," *East China Economic Management*, vol. 34, no. 6, pp. 10–18, 2020.

[54] G. Wang, "New trends in migration and urbanization in China: a preliminary investigation based on the seventh census data," *Population & Economics*, vol. 5, pp. 36–55, 2021.

Hindawi
Scientific Programming
Volume 2022, Article ID 9616454, 10 pages
https://doi.org/10.1155/2022/9616454

Research Article

Design and Numerical Simulation of an Odometer Wheel Used in an Ultrasonic In-Line Inspection Tool

Liangxue Cai,[1,2,3] **Liang Guan**,[1] **Xu Qin**,[1] **Lipeng Chen**,[1] **and Guangli Xu**[1,3]

[1]*School of Oil & Natural Gas Engineering, Southwest Petroleum University, Chengdu, China*
[2]*Shandong Key Laboratory of Oil & Gas Storage and Transportation Safety, Qingdao, China*
[3]*Oil & Gas Fire Protection Key Laboratory of Sichuan Province, Chengdu, China*

Correspondence should be addressed to Liangxue Cai; cailiangxue-184@163.com and Guangli Xu; 530xugl@163.com

Received 1 May 2022; Accepted 11 June 2022; Published 28 June 2022

Academic Editor: Xiaofeng Xu

An odometer wheel is used to measure forward distance for a piezoelectric ultrasonic device, and it plays a key role in locating defections. Buckling and skidding are its main problems in applications. Based on a self-developed in-line inspection tool, an odometer wheel was designed according to mechanical design principles. A finite element model on scale of 1 : 1 was built to simulate mechanical performance of the odometer wheel. Results show that the maximum Mises stress under the worst condition is far less than the elastic limit of chosen stainless steel, and deformations of swing arms are slight. Acting forces under different conditions also meet requirements of antiskid. It indicates that the designed object has good performance in terms of structure stability and skid prevention.

1. Introduction

Along with the rapid development of oil and gas pipeline network in China, upgrading from digitization to intelligentization is an apparent trend in pipeline constructions and managements [1]. Nondestructive testing (NDT) can supply pipe's basic data, and it plays an important role in developing intelligent pipelines. The ultrasonic in-line inspection technology is an efficient way to detect cracks [2] which appear and grow in pipes exposed to various environmental media [3] and loading conditions [4]. Several types of ultrasonic inspection tools are in progress of developing. A typical piezoelectric ultrasonic device usually contains ultrasonic probe array, data acquisition system (DAS), odometer wheel, driving cups, and battery [5]. An odometer wheel is used to locate flaws during in-line inspection [6], and it plays an important role in processing inspection data [7]. Structural stability and skid prevention are essential factors to guarantee an odometer wheel's normal performance, and hence they should be analyzed particularly in designing an odometer wheel.

Zang et al. [8] studied the reason why an internal detector's odometer wheel skids, and they suggested that the diameter of an odometer wheel should be as small as possible on the premise that an odometer wheel meets other requirements. Li et al. [9] analyzed the kinematics of an odometer wheel through numerical simulation. It is found that the odometer wheel causes a larger odometer error when the speed of its roller's rotation speed is higher when it passes through a girth weld. Xie et al. [10] investigated the preprocessing algorithm of odometer wheel positioning data, and they proposed a preprocessing algorithm for in-line detector odometer wheel positioning data based on the Kalman filter. The accuracy of locating a detector was improved effectively [11]. The abovementioned studies mainly focus on improving the performance of an odometer wheel's locating function, while the structural stability of an odometer wheel as well as its antiskid performance [12] is not analyzed considering extreme conditions during in-line inspection [13].

Based on parameters of a self-developed piezoelectric ultrasonic device for detecting pipes with diameter of 219 mm, an overall structure of a relative odometer wheel was designed according to mechanical design principles and actual application conditions. A finite element model was built using ANSYS to analyze its structural stability under

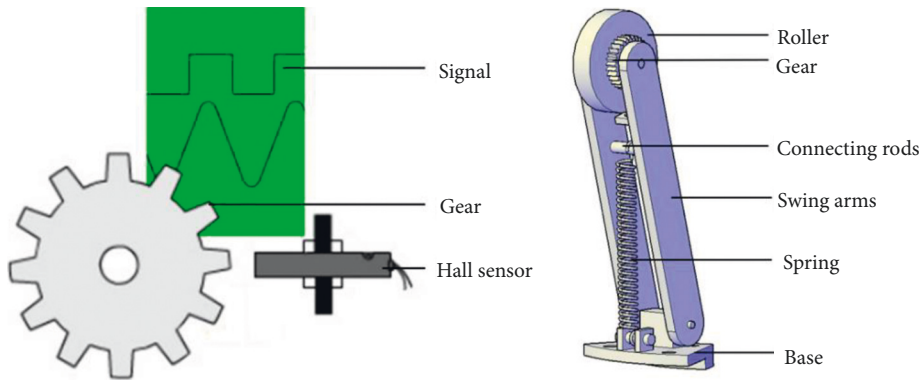


Figure 1: Structure of an odometer wheel.

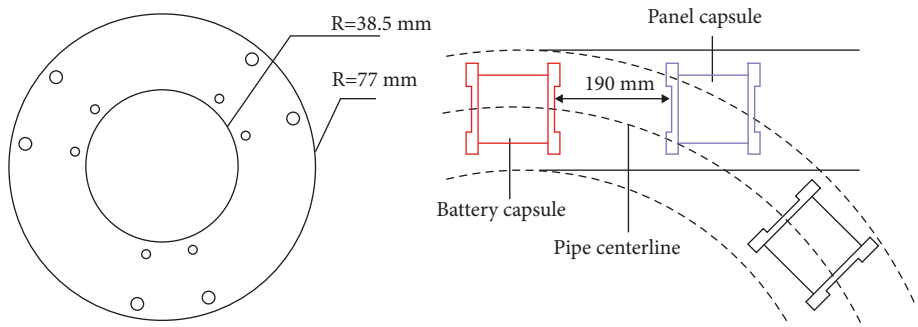


Figure 2: Structure of a battery capsule's end.

four kinds of application conditions. Acting forces on a pipe's inner wall by the roller were discussed to check its antiskid performance.

2. Structural Design of an Odometer Wheel

An ultrasonic in-line inspection tool is pushed forward by pressure difference between its front and back in oil and gas pipes. As a spring is used in an odometer wheel, it pulls the swing arm outward so that a roller installed at an end of the swing arm can press tightly against the inner wall of a pipe. A gear that rotates synchronously is set on one side of the roller, and a Hall sensor is used to capture the rotation signal of a gear [14], as shown in Figure 1. After the signal being processed and stored, it is converted into a required odometer signal, and finally we can get the locating data.

A typical odometer wheel mainly contains swing arm, roller, gear, mounting base, spring, sensor, connecting rods, etc. [15]. No buckling of an odometer wheel and roller skidding on the inner wall of a pipe is the premise of an odometer wheel's good performance [16]. The self-developed piezoelectric ultrasonic device size is available for pipes $\Phi 219 \times 7.5$ mm, and its minimum pipeline bend radius is $3D \times 90^{\circ}$. An odometer wheel is placed at the end of a battery capsule, and Figure 2 shows its structure and sizes. The mounting base is connected with battery capsule's end face and an annular plate fixing a driving cup. In constructing projects of oil and gas pipelines, the usual minimum radius of pipelines is recommended as 5D [17], and it should not be less than 4D [18], which are 1095 mm and 876 mm for 219 mm pipes, respectively.

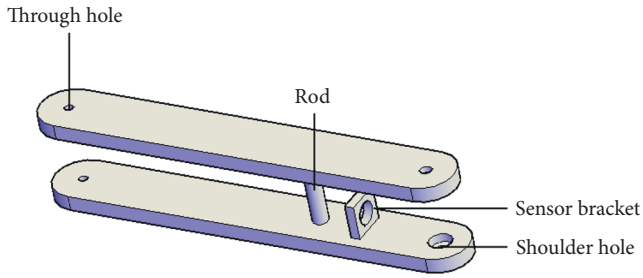


Figure 3: Structure of swing arms.

2.1. Swing Arm. A Swing arm is designed to keep tight contact between a roller and a pipe's inner wall during in-line inspection. It is a pendulum component and can be damaged with high probability in applications. In order to enhance the reliability, a kind of double swing arms structure is adopted to improve its stability in design (Figure 3). Structures of two swing arms are same, and they are set symmetrically. Two rods located at one end and middle part of arms are used to connect these two arms. The latter rod serves as a fulcrum for the spring as well. An axle is set at one end to support the roller and gear.

The free space between two adjacent capsules and extreme conditions in-situ applications of an in-line inspection tool are key factors to determine the length of swing arms.

For the piezoelectric ultrasonic device, the free space's axial length is 190 mm and the extreme loading condition occurs when it passes through an outer side accompanied with a dent (depth 6 mm) of a 3D elbow. Besides, stricter performance is required for a spring when the length of a swing arm decreases. Hence, the length of swing arms was designed as 150 mm. Considering requirements of strength and stiffness, the width and thickness of a swing arm were set as 20 mm and 5 mm, respectively. Detailed strength and stability analysis were conducted by numerical simulation in following contents.

In view of swing arms' smooth rotation, the end of a swing arm jointed with a mounting base was set as a circular arc with diameter of 20 mm. The structure of another end was same. A through hole and a shoulder hole were perforated at centers of the above two arcs, respectively. A bracket was welded on a swing arm to fix a Hall sensor. The sensor probe faces the gear side to capture rotation signals of a roller.

2.2. Roller and Gear. The function of a roller and gear is to measure forward distances of an inspection device using a periodic pulse signal generated by a Hall sensor when a gear rotates. Meanwhile, the signal is also treated as the initial point in DAS to start repetitive detection. When a detector moves forward at a certain velocity, small diameter of a roller leads to high rotation speed. Consequently, possibility of a roller's abrasion and skid rises rapidly. In addition, considering that the axial resolution of self-developed piezoelectric ultrasonic device is 5 mm and the free space between two adjacent capsules, the roller's diameter is determined as 47.75 mm. The number of a gear's teeth (n) can be calculated by

$$n = \frac{\pi D}{A}, \tag{1}$$

where D is the diameter of a roller and A is the axial resolution. Then, 30 teeth were obtained for the designed gear. The edge of a gear cannot contact with a pipe's inner wall when a roller rolls forward. Hence, a gear's diameter is required to be less than a roller's diameter. On the other hand, large pitch which leads to big diameter of a gear is needed to guarantee a Hall sensor's normal performance. Therefore, the gear's diameter is determined as 25 mm, as shown in Figure 4.

Skid prevention plays a key role in improving the accuracy of measuring distances [8]. Increasing the contact area and frictional coefficient between a roller and a pipe's inner wall can be considered in design. As the inner diameter of pipe segments used in ultrasonic testing experiments was 204 mm, the cross section of roller's contact surface along radial direction is designed as an arc with radius of 102 mm. Meanwhile, roller's contact surface is knurled.

2.3. Mounting Base. The mounting base in this project functions in three aspects: a bracket to install swing arms, a pivot to fix a spring, and a connector to fasten the odometer wheel on the battery capsule. Based on the structure of capsule's end shown in Figure 2, a block projected as an annular sector was designed. Its outer radius and inner radius are 77 mm and 38.5 mm, respectively. The bottom is step-shaped to match the structure of capsule's end. The block is fixed on the end face by four bolts, of which two bolts are connected to the annular plate and other bolts to the end of capsule. The built mounting base is shown in Figure 5.

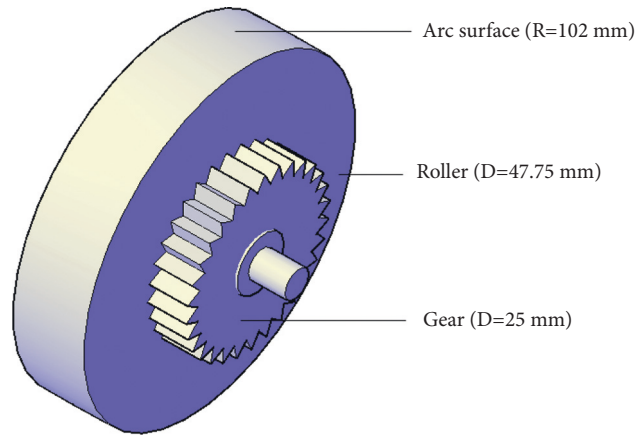


Figure 4: Structure of a roller and gear.

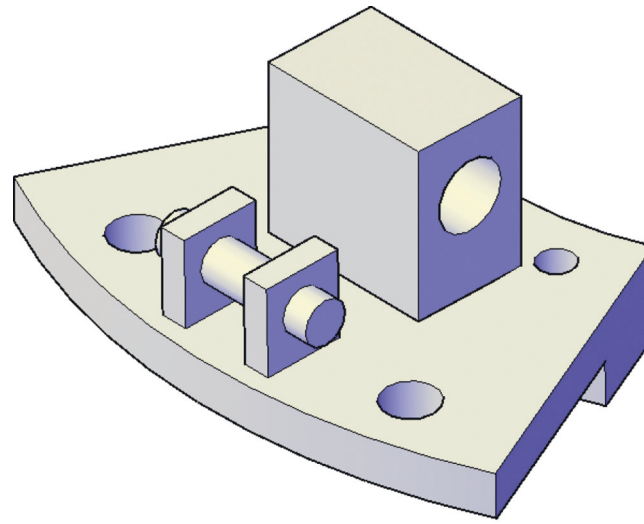


Figure 5: Structure of a mounting base.

2.4. Spring. When a detector moves forward in a pipe, the acting force between the roller and the inner wall is determined by a spring. A steady acting force with a suitable value can enhance the ability of skid prevention. During in-line inspection using the piezoelectric ultrasonic device, the spring acts under three typical conditions, i.e., the odometer wheel passes through (1) an inner side of a 3D elbow, (2) straight pipe, and (3) an outer side accompanied with a dent (depth 6 mm) of a 3D elbow, as shown in Figure 6. The minimum and maximum lengths of the spring occur in conditions (1) and (3) are 77 mm and 82 mm, respectively. In a certain condition, the minimum spring force applied on the odometer wheel is determined by free length and spring rate while the spring force's variation is determined by

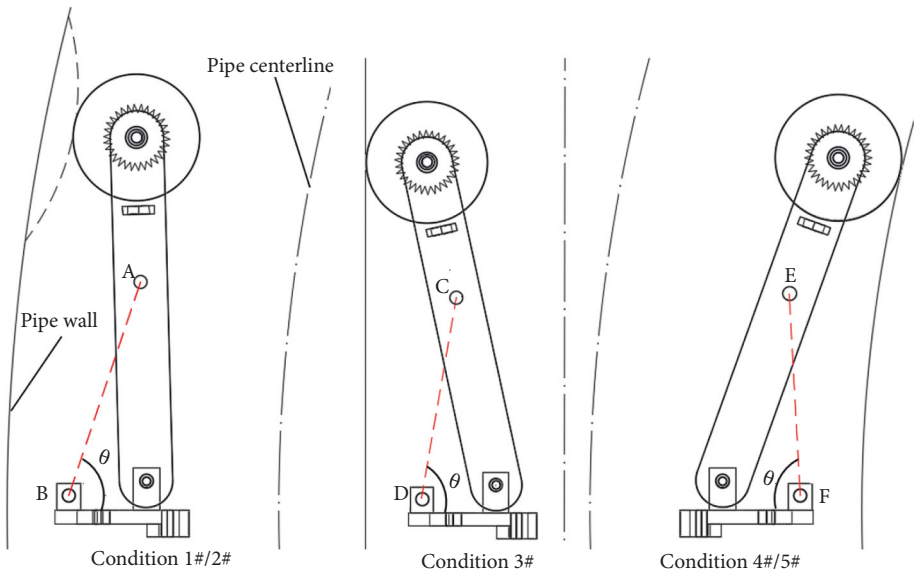


Figure 6: Sketch of typical conditions passed by an odometer wheel.

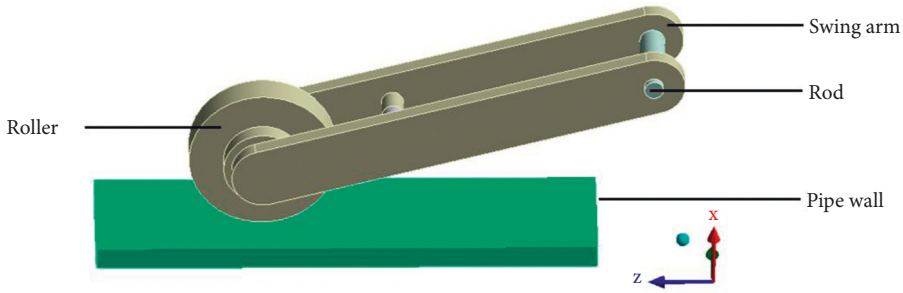


Figure 7: Physical model of the odometer wheel.

spring rate. Hence, four kinds of spring were compared with free length and spring rate of (1) 62 mm and 8 N/mm, (2) 65 mm and 10 N/mm, (3) 67 mm and 12 N/mm, and (4) 69 mm and 15 N/mm, respectively. The minimum spring force is set as 120 N in design and the above four springs meet this requirement.

3. Finite Element Model of the Odometer Wheel

Analyzing structure stability and skid prevention of the designed odometer wheel is necessary for its good performance. Maximum stress and deformation under extreme loading conditions are key criteria for structure stability while acting force between a roller and pipe's inner wall under extreme conditions is a critical factor in assessing skid prevention. The finite element software ANSYS is powerful and convenient to conduct mechanical simulation of this object with complex structure.

3.1. Physical Model. Based on the designed odometer wheel, a physical model on the scale of 1 : 1 is built as shown in Figure 7. As mounting base, Hall sensor and its bracket, gear have little affections on the odometer wheel's mechanical characteristics, the following simplifications are assumed: (1) the rod connecting swing arms and mounting base is treated as a fulcrum and mounting base was ignored, (2) Hall sensor and its bracket were ignored, and (3) the gear was simplified as a cylinder with diameter equal to the diameter of its point circle. The spring was replaced by an equivalent pull force. A pipe wall with axial length of 160 mm, circumferential length of 40 mm, and thickness of 7.5 mm was used. The values of structure's sizes are listed in Table 1.

Table 1: Parameters of physical model for the odometer wheel (units: mm).

Components	Parameters	Values
Swing arms	Length	150
	Width	20
	Thickness	5
	Arc's diameter	20
Roller	Diameter	47.75
	Thickness	10
	Contact surface's radius	102
Gear	Diameter	25
	Thickness	6
Pipe wall	Axial length	160
	Circumferential length	40
	Thickness	7.5

Table 2: Input parameters of material properties.

Components	Mark	Yield Strength/MPa	Tangent modulus (MPa)	Elastic modulus (GPa)	Poisson ratio
Pipe wall	L245	245	777	203	0.3*
Odometer wheel	06Cr19Ni10	220	752	193	0.3*
	2014-T4	230	1429	71	0.3

*Value of Poisson ratio is referred to data given by Su et al. [22].

3.2. Material Properties. The object in simulation is made of steel. The metal has an initial elastic region in which deformation is proportional to the load, while an irrecoverable plastic strain occurs when stress exceeds elastic limit. In order to simulate the mechanical behaviour of steel accurately during operation of the odometer wheel, the bilinear isotropic hardening model was used to describe the material's stress-strain curve. In this model, the curve's initial slope is elastic modulus of the steel. Beyond the yield stress, the plastic strain develops and the stress and strain continue along a line with a slope equal to tangent modulus.

Steel pipe L245 [19] was used in ultrasonic testing experiments. There are two options for the material of the mileage wheel: stainless steel 06Cr19Ni10 [20] and 2014-T4 aluminum alloy [21]. Initial parameters for these three materials used in bilinear isotropic hardening model were calculated according to their properties, and relative values used in simulations are listed in Table 2.

3.3. Meshing. Meshing is the basis of finite element analysis, replacing the original continuum with a collection of finite elements. Element SOLID186 was used to simulate the steel. An accurate calculation can be got by meshing densely. In view of the odometer wheel is a main object in simulation analysis, its size of meshing is set to overall meshing size. Adopting the Hex Dominant method to mesh is beneficial to keep the grid's regularity for the physical model. Therefore, the overall meshing size was set to 2 mm, and the Hex Dominant method was adopted to mesh. The result of meshing was 72821 nodes and 17319 elements, as shown in Figure 8.

3.4. Boundary Conditions and Loads. There are two contact interactions in simulation, i.e., interaction between the roller and pipe wall and that between swing arms and the rod connected to mounting base. Both problems are flexible-to-flexible contact. Contact unit CONTA174 and target unit TARGE170 were chosen to form surface-to-surface three dimensional contact pairs. The Lagrange frictional contact logic of face to face was defined in contact interactions, and the contact behavior was set asymmetrically. The pipe inner wall and the through hole's wall in swing arms were set as the target surface. Rolling surface of the roller and side face of the rod were set as the contact surface.

A pipe is immobile during in-line inspection and fixed constraints were applied to the pipe's outer surface to make the model's boundary closer to actual operating environments. As the rod installed in the mounting base cannot rotate, full constraints were applied to the side face of the rod.

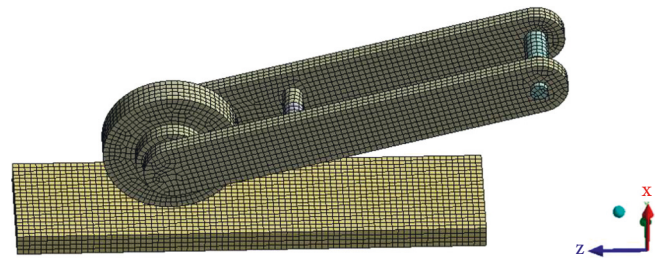


Figure 8: Odometer wheel meshing.

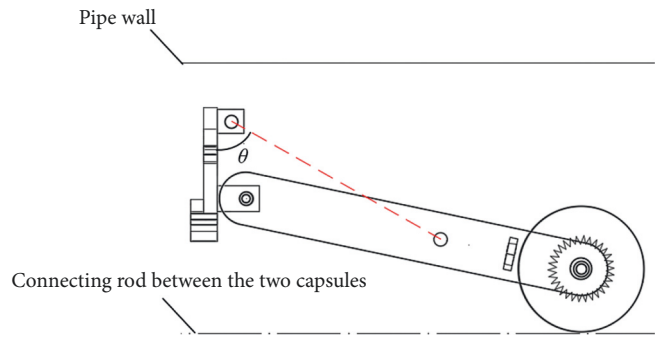


Figure 9: The extreme condition.

The profile of oil and gas pipeline is determined by terrain, and there usually are a large amount of elbows. The recommended minimum radius of pipelines is 4D [18] while the piezoelectric ultrasonic device size is available for 3D. Moreover, there are various kinds of defections existing in oil and gas pipelines and dents have the greatest influence on the operating performance of an odometer wheel. As the diameter of pipe segments used in ultrasonic testing experiments is less than 300 mm, the maximum depth of dents allowable in projects is 6 mm [23]. In order to assess the designed odometer wheel's performance, five typical conditions and an extreme condition were taken into account in numerical analysis. The typical conditions include passing through outer side of a 3D/4D elbow with a 6 mm dent, straight pipe, and inner side of a 3D/4D elbow as shown in Figure 6. Besides, the extreme loading condition for the odometer wheel is that the roller and connecting rod between a battery capsule and a panel capsule are coincident, as shown in Figure 9. The spring's pull force under a certain condition was calculated theoretically and was applied to the rod located at the middle part of swing arms directly. Initial pull forces used in simulations are listed in Table 3, where the

Table 3: Input values of initial loads.

Conditions	Outer side with a 6 mm dent		Straight pipe	Inner side		Extreme condition
	3D	4D		3D	4D	
Serial number	1#	2#	3#	4#	5#	6#
1# spring's pull force (F/N)	152	144	136	120	128	232
2# spring's pull force (F/N)	160	150	140	120	130	260
3# spring's pull force (F/N)	168	156	144	120	132	288
4# spring's pull force (F/N)	180	165	150	120	135	330
Pull force's direction (θ/°)	76	76	80	82	81	61

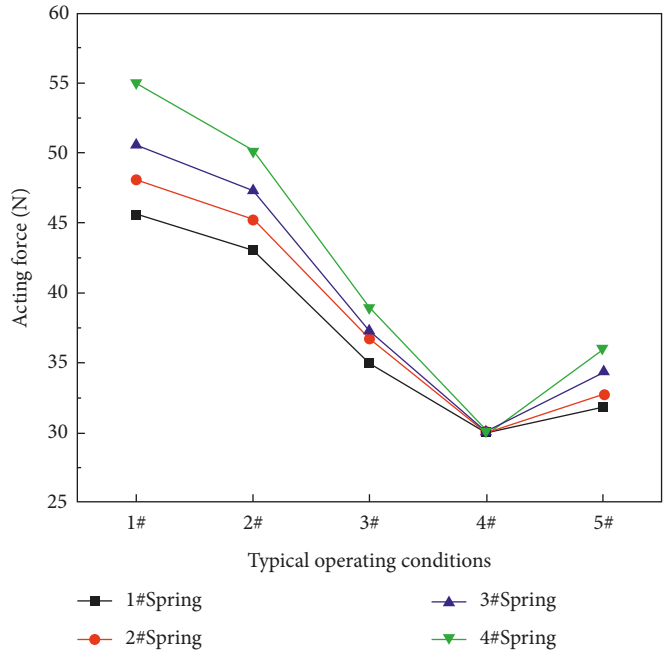
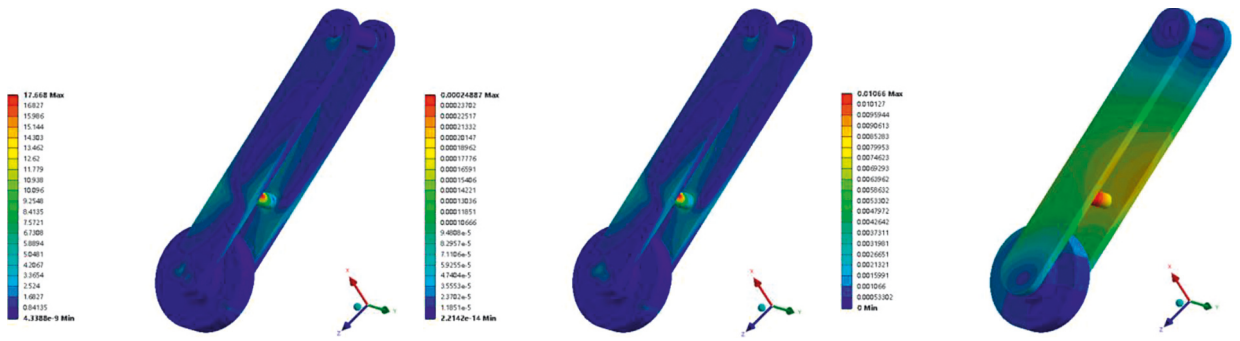


Figure 10: Acting forces under different conditions.



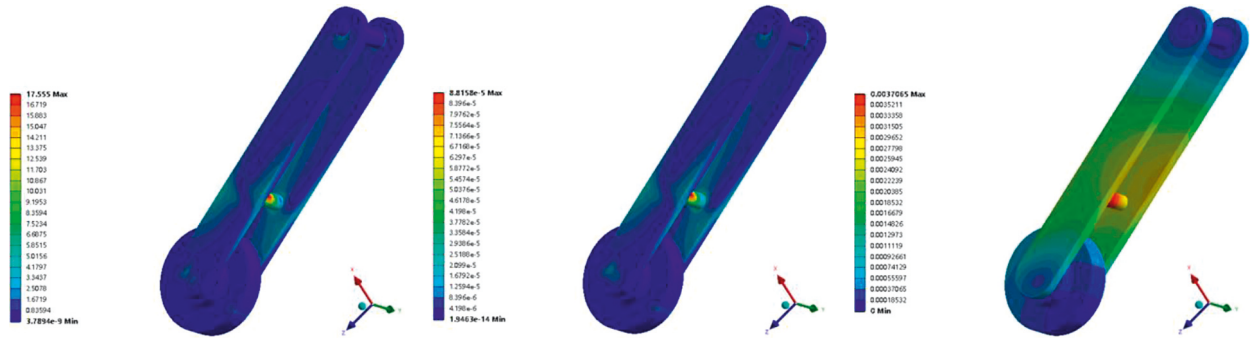
(a) (b) (c)

Figure 11: Mises stress, strain, and deformation cloud diagram of 2014-T4 aluminum alloy (1#).

pull force direction which is the angle between center axis of spring and mounting base surface is denoted as θ.

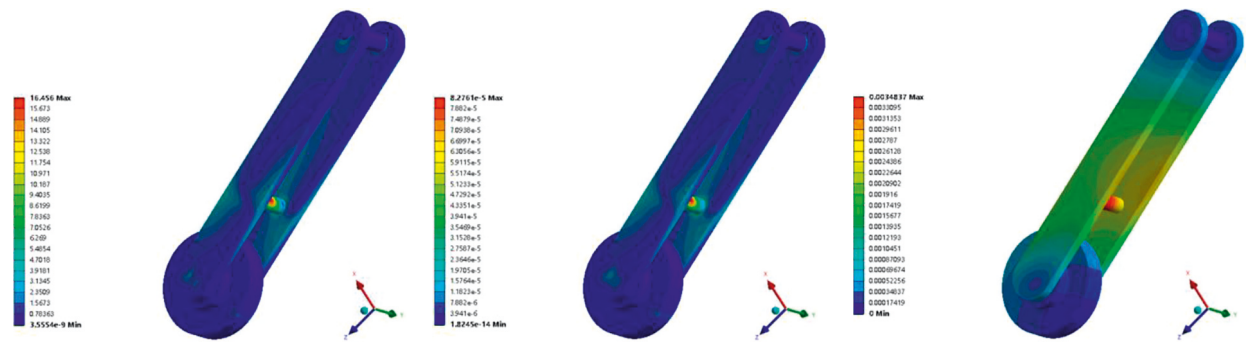
4. Analysis of Simulation Results

4.1. Antiskid Performance Analysis. Low stiffness coefficient of a spring, powdery dirt in pipelines, nonuniform velocity of a detector can lead to skid problem of an odometer wheel. The essential reason is low frictional coefficient and acting force between a roller and pipe's inner wall. The roller's contact surface is knurled to increase frictional coefficient. The dynamic characteristics of acting force under different conditions with different springs are presented in Figure 10. A suitable spring should meet two principles, i.e., the acting



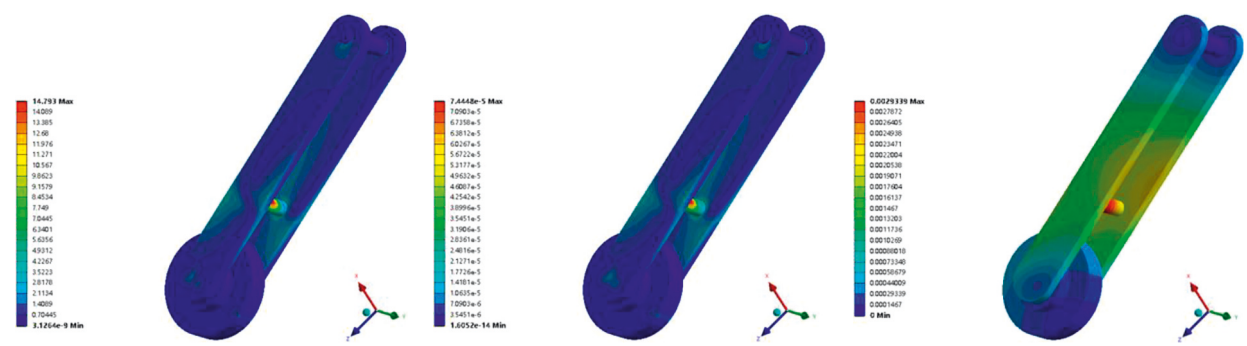
(a) (b) (c)

Figure 12: Mises stress, strain, and deformation cloud diagram of 06Cr19Ni10 (1#).



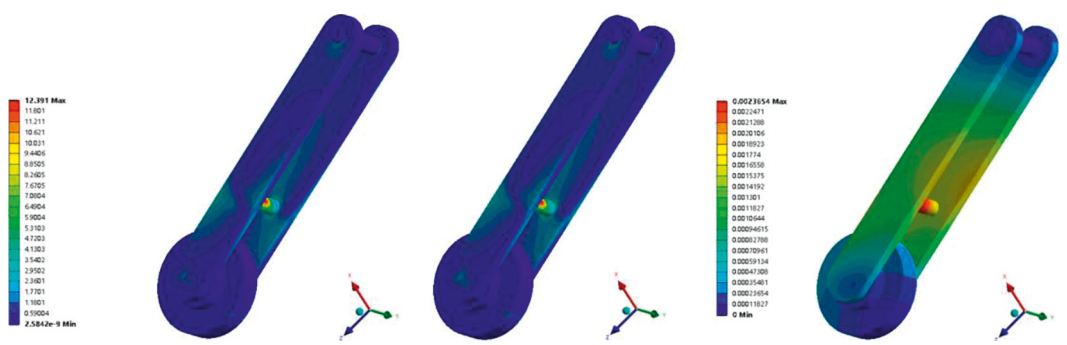
(a) (b) (c)

Figure 13: Mises stress, deformation, and strain cloud diagram (2#).



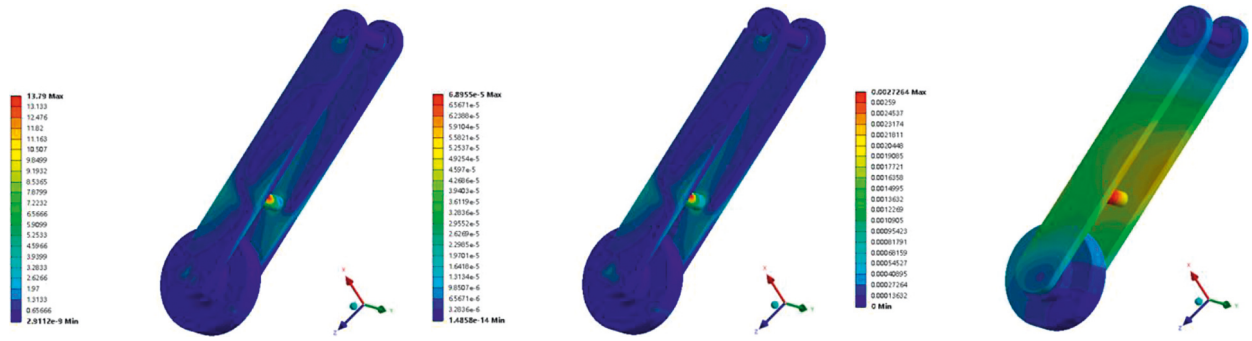
(a) (b) (c)

Figure 14: Mises stress, deformation, and strain cloud diagram (3#).



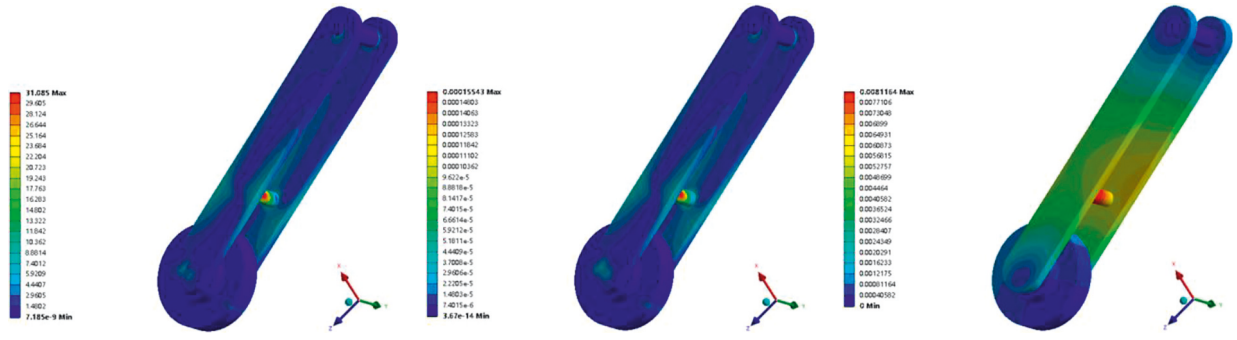
(a) (b) (c)

Figure 15: Mises stress, deformation, and strain cloud diagram (4#).



(a) (b) (c)

Figure 16: Mises stress, deformation, and strain cloud diagram (5#).



(a) (b) (c)

Figure 17: Mises stress, deformation, and strain cloud diagram (6#).

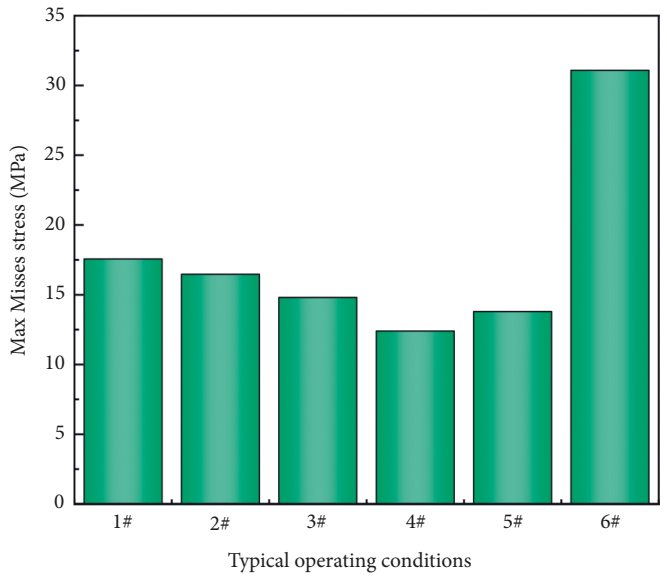


Figure 18: Maximum Mises stresses in six conditions.

force meets requirements for antiskid and its variation under various working conditions is small. The variation of 3# spring and 4# spring is relatively large, and the difference between maximum value and minimum value exceeds 50% of the minimum value. Although 1# spring has a small change in various typical operating conditions, the acting force under 1# working condition is only 45 N. Hence, 2# spring was chosen in design. The maximum value of acting

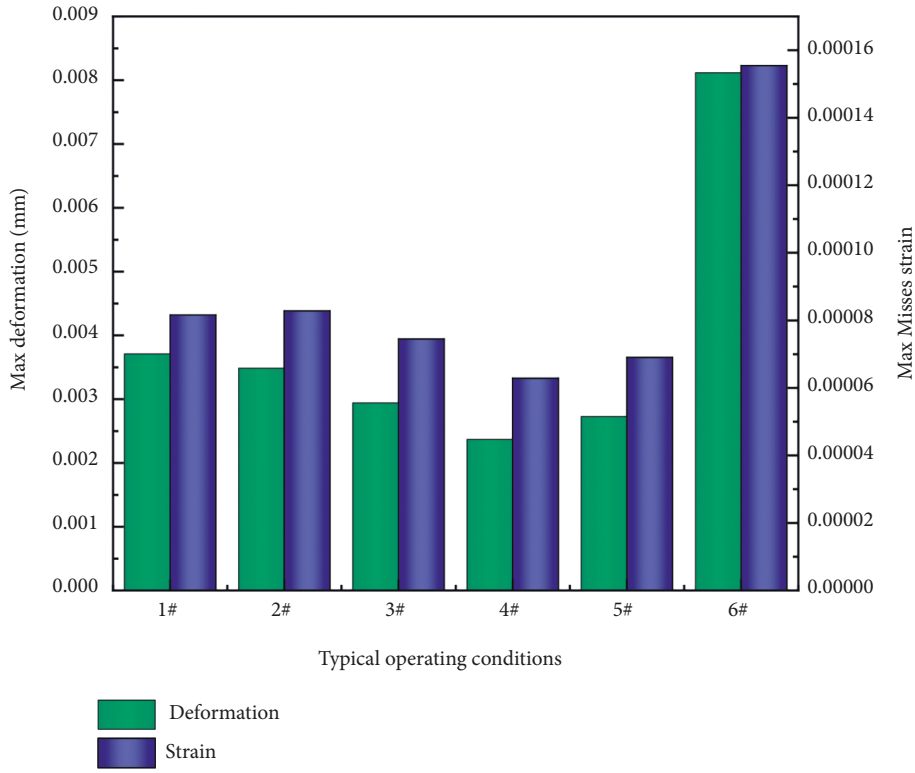


Figure 19: Maximum Mises strain and deformations in six conditions.

force is 48.08 N, and the minimum value is 30.03 N in five working conditions. It indicates that the design odometer wheel meets requirements of antiskid.

4.2. Material of Odometer Wheel Analysis. In order to determine the appropriate material, odometer wheels made of stainless steel 06Cr19Ni10 and 2014-T4 aluminum alloy were simulated and analyzed in five working conditions and the extreme condition. Taking the 1# working condition as an example, as shown in Figures 11 and 12, the Mises stress of all working conditions of the two materials are similar. The strain and deformation of aluminum alloy are larger than those of stainless steel. Considering the self-developed piezoelectric ultrasonic device was made of stainless steel, stainless steel 06Cr19Ni10 is selected in order to keep the uniformity and stability of a device's material.

4.3. Structural Stability Analysis. Figures 13 and 14 present the Mises stress, strain, and deformation. Figure 15 shows the cloud diagrams of the odometer wheel in all conditions. In all working conditions, maximum values of Mises stress (Figure 16), Mises strain (Figure 17), and total deformation occur at the rod in middle part of swing arms, which are compared in Figures 18 and 19. Corresponding values for condition 1# with the worst mechanical state and extreme condition 6# are (17.55 MPa, 0.0000882, 0.00371 mm), (31.09 MPa, 0.000155, 0.00812 mm), respectively. Results show that maximum Mises stresses in condition 1# and condition 6# are far less than the elastic limit of stainless steel 220 MPa. The deformation in extreme condition has little influence on the odometer wheel's normal performance, and it can be ignored. It indicates that the designed odometer wheel has good performance in term of structure stability.

5. Conclusion

The piezoelectric ultrasonic in-line inspection technology is an effective way to conduct detection of oil and gas pipelines. Based on a self-developed piezoelectric ultrasonic device, mechanical design principles were discussed and an odometer wheel was designed. The material of odometer was chosen stainless steel 06Cr19Ni10. The free length of spring and spring rate were chosen 65 mm and 10 N/mm. Five typical conditions under which the odometer wheel works were summarized according to in-situ applications. A finite element model was built by the software ANSYS, and the mechanical performance of the odometer wheel were analyzed. The maximum Mises stress (31.09 MPa) under extreme condition 6# is far less than the elastic limit (220 MPa) of chosen stainless steel. Maximum deformation and Mises strain of the odometer wheel are 0.00812 mm and 0.000155. Good performance of structure stability is shown for the designed odometer wheel. The acting force under five working conditions varies from 30.03 N to 48.08 N. The designed object also meets requirements of antiskid.

Data Availability

The data used to support the findings of this study are included within the article.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was financially supported by Fundamental Research Funds for the Central Universities (19CX05007A) and the National Natural Science Foundation of China (No. 51606160).

References

[1] J. J. Zhang, H. Su, and P. Gao, "Resilience-based supply assurance of natural gas pipeline networks and its research prospects," *Acta Petrolei Sinica*, vol. 41, no. 12, pp. 1665–1678, 2020.

[2] O. Taylan, M. A. Sattari, I. Elhachfi Essoussi, and E. Nazemi, "Frequency domain feature extraction investigation to increase the accuracy of an intelligent non-destructive system for volume fraction and regime determination of gas-water-oil three-phase flows," *Mathematics*, vol. 9, no. 17, Article ID 2091, 2021.

[3] H. Zhang, S. Dong, J. Ling, L. Zhang, and B. Cheang, "A modified method for the safety factor parameter: the use of big data to improve petroleum pipeline reliability assessment," *Reliability Engineering & System Safety*, vol. 198, Article ID 106892, 2020.

[4] O. Taylan, M. Abusurrah, S. Amiri, E. Nazemi, E. Eftekhari-Zadeh, and G. H. Roshani, "Proposing an intelligent dual-energy radiation-based system for metering scale layer thickness in oil pipelines containing an annular regime of three-phase flow," *Mathematics*, vol. 9, no. 19, Article ID 2391, 2021.

[5] M. S. Chowdhury and M. F. Abdel-Hafez, "Pipeline inspection gauge position estimation using inertial measurement unit, odometer, and a set of reference stations," *ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part B: Mechanical Engineering*, vol. 2, no. 2, 2016.

[6] X. Zhu, X. Li, C. Zhao, S. Zhang, and S. Liu, "Dynamic simulation and experimental research on the motion of odometer passing over the weld," *Journal of Natural Gas Science and Engineering*, vol. 30, pp. 205–212, 2016.

[7] D. Zelmati, O. Bouledroua, Z. Hafsi, and M. B. Djukic, "Probabilistic analysis of corroded pipeline under localized corrosion defects based on the intelligent inspection tool," *Engineering Failure Analysis*, vol. 115, Article ID 104683, 2020.

[8] Y. X. Zang, C. Qiu, T. H. Hu, and H. Yang, "Causes of slippage of odometer wheels in pipeline inline detectors," *J. Oil & Gas Storage and Transportation,* vol. 35, no. 3, pp. 306–309, 2016.

[9] X. L. Li, J. Z. Chen, Y. L. Ma, R. X. He, and T. Meng, "ADAMS-Based kinematics analysis of the odometer passing over girth weld," *J. China Petroleum Machinery,* vol. 47, no. 4, pp. 118–123, 2019.

[10] Y. Xie, D. Gu, X. Wang et al., "A smart healthcare knowledge service framework for hierarchical medical treatment system," *Healthcare*, vol. 10, no. 1, p. 32, 2021.

[11] H. Xu, Z. X. Li, Y. Y. Li, and L. H. Gong, "Study on the error analysis of pipeline inspection robot mileage wheel localization," *Advanced Materials Research*, vol. 706-708, pp. 1171–1174, 2013.

[12] X. Wu, S. J. Jin, Y. B. Li, Y. K. Xiao, and H. M. Zhao, "Above-ground marker system of pipeline internal inspection instrument based on geophone array," *Journal of Nanotechnol Precision Engineering*, vol. 8, pp. 553–558, 2010.

[13] Y. Li, S. Liu, D. J. Dorantes-Gonzalez, C. Zhou, and H. Zhu, "A novel above-ground marking approach based on the girth weld impact sound for pipeline defect inspection," *Insight - Non-Destructive Testing and Condition Monitoring*, vol. 56, no. 12, pp. 677–682, 2014.

[14] Z. Wang, J. Tan, and Z. Sun, "Error factor and mathematical model of positioning with odometer wheel," *Advances in Mechanical Engineering*, vol. 7, no. 1, Article ID 305981, 2015.

[15] X. Li, S. Zhang, S. Liu, Q. Jiao, and L. Dai, "An experimental evaluation of the probe dynamics as a probe pig inspects internal convex defects in oil and gas pipelines," *Measurement*, vol. 63, pp. 49–60, 2015.

[16] X. Li, S. Zhang, S. Liu, X. Zhu, and K. Zhang, "Experimental study on the probe dynamic behaviour of feeler pigs in detecting internal corrosion in oil and gas pipelines," *Journal of Natural Gas Science and Engineering*, vol. 26, pp. 229–239, 2015.

[17] GB 50253-2014, *Code for Design of Oil Transportation Pipeline Engineering*, China Planning Press, Beijing, China, 2014.

[18] GB 50369-2014, *Code for Construction and Acceptance of Oil and Gas Long-Distance Transmission Pipeline Engineering*, China Planning Press, Beijing, China, 2014.

[19] Gb/T 9711-2017, *Petroleum and Natural Gas Industries-Steel Pipe for Pipeline Transportation Systems*, China Planning Press, Beijing, China, 2017.

[20] Gb/T 24511-2017, *Stainless Steel and Heat Resisting Steel Plate, Sheet and Strip for Pressure Equipments*, China Planning Press, Beijing, China, 2017.

[21] Gb/T 6892-2015, *Wrought Aluminium and Aluminium Alloys Extruded Profiles for General Engineering*, China Planning Press, Beijing, China, 2012.

[22] Y. Su, J. Li, B. Yu, Y. Zhao, and J. Yao, "Fast and accurate prediction of failure pressure of oil and gas defective pipelines using the deep learning model," *Reliability Engineering & System Safety*, vol. 216, Article ID 108016, 2021.

[23] Gb/T. 50251-2015, *Code for Design of Gas Transmission Pipeline Engineering*, China Planning Press, Beijing, China, 2014.