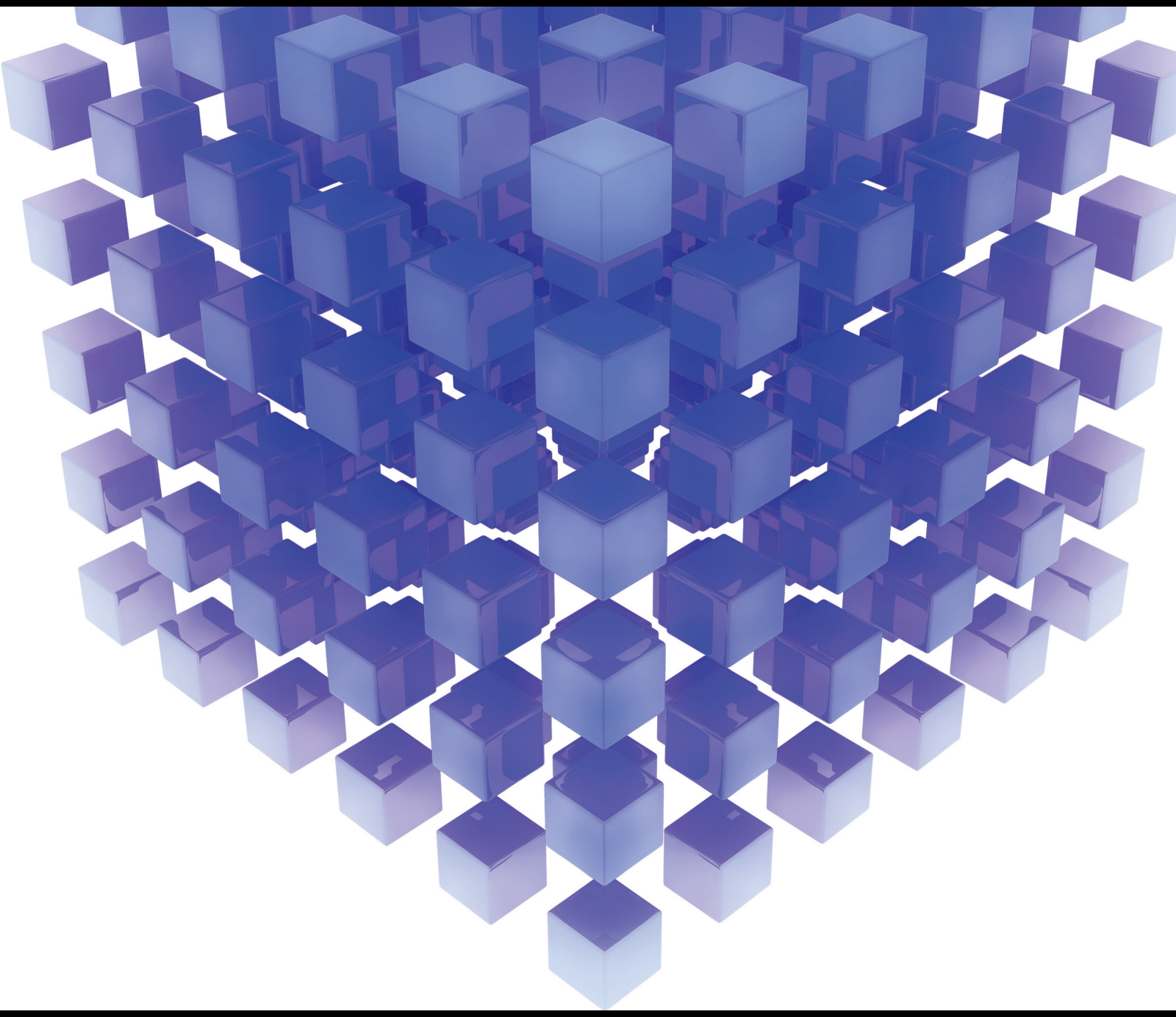
Mathematical Problems in Engineering

# Computational Intelligence in Complex Decision Making

Lead Guest Editor: Shouzhen Zeng
Guest Editors: Tomas Balezentis, Chonghui Zhang, and Dalia Streimikiene



# Computational Intelligence in Complex Decision Making

Mathematical Problems in Engineering

# Computational Intelligence in Complex Decision Making

Lead Guest Editor: Shouzhen Zeng
Guest Editors: Tomas Balezentis, Chonghui Zhang, and Dalia Streimikiene

# Contents

Hindawi
Mathematical Problems in Engineering
Volume 2023, Article ID 9759528, 1 page
https://doi.org/10.1155/2023/9759528

*Retraction*

# Retracted: Probabilistic Hesitant Fuzzy Methods for Prioritizing Distributed Stream Processing Frameworks for IoT Applications

**Mathematical Problems in Engineering**

Received 17 October 2023; Accepted 17 October 2023; Published 18 October 2023



This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

(1) Discrepancies in scope
(2) Discrepancies in the description of the research reported
(3) Discrepancies between the availability of data and the research described
(4) Inappropriate citations
(5) Incoherent, meaningless and/or irrelevant content included in the article
(6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

## References

[1] Z. Lin, C. Huang, and M. Lin, "Probabilistic Hesitant Fuzzy Methods for Prioritizing Distributed Stream Processing Frameworks for IoT Applications," *Mathematical Problems in Engineering*, vol. 2021, Article ID 6655477, 12 pages, 2021.

Hindawi
Mathematical Problems in Engineering
Volume 2021, Article ID 6688750, 5 pages
https://doi.org/10.1155/2021/6688750

*Research Article*

# A Consumer Behavior Prediction Model Based on Multivariate Real-Time Sequence Analysis

**Lin Guo**,[1] **Ben Zhang**,[2] **and Xin Zhao**[1]

[1]*School of Economics and Management, Changchun University of Science and Technology, Jilin 130022, China*
[2]*College of Engineering, Northeastern State University, Boston 02115, MA, USA*

Correspondence should be addressed to Lin Guo; guolin@cust.edu.cn

Received 8 December 2020; Accepted 19 May 2021; Published 26 May 2021

Academic Editor: Tomas Balezentis



With the rapid development of online finance and social networks, a large amount of behavioral data is stored on the Internet, which can fully reflect the shopping tendencies and habits of real users. Using big data to analyze consumer behavior is more scientific and accurate than the traditional sampling survey method. Internet consumption behavior data are time series data. Therefore, this paper proposes a method of analyzing behavioral sequence data, which learns personal consumption interests and habits, and finally predicts payment behavior. The experiments compare the execution effect of different algorithms on multiple databases and verify the feasibility and effectiveness of the proposed algorithm SeqLearn.

## 1. Introduction

Internet e-commerce platforms contain a wealth of information about events, relationships, and attitudes. A series of technologies such as text mining, statistical theory, association analysis, and visualization are used to realize sentiment analysis, information extraction, and user influence analysis. Learning consumer behavior can help analyze the characteristics of consumers, the relationship between products, and so on. Therefore, it is a very valuable research topic to construct consumption structures according to the consumption behavior records of different consumers. The algorithm SeqLearn proposed in this paper extracts consumers' consumption interests and habits by analyzing the behavior data of e-commerce platforms and calculates the comprehensive probability to predict the next payment behavior. Forecast results can be used for product recommendations, advertising placement, and other applications.

Online shopping is a process in which users browse and search for related products to complete shopping or other related tasks in a virtual shopping environment. Payment decision is the integration and unity of consumer demand, motivation, activity, and afterthought. By analyzing the network data, the potential connections behind a series of consumer behaviors can be mined out and the prediction function can be realized. The application of this method was quick and effective to identify and segment consumer groups and facilitated the mapping of the differences among these groups and the comparison of the consumption behavior expressed by consumers on different markets [1]. In the field of social network analysis, users are defined as nodes. Relationships between nodes can be abstractly represented by edges. Common techniques used to analyze relationships between nodes include causal inference [2–4] and Bayesian network [5, 6]. Different strategies are used for different purposes and needs, such as emotion-oriented analysis, information extraction, and user influence analysis. For example, Bollen analyzed the sentiment in tweets and found that the creation and dissemination of online information are closely related to stock market, future commodity prices, and major social events [7]; Asur of HP LABS successfully predicted the box office using Twitter data [8]. Based on the big data of social networks, González-Bailón [9] studied the dynamics of the protest movements and revealed its influences on political trends.

Group interaction, information dissemination, and other behaviors on the Internet can affect politics, economy, and society. Network data analysis methods are divided into

the following aspects: (1) analyzing the characteristics of nodes [10]; (2) analyzing relationships (such as advertising push relationship [11], intimate relationship [12], and other types of relationships [13, 14]) between nodes according to sociology [15–18]; (3) analyzing the law of information transmission; and (4) group behavior studies (such as topology analysis, normalization analysis, modular analysis, random data flow analysis, node clustering, and classification) [19–21].

The Internet has become a new media, which makes users have a great sense of participation. The convenient and efficient mode of communication is a driving force for Internet business to analyze and mine network data. A recommendation model based on statistical modelling is used to assist consumers facing choice overload by predicting their interests and consumption behaviors [22]. However, due to the complexity of network data, it is a huge challenge for any enterprise or operator to mine valuable information with commercial value efficiently and quickly. The development of the Internet has brought about tremendous changes in consumption concepts, consumption patterns, and consumer conditions.

In this paper, a large number of real network data are collected to support the analysis of consumer behavior. We study personal consumption interests and consumption habits and finally predict the next payment behaviors. Forecast results can be used for product recommendations, advertising placement, and other applications. In the experimental part, online consumer behavior data, such as Ali Data (an online shopping website in China), are used to compare the execution effects of different algorithms on different databases and verify the feasibility and effectiveness of SeqLearn.

The remainder of this paper is organized as follows. Section 2 proposes the predictive model of consumer behavior. Section 3 introduces the experimental results. Section 4 states the conclusions of research achievements.

## 2. Consumer Behavior Prediction Model

Online consumption data can reflect the actual consumption trends and consumption changes of users. This paper proposes a sequence analysis algorithm to study personal consumption preferences and consumption habits. Personal consumption preferences can be used to predict what consumers will buy over a period of time. Consumption habits reflect a time cycle, that is, how long it takes for consumers to pay attention to a certain commodity before they buy it.

The behaviors of Internet users can be divided into four categories: click, collect, add-to-cart, and payment. Among these four behaviors, payment data are considered outdated data, which only reflects the user's past shopping preferences and is used as validation data to verify the accuracy of the algorithm. Add-to-cart and collect behaviors are regarded as equal. They have the same effect on predicting shopping behavior. The amount of browsing behavior data is huge, and its correlation with user interests is weak. Experiments show that the accuracy of our prediction algorithm will not be significantly improved by analyzing the browsing behavior data. Therefore, this paper does not consider browsing behavior data and obtains accurate analysis results with as little computational effort as possible. In summary, this paper only considers purchase behavior, add-to-cart behavior, and collect behavior. In order to verify the feasibility of the algorithm, we analyze a large amount of user data. It is found that these three behaviors are strongly correlated and often appear in a logical order. This proves that using only these three behavioral data to predict consumption trends is feasible.

In order to analyze the continuously updated behavioral data sequence and prevent the impact of stale data on analysis accuracy, the length of the dynamic sequence to be analyzed is selected as $n$. We transform the training sequence into a fixed-length sequence $s = (s_1, s_2, \ldots, s_n)$, where $n$ represents the maximum length. If the sequence length is greater than $n$, then the nearest $n$ actions are considered. If the sequence length is less than $n$, we will repeatedly add a "padding" item to the left until the length is $n$.

The most important task of sequential analysis is to learn each user's personal consumption preferences and consumption habits based on historical data. The prediction formula is shown as follows:

$$P\left(\widehat{i}, \widehat{a} \mid S_i^u, S_a^u\right) = P\left(\widehat{i} \mid S_i^u\right) \times P\left(\widehat{a} \mid S_a^u\right). \tag{1}$$

$P(\widehat{i}, \widehat{a} \mid S_i^u, S_a^u)$ calculates the probability that user $u$ performs action $\widehat{a}$ on item $\widehat{i}$, where $\widehat{i}$ represents the predicted next item and $\widehat{a}$ represents the predicted user action. $S_a^u$ represents the historical access record of user $u$. $S_i^u$ represents the historical action sequence of user $u$.

The following sections introduce the formulas for learning personal consumption interests and habits.

*2.1. Personal Consumption Preference Analysis.* In a long time, the consumer's consumption trend is constantly changing, but in a short time, the consumption trend is stable. By analyzing user behavior over a period of time, we can find out consumers' consumption interests, which reflect the users' recent consumption trends. The following formula illustrates the process of quantifying user interests to predict user behaviors. $P(\widehat{i} \mid S_i^u)$ in formula (1) reflects the consumption interest of a consumer over a period of time. The probability function formula is as follows:

$$P\left(\widehat{i} \mid S_i^u\right) = \frac{P\left(\widehat{i}, S_i^u\right)}{P\left(S_i^u\right)}, \tag{2}$$

where $S_i^u$ represents the sequence of access items for user $i$. If the length of the sequence is less than $n$, it is automatically filled. The number of occurrences of each item in the sequence is denoted by $x$. $x$ corresponds to the Poisson distribution. $x$ is related to the number of times users use online media and the amount of information generated [23], so $x \sim$ Poisson $(\lambda)$, where $\lambda = ev \times c$, in which $ev$ represents the number of times a user uses network media in time $t$ and $c$ represents the total amount of data generated in a certain

consumer interest area within time $t$. From this, the following probability density formula can be derived:

$$P\left(S_i^u\right) = \prod_{k=1}^{n} P\{\text{item}_k = x\} = \prod_{\substack{k=1 \\ \text{item}_k = x}}^{n} e^{-ev\times c} \frac{ev \times c^x}{x!}. \quad (3)$$

$P(\widehat{i}, S_i^u)$ in formula (2) represents the probability that $\widehat{i}$ and $S_i^u$ appear at the same time. The specific formula is shown as follows:

$$P\left(\widehat{i}, S_i^u\right) = \frac{m+1}{n}, \quad (4)$$

where $n$ represents the sequence length and $m$ represents the number of elements that appear in both $S_i^u$ and the sequence being studied. Substituting formula (3) and formula (4) into formula (2), the following formula is obtained:

$$P\left(\widehat{i} \mid S_i^u\right) = \frac{m+1}{n} \prod_{\substack{k=1 \\ \text{item}_k = x}}^{n} e^{-ev\times c} \frac{ev \times c^x}{x!}. \quad (5)$$

*2.2. Consumption Habit Analysis.* The consumption habits of consumers are different, which can be roughly divided into impulse type and conservative type. Impulse consumers refer to consumers who purchase goods intuitively and impulsively under the influence of urgent purchase psychology. Conservative consumers have strict attitudes towards consumption. They are sensitive and wait-and-see attitudes towards products and are sensitive to prices. By comparing these two types of consumers, impulsive consumers will be influenced by the first suitable product and make quick purchase decisions instead of repeated selection and comparison. However, conservative consumers have long consumption cycles and need to compare many products before buying.

The consumption habits of any consumer are different. By studying historical data, the consumption habits of different consumers can be derived. The probability formula for predicting the next action of consumers is shown as follows:

$$P\left(\widehat{a} \mid S_a^u\right) = \frac{P\left(\widehat{a}, S_a^u\right)}{P\left(S_a^u\right)}, \quad (6)$$

$$P\left(\widehat{a}, S_a^u\right) = \frac{t}{n}. \quad (7)$$

Formula (7) calculates the frequency of purchase behavior, where $n$ is the length of the sequence being studied and $t$ is the number of purchases.

$P(S_a^u)$ satisfies binomial distribution B $(n, p)$, and let $y$ be the number of times the purchase behavior A occurs in $n$ user actions, so $y$ is a random variable. $y$ is equal to 0, 1, 2, ..., $n$. Set $P(A) = p$, $P(\overline{A}) = 1 - p$. So the distribution law of $P(S_a^u)$ is shown as follows:

$$P\left(S_a^u\right) = P\{S_a^u = k\} = C_n^k p^k (1-p)^{n-k}, \quad (8)$$

where $k$ is the number of times the purchase behavior occurs in sequence $S_a^u$. Then, formula (7) and formula (8) are substituted into formula (6), and the following formula can be obtained:

$$P\left(\widehat{a} \mid S_a^u\right) = \frac{t}{nC_n^k p^k (1-p)^{n-k}}, \quad k \in [0, n]. \quad (9)$$

Finally, by substituting formula (5) and formula (9) into formula (1), the following formula can be obtained:

$$P\left(\widehat{i}, \widehat{a} \mid S_i^u, S_a^u\right) = \frac{t(m+1)}{n^2 C_n^k p^k (1-p)^{n-k}} \prod_{\substack{k=1 \\ \text{iterm}_k = x}}^{n} e^{-ev\times c} \frac{ev \times c^x}{x!}. \quad (10)$$

## 3. Experiment

The parameters used in the experiment are shown in Table 1.

Datasets used in the experiment are as follows:

(1) Ali data (https://tianchi.aliyun.com/dataset/dataDetail?dataId=46): this dataset is provided by Alibaba Group. It possesses a wealth of user data, such as user location information and access time. This dataset contains the following attributes: user ID, item IDs, types of behaviors (including click, collect, add-to-cart, and payment; the corresponding values are 1, 2, 3, and 4, respectively), locations, the category ID, and time.

(2) EP data: it contains 15,890,209 pieces of data collected from http://www.dianping.com/ in August 2018. The dataset contains the following attributes: shop_id (unique), province, city, city_id, area, big_cate, big_cate_id, small_cate, small_cate_id, service_rating, all_remarks, very_good_remarks, good_remarks, common_remarks, bad_remarks, and very_bad_remarks.

The algorithms used in the experiment are as follows:

(1) Statistical learning: statistical learning techniques are the tools we use to understand data, which are divided into supervised learning and unsupervised learning. Broadly speaking, statistical learning builds statistical models based on one or more inputs to predict or estimate outputs.

(2) EC-Structure [23]: this algorithm grabs consumption data of the e-commerce platform, analyzes consumption structures, and predicts consumption behaviors according to consumption structures. The algorithm integrates multiple dimensions of network data to comprehensively study consumer behavior.

In order to show the relationship between user behavior and payment behavior, describes the behavior data of an Ali

Table 1: Parameter description.

| | |
|---|---|
| *tp* | The number of behaviors that are correctly predicted |
| *tn* | The number of the correctly predicted behaviors that will not occur |
| *Fn* | The number of incorrectly predicted behaviors that will not occur |
| *Fp* | The number of incorrectly predicted consumption behaviors that will not occur |
| ACC | A parameter to measure the accuracy of prediction results |
| *P* | Precision |
| *R* | Recall |
| FB | Synthesis of *P* and *R* |

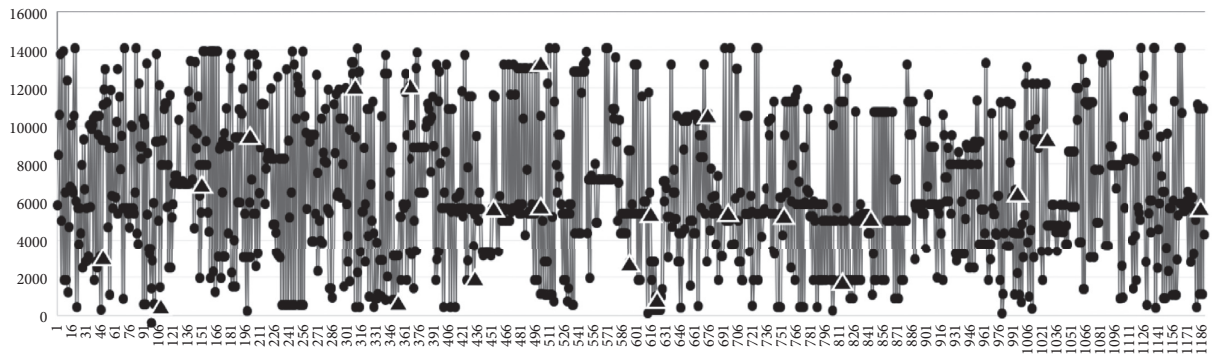



Figure 1: Predict consumer behavior based on Ali data.

Table 2: Confusion matrix (Top@10).

| | | Actual | |
|---|---|---|---|
| | | Positive | Negative |
| Predicted | Positive | 132910 | 20137 |
| | Negative | 28193 | 213461 |

Table 3: Comparison of algorithm execution effects.

| Dataset | Top@*N* | Statistical algorithm | | | | EC-Structure | | | | SeqLearn | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | *P* | *R* | FB | ACC | *P* | *R* | FB | ACC | *P* | *R* | FB | ACC |
| Ali data | Top@5 | 0.64 | 0.70 | 0.67 | 0.60 | 0.78 | 0.70 | 0.74 | 0.75 | 0.80 | 0.81 | 0.80 | 0.77 |
| | Top@10 | 0.72 | 0.77 | 0.74 | 0.70 | 0.83 | 0.82 | 0.82 | 0.80 | 0.86 | 0.88 | 0.87 | 0.85 |
| EP data | Top@5 | 0.62 | 0.60 | 0.61 | 0.61 | 0.81 | 0.70 | 0.75 | 0.80 | 0.82 | 0.86 | 0.84 | 0.81 |
| | Top@10 | 0.73 | 0.77 | 0.75 | 0.72 | 0.83 | 0.85 | 0.84 | 0.81 | 0.88 | 0.90 | 0.89 | 0.87 |

user for one month. It visually shows the sequential relationship between different user behaviors.

In Figure 1, the abscissa indicates the order of appearance of the behavior data, and the ordinate indicates the consumption category number. The dots in Figure 1 represent the behaviors of collect, add-to-car, and payment. The triangles represent the predicted payment behaviors based on historical data. It can be seen from Figure 1 that there is a certain logical connection between users' online behavior and payment behavior, and it is verified that the method of integrating consumer interests and consumption habits can accurately predict the next payment behavior.

Table 2 illustrates the accuracy of the consumer behavior prediction algorithm SeqLearn (Top@*N*: among the results calculated by the algorithm, the first *N* are selected for precision analysis). The horizontal quantity represents the actual value, and the vertical quantity represents the predicted value. In order to measure the accuracy of the algorithm, the following indicators are selected for measurement: precision $P = tp/(tp + fp)$; recall $R = tp/(tp + fn)$; FB parameters $\mathrm{FB} = P \times R \times 2/(P + R)$, where *tp* is the number of behaviors that are correctly predicted, *tn* is the number of the correctly predicted behaviors that will not occur, *fn* is the number of incorrectly predicted behaviors that will not occur, and *fp* is the number of incorrectly predicted consumption behaviors that will not occur. In addition, parameter $\mathrm{ACC} = tp/(tp + tn + fn + fp)$ is used to measure the prediction accuracy of the algorithms.

Table 3 displays the performance comparison of statistical learning, EC-Structure, and SeqLearn using different datasets. According to Table 3, we can conclude that SeqLearn has the most stable execution effect and the best results according to FB and ACC.

## 4. Conclusion

If a questionnaire survey is used for research, problems such as low sample coverage, long data survey cycle, and lagging survey results will arise. In this paper, a large number of real network data is collected to support the analysis of consumer behavior. The results of this method are accurate. Internet consumption behavior data are time series data. Therefore, this paper adopts the method of time series analysis to study personal consumption interests and consumption habits and finally predicts next payment behaviors. The prediction results can be used in areas such as product recommendation and advertising push. The experiment shows the process of predicting consumer behavior and verifies the feasibility and efficiency of the SeqLearn algorithm on multiple datasets. In the future research, we hope to study the characteristics of long-term consumption and short-term consumption, so that we can accurately calculate the parameters of different types of goods in the recommendation method.

## 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 Youth Program of The National Social Science Fund of China (Project name: Research on online behavior pattern of customers and multidimensional customer insight method under big data; Grant no.19CGL024).

## References

[1] K. Migda-Najman, K. Najman, and S. Badowska, "The GNG neural network in analyzing consumer behaviour patterns: empirical research on a purchasing behaviour processes realized by the elderly consumers," *Advances in Data Analysis and Classification*, vol. 14, pp. 1–36, 2020.

[2] J. Pearl, "Causal inference in statistics: an overview," *Statistics Surveys*, vol. 3, pp. 96–146, 2009.

[3] A. Belloni, V. Chernozhukov, I. Fernndez-Val, and C. Hansen, "Program evaluation and causal inference with high-dimensional data," *Econometrica*, vol. 85, no. 1, pp. 233–298, 2017.

[4] G. Hong, "A review of explanation in causal inference: methods for mediation and interaction," *Journal of Educational and Behavioral Statistics*, vol. 42, pp. 34–53, 2017.

[5] S. Mei, J. Zhu, and X. Zhu, "Robust regbayes: selectively incorporating first-order logic domain knowledge into Bayesian models," in *Proceedings of the International Conference on Machine Learning*, Beijing, China, June 2014.

[6] A. M. Oliveira and O. L. V. Costa, "An iterative approach for the discrete-time dynamic control of Markov jump linear systems with partial information," *International Journal of Robust and Nonlinear Control*, vol. 30, no. 2, 2020.

[7] J. Bollen, H. Mao, and X. Zeng, "Twitter mood predicts the stock market," *Journal of Computational Science*, vol. 2, no. 1, pp. 1–8, 2011.

[8] S. Asur and B. A. Huberman, "Predicting the future with social media," in *Proceedings of the Web Intelligence and Intelligent Agent Technology(WI-IAT)*, IEEE, Toronto, Canada, September 2010.

[9] S. González-Bailón, J. Borge-Holthoefer, A. Rivero et al., "The dynamics of protest recruitment through an online network," *Scientific Reports*, vol. 1, p. 197, 2011.

[10] L. Backstrom and J. Kleinberg, "Romantic partnerships and the dispersion of social ties:a network analysis of relationship status on facebook," in *Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work and Social Computing*, pp. 831–841, Baltimore, MA, USA, February 2014.

[11] C. Wang, J. Han, and Y. Jia, "Mining advisor-advisee relationships from research publication networks," in *Proceedings of the 16th ACMSIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 203–212, Washington, DC, USA, July 2010.

[12] L. Aggeborn, N. Lajevardi, K.-O. Lindgren, P. Nyman, and S. Oskarsson, "Parents, peers, and politics: the long-term effects of vertical social ties," *Quarterly Journal of Political Science*, vol. 15, no. 2, pp. 221–253, 2020.

[13] Y. Sun, R. Barber, and M. Gupta, "Co-author relationship prediction in heterogeneous bibliographic networks," in *Proceedings of the International Conference on Advances in Social Networks Analysis and Mining*, pp. 121–128, Kaohsiung, Taiwan, July 2011.

[14] P. Rozenshtein, N. Tatti, and A. Gionis, "Inferring the strength of social ties: a community-driven approach," *Data Structures and Algorithms*, 2019, https://arxiv.org/abs/1902.01832.

[15] K. Kamath, K. Kamath, and A. Sharma, "Detecting strong ties using network motifs," in *Proceedings of the International Conference on World Wide Web Companion*, pp. 983–992, Perth, Australia, April 2017.

[16] K. Farrahi and K. Zia, "Trust reality-mining: evidencing the role of friendship for trust diffusion," *Human-centric Computing and Information Sciences*, vol. 7, no. 1, 2017.

[17] S. Otte, T. Schmitt, and K. Friston, "Inferring adaptive goal-directed behavior within recurrent neural networks," in *Proceedings of the International Conference on Artificial Neural Networks*, Alghero, Italy, September 2017.

[18] R. Azizur, Q. Tie, and N. Zhaolong, "Social acquaintance based routing in vehicular social networks," *Future Generation Computer Systems*, vol. 93, pp. 751–760, 2019.

[19] X. Li, B. Kao, Z. Ren, and D. Yin, "Spectral clustering in heterogeneous information networks," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 33, no. 1, pp. 4221–4228, Honolulu, HI, USA, January 2019.

[20] Y. Song, J. Lu, H. Lu, and G. Zhang, "Fuzzy clustering-based adaptive regression for drifting data streams," *IEEE Transactions on Fuzzy Systems*, vol. 28, no. 3, pp. 544–557, 2020.

[21] L. Ma, J. Li, Q. Lin, M. Gong, C. A. Coello Coello, and Z. Ming, "Reliable link inference for network data with community structures," *IEEE Transactions on Cybernetics*, vol. 49, no. 9, pp. 3347–3361, 2019.

[22] M. R. Haddad and H. Baazaoui, "An adaptive and interactive recommendation model for consumers' behaviours prediction," *International Journal of Data Mining, Modelling and Management*, vol. 10, no. 1, p. 89, 2018.

[23] L. Guo and D. Z. Zhang, "EC-Structure: Establishing consumption structure through mining e-commerce data to discover consumption upgrade," *Complexity*, vol. 2019, Article ID 6543590, 8 pages, 2019.

Hindawi
Mathematical Problems in Engineering
Volume 2021, Article ID 6664413, 8 pages
https://doi.org/10.1155/2021/6664413

*Research Article*

# Prediction of Short-Time Rainfall Based on Deep Learning

**Dechao Sun, Jiali Wu, Hong Huang, Renfang Wang, Feng Liang, and Hong Xinhua**

*College of Big Data and Software Engineering, Zhejiang Wanli University, Ningbo, China*

Correspondence should be addressed to Renfang Wang; renfang_wang@126.com and Hong Xinhua; hongxinhua@zwu.edu.cn

Received 17 December 2020; Revised 24 February 2021; Accepted 10 March 2021; Published 31 March 2021

Academic Editor: Chonghui Zhang



Short-time heavy rainfall is a kind of sudden strong and heavy precipitation weather, which seriously threatens people's life and property safety. Accurate precipitation nowcasting is of great significance for the government to make disaster prevention and mitigation decisions in time. In order to make high-resolution forecasts of regional rainfall, this paper proposes a convolutional 3D GRU (Conv3D-GRU) model to predict the future rainfall intensity over a relatively short period of time from the machine learning perspective. Firstly, the spatial features of radar echo maps with different heights are extracted by 3D convolution, and then, the radar echo maps on time series are coded and decoded by using GRU. Finally, the trained model is used to predict the radar echo maps in the next 1-2 hours. The experimental results show that the algorithm can effectively extract the temporal and spatial features of radar echo maps, reduce the error between the predicted value and the real value of rainfall, and improve the accuracy of short-term rainfall prediction.

## 1. Introduction

Short-term heavy precipitation is a kind of weather process with sudden heavy rainfall, short precipitation time, and large precipitation. Every year, natural disasters caused by it emerge endlessly, which seriously threaten people's life and property safety. It is of great significance to carry out early warning and near forecast for disaster prevention and mitigation.

Radar echo maps extrapolation technology is the main technical means of precipitation nowcasting [1]. According to the observed echo maps, the echo's intensity distribution and the echo body's moving speed and direction (such as the rainfall area) are determined. By linear or nonlinear extrapolation of the echo body, the future radar echo maps can be predicted. At present, radar echo extrapolation technology is mainly divided into four categories: single body centroid-based method, cross-correlation-based method, optical flow-based method, and machine learning-based method. Through the recognition and analysis of radar echoes maps, the single body centroid-based method can obtain the features of thunderstorm cells, such as the thunderstorm center, the thunderstorm volume, and the weight center of reflectivity factor and extrapolate features of these thunderstorm movements to make convective near prediction [2–6]. Single body centroid-based method is suitable for tracking isolated, large, and strong echo monomer or monomer group, but the tracking success rate is low when the echo is scattered, merged, or split.

Cross-correlation-based method calculates the optimal spatial correlation of different regions of radar echo at adjacent time to determine the moving vector characteristics of echo maps and extrapolates the position of radar echo at future time [7–12]. This method is intuitive, simple, and easy to implement. When the shape, moving speed, and direction of radar echo image change gently, cross-correlation-based method can achieve better results. However, it is difficult to ensure the accuracy of the tracking for the strong convective weather process with rapid change of echo image because it is only simple to calculate the correlation coefficient.

Optical flow-based method is a tracking method in the field of computer vision. When there is relative motion between the observed target and the sensor, the motion of the brightness mode observed is called optical flow. This method has also been applied in the meteorological field [13–18]. The optical flow based method can get a better overall movement trend for the heavy convective precipitation. However, this method is required to follow the

invariance assumption, and radar echo has generation and elimination evolution to a certain extent. Therefore, the nonconservation of reflectivity factor leads to extrapolation error, and the error for fast-moving echo is large.

Machine learning-based method uses its self-learning ability to obtain some hidden features of echo changes and shows good memory and association ability [19, 20]. It has been applied as classification model and numerical prediction model in weather forecast, showing the potential and broad prospects of applying neural network model to radar echo extrapolation [21, 22]. In particular, it has recently used deep learning to process meteorological big data, showing strong technical advantages and performance, which has received great attention from the industry [23–25].

By transforming the reflectivity factor into a gray image, the prediction of short-term and imminent rainfall is transformed into a video prediction problem [20]. RNN model, 2D CNN model, and 3D CNN model are three common network structures for video prediction. Attali and Montanvert [23] proposed the first RNN-based video prediction model by using convolutional RNN to encode observation frames; Dey and Zhao [26] proposed the LSTM based codec network model, using one LSTM to encode input frames and another LSTM to predict future frames; Tam and Heidrich [24] used ConvLSTM to replace LSTM to better capture temporal and spatial correlation. Giesen et al. [27] and Faraj et al. [28] extend the model in [20] to predict the transformation of the input frame instead of directly predicting the original pixel. Shen et al. [29] use RNN to capture temporal motion and CNN to capture spatial features to predict video sequences. The 2D CNN network model proposed in [30] regards video frames as multichannel. At present, the application of deep learning technology in related fields is in full swing, but its application in weather prediction is still in its infancy.

In order to capture well the spatiotemporal correlations and solve the problem that low accuracy of short-term rainfall prediction, this paper proposes a Conv3D-GRU model for radar echo short-term and imminent rainfall prediction. We extend the idea of 2D convolution to 3D convolution which is introduced to extract spatial dimension features of radar echo images at different heights. Then, the GRU network is built to extract the time dimension features. Finally, we build an end-to-end trainable model to realize the short-term rainfall forecast in the future. When evaluated on the radar echo dataset that our Conv3D-GRU model consistently outperforms both the Conv2D and the Conv2D-GRU, which can effectively improve the accuracy of short-term rainfall prediction.

## 2. Preliminaries

Short-term precipitation forecast takes the past radar echo extrapolation maps data as input and outputs radar echo extrapolation maps sequence in the future 1-2 hours, which can be summarized as a time series prediction problem.

RNN is known as temporal neural network, which can be used to process data containing time series. The gradient of RNN is easy to disappear in the long network, which makes the learning of RNN model difficult. Long short-term memory (LSTM) network can deal with the above problems. Gated recurrent unit (GRU) is a kind of gated cyclic unit structure and an improved network model of LSTM. In order to improve the training effect, we use the GRU recurrent neural network to learn the features of radar echo sequence data.

LSTM is composed of forgetting gate, input gate, and output gate to form a memory unit to filter input information, while GRU neural network improves the design of "gate," which combines input gate and forgetting gate of LSTM neural network into an update gate $Z_t$. The LSTM is optimized from three gates to two gates. Like LSTM memory unit, GRU is composed of many neural units, each of which is also a complex "gate" structure. Among them, the operation in GRU neural unit can be expressed as follows:

$$\begin{aligned} Z_t &= \sigma\left(W_z \cdot \left[h_{t-1}, X_t\right]\right), \\ r_t &= \sigma\left(W_r \cdot \left[h_{t-1}, X_t\right]\right), \\ \widetilde{h}_t &= \tanh\left(W_{\widetilde{h}_t} \cdot \left[r_t * h_{t-1}, X_t\right]\right), \\ h_t &= \left(1 - Z_t\right) * h_{t-1} + Z_t * \widetilde{h}_t. \end{aligned} \tag{1}$$

From formula (1), it can be seen that each neural unit in the network has a dependency relationship with each other, and each neural unit participates in the decision of information screening. Weight of update door $Z_t$ is expressed as $W_z$. The output of the previous neural unit is $h_{t-1}$, and input of the current neural unit is $X_t$. The sigmoid activation function is represented by $\sigma$. Add the input of the previous neural unit $h_{t-1}$ and the current neural unit $X_t$, multiply with $W_z$ to get the update gate $Z_t$, and then use sigmoid function to operate. When the value is larger, the information of the current neural unit will be retained more, and the information of the previous neural unit will be ignored more in this process.

## 3. Materials and Methods

Although GRU has been proved to be able to deal with the long-range dependence problem well, the input of the network is a one-dimensional vector, and the radar echo maps used in this paper are three-dimensional images. If the GRU is directly applied, the image must be transformed into one-dimensional vector, which will undoubtedly lose a lot of space information of radar maps. To deal with this problem, a network model based on Conv3D-GRU is proposed in this paper. The network can receive the 3D images as input data, so as to better retain the spatial characteristics of radar map. As shown in Figure 1, the model is composed of 3D feature extraction module and GRU-based coding prediction module. The former mainly uses 3D convolution to extract radar echo maps features of different heights at a certain time, while the latter mainly encodes the radar images on time series and predicts the future radar images.

*3.1. Spatial Feature Extraction of 3D Convolutional Neural Network.* In order to extract the spatial features of radar echo maps, we stack radar maps of different heights on a

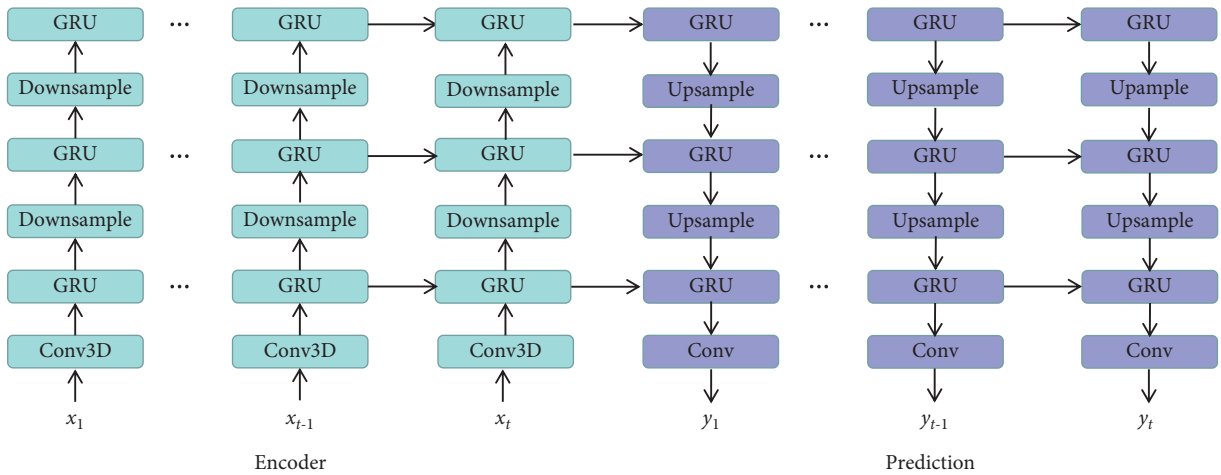



Figure 1: Conv3D-GRU network structure.

certain time $t$ to form a cube and then use 3D convolution to fuse the spatial features in the cube. 3D convolution is an extension of 2D convolution neural network on the dimension, including convolution layer, pooling layer, activation function layer, and fully connected layers. 3D convolution improves the convolution kernel size and pooling layer filter size in the network structure to three-dimensional. At the same time, multidimensional image can also be directly used as the input of neural network. 3D maximum pooling layer selects the maximum pooling method, and the received convolution layer output is a cube data. In fully connected layers, the neurons are connected with all the neurons in the adjacent layer. The neuron vector obtained from the feature space is taken as the input of the full connection layer. The input eigenvector is processed by matrix multiplication, and output is taken as the input of the next GRU module.

The output of 3D convolution is expressed as $f_{ij}^{xyz}$, where $x$ and $y$ represent the spatial dimension of the input image and $z$ represents the time dimension, $\sigma$ represents the activation function, $b_{ij}$ is offset function of the $j$-th feature map in layer $i$, $P_i$, $Q_i$, and $R_i$ represent the size of convolution kernel, $W_{ijm}^{qpr}$ represents the weight of the $(p, q, r)$ neuronal connections in the $M$th feature, $V_{(i-1)m}^{(x+p)(y+q)(z+r)}$ indicates the dimension size of the input information, and $p$, $q$, and $r$ are convolution values, respectively. The 3D convolution process can be expressed as follows:

$$f_{ij}^{xyz} = \sigma\left(b_{ij} + \sum_{m}\sum_{p=0}^{P_{i-1}}\sum_{q=0}^{Q_{i-1}}\sum_{r=0}^{R_{i-1}} W_{ijm}^{qpr} V_{(i-1)m}^{(x+p)(y+q)(z+r)}\right). \quad (2)$$

The loss function of 3D convolution neural network is constructed as follows:

$$\text{Loss} = -\frac{1}{N}\sum_{a=1}^{N}\sum_{c=0}^{C-1} \text{indicatory}^{b} = cP\left(\tilde{y}^{(b)} = c|I^{(b)};\theta\right), \quad (3)$$

where $I^{(b)}$ represents the 3D input vector, $y^{(b)}$ represents the corresponding label, $\tilde{y}^{(b)}$ represents the prediction output, $\theta$ is all parameters, indicatory$^b$ is the indicator function, and $P(\tilde{y}^{(b)} = c|I^{(b)};\theta)$ represents the estimated probability that $I^{(b)}$ belongs to classification $c$.

Table 1 gives the 3D convolution network structure in this paper which consists of one input layer, four 3D convolution layers, and four fully connected layers.

*3.2. GRU Coding: Prediction Module.* As shown in Figure 2, the prediction network of GRU coding-prediction module consists of two modules, namely, (1) GRU coding module, which is used to extract the temporal characteristics of input time series; (2) GRU prediction module, which can predict the radar echo map in the future based on the time series characteristics obtained by the coding module.

According to the time characteristic sequence of the 3D convolution input, the GRU time series prediction network can predict the output sequence of the future period. Firstly, the sequence output of 3D convolution is used as the input time series of the GRU coding module, and then the input data is encoded into a fixed-size state vector, so as to complete the extraction of the time series features of the input time series. At this time, the information of the entire input sequence will be stored in the cell state of the GRU neuron. After that, the GRU prediction module takes the cell states of the abovementioned neurons as the initial state of the module's cells and generates the prediction output sequence of the future time period one by one based on the time series features obtained by the GRU coding module.

After encoding, the extracted radar map feature information is retained in the neural unit $h_t$ of the GRU model. The initial neural unit state $h_t$ of the prediction module is copied from the state of the previous encoder module. According to the characteristic information obtained by the encoder module, the prediction module is responsible for the prediction of rainfall in a short time in the future. The final output sequence corresponds to the input moment one to one. The parameters in the encoding phase are shared, and the parameters in the decoding phase are shared, but the parameters between encoding and decoding are not shared, so the model learns two sets of parameters.

Table 1: 3D convolution network structure.

| Number of layers | Kernel size | Number of kernels | Output |
|---|---|---|---|
| Data | — | — | [5, 101, 101, 3] |
| Conv3D_1 | $5\times1\times1$ | 8 | [5, 101, 101, 8] |
| Maxpooling3D_1 | $3\times2\times2$ | — | [5, 50, 50, 8] |
| Conv3D_2 | $3\times5\times5$ | 16 | [5, 50, 50, 16] |
| Maxpooling3D_2 | $1\times2\times2$ | — | [5, 25, 25, 16] |
| Conv3D_3 | $3\times3\times3$ | 32 | [5, 25, 25, 32] |
| Maxpooling3D_3 | $1\times2\times2$ | — | [5, 12, 12, 32] |
| Conv3D_4 | $3\times3\times3$ | 32 | [5, 12, 12, 32] |
| Maxpooling3D_4 | $1\times2\times2$ | — | [5, 6, 6, 32] |
| Flattern | — | — | 5760 |
| Dense_1 | — | — | 1024 |
| Dense_2 | — | — | 512 |
| Dense_3 | — | — | 256 |
| Dense_4 | — | — | 128 |

## 4. Experiment and Result Analysis

*4.1. Experimental Environment and Dataset.* In order to verify the validity of our Conv3D-GRU model in rainfall prediction on radar echo images, this study will combine the weather radar intensities collected in Ningbo Meteorological Station for training and testing. The environment developed on Win10 OS with 128 G RAM and 2 GHz CPU. CPU processor model is Inter (R) Xeon(R) CPU E5-2683 v3, and deep learning framework is Pytorch-GPU(1.3.1). As rainfall events occur sparsely, we select the rainy days based on the rain barrel information from the radar echo image to build our final dataset. The resolution of radar echo image is $480\times480$ pixels. When preprocessing, we transform the intensity values $Z$ to gray-level pixels $P$ by setting $P = ((Z - \min\{Z\})/(\max\{Z\} - \min\{Z\}))$ and crop the radar maps to $101\times101$ pixels. The weather radar data is recorded every 6 minutes, so there are 240 frames per day to predict the rainfall of radar echo maps in the next two hours. Firstly, the original radar echo map is processed into gray image by linear transformation. Since the original radar echo map has noise interference in the acquisition process, the bilinear filter is used to filter the image to reduce the impact of noise on training and evaluation. Original radar echo images and filtered images are shown in Figure 3.

*4.2. Evaluated Algorithm.* In the process of experimental training, the network structure of Conv3D-GRU was constructed, and the network weight parameters were initialized. Three radar map sequences of different heights were used for size normalization, and frame data was used as the video stream for input. Each altitude receives 5 frames as input and 20 frames output as prediction value. The input radar echo image sequence is processed by 3D convolution network to obtain the spatial dimension features. The time series radar echo features are generated by GRU network, and the convolution decoding module is used to obtain the output of the image sequence. Finally, the heavy rainfall prediction in the next two hours is realized.

In order to evaluate the performance of the algorithm in this paper, critical success index (CSI), Heidke skill score (HSS), mean square error (MSE), mean absolute error (MAE), balanced mean square error (B-MSE), and balanced mean absolute error (B-MAE) were used as evaluation indexes of the model. The score was calculated by using rainfall thresholds of 0.5, 2, 5, 10, and 30. The main formulas are as follows:

$$\begin{aligned}
\mathrm{CSI} &= \frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FN}+\mathrm{FP}},\\
\mathrm{HSS} &= \frac{\mathrm{TP}\times\mathrm{TN}-\mathrm{FN}\times\mathrm{FP}}{(\mathrm{TP}+\mathrm{FN})(\mathrm{FN}+\mathrm{TN})+(\mathrm{TP}+\mathrm{FP})(\mathrm{FP}+\mathrm{TN})},\\
\mathrm{MSE} &= \frac{1}{N}\sum_{i}^{N}(y_i'-y_i)^2,\\
\mathrm{MAE} &= \frac{1}{N}\sum_{i=1}^{N}\left|y_i'-y_i\right|.
\end{aligned} \tag{4}$$

Among them, TP indicates (prediction = 1, truth = 1),FP indicates (prediction = 1, truth = 0), TNindicates (prediction = 0, truth = 0), and FN indicates (prediction = 0, truth = 1). When the values of CSI and HSS are higher, the probability of heavy rainfall is higher. $y_i'$ represents the predicted value, $y_i$ represents the true value, and $N$ represents the number of samples in the test set. The smaller the value of MSE and MAE is, the smaller the error between the predicted value and the true value of rainfall is, and the better the performance of the model is.

*4.3. Analysis of Experimental Results.* In the process of training, Adam optimizer was used to optimize. The batch size was set to 8, the learning rate was set to 0.0001, and the momentum was 0.5. In the experiment of this paper, the radar maps of the first five moments were selected as the input to realize the rainfall prediction of the next 20 moments. Figure 4 shows the rainfall prediction results of radar echo maps, in which group (a) represents the time series of radar echo maps obtained within half an hour; (b) is the real output image sequence (within the next 2 hours); (c), (d), and (e) are three groups of experiments, respectively, which show the predicted output of Conv2D, Cov2D-GRU, and Conv3D-GRU algorithms on radar echo maps sequence in the next 2 hours. The experimental results show that the short-term rainfall prediction of this algorithm is more clear, which can better obtain the features of radar map time and space dimensions at different heights, more accurately predict the future rainfall contour, and use the radar map information to realize the rainfall forecast in the next 2 hours.

*4.4. Evaluation Results.* In order to further verify the performance of the proposed algorithm for short-term heavy rainfall prediction, we compared Conv3D-GRU with other algorithms. It can be seen from Table 2 that

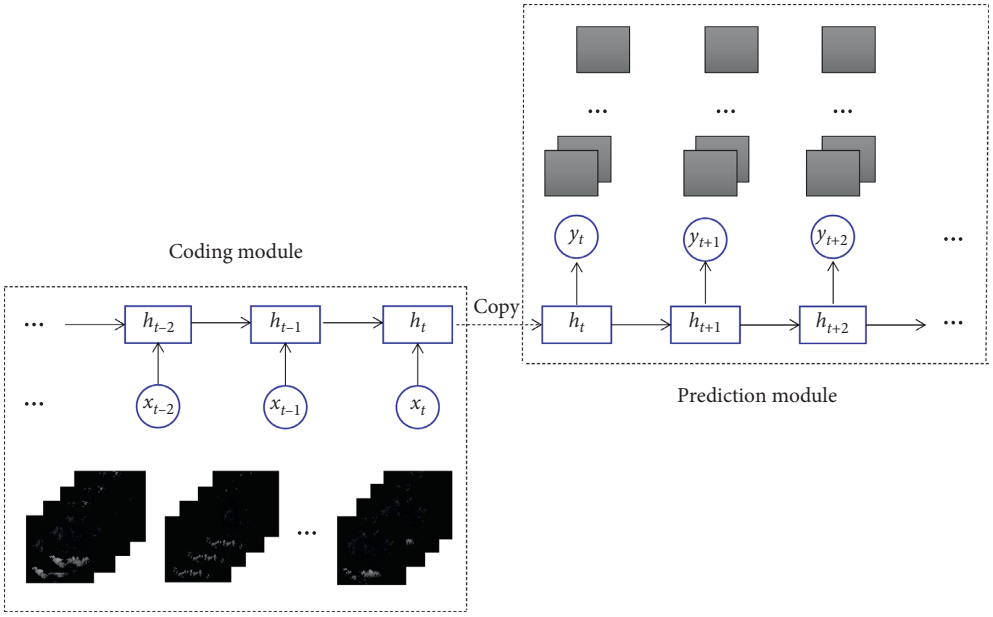
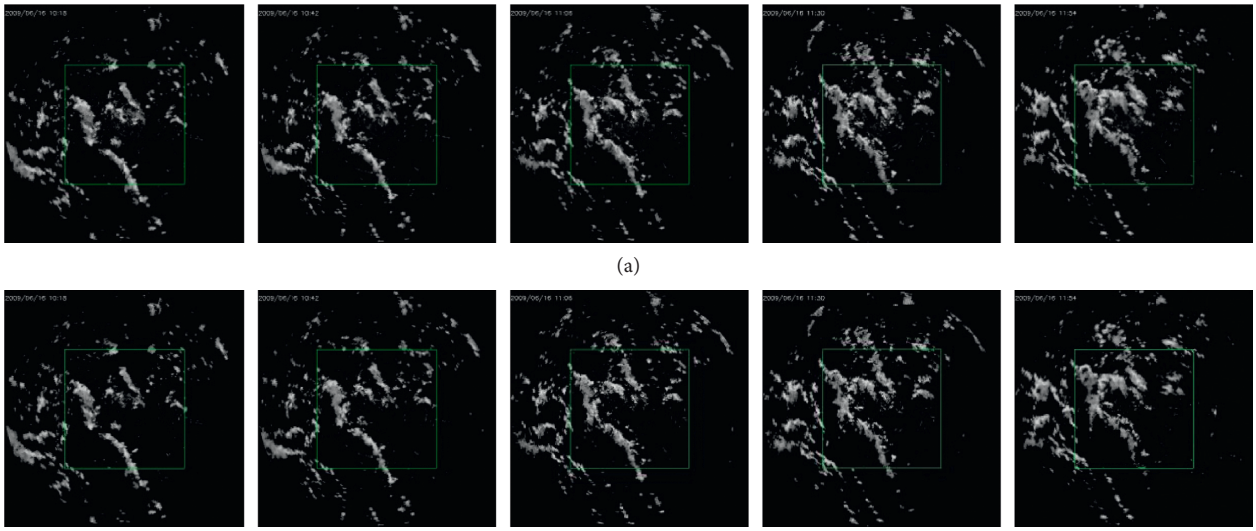


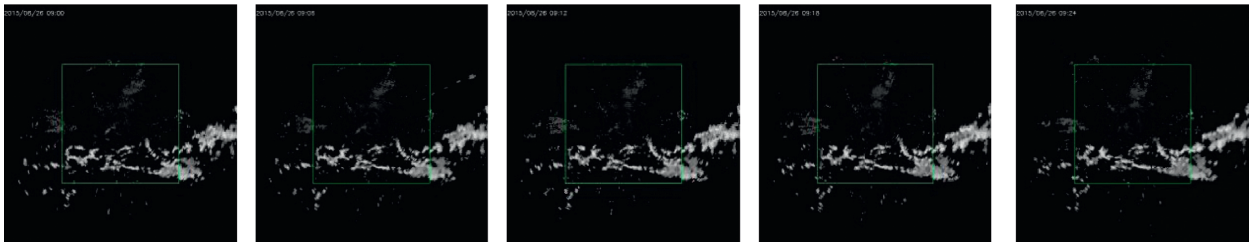

Figure 2: Structure diagram of coding-prediction module.



(a)

(b)

Figure 3: Filtering for image processing. (a) Noise radar echo images. (b) Filtered images.



(a)

Figure 4: Continued.

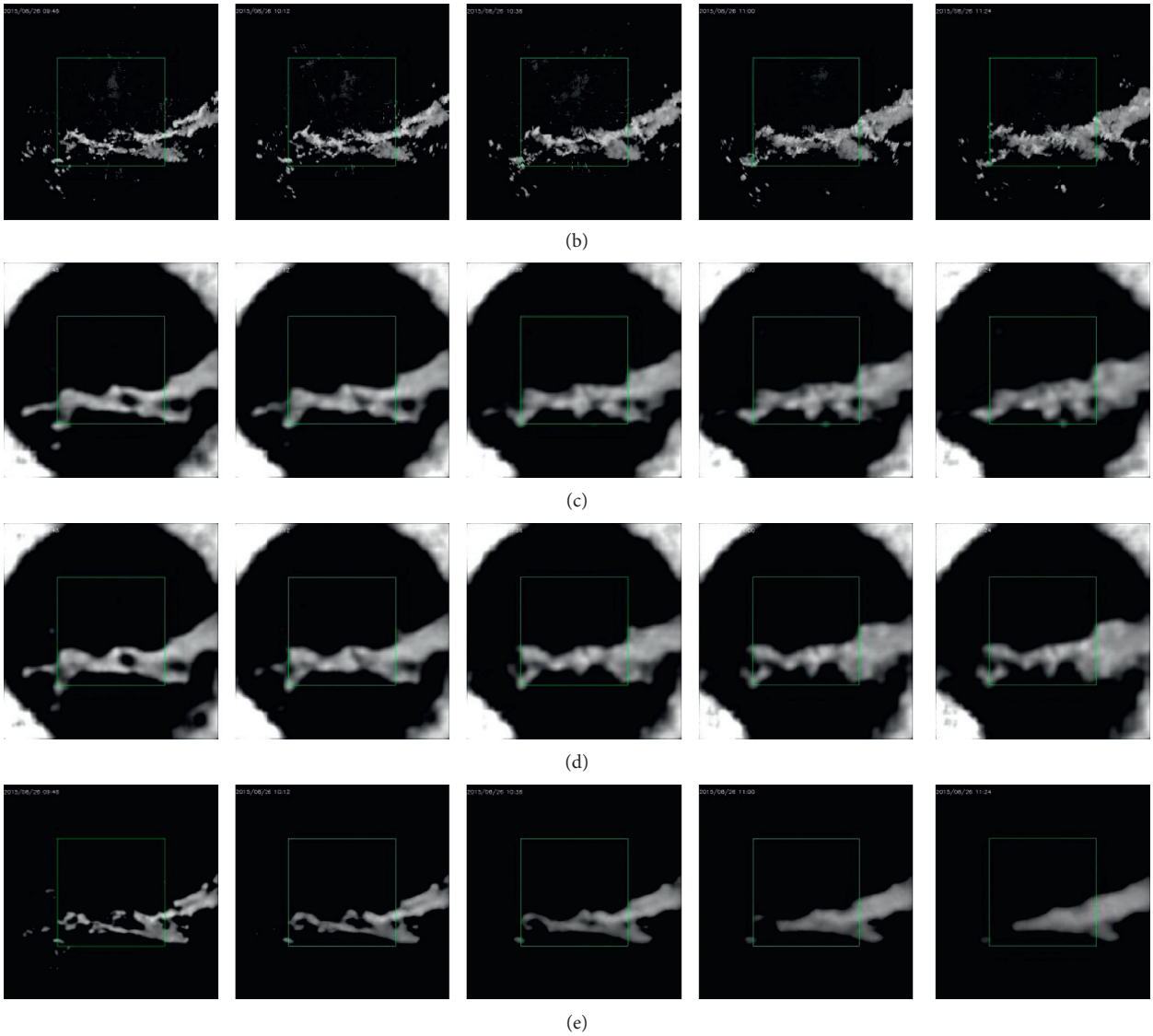



Figure 4: Radar echo map rainfall prediction. (a) Input image. (b) Real value images (i). (c) Conv2D rainfall forecast. (d) Conv2D-GRU rainfall forecast. (e) Conv3D-GRU rainfall forecast in the next two hours.

Table 2: Rainfall threshold evaluation results.

| Algorithms | CSI | | | | | HSS | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | $r \geq 0.5$ | $r \geq 2$ | $r \geq 5$ | $r \geq 10$ | $r \geq 30$ | $r \geq 0.5$ | $r \geq 2$ | $r \geq 5$ | $r \geq 10$ | $r \geq 30$ |
| Conv2D | 0.5048 | 0.4321 | 0.3352 | 0.2384 | 0.1053 | 0.6325 | 0.5743 | 0.4839 | 0.3611 | 0.1857 |
| Conv2D-GRU | 0.5083 | 0.4406 | 0.3421 | 0.2396 | 0.1092 | 0.6348 | 0.5803 | 0.4848 | 0.3689 | 0.1859 |
| Conv3D-GRU | 0.5231 | 0.4518 | 0.3607 | 0.2726 | 0.1616 | 0.6466 | 0.5915 | 0.5062 | 0.4091 | 0.2654 |

"$r \geq \tau$" represents the rainfall threshold of skill score at $\tau$ mm/h, and each cell represents the average score of the next 20 frames. Through comparison, it is found that the average score of this algorithm on CSI and HSS is better than that of Conv2D and Conv2D-GRU algorithms, indicating that the Conv3D-GRU proposed in this paper has achieved good results in extracting radar image features. Moreover, it can make full use of the correlation between the current frame information and the historical frame and the future frame time information, which improves the accuracy of short-term heavy rainfall prediction to a certain extent.

It can be seen from Table 3 that the four indicators of Conv3D-GRU are better than Conv2D and Conv2D-GRU. MSE, MAE, B-MSE, and B-MAE are smaller, which indicates that the error between the predicted value and the real value of rainfall is smaller, and the rainfall prediction of the model is more accurate and has higher performance.

Table 3: Score of Conv3D-GRU model on radar map test set.

| Algorithms | MSE | MAE | B-MSE | B-MAE |
|---|---|---|---|---|
| Conv2D | 2629 | 7295 | 7525 | 18317 |
| Conv2D-GRU | 2582 | 7224 | 7454 | 18180 |
| Conv3D-GRU | 2596 | 6790 | 5911 | 14445 |

## 5. Conclusion

In this paper, we have proposed a Conv3D-GRU model for short-term rainfall prediction of radar echo image to improve the accuracy of regional rainfall prediction.

We construct the Conv3D network to extract the spatial dimension features of radar images at different heights and then build the encoder module to obtain the spatiotemporal feature sequence of radar images. The GRU neural network is introduced to extract the time dimension information. Finally, the prediction module is used to realize the radar echo rainfall prediction in the next 2 hours. Our experimental validation shows that the Conv3D-GRU model consistently outperforms both the Conv2D and the Conv2D-GRU algorithm, and the Conv3D-GRU can capture well the spatiotemporal correlations. The redundancy of spatial data is reduced, and the accuracy of rainfall prediction is improved. Although the Conv3d-GRU model in this paper has achieved satisfactory results in short-term regional rainfall prediction, some remaining research work can be carried out to further improve our method performance. Since the limitation of rainfall timeliness within a certain hour and the influence of single meteorological factors, it is inevitable that there are shortcomings. For future work, we plan to optimize the structure of the constructed neural network and learn more spatiotemporal feature sequence by increasing the network depth of convolution encoder. We will also combine with other meteorological features, such as temperature, wind field, and other information.

## 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

This work was supported in part by the Project of the Science and Plan for Zhejiang Province (nos. LGF19F020008, LGF21F020023, and LGF21F020022), NingBo Science and the Technology Project (no. 2019C50008), Ningbo Natural Science Foundation (nos. 202003N4320, 202003N4324, and 202003N4321), and Zhejiang Education Department Project (no. Y202045246).

## References

[1] H. Blum, "A transformation for extracting new descriptors of shape," *Models for the Perception of Speech and Visual Form*, vol. 19, no. 5, pp. 362–380, 1967.

[2] S. Bouix, J. C. Pruessner, D. Louis Collins, and K. Siddiqi, "Hippocampal shape analysis using medial surfaces," *NeuroImage*, vol. 25, no. 4, pp. 1077–1089, 2005.

[3] R. Dorado, "Medial axis of a planar region by offset self-intersections," *Computer-Aided Design*, vol. 41, no. 12, pp. 1050–1059, 2009.

[4] W. Shen, Y. Wang, X. Bai, H. Wang, and L. Jan Latecki, "Shape clustering: common structure discovery," *Pattern Recognition*, vol. 46, no. 2, pp. 539–550, 2013.

[5] N. Amenta, S. Choi, and R. K. Kolluri, "The power crust, unions of balls, and the medial axis transform," *Computational Geometry*, vol. 19, no. 2-3, pp. 127–153, 2001.

[6] Z. Mu, S. Zeng, and P. Wang, "Novel approach to multi-attribute group decision-making based on interval-valued Pythagorean fuzzy power Maclaurin symmetric mean operator," *Computers & Industrial Engineering*, vol. 155, Article ID 107049, 2021.

[7] B. Miklos, J. Giesen, and M. Pauly, "Discrete scale axis representations for 3D geometry," *ACM Transactions on Graphics (TOG)*, vol. 29, no. 4, pp. 101–109, 2010.

[8] P. Li, B. Wang, F. Sun, X. Guo, C. Zhang, and W. Wang, "Q-MAT," *ACM Transactions on Graphics*, vol. 35, no. 1, pp. 1–16, 2015.

[9] Z. P. Ji, L. Liu, Y. Wang, and D. Zhang, "Approximated skeleton extraction of planar shape with feature preservation," *Journal of Computer-Aided Design & Computer Graphics*, vol. 22, no. 7, pp. 1110–1115, 2010.

[10] F. Aurenhammer, "Power diagrams: properties, algorithms and applications," *SIAM Journal on Computing*, vol. 16, no. 1, pp. 78–96, 1987.

[11] L. Lam, S.-W. Lee, and C. Y. Suen, "Thinning methodologies-a comprehensive survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 14, no. 9, pp. 869–885, 1992.

[12] S. Zeng, Y. Hu, and X. Xie, "Q-rung orthopair fuzzy weighted induced logarithmic distance measures and their application in multiple attribute decision making," *Engineering Applications of Artificial Intelligence*, vol. 100, Article ID 104167, 2021.

[13] H. Zhu and Y. Liu, "Multi-resolutional medial axis generation for solid models," *Journal of Computer-Aided Design & Computer Graphics*, vol. 6, pp. 1110–1119, 2015.

[14] M. Foskey, M. C. Lin, and D. Manocha, "Efficient computation of a simplified medial axis," *Journal of Computing and Information Science in Engineering*, vol. 3, no. 4, pp. 274–284, 2003.

[15] G. Hirota, R. Maheshwari, and M. C. Lin, "Fast volume-preserving free-form deformation using multi-level optimization," *Computer-Aided Design*, vol. 32, no. 8-9, pp. 499–512, 2000.

[16] R. Kimmel, D. Shaked, N. Kiryati, and A. M. Bruckstein, "Skeletonization via distance maps and level sets," *Computer Vision and Image Understanding*, vol. 62, no. 3, pp. 382–391, 1995.

[17] K. Siddiqi, S. Bouix, A. Tannenbaum, and S. W. Zucker, "Hamilton-Jacobi skeletons," *International Journal of Computer Vision*, vol. 48, no. 3, pp. 215–231, 2002.

[18] O. Aichholzer, W. Aigner, F. Aurenhammer, T. Hackl, B. Jüttler, and M. Rabl, "Medial axis computation for planar free-form shapes," *Computer-Aided Design*, vol. 41, no. 5, pp. 339–349, 2009.

[19] S. Zeng, Y. Hu, T. Balezentis, and D. Streimikiene, "A multi-criteria sustainable supplier selection framework based on neutrosophic fuzzy data and entropy weighting," *Sustainable Development*, vol. 28, no. 5, pp. 1431–1440, 2020.

[20] J. Wang, S. Zeng, and C. Zhang, "Single-valued neutrosophic linguistic logarithmic weighted distance measures and their application to supplier selection of fresh aquatic products," *Mathematics*, vol. 8, no. 3, p. 439, 2020.

[21] M. Ramanathan and B. Gurumoorthy, "Constructing medial axis transform of planar domains with curved boundaries," *Computer-Aided Design*, vol. 35, no. 7, pp. 619–632, 2003.

[22] N. Amenta and M. Bern, "Surface reconstruction by voronoi filtering," *Discrete & Computational Geometry*, vol. 22, no. 4, pp. 481–504, 1999.

[23] D. Attali and A. Montanvert, "Computing and simplifying 2D and 3D continuous skeletons," *Computer Vision and Image Understanding*, vol. 67, no. 3, pp. 261–273, 1997.

[24] R. Tam and W. Heidrich, "Shape simplification based on the medial axis transform," in *Proceedings of the 14th IEEE Visualization, 2003 (VIS 2003)*, pp. 481–488, IEEE, Washington, DC, USA, October 2003.

[25] D. Attali and A. Montanvert, "Modeling noise for a better simplification of skeletons,"vol. 3, pp. 13–16, in *Proceedings of the International Conference on Image Processing*, vol. 3, pp. 13–16, IEEE, Lausanne, Switzerland, September 1996.

[26] T. K. Dey and W. Zhao, "Approximate medial axis as a Voronoi subcomplex," *Computer-Aided Design*, vol. 36, no. 2, pp. 195–202, 2004.

[27] J. Giesen, B. Miklos, M. Pauly et al., "The scale axis transform," in *Proceedings of the Twenty-Fifth Annual Symposium on Computational Geometry*, pp. 106–115, ACM, Aarhus, Denmark, June 2009.

[28] N. Faraj, J. M. Thiery, and T. Boubekeur, "Progressive medial axis filtration," in *Proceedings of the SIGGRAPH Asia 2013 Technical Briefs*, ACM, Hong Kong, China, November 2013.

[29] W. Shen, X. Bai, X. Yang, and L. J. Latecki, "Skeleton pruning as trade-off between skeleton simplicity and reconstruction error," *Science China Information Sciences*, vol. 56, no. 4, pp. 1–14, 2013.

[30] Y. Zhu, F. Sun, Y.-K. Choi, B. Jüttler, and W. Wang, "Computing a compact spline representation of the medial axis transform of a 2D shape," *Graphical Models*, vol. 76, no. 5, pp. 252–262, 2014.

Hindawi
Mathematical Problems in Engineering
Volume 2021, Article ID 6632960, 9 pages
https://doi.org/10.1155/2021/6632960

*Research Article*

# Improvement of CT Target Scanning Quality for Pulmonary Nodules by PDCA Management Method

**Dongquan Liu**,[1] **Shaojun Zhu**,[2] **Bangquan Liu**,[3] **Dechao Sun**,[4] **and Fangqin Fei**[5]

[1]*Radiology Department, Ninghai First Hospital Medicare and Health Group, Ningbo, China*
[2]*College of Information Engineering, Huzhou University, Huzhou, China*
[3]*College of Digital Technology and Engineering, Ningbo University of Finance & Economics, Ningbo, China*
[4]*College of Electronics and Computer, Zhejiang Wanli University, Ningbo, China*
[5]*Department of Nursing, The First People's Hospital of Huzhou, Huzhou, Zhejiang, China*

Correspondence should be addressed to Fangqin Fei; feifangqin@139.com

Received 15 December 2020; Revised 6 February 2021; Accepted 6 March 2021; Published 19 March 2021

Academic Editor: Chonghui Zhang



High CT image quality is an important guarantee for doctors to correctly diagnose pulmonary nodules. The aim of this study was to explore the application value of PDCA management method in improving the quality of CT target scanning for pulmonary nodules. We identified 480 patients' CT image with at least one pulmonary nodule admitted in Ninghai First hospital from September 1st, 2018, to April 30th, 2019. 240 CT images are carried out by the conventional target scanning method, and we analyzed the reasons for the low quality of some CT target scanning images of pulmonary nodules in the radiology department of our hospital. We established a new process of CT target scanning for pulmonary nodules based on the PDCA method and then tested 240 patients who were checked after January 1st, 2019. The excellent rate of CT target scanning image of pulmonary nodules in our department increased from 60.0% to more than 90.0%. The patients' satisfaction with the examination was significantly higher than that without the implementation of PDCA management. The research result indicated that the process of CT target scanning image, postprocessing reconstruction, and numerical measurement of pulmonary nodules can be improved by standardized PDCA cycle, which benefits effectively improving the theoretical and operational skills of radiologists and significantly improving the image quality rate of CT target scanning of pulmonary nodules.

## 1. Introduction

Delivering improvements in the quality and safety of clinical diagnosis of pulmonary nodules remain an international challenge [1]. In recent years, quality improvement (QI) methods such as plan-do-check-act (PDCA) cycles [2] have been used in an attempt to drive such improvements. CT target scanning is a key process during clinical diagnosis of pulmonary nodules, and the image quality of CT target scanning of pulmonary nodules is one of the major factors affecting the radiologist's judgment of patients' condition and directly related to the doctor's clinical treatment and patient's rehabilitation [3]. CT target scanning of pulmonary nodules includes the following steps: scanning appointment process, saving and utilization of scanning results, operation specification of doctors in the diagnosis group, respiratory training of patients, and setting of scanning range, which involves the cooperation of registrants, technicians, diagnostic doctors, and patients at various levels [4]. Key components of the imaging process, such as patient scheduling and protocol selection, must occur before image generation, diagnostic interpretation, and report generation can be performed [5–7].

PDCA management method ensures the effect and quality of management by formulating activity methods and steps strictly [8–10]. At present, lung cancer is one of the malignant tumors that endanger human life and health. Its morbidity and mortality are at the top of the list [11]. According to the statistical analysis of cancer epidemiology in 2013, the 5-year survival rate of lung cancer is only 15%

[12]. The survival rate of lung cancer can be significantly improved by surgical resection in time. Therefore, early diagnosis of lung cancer can improve the survival rate of patients and reduce the rate of nodule thoracotomy with lung benign [13].

The PDCA management and the statistical methods [14, 15] are always used to control the occurrence of the problems by analyzing the problems and the causes of the problems [16]. With the development of multislice spiral CT technology and the rapid development of scanning technology in lung cancer screening, the detection rate of pulmonary nodules has significantly increased. However, the focus of pulmonary nodule is usually small, and conventional scanning cannot clearly show the morphological features of the focus. Some phenomena increase the rate of misdiagnosis, such as volume effect after calcification of nodules and solid nodules presented as ground glass nodules. The CT target scan of pulmonary nodules can clearly show the signs of vacuole, lobulation, air bronchus, and margin [17, 18].

The purpose of our study was to apply the advantages of PDCA cycle management to the quality control management of CT target scanning of pulmonary nodule, which improves the image quality significantly.

## 2. Materials and Methods

*2.1. Materials.* The prospective study was approved by the ethical committee of Ninghai First Hospital. In this study, we implemented PDCA management method on January 1, 2019, and selected 60 patients' image of CT target scanning of pulmonary nodules per month from September 1st, 2018, to April 30th, 2019. There was no significant statistical difference between the selected data ($P < 0.05$), and the results were comparable.

*2.2. Methods.* Before implementing the PDCA management method, the routine method is used to manage CT target scanning of pulmonary nodules; since January 2019, the PDCA circulation method is used to manage it and the PDCA project improvement and implementation team are established. All the staff members of the department participated in the team. The director of the department was appointed as the team leader and responsible for the implementation, training, and supervision of the members of the team. The group meeting is held every week.

*2.2.1. Planning (P).* We established a PDCA project improvement implementation team with department director as the leader and organized all members of the department to brainstorm the reasons for the poor image quality of CT target scanning part of pulmonary nodules, the nonuniform standard of image postprocessing and reconstruction, and the nonstandard numerical measurement. The fishbone diagram of poor quality of CT target scan of pulmonary nodules is shown in Figure 1, and the detail plans are shown in Table 1.

*2.2.2. Do (D).* The improvement measures for the noncompliance with the diagnostic specifications (the first reason shown in Table 1) are as follows:

(1) The director of the department organizes all the staff members to study the "CT scan specification of pulmonary nodules" and "imaging diagnosis specification of pulmonary nodules" regularly.

(2) According to the standard, we hold a seminar on the improvement of CT target scanning image of pulmonary nodules and put forward the improving measures and that can be used to solve existing problems on CT target scanning of pulmonary nodules, as follows:

  (i) In order to improve the problem of nonstandard reconstruction image caused by the unreasonable FOV setting of thinning image, the scanning technician selects the upper and lower +3 cm nodules according to the upper and lower ranges of the specifications, and one lung field and half mediastinum are selected for thinning in the left and right ranges. The reconstructed image was not standard due to the unreasonable FOV setting of thinning image.

  (ii) The measures to strengthen the management of scanning plan to improve the problem caused by wrong position selection are as follows: on the day of appointment scanning, the registration personnel shall fill in the basic information of the patient and send it to the doctor of the diagnosis group to mark the location of the nodule; the scanning plan sheet shall be transferred to the CT operation platform for storage; on the day of scanning, the technician shall determine the location of the nodule scanning according to the location marked on the plan sheet and carry out accurate scanning. The lateral position shall be adopted for the CT target scanning of the pulmonary nodule according to the location of the nodule. According to the location of the nodule, the position can be adjusted to recline, supine, and prone so that the nodule is located at the center of the radiation field at the top of the chest.

  (iii) The following measures are taken to solve the problem of blur artifacts caused by poor breath holding: the simple and practical method is adopted so that the patients can carry out breathing training in an understanding way, and it can be demonstrated on the spot, establishing the corresponding dialect training mechanism for county hospital, the voice prompts of inspiratory breath holding operation added to the scanning console, and some certain time inspiratory live demonstration for elderly patients; respiratory training was conducted in the scanning interval of the last patient for improving scanning efficiency.

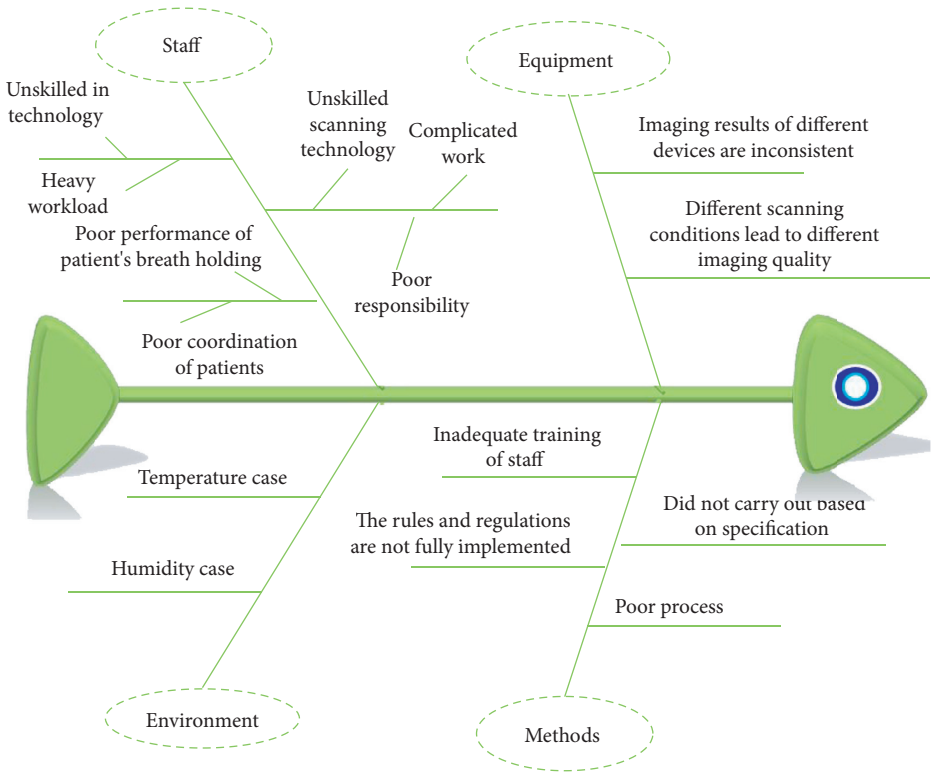



Figure 1: Fishbone diagram of poor quality of CT target scan of pulmonary nodules.

Table 1: Improvement planning.

| Why | What | How | When | Where | Who |
|---|---|---|---|---|---|
| Implementation is not in accordance with specifications | Scanning and reconstruction according to CT scanning standard of pulmonary nodules | Learning CT scan standard of pulmonary nodules | 2018.9–2018.10 | Radiology classroom | All staff |
| | | Discuss the problems of CT target scanning of pulmonary nodules and propose improvement measures | 2018.9–2018.10 | Radiology classroom | All staff |
| | | Checking the theoretical knowledge of staff and implementation of CT scan standard for pulmonary nodules | 2018.9–2018.10 | Radiology classroom | All staff |
| Poor process | Improve the process of appointment, scanning, and reconstruction | Asking about routine lung scanning results when booking | 2018.9–2018.12 | Booking windows | Registrars |
| | | Technician scans according to standard process | 2018.9–2018.12 | Toshiba CT scan console | Scanning technician |
| | | The diagnosis doctor reconstructs the image according to the process of image reconstruction | 2018.9–2018.12 | The reconstruct computer | Diagnostician |
| Unskilled reconstruction technology | Reconstruction technology training | One to one training for senior doctors and junior doctors | 2018.9–2018.10 | Radiology classroom | Senior doctors |
| | | Check the technical ability of scanning image reconstruction | 2018.9–10 | Rebuild console | Department director |

(iv) The unreasonable improvement measures for the measurement of nodule diameter surface are as follows: measure the maximum diameter and its vertical diameter at the maximum level of transverse section, and calculate the average diameter; if there is postprocessing, measure the maximum diameter and its vertical diameter at the second large section on MPR and calculate the average value. We measure them on fc52 with a thickness of 3 mm for effect comparison.

(v) In order to solve the problem of irregular measurement of CT value of nodule, the CT value was measured on fc03, superimposed with 1 mm slice thickness image, and the CT value of lung background was measured for comparison.
(vi) In order to solve the problem of image distortion caused by MIP or MINP reconstruction, we unify the image and select the average menu for reconstruction.

The measures for testing the theoretical knowledge and CT scanning procedures of pulmonary nodules are as follows:

(i) Using "Ding Ding" app to check theoretical knowledge.
(ii) Detecting the scanning image quality after the improvement measures, and trace each nonstandard image to the responsible employee.
(iii) Retraining and assessment of unqualified employees.

The corresponding measures taken due to poor process (the second reason shown in Table 1) are as follows: improve the appointment process of CT target scanning of pulmonary nodule:

(i) Improve the appointment process of CT target scanning of pulmonary nodule: if the routine CT is performed in our hospital, the CT target scanning of pulmonary nodule can be performed directly; if the routine CT is performed in the external hospital, the CT target scanning of pulmonary nodule can also be performed in our department with the CT film and report sheet; if only the chest plain film is performed, the CT target scanning of pulmonary nodule cannot be performed in our department, and the patient should be examined by routine CT scanning before deciding whether to have a CT target scanning.
(ii) Attach the flow chart of CT target scanning (shown in Figure 2) to the box where the target scanning plan is stored, and the technician should process it according to the CT target scanning flow of pulmonary nodule strictly.
(iii) Attach the flow chart of CT target image reconstruction of pulmonary nodules (shown in Figure 3) on the table of the image-reconstruction computer where the target scanning plan is stored, and the diagnoses should reconstruct it according to the reconstruction procedure.

For problem of the unskilled technician (the third reason shown in Table 1), the senior doctors will conduct one-to-one guidance and training for the junior doctors, and the regular assessment will be conducted after the training.

*2.2.3. Check (C).* When they viewed the CT images, the readers were asked to make a diagnosis, evaluate the quality of CT findings, and assign scores using a 100-point score as follows:

(1) Each CT image was subjectively evaluated and graded using a 100-point score, and the detailed scoring items are as follows:
    (i) Scanning position selection (10 points).
    (ii) Respiratory motion artifact (20 points).
    (iii) Scanning range (10 points).
    (iv) Thinning image FOV (10 points).
    (v) Measurement of nodule diameter and surface (20 points).
    (vi) Measurement of nodule CT value + measurement of lung background CT value (20 points).
    (vii) Pulmonary nodule reconstruction mode (5 points).
    (viii) Image enlargement and filling (5 points).
(2) Diagnostic confidence: "90–100" indicated excellent, and "80–90" indicated full good.

*2.2.4. Action (A).* The actions are as follows:

(i) According to the norms of pulmonary nodule scanning and diagnosis, the technician scans the pulmonary nodule to narrow the scanning range and POV value, and the diagnostic physician measures the CT value of ground glass nodule during reconstruction.
(ii) We analyze the unqualified target scanning image to find out the problems and improve continuously every month.
(iii) Administrators need to gather the clinical needs and suggestions every quarter, and the team needs to rectify it and feedback in time.
(iv) Holding training in theory and technology for employees every year.

### 2.3. Observed Indexes

*2.3.1. Image Excellent Rate.* The image quality was evaluated by radiologists from the aspects of noise, artifact, contrast, and resolution. The image excellent rate is given by

$$r = \frac{m}{n}, \tag{1}$$

where $m$ is the number of high-quality CT scanning image and $n$ is the total number of CT scanning image.

*2.3.2. Inspection Time.* In this study, we identified patients with at least one pulmonary nodule who came to our hospital. Two sets of images with 480 images were selected the final analysis. The one set images were selected from September 1st, 2018, to December 31th, 2018 (60 images per month), which were performed using routine CT scanning. The other set images were performed under CT target scanning from January 1st, 2019, to April 30th, 2019 (60 images per month). For each month, we evaluate the quality of CT findings and assign scores using the 100-point method, which is introduced in Section 2.2.3.

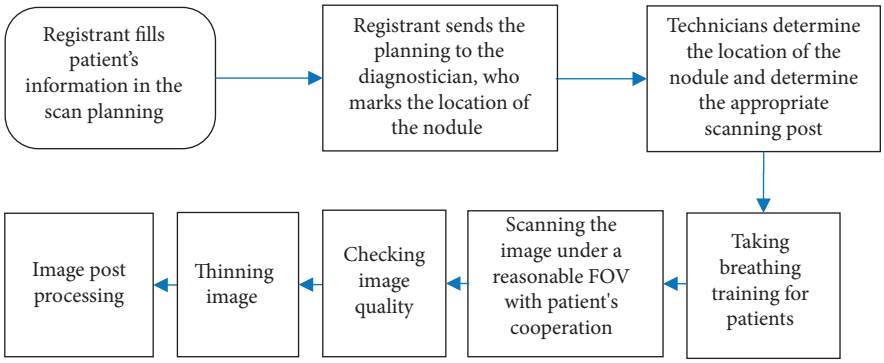



Figure 2: Flow chart of CT target scanning of pulmonary nodules.

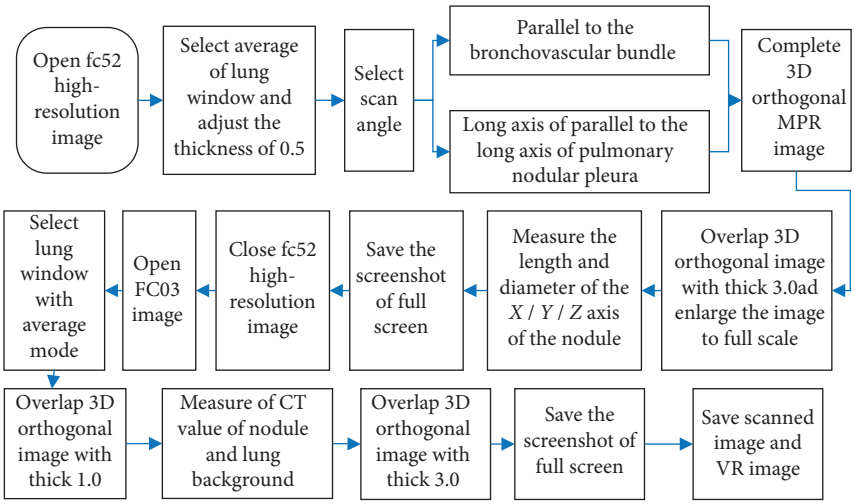



Figure 3: Flow chart of CT target image reconstruction of pulmonary nodules.

*2.3.3. Patient Satisfaction.* The questionnaire survey was used to count the satisfaction of patients with our method. Inclusion criteria were as follows: operation process, service attitude, examination environment, health education, and inspection report and graded using a 20-point score according to a previous report. The total score is 100 points, and the score is directly proportional to the patient's satisfaction.

*2.4. Statistical Analysis.* Data analysis was performed using the SPSS21.0 software package (SPSS Inc. Chicago, IL, United States). The diagnosis results, treatment recommendation, and diagnosis accuracy were determined using McNemar tests. The Wilcoxon signed-rank test was used to evaluate the image quality scores of the findings and confidence of readers. A $P$ value of less than 0.05 was considered statistically significant.

## 3. Results

*3.1. Comparison of Image Excellent Rate.* As a result, 480 patients were included the final analysis, and the detail results of CT image quality are shown in Table 2 (all $P < 0.05$).

*3.2. Comparison of Inspection Time.* Before the implementation of PDCA management, the CT target scan image score of pulmonary nodules is shown in Table 3. From September to December in 2018, the excellent rates of CT images were 60.0%, 76.6%, 83.3%, and 86.3%, respectively.

The excellent rates of CT images with PDCA management have significantly higher score than without PDCA management. From January to April of 2019, the excellent rates of CT images were 90%, 95.0%, 93.3%, and 96.6%, respectively (shown in Table 4).

*3.3. Comparison of Patient Satisfaction.* After the implementation of PDCA management, the patients' satisfaction with the examination was significantly higher than that without the implementation of PDCA management, and the $P$ value of less than 0.05 was considered statistically significant; the results are shown in Table 5.

*3.4. Comparison of CT Image.* Figures 4 and 5 show two representative cases. In cross-sectional images and corresponding MPR images, the edge of nodules and solid internal components were shown more clearly with PDCA management than that without PDCA management.

Table 2: Comparison of excellent rate before and after PDCA methods.

| | Noise | Artifact | Contrast | Resolution |
|---|---|---|---|---|
| Without PDCA | 73.16 ± 5.24 | 76.42 ± 4.32 | 76.16 ± 5.24 | 90.16 ± 5.24 |
| With PDCA | 93.16 ± 2.24 | 94.42 ± 2.32 | 92.16 ± 3.04 | 94.16 ± 2.24 |
| $t$ value | 4.81 | 6.95 | 5.22 | 8.13 |
| $P$ value | <0.05 | <0.05 | <0.05 | <0.05 |

Table 3: The excellent rates of CT images without PDCA management.

| Times | 2018.9 | 2018.10 | 2018.11 | 2018.12 |
|---|---|---|---|---|
| Scanning position selection (10 points) | 6 | 7.7 | 8.4 | 8.5 |
| Respiratory motion artifact (20 points) | 11.5 | 15.6 | 17.1 | 17.5 |
| Scanning range (10 points) | 6.3 | 7.8 | 8.4 | 8.7 |
| Thinning image FOV (10 points) | 5.8 | 7.5 | 8.3 | 8.8 |
| Measurement of nodule diameter and surface (20 points) | 12.6 | 15.8 | 17.4 | 17.8 |
| Measurement of nodule CT value + measurement (20 points) | 12.4 | 15.3 | 17.5 | 18 |
| Pulmonary nodule reconstruction mode (5 points) | 2.8 | 3.7 | 4.1 | 4.3 |
| Image enlargement and filling (5 points) | 3.1 | 3.8 | 4.2 | 4.4 |
| Excellent rates | 60.0% | 76.6% | 83.3% | 86.3% |

Table 4: The excellent rates of CT images with PDCA management.

| Times | 2019.1 | 2019.2 | 2019.3 | 2019.4 |
|---|---|---|---|---|
| Scanning position selection (10 points) | 9.2 | 9.5 | 9.3 | 9.6 |
| Respiratory motion artifact (20 points) | 18.2 | 19.1 | 18.7 | 19.3 |
| Scanning range (10 points) | 9.1 | 9.4 | 9.2 | 9.5 |
| Thinning image FOV (10 points) | 8.9 | 9.3 | 9.3 | 9.4 |
| Measurement of nodule diameter and surface (20 points) | 18.1 | 18.8 | 18.9 | 19.2 |
| Measurement of nodule CT value + measurement (20 points) | 18.1 | 19.3 | 19.1 | 19.4 |
| Pulmonary nodule reconstruction mode (5 points) | 4.1 | 4.7 | 4.6 | 4.8 |
| Image enlargement and filling (5 points) | 4.0 | 4.8 | 4.7 | 4.9 |
| Excellent rates | 90.0% | 95.0% | 93.3% | 96.6% |

Table 5: Comparison of patient satisfaction without and with PDCA methods.

| | Operation process | Service attitude | Examination environment | Health education | Inspection report |
|---|---|---|---|---|---|
| Without PDCA | 15.2 ± 1.3 | 15.5 ± 2.2 | 14.2 ± 2.1 | 13.0 ± 2.4 | 15.6 ± 0.7 |
| With PDCA | 18.2 ± 1.4 | 17.8 ± 2.3 | 18.1 ± 1.7 | 17.4 ± 2.4 | 17.1 ± 0.6 |
| $t$ value | 4.16 | 1.76 | 3.42 | 3.52 | 9.85 |
| $P$ value | <0.05 | <0.05 | <0.05 | <0.05 | <0.05 |

## 4. Discussion

The PDCA method originates from industry and Walter Shewhart and Edward Deming's articulation of iterative processes which eventually became known as the four stages of PDCA [2]. PDCA cycles offer a supporting mechanism for iterative development and scientific testing of improvements in medical diagnosis management [19]. In clinical examination, the image quality obtained by CT target scanning is often not high, thus affecting the diagnosis of patients [20–22]. However, the excessive and large region CT examination will increase the risk of high cumulative radiation doses. Therefore, it is very important to manage the quality of CT target scan scientifically and effectively. Compared with the standard of technical and diagnostic, the quality of some images of CT target scanning of pulmonary nodules in our department is not high, the standards of image post-processing and reconstruction are not unified, and the numerical measurement is not standardized.

In this study, PDCA management method was carried out in CT target scanning of pulmonary nodules in our department. In the planning stage (P), we analyzed the main reasons for the low excellent rate of pulmonary nodule image of CT target scanning in our department. The management plans are made to improve the quality of CT target scanning. The medical staff members were trained according to the operation and diagnosis criterion of CT target scanning of pulmonary nodule. In the implementation stage (D), the formulated measures were implemented strictly and all staff members carried out self-inspection actively. In the check

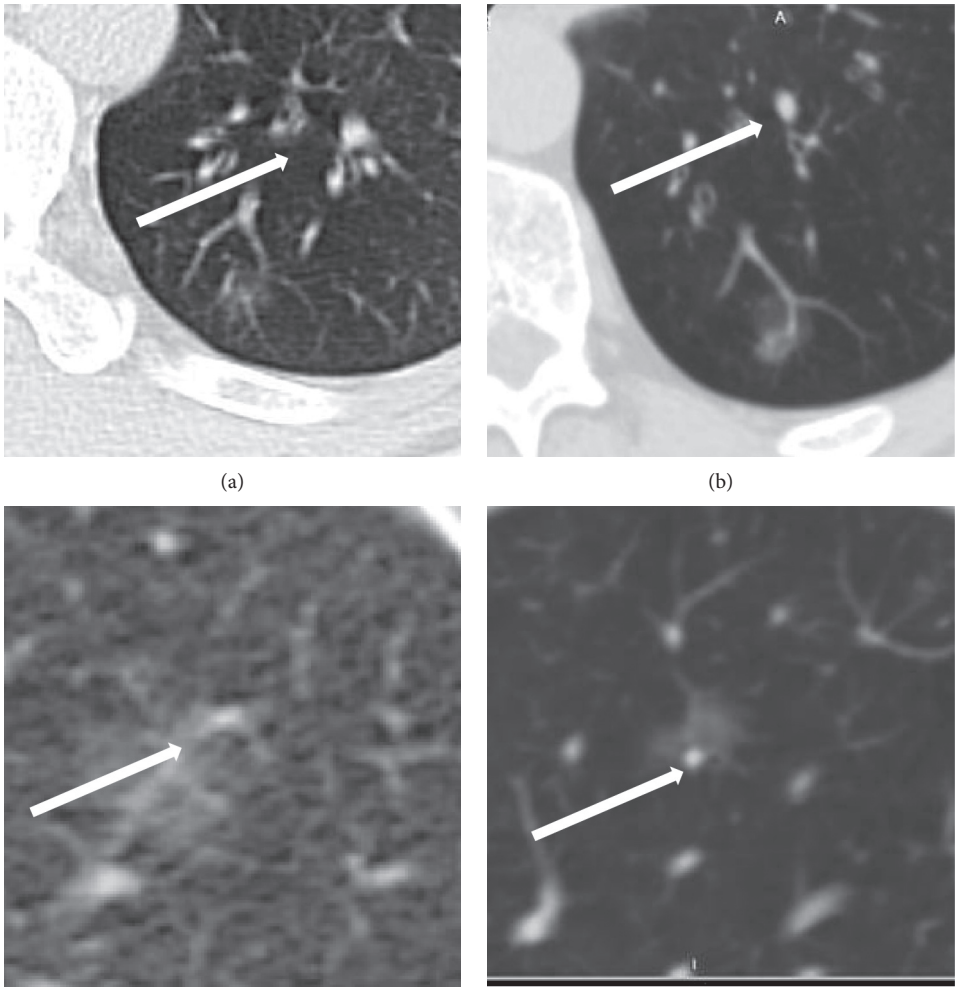



Figure 4: A 55-year-old male patient. The nodule is located in the dorsal segment of the right lower lobe of the lung: (a) cross-sectional without PDCA management; (b) cross-sectional with PDCA management; (c) coronal without PDCA management; (d) coronal with PDCA management.

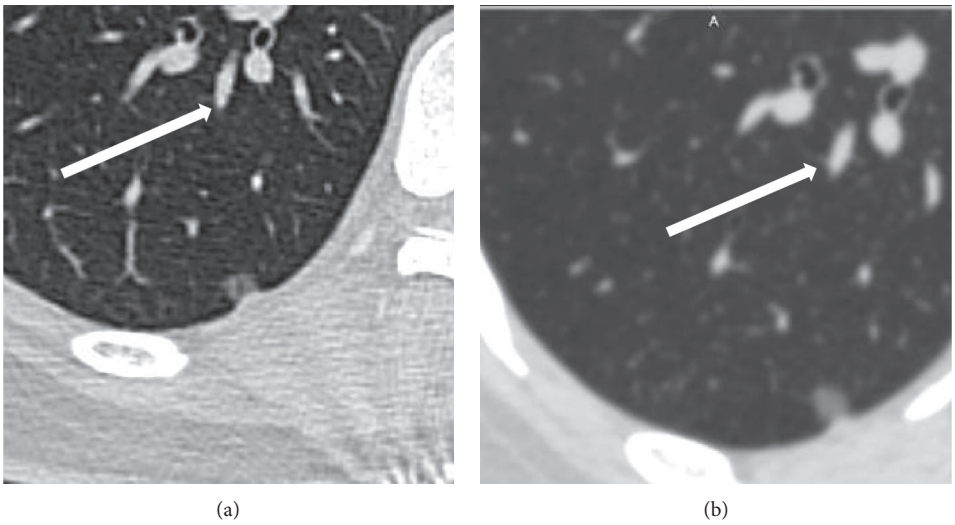
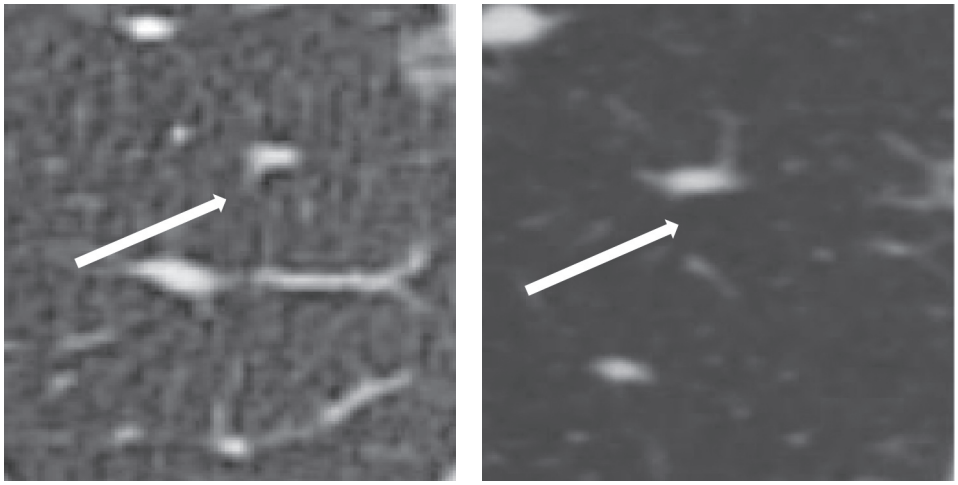



Figure 5: Continued.



(c) (d)

Figure 5: A 33-year-old male patient. The nodule (arrow) is located in the right lung apex: (a) cross-sectional without PDCA management; (b) cross-sectional with PDCA management; (c) coronal without PDCA management; (d) coronal with PDCA management.

stage (C), the implementation of the measures shall be strictly supervised, and the corresponding operation results shall be spot checked to form the quality table of target scanning of pulmonary nodule. We also summarized and analyzed the existing deficiencies and recorded all rectification measures. In the action stage (A), we analyzed the unqualified target scanning images every month to find out problems and improve them continuously. The problems are found to be continuously improved and incorporated into the next cycle management. Then, we put the new problems into the next cycle management and make a new plan to gradually improve and perfect the procedure.

The results indicated that CT target scanning with PDCA management method can provide a higher image excellent rate compared to without PDCA management method. The application of scanning plan sheet can shorten the examination time and make the localization of scanning nodule more accurate. It can improve the efficiency and quality of the examination process to strengthen the breath holding training of patients according to the CT scan standard of pulmonary nodules. The measurement of the nodules of ground-glass and CT value of lung background can improve the quality of reconstruction image through the use of PDCA cycle to effectively control management of CT target scanning of pulmonary nodule, reduce the radiation dose of patients, optimize the patient's examination process, improve the patient's satisfaction, and achieve good results.

## Data Availability

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

## Ethical Approval

This article does not contain any studies with human participants or animals performed by any of the authors.

## Disclosure

Dongquan Liu and Shaojun Zhu should be regarded as the co-first authors.

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## Authors' Contributions

Dongquan Liu and Shaojun Zhu contributed to the work equally.

## Acknowledgments

This work was supported by the NSF of Zhejiang (Grant nos. LY19F020001 and LQ18F010008), Social Sciences Project of Zhejiang (Grant no. 21NDJC021Z), Science and Technology Plan Project of Zhejiang (Grant no. LGG18F020001), NSF of Ningbo City (Grant nos. 202003N4072, 2017A610115, and 2018A610092), NSF of Huzhou City (Grant no. 2016YZ02), and Peiying Project of Ninghai First Hospital.

## References

[1] M. J. Taylor, C. Mcnicholas, C. Nicolay, A. Darzi, D. Bell, and J. E. Reed, "Systematic review of the application of the plan-do-study-act method to improve quality in healthcare," *BMJ Quality & Safety*, vol. 23, no. 4, pp. 290–298, 2014.

[2] W. E. Deming, *Out of the Crisis*, Massachusetts Institute of Technology Center for Advanced Engineering Study XIII, Cambridge, MA, USA, 1986.

[3] B. Lei, W. Hao, W. Zhu, and K. Yan, "Air-ground rotary scanning imaging system for ground vehicles based on passive infrared detection," *IOP Conference Series Materials Science and Engineering*, vol. 799, pp. 12–15, 2020.

[4] J. E. Neggers, B. Kwanten, T. Dierckx et al., "Target identification of small molecules using large-scale CRISPR-Cas

mutagenesis scanning of essential genes," *Nature Communications*, vol. 9, no. 1, p. 754, 2018.

[5] D. R. Enzmann and D. F. Schomer, "Analysis of radiology business models," *Journal of the American College of Radiology*, vol. 10, no. 3, pp. 175–180, 2013.

[6] B. V. Wessman, A. K. Moriarity, V. Ametlli, and D. J. Kastan, "Reducing barriers to timely MR imaging scheduling," *RadioGraphics*, vol. 34, no. 7, pp. 2064–2070, 2014.

[7] S. Zeng, Y. Hu, and X. Xie, "Q-rung orthopair fuzzy weighted induced logarithmic distance measures and their application in multiple attribute decision making," *Engineering Applications of Artificial Intelligence*, vol. 100, Article ID 104167, 2021.

[8] G. D. Rubin, "Lung nodule and cancer detection in computed tomography screening," *Journal of Thoracic Imaging*, vol. 30, no. 2, pp. 130–138, 2015.

[9] C. Zhang, W. Su, S. Zeng, T. Balezentis, and E. Herrera-Viedma, "A two-stage subgroup decision-making method for processing large-scale information," *Expert Systems with Applications*, vol. 171, no. 3, Article ID 114586, 2021.

[10] M. K. Bader, S. Palmer, C. Stalcup, and T. Shaver, "Using a focus-PDCA quality improvement model for applying the severe traumatic brain injury guidelines to practice: process and outcomes," *The Online Journal of Knowledge Synthesis for Nursing*, vol. 9, 2002.

[11] S. M. Lee, C. M. Park, J. M. Goo, H.-J. Lee, J. Y. Wi, and C. H. Kang, "Invasive pulmonary adenocarcinomas versus preinvasive lesions appearing as ground-glass nodules: differentiation by using CT features," *Radiology*, vol. 268, no. 1, pp. 265–273, 2013.

[12] F. Foley, S. Rajagopalan, S. M. Raghunath et al., "Computer-aided nodule assessment and risk yield risk management of adenocarcinoma: the future of imaging?" *Seminars in Thoracic and Cardiovascular Surgery*, vol. 28, no. 1, pp. 120–126, 2016.

[13] P. Cirujeda, Y. Dicente Cid, H. Muller et al., "A 3D Riesz-covariance texture model for prediction of nodule recurrence in lung CT," *IEEE Transactions on Medical Imaging*, vol. 35, no. 12, pp. 2620–2630, 2016.

[14] J. Wang, S. Zeng, and C. Zhang, "Single-valued neutrosophic linguistic logarithmic weighted distance measures and their application to supplier selection of fresh aquatic products," *Mathematics*, vol. 8, no. 3, p. 439, 2020.

[15] S. Zeng, Y. Hu, T. Balezentis, and D. Streimikiene, "A multi-criteria sustainable supplier selection framework based on neutrosophic fuzzy data and entropy weighting," *Sustainable Development*, vol. 28, no. 5, pp. 1431–1440, 2020.

[16] C. Young, D. Burch, J. Childress, C. Kucharski, E. A. Kazerooni, and M. S. Davenport, "Bacterial contamination CT equipment: use of ATP detection and culture results to target quality improvement," *Academic Radiology*, vol. 24, no. 8, pp. 923–929, 2017.

[17] E. F. Patz Jr, E. Greco, C. Gatsonis, P. Pinsky, B. S. Kramer, and D. R. Aberle, "Lung cancer incidence and mortality in national lung screening trial participants who underwent low-dose CT prevalence screening: a retrospective cohort analysis of a randomised, multicentre, diagnostic screening trial," *The Lancet Oncology*, vol. 17, no. 5, pp. 590–599, 2016.

[18] K. A. Khan, P. Nardelli, A. Jaeger, C. O'Shea, P. Cantillon-Murphy, and M. P. Kennedy, "Navigational bronchoscopy for early lung cancer: a road to therapy," *Advances in Therapy*, vol. 33, no. 4, p. 580, 2016.

[19] E. B. Stein, P. S. Liu, E. A. Kazerooni, K. Barber, and M. S. Davenport, "Reducing variability in orthogonal reformatted image quality associated with axial long-$z$-axis CT angiography," *AJR: American Journal of Roentgenology*, vol. 207, no. 6, pp. 1360–1365, 2016.

[20] Y. Gu, V. Kumar, L. O. Hall et al., "Automated delineation of lung tumors from CT images using a single click ensemble segmentation approach," *Pattern Recognition*, vol. 46, no. 3, pp. 692–702, 2013.

[21] S. P. Raman, J. L. Schroeder, P. Huang et al., "Preliminary data using computed tomography texture analysis for the classification of hypervascular liver lesions: generation of a predictive model on the basis of quantitative spatial frequency measurements–a work in progress," *Journal of Computer Assisted Tomography*, vol. 39, no. 3, pp. 383–395, 2015.

[22] D. H. Sterman, T. Keast, L. Rai et al., "High yield of bronchoscopic transparenchymal nodule access real-time image-guided sampling in a novel model of small pulmonary nodules in canines," *Chest*, vol. 147, no. 3, pp. 700–707, 2015.

Hindawi
Mathematical Problems in Engineering
Volume 2021, Article ID 6688110, 14 pages
https://doi.org/10.1155/2021/6688110

*Research Article*

# Ecological Security Evaluation of Marine Ranching Based on DEMATEL-Fuzzy Comprehensive Evaluation

**Yuan-Wei Du**,[1,2] **Jing Fang**,[1] **and Ping Wang**[3]

[1]*Management College, Ocean University of China, Qingdao 266100, China*
[2]*Marine Development Studies Institute of OUC, Key Research Institute of Humanities and Social Sciences at Universities, Ministry of Education, Qingdao 266100, China*
[3]*Business School, Qingdao University, Qingdao 266100, China*

Correspondence should be addressed to Ping Wang; 6240821@qq.com

Received 17 December 2020; Revised 7 January 2021; Accepted 31 January 2021; Published 25 February 2021

Academic Editor: Shouzhen Zeng



Marine ranching plays an integral role in providing fishery resources. Given the critical importance of protecting the marine ecological system, the ecological security evaluation of marine ranching is significant. However, the existing research rarely involves marine ranching ecological security (MRES) evaluation and does not provide a complete evaluation index system. The purpose of the present study was to estimate MRES through several steps and evaluate it scientifically by fully considering the relationship between the factors. First, the evaluation index system was structured using the Driver-Pressure-State-Impact-Response (DPSIR) framework by following certain principles. Second, the study applied the Decision Making Trial and Evaluation Laboratory (DEMATEL) method and fuzzy comprehensive evaluation (FCE) method to assess MRES. Third, a case study of a marine ranching in the Shandong Province was discussed to demonstrate the applicability of the proposed method. This also illustrated the steps of the evaluation. In this case study, the result of the MRES evaluation was graded as excellent. The results demonstrate that using these methods to evaluate MRES can account for the complex relationships between the factors and the cognitive ability of the experts and thus integrate the experts' information comprehensively, which adds to the credibility of the evaluation results.

## 1. Introduction

The continuous growth of the human population calls for high-quality protein. There has been an increasing demand for marine products, which has stressed fishery production. Severe problems, such as the deterioration of the coastal environment and the rapid decline in biological resources, have materialized along with the intervention of human activities [1]. These issues have made people realize the value of protection of marine bioresources and fostered the idea of marine ranching, which can enable the artificial production of economic marine life. The importance of marine ranching lies in part in its potential to increase the harvest capacity of many species and to carry out planned and purposeful production of the marine resources.

A marine ranch is concerned with the sustainable development of marine ecology, marine economy, and fishery production. Several countries, such as China [2], Japan [3], Australia [4], and Norway [5], consider marine ranching as the most promising method to increase fishery resources and stock, and promote sustainable development of fishery resources. In recent years, many researchers have noticed the important role of marine ranching. These inquiries usually study the following three aspects. First, the concept of marine ranching is defined in a variety of forms in different studies. For example, Zhang et al. held the view that marine ranching is a type of mariculture industry that adopts scientific principles to cultivate and manage fishery resources [6]. Grant et al. regarded that marine ranching involves releasing hatchlings into the wild. However, it is desirable to capture them before they breed or mate with wild individuals

[7]. Kitada explained that the purpose of marine ranching is to catch (to the extent possible) all juvenile fish released in the fishing area, which can be managed by fishermen [3]. Second, several researchers have explored the effective allocation method based on the breeding species of marine ranching. For example, Hadas et al. focused on two methods to evaluate sponge cultivation in marine ranching along with environmental factors, with the aim of improving the quality of life and the yield of the sponge [8]. Taylor et al. explored an effective way to dispose of artificially cultured sea cucumbers under the environmental conditions in northern Australia [4]. Liang et al. showed the natural process-dependent effects of shellfish cultures [9]. Third, marine ranching has different impacts on economy, resources, and environment, among others. For example, Moksness and Støle used the net present value approach to calculate the profitability of marine ranching and explored the factors that affect the profitability of marine ranching activities [5]. Kim et al. adopted a single-region model and a multi-region input-output model to calculate the intraregion and interregion economic impacts of marine ranching [10]. Lee and Zhang applied ecological models (Ecopath, Ecosim, and Ecospace) to evaluate the impact of marine ranching activities on the ecosystem [11]. Grant pointed out that mariculture, including marine ranching, may affect wild populations in several ways and lead to dynamic genetic risks [7].

Ecological security is playing an increasingly significant role in national security and social stability. Ecological security is a complicated system involving natural, landscape, and socioeconomic factors [12]. Ecological security evaluation, which refers to a description of the quality of an ecological environment system and its security status, is a popular research topic [13, 14]. Previously, researchers mainly studied the ecological security evaluation of a region [15], city [16], marine body [17], basin [18], or land [19] from the perspective of ecological early warnings [20], ecological risk [21], or landscape pattern [22]. Although various evaluation approaches have been adopted to address the issue of ecological security in the above fields, the evaluation index systems are constructed by following the framework of the Pressure-State-Response (PSR) model or its extensions [20, 23], and evaluation methods are frequently constructed from multiple criteria decision-making (MCDM) [24–28]. For example, Li et al. proposed a Decision Making Trial and Evaluation Laboratory (DEMATEL) method for probabilistic linguistic term sets to analyze relations of criteria and find key factors in the evaluation system [29]. Ruan et al. proposed an evaluation model of tourism ecological security based on the Driver-Pressure-State-Impact-Response (DPSIR) and data envelopment analysis (DEA) methods in order to evaluate the quality of tourism ecological security from the perspective of efficiency [30]. Han et al. used a fuzzy comprehensive evaluation (FCE) method to calculate, analyze, and evaluate the urban ecological security level from 2003 to 2012 and used the entropy weight method to calculate the index weight [31]. Jiang established an FCE model of urban ecological security and used membership degree to express the evaluation and description of ecological security [32]. Solovjova presented an ecological risk assessment method for the marine ecosystem and a mathematical model under the joint action of natural and human factors [33]. Li and Wei proposed an emergency decision-making style based on D-S evidence theory [34]. Zeng et al. indicated that the information for business decisions is often vague and imprecise, and developed a fuzzy methodology for sustainable supplier selection based on fuzzy information [35].

Many researchers have focused their attention on the problem of marine ranching, and they have carried out valuable research on the above-discussed three aspects. With respect to ecological security, scholars have also reported results in several areas, including regional ecological security and ecological security evaluation. The abovementioned research has expanded the theory of marine ranching and the methods of ecological security. However, few scholars have paid attention to the problem of marine ranching ecological security (MRES). Although previous studies have focused on MRES and its influence paths, they have not paid attention to how to evaluate it [36]. MRES is the overall balance between the resource structure of the marine ranching artificial ecosystem and the marine environment. It maintains environmental protection, resource conservation, and sustainable output of fishery. MRES is significant for the ecological benefit and sustainable development of marine ranching. In particular, MRES can affect the sustainable development of marine resources under the condition of satisfying three aspects, namely, resource supply, ecological environment, and human needs. A safe marine ranching ecological system is conducive to improving the quality of resource supplies and the restoration ability of the ecological environment, thus helping to maximize ecological benefits while ensuring economic benefits and social benefits. The motivation of the present paper is to propose an evaluation mechanism for MRES based on DEMATEL and FCE. The proposed approach focuses on solving the following aspects of these problems:

The first is how to construct an evaluation index system for MRES. The marine environment is a complex and dynamic system, and MRES results from numerous interactions. Thus, there is a need to construct a systematic evaluation index system for evaluating it. It is worth noting that recent studies [37] have found several advantages in applying the PSR framework to coastal or marine systems studies. This framework does not reflect the Driver ($D$) and Impact ($I$) factors. The DPSIR framework was proposed to allow wide application; it is more conducive for identifying the root causes of pressure on ecosystems [38]. Furthermore, the DPSIR framework is more systematic and comprehensive compared with the PSR [1]. Traditionally, the DPSIR framework, which has the potential to describe the relationship between human activities and ecological environmental problems, has been used to study the processes of complex coastal systems and marine management [39]. It is recognized as mature and appropriate for coastal systems. Consequently, in this paper, the MRES evaluation index system is set and organized based on the DPSIR framework and its five aspects.

The second one is how to reflect influence relationships between evaluation indices in MRES. Marine ranching is an

artificial ecosystem, and its MRES is a functional state produced by the interaction of internal factors. According to the complex system theory [40], the factor state in the system and the structural relationship between the factors jointly determine the function of the system. Speaking concretely, the performance value of a particular marine ranching enterprise (MR-E) of each evaluation index represents the factor state, and the influence between the evaluation indices represents the structure of the factors. Thus, MRES is the overall performance of the system function. If we only consider the state of the factors and ignore the relationship between the factors in the process of MRES evaluation, then the influence of system structure on system function is neglected, and the overall MRES results of the evaluation objects are flawed. With regard to this, an effective evaluation method should be used to enhance decision quality. Fortunately, the DEMATEL method is powerful for solving MCDM problems. Unlike other MCDM methods, the DEMATEL method can describe the strength of the relationship between factors. Understanding the causes and the influencing factors is conducive to comprehensive decision-making by stakeholders [41]. It is an effective method for factor analysis and identification, which refers to making full use of experts' experience and knowledge to analyze complex systems; it also reveals the causal relationship between the components [42]. With rich expert knowledge, the interrelation and overall influence of various factors can be analyzed better to identify the structural relationships within the complex system, which is again conducive to comprehensive decision-making by stakeholders. Furthermore, it provides methodical guidance for solving the complex system problem of MRES evaluation [42, 43]. Considering the unique advantages of the approach, the present paper utilizes the DEMATEL method to reflect the influence of relationships of the evaluation indices in MRES.

The third one is how to make evaluations for MRES by combining subjective experience and objective data. A realistic evaluation of marine ranching is generally carried out not only through an MR-E submitting self-evaluation reports on specific issues but also by the government organizing several experts in relevant fields to conduct a group evaluation (MRES should not be an exception). For example, the annual evaluation method for China's national marine ranching demonstration zones stipulates that the management and maintenance personnel of the demonstration zone shall organize and summarize the relevant information of the marine ranching that year and then write its annual work report. The provincial department in charge of fishery must utilize more than five experts to make a comprehensive evaluation of the construction and operation of the demonstration area based on the annual work report and relevant information. This paper holds that the above evaluation process has significant reference value for MRES evaluation, but the following two issues should also be paid attention to in the implementation process. Firstly, in the self-evaluation report, the MR-E should be required to provide objective data on the lowest indices to the extent possible (called MR-E self-evaluation). The MR-E is the direct operator of marine ranching and can be reasonably expected to provide accurate data. For example, the MR-E can provide data on the marine environment, biological resources, and other indices based on the lowest level indices such as target biomass and biodiversity index. As the MR-E is required to only provide objective data for the bottom indices rather than make evaluations, it is conducive for the reduction of self-interest behavior. Secondly, although experts should be invited to carry out MRES evaluation based on the underlying objective data, one must also rationally examine the experts' evaluation ability. The knowledge backgrounds of experts are different, and their judgments on the same problem may also differ. If the evaluation results of individual experts are simply weighted and averaged as group results, the accuracy of evaluation results may be affected owing to information loss. While experts can evaluate relatively simple problems, it is difficult to judge relatively complex problems because there are complicated correlation and influence relations between factors. For example, based on the previous two bottom-level index data samples, experts can easily judge the performance of the evaluation object (MR-E) on the State (S). However, because of the complex action mechanism among the five dimensions of DPSIR, it is difficult to make a comprehensive judgment on the overall performance of the MR-E. Therefore, we should aim to ensure that the problem of direct evaluation by experts is simple enough in the evaluation process. The FCE method presents qualitative results in a quantitative form based on the theory and mathematics of the membership degree of the model. It uses fuzzy mathematics to make an overall evaluation of things or objects subject to various factors [44, 45]. To a certain extent, it avoids the possibility that the evaluation results may be influenced by the subjective feelings of experts, thus further improving the objectivity and credibility of the evaluation results [46]. This approach, which provides a comprehensive and scientific method for MRES, can solve issues that are fuzzy and difficult to quantify and is suitable for solving various nondeterministic problems [47]. Therefore, this paper combines MR-E self-evaluation with expert evaluation and constructs an MRES evaluation method by introducing the FCE method.

The structure of the paper is as follows. The evaluation index system of MRES based on the DPSIR model is presented in Section 2. The evaluation mechanism is proposed in Section 3. In Section 3, DEMATEL-FCE methods are applied to expound the evaluation steps of MRES evaluation. An illustrative case is in Section 4 to demonstrate the applicability of the proposed method. Section 5 concludes the paper.

## 2. Evaluation Index System

The index system is the basis of MRES evaluation. Its construction may impact the effectiveness of the evaluation results. As mentioned in section 1, MRES is the overall performance of the system function, and it is related to aspects such as marine environment, marine resources, and their interaction mechanism. This study mainly constructs the evaluation index system of MRES from the three

objectives of biosafety, environmental safety, and ecosystem safety. These three types of factors have mutual influence and functions, which are affected by socioeconomic activities. To fully reflect the action mechanism between the factors, this paper constructs the MRES evaluation index system according to the DPSIR framework.

The steps to construct the index system are as follows. (1) Collect possible indicators. The literature provides the basis for the selection of indicators [36, 48–51]. The annual evaluation system of the national marine ranching demonstration areas helps in index selection. Thus, several possible indicators are achieved by the above ways. (2) Sort out the indicators. Based on the DPSIR framework, classify the collected indicators along five dimensions (Driver, Pressure, State, Impact, and Response). (3) Determine the MRES evaluation index system. Through research and consultation with experts and following the principles of systematic, comprehensive, forward-looking, and data accessibility [52], the indices are continuously adjusted, eliminated, and supplemented, and 32 evaluation indices of 13 categories for MRES are finally determined, as shown in Figure 1. For convenience, in this paper, the first-level indicator is expressed by factor ($C_i$, $i = 1, \ldots, I$), the second-level indicator by element ($C_{ij}$, $i = 1, \ldots, I$, $j = 1, \ldots J$), and the third-level indicator by index ($C_{ijk}$, $i = 1, \ldots, I$, $j = 1, \ldots, J$, $k = 1, \ldots, K$). The hierarchy of the indicators can be defined as factors on level 1, elements on level 2, and indices on level 3. To simplify the evaluation problem and enhance the reliability of the evaluation information in the actual evaluation process, the indices on level 3 are regarded as the interpretation of the upper-level index, and thus, the established indices on level 3 are the detailed explanations of the elements on level 2. In other words, the indices on level 3, which are a further explanation of the elements on level 2, are helpful for a comprehensive understanding of the current situation of the MRES. It provides clarity of the elements on level 2:

(1) Driver ($C_1$) is the core of factors such as high living standards and ecological security. In the present study, the Driver mainly describes the potential factors that cause the change of MRES. The Driver factor is divided into three types of elements: policy driver ($c_{11}$), economic driver ($c_{12}$), and responsibility driver ($c_{13}$). Among them, policy support ($c_{111}$) and the laws and regulations system ($c_{112}$) influence the policy driver situation. That is, the formulation, improvement, and implementation of policies play a vital role in maintaining MRES. They provide a significant basis for decision-making regarding the marine ranching production mode. In addition, the special financial fund input ($c_{121}$), government subsidies ($c_{122}$), and the economic incentives of the profit margin ($c_{123}$) of the MR-E itself, which are the economic factors inside and outside the enterprises, are also powerful driving forces for the MR-E to carry out ecological security maintenance. To some extent, the enterprise's ecological awareness ($c_{131}$) also subtly influences decision-making. The proportion of investment in environmental protection funds reflects the importance that the MR-E assigns to eco-environmental protection. Media messaging would also benefit the MR-E, i.e., the wider its media influence, the greater are the potential incentives for an MR-E, which may prompt the MR-E to make more positive changes in order to maintain the reputation of the marine ranching brand.

(2) Pressure ($C_2$) is mainly the result of one or more activities. It can cause changes in the natural system (State). The Pressure factor includes two types of elements: development pressure ($c_{21}$) and environmental pressure ($c_{22}$). The application efficiency of policies and funds affects the output of marine ranching construction and is mainly reflected in the numbers of bottom sowing and proliferation and release ($c_{211}$), seaweed field and seagrass bed transplant cultivation ($c_{212}$), and artificial reef construction and maintenance ($c_{213}$). Moreover, the degree of greenness of the farming method ($c_{214}$) affects the status of the MRES. The type of cage input affects the aquatic ecology and operational efficiency of marine ranching. Modern large-scale, intelligent deep-water cages and large-scale ecological fences can be considered efficient. Water pollution ($c_{221}$) may occur during construction and routine maintenance, and uncontrollable natural disasters ($c_{222}$) may also change the state of marine ecological security.

(3) State ($C_3$) relates to changes in the present natural situation as a result of Driver ($C_1$) and Pressure ($C_2$). In the present study, the State factor includes two types of elements: marine environment ($c_{31}$) and biological resources ($c_{32}$). According to the "GB 3097-1997 Seawater Quality Standards," it is possible to judge the standards reached by the water quality of the water in which the MR-E is located. Moreover, according to the "GB 18668-2002 Marine Sediment Quality Standards," the status of marine sediments in the sea area of the MR-E can also be judged. Water quality ($c_{311}$) and marine sediments ($c_{312}$) can reflect the quality of the marine environment. Target biological resources ($c_{321}$) and biodiversity index ($c_{322}$) reflect the level of biological resources to a certain extent, and they can be judged according to the "SC/T9417-2015 Technical Specifications for the Evaluation of the Conservation Effect of Artificial Reef Resources."

(4) Impact ($C_4$) is the outcome of changes in both natural and anthropogenic states [48]. It is mainly manifested in the final result of ecological security changes through Driver ($C_1$), Pressure ($C_2$), and State ($C_3$). The Impact factor can be measured through three types of elements: ecological impact ($c_{41}$), social impact ($c_{42}$), and economic impact ($c_{43}$). Specifically, the ecological change of marine ranching can be reflected by the improvement of water quality ($c_{411}$), the cultivation and restoration

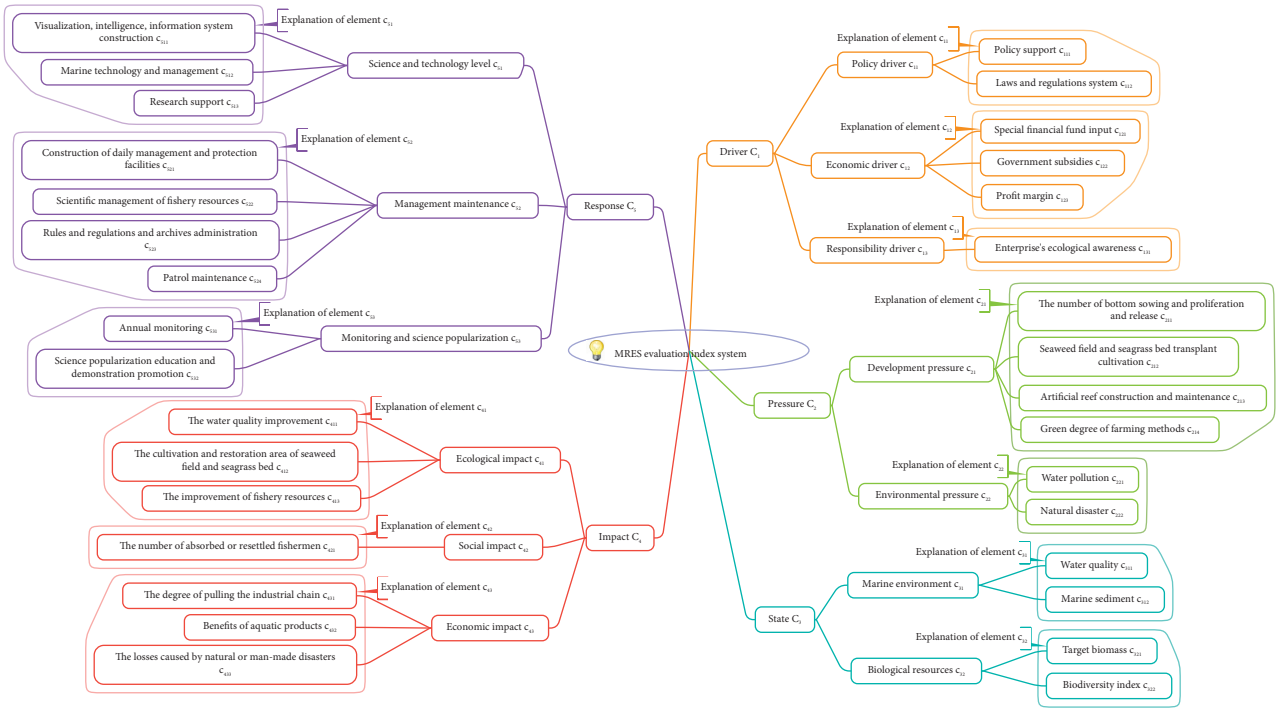



Figure 1: The evaluation index system for MRES.

area of seaweed field and seagrass bed ($c_{412}$), and the improvement of fishery resources ($c_{413}$) relative to the situation in the previous year. If the survival area of the seaweed field or seagrass bed transplanting accounts for more than one-fifth of the total area of marine ranching, it can be seen as having a good impact. The social impact is mainly reflected in the situation of employment in the surrounding areas, namely, the number of absorbed or resettled fishermen ($c_{421}$). On the economic side, if the MR-E creates a new mode of management that can be promoted more widely, it will attract many different enterprises and achieve cooperation, which can result in an economic effect. Based on this, the degree to which the development of the MR-E drives the relevant businesses of the upstream and downstream enterprises of the industrial chain and related enterprises associated with the industrial chain ($c_{431}$) can be taken as one of the indices of economic impact. The benefits of aquatic products ($c_{432}$) and the losses caused by natural or man-made disasters ($c_{433}$) can be regarded as positive or negative economic impacts, respectively.

(5) Response ($C_5$) arises with the changes of MRES. They are active countermeasures taken by the MR-E to prevent ecological deterioration, reduce resource waste, improve biodiversity, and adapt to unexpected changes of the environmental state. The Response factor is mainly reflected by three types of elements: science and technology level ($c_{51}$), management maintenance ($c_{52}$), and monitoring and science popularization ($c_{53}$). Specifically, technological innovation is a strong driving force for ecologically healthy development of marine ranching. It provides reliable guarantees for avoiding ecological risks and improving operational efficiency. In this paper, visualization, intelligence, information system construction ($c_{511}$), marine technology and management ($c_{512}$), and research support ($c_{513}$) are regarded as the aspects to measure the scientific and technological level of marine ranching. Daily management and maintenance can achieve timely stop-loss and reasonable management of resources. It is mainly reflected in the construction of daily management and protection facilities ($c_{521}$), scientific management of fishery resources ($c_{522}$), rules and regulations and archives administration ($c_{523}$), and patrol maintenance ($c_{524}$). In addition, the results of annual monitoring ($c_{531}$) play an important role in the development of MRES, including the number of monitoring activities on resources and the environment and the number of monitoring of technology reports. Science popularization and education ($c_{532}$) can help publicize and promote the achievements of marine ranching.

## 3. Evaluation Method for MRES

*3.1. Evaluation Mechanism.* MRES evaluation is a process to determine the MRES grade by extracting and integrating the performance data of an MR-E in the five dimensions of DPSIR. In order to make evaluations for the MRES by combining subjective experience and objective data, it is

necessary to carry out self-evaluation first, then expert evaluation, and finally comprehensive integration. (1) Self-evaluation is a process of obtaining the performance data of the MR-E on various underlying elements of MRES. As the MR-E is the main body of marine ranching operation, it has the best understanding of the actual situation of marine ranching. Therefore, it can provide objective underlying performance data. (2) Expert evaluation is a process to evaluate the performance of the MR-E on factors based on the objective data provided through the self-evaluation. Since the objective data gives the basis and the evaluation question is simple enough, the expert can independently provide scientific evaluation results. (3) Comprehensive integration is a process of integrating the evaluation information given by experts and the influence relationships among factors to determine the overall performance level of the evaluated MR-E on MRES.

In the evaluation of MRES in this study, three methods (DPSIR, DEMATEL, and FCE) are applied. DPSIR can not only provide an index system for self-evaluation and expert evaluation but also provide a basis for reflecting the influence relationships between factors. The role of DEMATEL is to determine the weights of the factors by combining the influence relationships between factors reflected by DPSIR. The function of FCE is to comprehensively integrate the evaluation results of the MR-E given by all experts on various factors and determine the overall performance grade of the MR-E on MRES.

The evaluation mechanism for MRES as described above is shown in Figure 2.

*3.2. Self-Evaluation and Expert Evaluation.* Self-evaluation is an objective description of each element of the marine ranching across all indices covered by the MR-E. Assume that the MR-E conducts a self-evaluation on the performance of marine ranching on $c_{ij}$, which includes index $c_{ij1}, \ldots, c_{ijK_{ij}}$. Then, the MR-E can describe the specific performance on $c_{ij}$ according to the objective and realistic situation in terms of resources, environment, management, and other aspects. The self-evaluation can be expressed as follows:

Question: How does the marine ranching perform on $c_{ij}$?

Answer: The performance of marine ranching on $c_{ij1}$ is $\text{fact}_{ij1}$, the performance on $c_{ij2}$ is $\text{fact}_{ij2}, \ldots$, the performance on $c_{ijK_{ij}}$ is $\text{fact}_{ijK_{ij}}$.

Therefore, the self-evaluation results given by the MR-E on $c_{ij}$ can be expressed as $X_{ij} = \{\text{fact}_{ij1}, \text{fact}_{ij2}, \ldots, \text{fact}_{ijK_{ij}}\}$, $i = 1, \ldots, 5$, $j = 1, \ldots, J_i$.

Taking development pressure ($c_{21}$) as an example, this element includes four indices (see Figure 1). This study aimed to determine the performance of marine ranching on $c_{21}$. Investigation of the MR-E revealed the following.

$\text{fact}_{211}$: on bottom sowing and proliferation and release ($c_{211}$), the annual number of proliferating and releasing streams or subsiding is 20 million/year. There is evidence that the target organisms of the enterprise's proliferating and releasing streams have established self-breeding populations in this sea area.

$\text{fact}_{212}$: on seaweed field and seagrass bed transplant cultivation ($c_{212}$), the MR-E has carried out the construction of a seaweed farm or seagrass bed that has been transplanted and cultivated. The planting area of this year is 1 hectare.

$\text{fact}_{213}$: on artificial reef construction and maintenance ($c_{213}$), the MR-E has undertaken an artificial reef construction project. The artificial reef release area this year is 163 ha, and the release reef types are proliferative reef and swim fishing reef. In the past three years, side sweep sonar has been used for artificial reef inspection.

$\text{fact}_{214}$: on degree of greenness of farming method ($c_{214}$), the production model that is adopted by the enterprise is a multi-nutrient-level comprehensive breeding mode, and the type of cage put into use in the marine ranching is an ordinary cage with 26 cages.

Therefore, the self-evaluation results on $c_{21}$ can be expressed as: $X_{21} = \{\text{fact}_{211}, \text{fact}_{212}, \text{fact}_{213}, \text{fact}_{214}\}$.

From the above self-evaluation process, it is not difficult to find that the MR-E only describes the facts and thus it is unlikely to lead to a self-promotional evaluation.

Expert evaluation is made on each element that the MR-E belongs to regarding its grade level based on the actual performance of the MR-E on the included indices. Suppose the set of grade levels is defined as $V = (V_1, \ldots, V_Q)$. For example, if $Q = 5$, then the set of grade levels can be defined as $V = (V_1, \ldots, V_5) = (\text{Excellent, Good, Average, Poor, Very Poor})$. For the performance of the MR-E on the element $c_{ij}$, expert $e_n$ can make an evaluation based on the fact $x_{ij}$ and give the assessment $M_{ij}^n$ as in

$$M_{ij}^n = \left( m_{ij}^{nq} \middle| \sum_q m_{ij}^{nq} = 1; m_{ij}^{nq} = 0, 1; q = 1, \ldots, Q \right) \quad i = 1, \ldots, 5 \; j = 1, \ldots, J_i, \; n = 1, \ldots, N, \tag{1}$$

$m_{ij}^{nq}$ describes whether the performance of the MR-E belongs to the grade level $V_q$ on the element $c_{ij}$ or not. Only one grade level is true in the assessment $M_{ij}^n$ and thus $m_{ij}^{nq} = 0, 1$ and $\sum_q m_{ij}^{nq} = 1$. Consequently, the performance of the MR-E on each factor can be described as a membership distribution by integrating those performances on its included elements. On factor $C_i$, the membership distribution of the MR-E can be calculated by

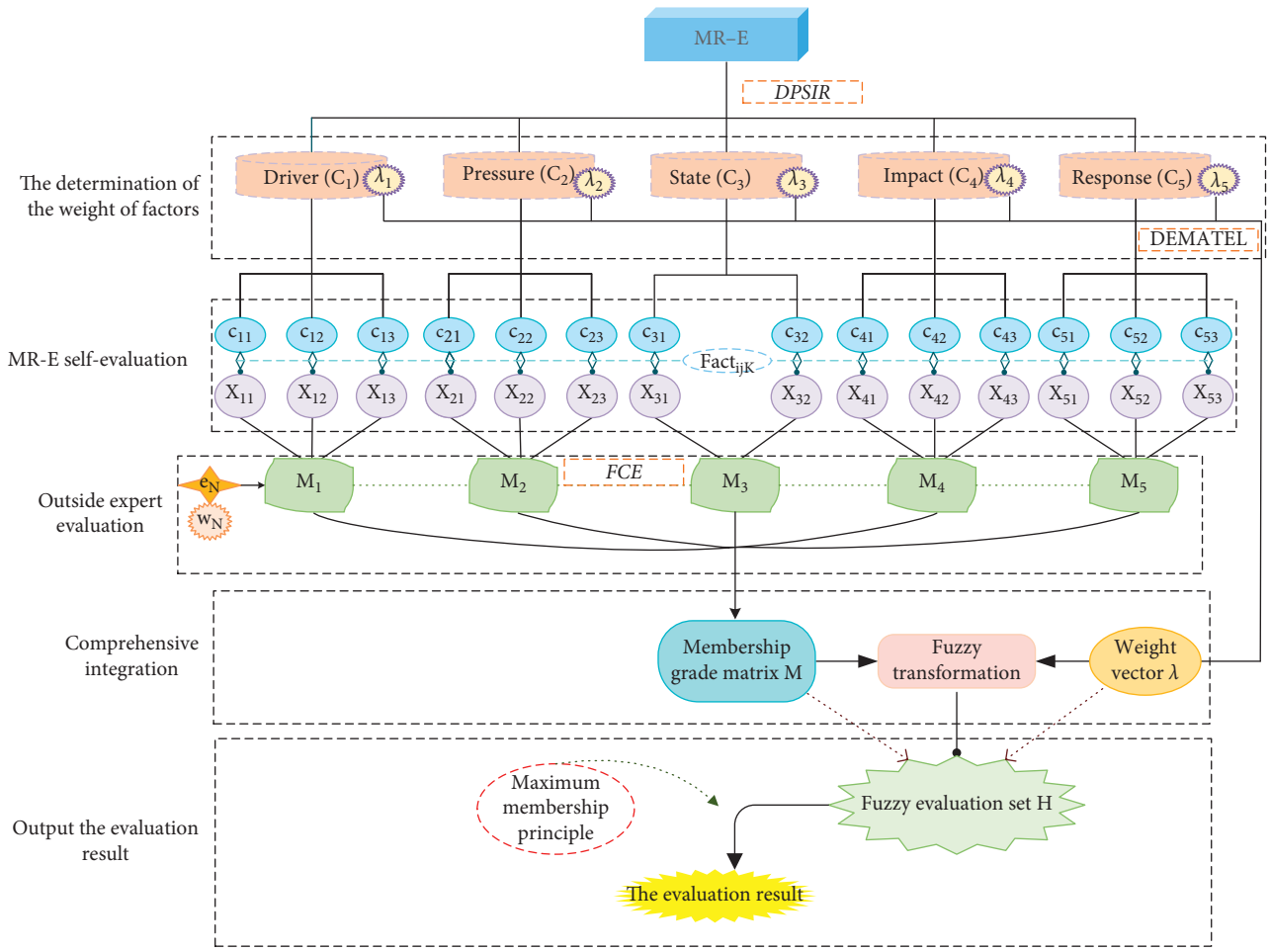



Figure 2: The evaluation mechanism for MRES.

$$M_i^n = \sum_{j=1}^{J_i} \frac{M_{ij}^n}{J_i} = \left( M_i^{nq} | = \sum_{j=1}^{J_i} \frac{M_{ij}^{nq}}{J_i}; q = 1, \ldots, Q \right), \quad i = 1, \ldots, 5,\ j = 1, \ldots, J_i,\ n = 1, \ldots, N. \tag{2}$$

Note that the evaluation ability of experts may be different owing to varied, subjective experiences. Therefore, weights are frequently employed to reflect the relative importance of the different judgments given by the experts [46, 53]. Suppose the weight of expert $e_n$ is $w_n$. The expert weight vector can be described as $W = (w_n | \sum_n w_n = 1, 0 \leq w_n \leq 1, n = 1, \ldots, N)$.

*3.3. Factor Weight Calculation.* As mentioned in Section 1, the MRES factors may interact with each other within the complex marine ranching system. The influence that the system structure has on the system function cannot be ignored. How to determine the weights of factors by considering the interactions between the factors is a crucial problem. Fortunately, it is well-known that DEMATEL is a popular MCDM technique that can find the direct and indirect causal relationships between the factors in complicated decision-making models [54, 55]. It has been successfully used to solve decision-making problems in a variety of situations. This study utilizes DEMATEL to determine the weights of the five factors. It is worth noting that DPSIR not only provides the five factors but also illuminates the direct-influence relationships between them, and thus, the provided relationships can be seen as the input information to DEMATEL. To avoid subjective arbitrariness as much as possible in the DEMATEL method, this study only considers whether the influencing relationship between the factors exists or not. In other words, the grade levels employed in DEMATEL are defined as 0 and 1. The specific steps are as follows.

Step 1: Establish an initial direct-relation (IDR) matrix. According to the relations provided in the DPSIR framework, the IDR matrix of the five factors

$(C_1, \ldots, C_5)$ can be established as $Z = [z_{ii'}]_{5\times5}$. If there is an influence relationship in the DPSIR framework $C_i \longrightarrow C_{i'}$, then let $z_{ii'} = 1$; otherwise, let $z_{ii'} = 0$.

Step 2: Normalize the IDR matrix. The IDR matrix is normalized by $z_{ii'}' = z_{ii'}/\max(\max_{1\le i'\le5}(\sum_i z_{ii'}), \max_{1\le i\le5}(\sum_{i'} z_{ii'}))$, and the normalized IDR matrix can be determined as $Z' = [z_{ii'}']_{5\times5}$.

Step 3: Compute the total relation matrix. To reflect the direct influence along with indirect influence, the total relation matrix is computed by $T = Z'(I - Z')^{-1} = [t_{ii'}]_{5\times5}$, where $I$ means the identity matrix.

Step 4: Determine the weights of factors. The weights of factors can be determined by computing the factor's influencing degree and influenced degree. The influencing degree of factor $C_i$ is computed by $f_i = \sum_{i'} t_{ii'}$, and the influenced degree is computed by $f_i' = \sum_{i'} t_{i'i}$. The prominence degree of factor $C_i$ is computed by $g_i = f_i + f_i'$, $i = 1, \ldots, 5$. The weight of factor $C_i$ can be determined by $\lambda_i = g_i/\sum_i g_i$, $i = 1, \ldots, 5$. The weights of factors are denoted by $\lambda = (\lambda_i | \sum_i \lambda_i = 1, 0 \le \lambda_i \le 1, i = 1, \ldots, 5)$.

*3.4. MRES Evaluation.* As mentioned in Subsections 3.2 and 3.3, each expert can evaluate the performance level $M_{ni}$ $(\forall i, n)$ of the MR-E on factor $C_i$ by considering the self-evaluation results $X_{i1}, \ldots, X_{iJ_i} (\forall i, j)$. In addition, DEMATEL can get the weights of factors as $\lambda = (\lambda_i | \sum_i \lambda_i = 1, 0 \le \lambda_i \le 1, i = 1, \ldots, 5)$ based on the DPSIR model. With the above information as the input, the FCE method is employed to determine the evaluation results of the MR-E on the five factors. The specific steps are as follows.

Step 1: Construct the fuzzy membership matrix. To describe the judgment information of all experts, it is necessary to construct an integrated fuzzy membership matrix. The membership function on factor $C_i$ can be obtained by integrating expert judgment and expert weight, which is shown as equation (3). The fuzzy membership matrix can be described as $M = (m_i^q)_{5\times Q}$.

$$M_i = \left( m_i^q | m_i^q = \sum_n w_n m_{ni}^q, \sum_q m_i^q = 1, 0 \le m_i^q \le 1, q = 1, \ldots, Q \right), \quad i = 1, \ldots, 5. \tag{3}$$

Step 2: Combine the factor weight $\lambda$ with the fuzzy relationship matrix $M$. To obtain the final result of the evaluated MR-E, a fuzzy operator $\circ$ should be selected to integrate the evaluation performance of the MR-E on different factors, i.e., $H = M \circ \lambda = (h_1, \ldots, h_Q)$. To fully use the decision information on all factors, this study suggests the weighted average operator to make combinations, i.e., $h_q = \sum_i \lambda_i m_i^q$, $q = 1, \ldots, Q$.

Step 3: Make comprehensive evaluation. Comprehensive membership vector $H$ represents the probability distribution of the comprehensive performance of the MR-E on all grade levels, and the comprehensive judgment can be obtained based on the maximum membership principle. Then, the $v_*$ corresponding to $h(v_*) = \max(h(v_1), \ldots, h(v_Q))$ is the final grade level to which the MR-E belongs.

*3.5. Summarization for the MRES Evaluation.* The process of the MRES evaluation is summarized as follows.

Step 1: Construct the MRES evaluation index system. Construct the evaluation index system for MRES as shown in Figure 1, which includes three levels, namely, the factor level, the element level, and the index level.

Step 2: Conduct self-evaluation. The MR-E is required to objectively evaluate the performance of marine ranching on $c_{ij}$ and give the fact set as $X_{ij} = \left\{\text{fact}_{ij1}, \text{fact}_{ij2}, \ldots, \text{fact}_{ijK_{ij}}\right\}$, $i = 1, \ldots, 5$, $j = 1, \ldots, J_i$.

Step 3: Undertake expert evaluation. According to the self-evaluation results $X_{i1}, \ldots, X_{iJ_i}$, expert $e_n$ with weight $w_n$ is required to evaluate the performance of the MR-E on the element $c_{ij}$ and give the assessment $M_{ij}^n$ as in equation (1). The membership distribution of the MR-E on factor $C_i$ is calculated by equation (2) and $M_i^n$ is determined, $i = 1, \ldots, 5$, $n = 1, \ldots, N$.

Step 4: Determine factor weights. The factor weight $\lambda = (\lambda_i | \sum_i \lambda_i = 1, 0 \le \lambda_i \le 1, i = 1, \ldots, 5)$ is determined by employing the DEMATEL method as in subsection 3.3.

Step 5: Conduct the MRES evaluation. Taking the membership distribution of the MR-E on factor $C_i (M_{ij}^n, i = 1, \ldots, 5, n = 1, \ldots, N)$ and factor weight $(\lambda)$ as inputs, the final grade level of the MR-E is determined by the FCE method as in Subsections 3.4.

The MRES evaluation process is shown in Figure 3.

## 4. Case Simulation Study

*4.1. Background.* Shandong Blue Ocean Technology Co., Ltd. (SBOT), located in Yantai, Shandong Province, was established in December 2011. It has a registered capital of 80 million yuan, total assets of 411.99 million yuan, 260 employees, and 160,000 mu of confirmed sea area. Its business scope is the research, development, and promotion of aquatic products technology and the breeding, processing, refrigeration, sales, and tourism of sea cucumbers and other sea treasures. SBOT is an MR-E, and its marine ranching has been made a national demonstration area. In this case

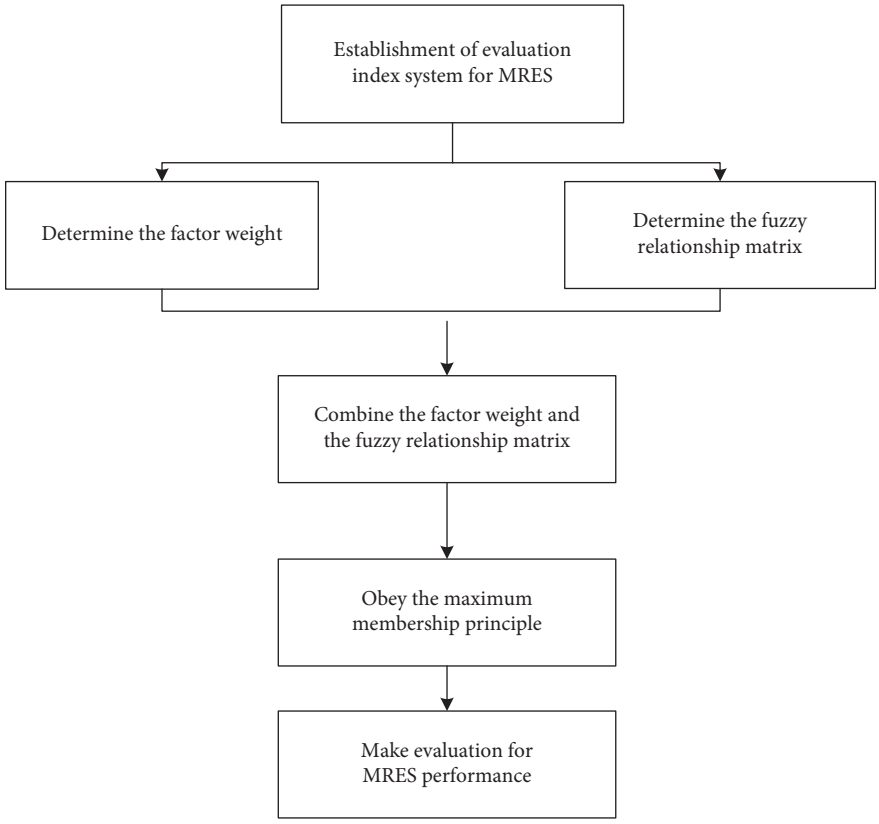



Figure 3: MRES evaluation framework.

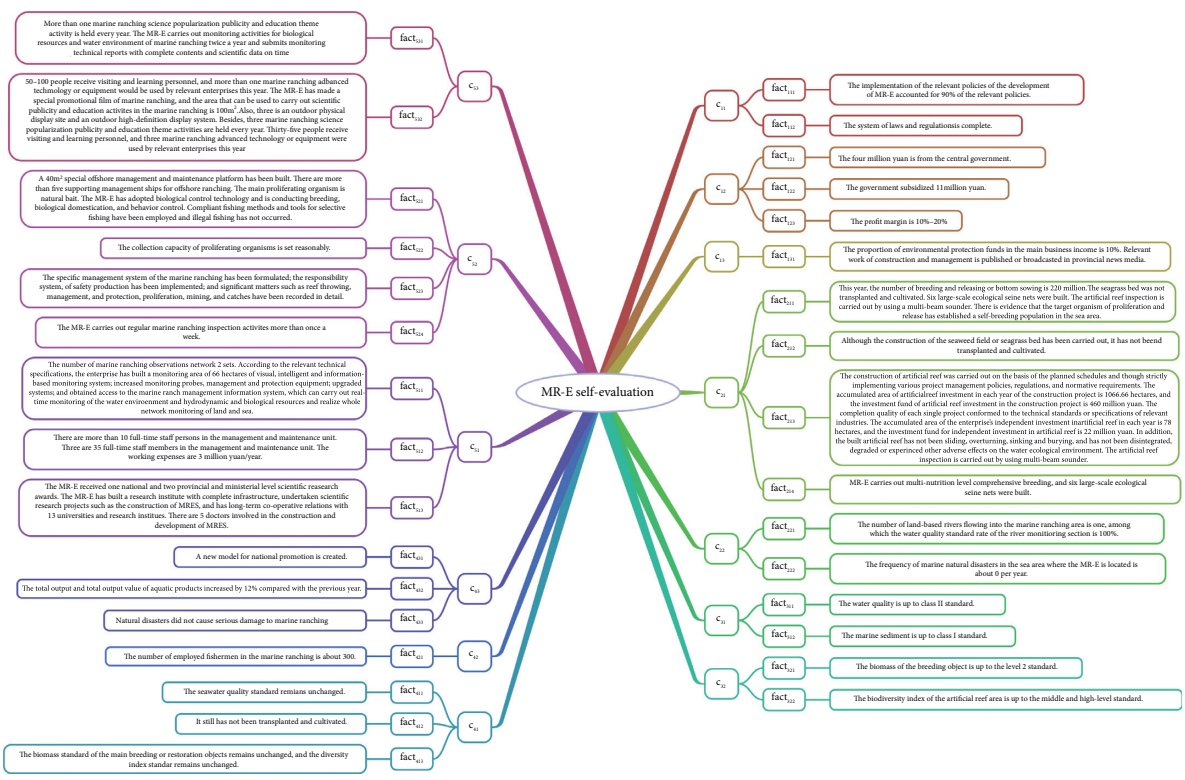



Figure 4: Self-evaluation by the MR-E.

survey, data collection was mainly accomplished through field visits, interviews, and questionnaires during September and October in 2019.

The MR-E is evaluated by five grade levels, i.e., Excellent ($v_1$), Good ($v_2$), Average ($v_3$), Poor ($v_4$), and Very Poor ($v_5$), so the set of grade levels is defined as $V = \{v_1, v_2, v_3, v_4, v_5\}$. We invited five experts $e_1, \ldots, e_5$ who are engaged in marine-ranching-related research at the Ocean University of China to participate in the MRES evaluation. Considering the differences in professional competence and other aspects of the five experts, the expert weight set is $W = (0.1, 0.25, 0.15, 0.2, 0.3)$.

*4.2. Self-Evaluation and Expert Evaluation.* The self-evaluation information plays an important role in the MRES evaluation of the MR-E. We invited the manager of SBOT to make MR-E self-evaluation. Following the same evaluation process mentioned earlier, $\mathrm{fact}_{ijK_{ij}}$ is the performance of $\mathrm{c}_{ijK_{ij}}$, and the specifics of the MR-E are as follows.

The performance of marine ranching on $\mathrm{c}_{ij}$ is shown in Figure 4.

Self-evaluation information of the MR-E that the MR-E manager provides according to the actual development status is as shown in Table 1. According to the self-evaluation information, the five experts invited in this paper provided the evaluation information $M_{ij}^n$ based on the fact $X_{ij}$ of the MR-E, as shown in Columns 4–8 in Table 1.

*4.3. Factor Weight Calculation*

(1) Confirm the direct-influence relationship between the five factors. DPSIR provides a possible influence relationship between its five factors, which can be regarded as $(\mathrm{C}_1, \mathrm{C}_5) \longrightarrow \mathrm{C}_2$, $(\mathrm{C}_2, \mathrm{C}_5) \longrightarrow \mathrm{C}_3$, $(\mathrm{C}_3, \mathrm{C}_5) \longrightarrow (\mathrm{C}_4)$, $\mathrm{C}_4 \longrightarrow \mathrm{C}_5$, $\mathrm{C}_5 \longrightarrow \mathrm{C}_1$. For example, as mentioned in Section 2, Driving ($\mathrm{C}_1$) may act directly on the environment and cause Pressure ($\mathrm{C}_2$) from changes in the marine environment, and Pressure ($\mathrm{C}_2$) can cause changes in the natural system and environmental State ($\mathrm{C}_3$). In order to avoid subjectivity, this paper follows the causality chain defined by DPSIR, and the DPSIR impossible influence mechanism is summarized in Figure 5.

(2) Establish the IDR matrix. The direct-influence matrix can be presented based on the causal relationship provided in the DPSIR framework. The results are shown in

$$Z = [z_{ii'}]_{5\times5} = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 1 & 1 & 1 & 1 & 0 \end{bmatrix}. \quad (4)$$

(3) Normalize the IDR matrix. Follow the steps mentioned above to obtain the normalized direct-influence matrix $Z'$ as expressed in equation (5), which shows the normalized direct-influence with their influence degrees of the five factors on each other. For example, we see that Pressure has an initial direct-influence on the State with an influence degree of 0.25, but it has no initial direct-influence on the remaining factors.

$$Z' = [z'_{ii'}]_{5\times5} = \begin{bmatrix} 0 & 0.25 & 0 & 0 & 0 \\ 0 & 0 & 0.25 & 0 & 0 \\ 0 & 0 & 0 & 0.25 & 0 \\ 0 & 0 & 0 & 0 & 0.25 \\ 0.25 & 0.25 & 0.25 & 0.25 & 0 \end{bmatrix}. \quad (5)$$

(4) Compute the total relation matrix. By a series of calculations, the total relations with their influence degrees between the five factors can be obtained. Equation (6) shows the total relations with their influence degrees between the five factors. For example, we see that Pressure has total relations with the Driver, Pressure, State, Impact, and Response, with influence degrees of 0.004, 0.005, 0.256, 0.068, and 0.017, respectively. It is worth noting that the influence degrees between equations (5) and (6) are diffident, because equation (5) only reflects the influence degree of direct relation, whereas equation (6) reflects both the direct and the indirect relations. Taking $\mathrm{C}_1 \longrightarrow \mathrm{C}_3$ for example, there is no direct-influence, but there is an indirect relation $\mathrm{C}_1 \longrightarrow \mathrm{C}_2 \longrightarrow \mathrm{C}_3$, which caused the total relation to be 0.064.

$$T = \begin{bmatrix} 0.001 & 0.251 & 0.064 & 0.017 & 0.004 \\ 0.004 & 0.005 & 0.256 & 0.068 & 0.017 \\ 0.017 & 0.021 & 0.022 & 0.273 & 0.068 \\ 0.068 & 0.085 & 0.089 & 0.091 & 0.273 \\ 0.273 & 0.341 & 0.358 & 0.362 & 0.091 \end{bmatrix}. \quad (6)$$

(5) Determine the weights of factors. Determine the weight of each factor according to the previous steps. Specifically, calculate $f_i$ and $f'_i$ of factor $\mathrm{C}_i$ and subsequently obtain the prominence $g_i$ and $\lambda_i$ of factor $\mathrm{C}_i$ (Table 2).

Then, the factor weight $\lambda = (\lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5) = (0.1123, 0.1690, 0.1909, 0.2270, 0.3008)$ is calculated by normalization of centrality. Thus, the weights of factors included in the five subsystems are ranked in the order: Response (0.3008) >Impact (0.2270) >State (0.1909) >Pressure (0.1690) >Driver (0.1123).

*4.4. MRES Evaluation*

(1) Construct the fuzzy membership matrix. Based on the quantitative self-evaluation information, five authoritative experts comprehensively evaluate factor $C_i$. The evaluation results are shown in columns 4–8 of Table 1. Then, the integrated fuzzy membership degree matrix $M$ can be obtained by

Table 1: MR-E self-evaluation and experts' evaluation information form.

| Factor on level 1 | Element on level 2 | Fact | $e_1$ | $e_2$ | $e_3$ | $e_4$ | $e_5$ |
|---|---|---|---|---|---|---|---|
| *Driver* $C_1$ | $C_{11}$ | $X_{11} = \{fact_{111}, fact_{112}\}$ | Excellent | Good | Average | Good | Excellent |
| | $C_{12}$ | $X_{12} = \{fact_{121}, fact_{122}, fact_{123}\}$ | Excellent | Excellent | Good | Good | Excellent |
| | $C_{13}$ | $X_{13} = \{fact_{131}\}$ | Good | Good | Average | Good | Good |
| *Pressure* $C_2$ | $C_{21}$ | $X_{21} = \{fact_{211}, fact_{212}, fact_{213}, fact_{214}\}$ | Good | Excellent | Excellent | Excellent | Excellent |
| | $C_{22}$ | $X_{22} = \{fact_{221}, fact_{222}\}$ | Good | Excellent | Excellent | Excellent | Excellent |
| *State* $C_3$ | $C_{31}$ | $X_{31} = \{fact_{311}, fact_{312}\}$ | Excellent | Good | Excellent | Excellent | Good |
| | $C_{32}$ | $X_{32} = \{fact_{321}, fact_{322}\}$ | Excellent | Excellent | Excellent | Excellent | Good |
| *Impact* $C_4$ | $C_{41}$ | $X_{41} = \{fact_{411}, fact_{412}, fact_{413}\}$ | Average | Excellent | Excellent | Good | Good |
| | $C_{42}$ | $X_{42} = \{fact_{421}\}$ | Excellent | Excellent | Excellent | Excellent | Excellent |
| | $C_{43}$ | $X_{43} = \{fact_{431}, fact_{432}, fact_{433}\}$ | Average | Excellent | Excellent | Good | Good |
| *Response* $C_5$ | $C_{51}$ | $X_{51} = \{fact_{511}, fact_{512}, fact_{513}\}$ | Excellent | Excellent | Good | Excellent | Excellent |
| | $C_{52}$ | $X_{52} = \{fact_{521}, fact_{522}, fact_{523}, fact_{524}\}$ | Excellent | Excellent | Good | Excellent | Excellent |
| | $C_{53}$ | $X_{53} = \{fact_{531}, fact_{532}\}$ | Excellent | Excellent | Excellent | Excellent | Excellent |

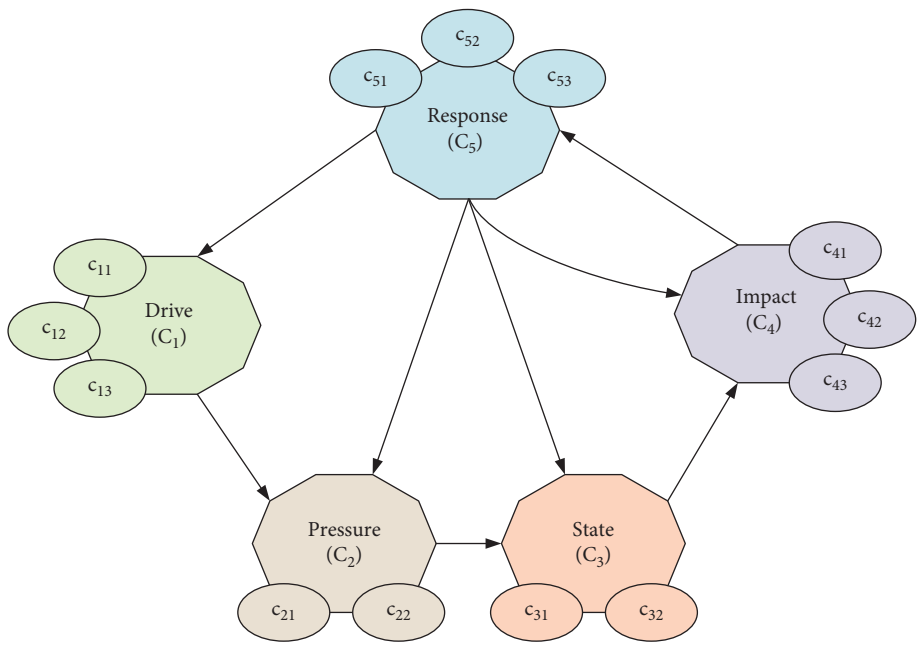



Figure 5: The DPSIR impossible influence mechanism.

Table 2: Results of DEMATEL.

| | Driver ($C_1$) | Pressure ($C_2$) | State ($C_3$) | Impact ($C_4$) | Response ($C_5$) |
|---|---|---|---|---|---|
| $f_i$ | 0.3376 | 0.3504 | 0.4015 | 0.6060 | 1.4239 |
| $f_i$ | 0.3632 | 0.7039 | 0.7891 | 0.8104 | 0.4526 |
| $g_i$ | 0.7007 | 1.0543 | 1.1906 | 1.4164 | 1.8765 |
| $\lambda_i$ | 0.1123 | 0.1690 | 0.1908 | 0.2270 | 0.3008 |

$$M = \begin{bmatrix} 0.35 & 0.55 & 0.1 & 0 & 0 \\ 0.9 & 0.1 & 0 & 0 & 0 \\ 0.33 & 0.67 & 0 & 0 & 0 \\ 0.6 & 0.33 & 0.07 & 0 & 0 \\ 0.9 & 0.1 & 0 & 0 & 0 \end{bmatrix}. \quad (7)$$

(2) Calculate the comprehensive evaluation value. The fuzzy operator $\circ$ is selected to integrate the weight vector $\lambda$ of the five factors with the membership matrix $M$, and finally, the results of FCE are obtained as $H = M \circ \lambda = (0.6613, 0.3116, 0.0271, 0.0000, 0.0000)$. It can be seen that the probability of the MR-E being rated as Excellent, Good, Average, Poor, and Very poor are 66.13%, 31.16%, 2.71%, 0.0%, and 0.0%, respectively.

(3) Make comprehensive evaluation. According to the principle of maximum membership degree, the evaluation result of the MRES is Excellent.

In summary, the DEMATEL method was used to determine the factor weight, and FCE was used to determine the comprehensive evaluation result from the factor weight and the evaluation value of the factor.

*4.5. Discussion.* Based on the case study, we can state that this method has the following three characteristics:

(1) The index system has a theoretical basis. The MRES index system is constructed according to the DPSIR framework (see Figure 1). Because it can reflect the concrete performance of marine ranching in five dimensions (including Driver, Pressure, State, Impact, and Response), it can systematically reflect the MRES level from the perspective of sustainable development. In the actual evaluation, the index system can be directly utilized for evaluation.

(2) The division of labor between the MR-E and the experts is clear. MR-E needs to objectively give its performance on the lowest indices in the MRES index system according to its actual situation (see Figure 4). Experts evaluate the overall performance of the MR-E on high-level factors according to the results of the self-evaluation of the MR-E (see Table 1). The MR-E and the experts can combine their advantages with a clear division of labor. The former can provide accurate decision-making information, and the latter can make scientific evaluation according to the decision-making information.

(3) The evaluation process is simple enough. Self-evaluation of the MR-E can be completed by filling out the self-evaluation report, and the content of the self-evaluation should reflect according to the index system in the self-evaluation report. Experts only need to evaluate the overall performance of the MR-E according to the self-evaluation results (see Table 1) without considering the complex mechanism of action in the MRES system. The mechanism is reflected by combining DPSIR, DEMATEL, and FCE (see Figure 4 and formulas (4)–(7)). Thus, a scientific evaluation of MRES is realized.

## 5. Conclusions

Marine ranching can enable marine food supply and effectively protect the marine ecosystem. This is one of the instruments of implementing a green development strategy. Moreover, MRES is related to national security, sea area safety, and life safety. However, few MRES evaluations have been included in the existing ecological security studies. For this reason, the present study took marine ranching as the research object, adopted three methods (DPSIR, DEMATEL, and FCE) to evaluate its ecological security, and provided an illustrative case to demonstrate the applicability of the proposed methods.

The main contributions of this study can be summarized in three aspects. First of all, the present study constructed an MRES evaluation index system. Based on field visits, interviews, questionnaires, and other ways, and by integrating social, economic, ecological, and other systems, the evaluation index system of the present study was established through the DPSIR method, which fills a research gap and provides a reference for researchers and practitioners. Secondly, this study proposed an evaluation method for MRES. The advantage of the integration of DEMATEL and FCE is that it avoids the influence of different factors and experts' decision-making ability on the evaluation results, which makes the evaluation results more scientific and rigorous. In addition, through MR-E self-evaluation and expert evaluation, the combination of subjective and objective evaluation makes the evaluation results more credible. Thirdly, the proposed method was applied to a real case. Taking SBOT as an example, the present study elaborated on the evaluation content, enriched the MRES evaluation framework, and demonstrated the operation process of applying the methods in practice. The grade level of the evaluation results set was determined in this study. Furthermore, the evaluation process has a certain guiding integral role in the development of MRES.

It is not difficult to find that Response ($C_5$) plays an important role in MRES through the weights of five factors. MR-E can take a series of positive measures to improve the quality of MRES and avoid the deterioration of ecological environment. The following managerial implications can be obtained. First, laws and regulations promulgated by the government can motivate MR-E to take positive actions—MR-E should be aware of the environmental consequences of their actions and constantly improve the level of scientific management. Second, MR-E should make full use of the technology and knowledge provided by universities and scientific research units to promote the transformation of scientific research achievements. Third, MR-E should enhance their awareness of ecological and environmental protection and create a harmonious and sustainable corporate culture atmosphere through a series of publicity activities and training.

Although the ecological benefit of marine ranching is acknowledged, it is worth noticing that improper management of MRES will bring greater potential environmental threats to the sea area. MRES evaluation is a long-term dynamic process. Therefore, it is necessary to consider the dynamic and forward-looking aspects of MRES evaluation. The establishment of a dynamic evaluation mechanism of MRES based on subjective experiences and objective data is a valuable research direction for future work.

## Data Availability

The 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 research was supported by the Major Program of National Social Science Foundation of China under Grant no. 18ZDA055.

## References


[1] Y. Du, B. Li, and X. Quan, "Construction and application of DPPD model for evaluating marine resources and environment carrying capacity in China," *Journal of Cleaner Production*, vol. 252, Article ID 119655, 2020.

[2] X. Zhou, X. Zhao, S. Zhang, and J. Lin, "Marine ranching construction and management in east China sea: programs for sustainable fishery and aquaculture," *Water*, vol. 11, no. 6, p. 1237, 2019.

[3] S. Kitada, "Economic, ecological and genetic impacts of marine stock enhancement and sea ranching: a systematic review," *Fish and Fisheries*, vol. 19, no. 3, pp. 511–532, 2018.

[4] A. L. Taylor, S. J. Nowland, M. N. Hearnden, C. A. Hair, and A. E. Fleming, "Sea ranching release techniques for cultured sea cucumber Holothuria scabra (Echinodermata: holothuroidea) juveniles within the high-energy marine environments of northern Australia," *Aquaculture*, vol. 465, pp. 109–116, 2016.

[5] E. Moksness and R. Støle, "Larviculture of marine fish for sea ranching purposes: is it profitable?" *Aquaculture*, vol. 155, no. 1–4, pp. 341–353, 1997.

[6] D. Zhang, Y. Cui, H. Zhou et al., "Microplastic pollution in water, sediment, and fish from artificial reefs around the Ma'an Archipelago, Shengsi, China," *Science of The Total Environment*, vol. 703, Article ID 134768, 2020.

[7] W. S. Grant, J. Jasper, D. Bekkevold, and M. Adkison, "Responsible genetic approach to stock restoration, sea ranching and stock enhancement of marine fishes and invertebrates," *Reviews in Fish Biology and Fisheries*, vol. 27, no. 3, pp. 615–649, 2017.

[8] E. Hadas, M. Shpigel, and M. Ilan, "Sea ranching of the marine sponge Negombata magnifica (Demospongiae, Latrunculiidae) as a first step for latrunculin B mass production," *Aquaculture*, vol. 244, no. 1–4, pp. 159–169, 2005.

[9] Y. Liang, Z. Zhao, G. Zhang, S. Wang, A. Wan, and Q. Liu, "Distinguishing nutrient-depleting effects of scallop farming from natural variabilities in an offshore sea ranch," *Aquaculture*, vol. 518, Article ID 734844, 2020.

[10] D.-H. Kim, C. K. Seung, and Y.-I. Seo, "Multi-regional economic impacts of recreational fisheries: analysis of small sea ranch in gyeong-nam Province, korea," *Marine Policy*, vol. 84, pp. 90–98, 2017.

[11] S. I. Lee and C. I. Zhang, "Evaluation of the effect of marine ranching activities on the tongyeong marine ecosystem," *Ocean Science Journal*, vol. 53, no. 3, pp. 557–582, 2018.

[12] Z.-T. Li, M.-J. Yuan, M.-M. Hu, Y.-F. Wang, and B.-C. Xia, "Evaluation of ecological security and influencing factors analysis based on robustness analysis and the BP-DEMALTE model: a case study of the Pearl River Delta urban agglomeration," *Ecological Indicators*, vol. 101, pp. 595–602, 2019.

[13] H. Hao, C. Bin, M. Zhiyuan et al., "Assessing the ecological security of the estuary in view of the ecological services- A case study of the Xiamen Estuary," *Ocean & Coastal Management*, vol. 137, pp. 12–23, 2017.

[14] D. Liu and Q. Chang, "Ecological security research progress in China," *Acta Ecologica Sinica*, vol. 35, no. 5, pp. 111–121, 2015.

[15] J. Tian and G. Gang, "Research on regional ecological security assessment," *Energy Procedia*, vol. 16, pp. 1180–1186, 2012.

[16] J.-Z. Gong, Y.-S. Liu, B.-C. Xia, and G.-W. Zhao, "Urban ecological security assessment and forecasting, based on a cellular automata model: a case study of Guangzhou, China," *Ecological Modelling*, vol. 220, no. 24, pp. 3612–3620, 2009.

[17] M. Elliott, S. J. Boyes, S. Barnard, and Á. Borja, "Using best expert judgement to harmonise marine environmental status assessment and maritime spatial planning," *Marine Pollution Bulletin*, vol. 133, pp. 367–377, 2018.

[18] Y. Wang and J. Pan, "Building ecological security patterns based on ecosystem services value reconstruction in an arid inland basin: a case study in Ganzhou District, NW China," *Journal of Cleaner Production*, vol. 241, Article ID 118337, 2019.

[19] Y. Hua, M. Yan, and D. Limin, "Land ecological security assessment for Bai autonomous prefecture of dali based using PSR model--with data in 2009 as case," *Energy Procedia*, vol. 5, pp. 2172–2177, 2011.

[20] Y. Chen and J. Wang, "Ecological security early-warning in central Yunnan Province, China, based on the gray model," *Ecological Indicators*, vol. 111, Article ID 106000, 2020.

[21] S. Li, W. Xiao, Y. Zhao, and X. Lv, "Incorporating ecological risk index in the multi-process MCRE model to optimize the ecological security pattern in a semi-arid area with intensive coal mining: a case study in northern China," *Journal of Cleaner Production*, vol. 247, Article ID 119143, 2020.

[22] L. Ma, J. Bo, X. Li, F. Fang, and W. Cheng, "Identifying key landscape pattern indices influencing the ecological security of inland river basin: the middle and lower reaches of Shule river basin as an example," *Science of the Total Environment*, vol. 674, pp. 424–438, 2019.

[23] X. Bai and J. Tang, "Ecological security assessment of tianjin by PSR model," *Procedia Environmental Sciences*, vol. 2, pp. 881–887, 2010.

[24] Z. Mu, S. Zeng, and P. Wang, "Novel approach to multi-attribute group decision-making based on interval-valued pythagorean fuzzy power Maclaurin symmetric mean operator," *Computers & Industrial Engineering*, Article ID 107049, 2020.

[25] J. Wang, S. Zeng, and C. Zhang, "Single-Valued neutrosophic linguistic logarithmic weighted distance measures and their application to supplier selection of fresh aquatic products," *Mathematics*, vol. 8, no. 3, p. 439, 2020.

[26] X. J. Gou, Z. S. Xu, W. Zhou, and E. Herrera-Viedma, "The risk assessment of construction project investment based on prospect theory with linguistic preference orderings," *Economic Research-Ekonomska Istraživanja*, pp. 1–23, 2020.

[27] X. Gou, Z. Xu, and W. Zhou, "Managing consensus by multi-stage optimization models with linguistic preference orderings and double hierarchy linguistic preferences," *Technological and Economic Development of Economy*, vol. 26, no. 3, pp. 642–674, 2020.

[28] X. J. Gou, H. C. Liao, Z. S. Xu, and F. Herrera, "Probabilistic double hierarchy linguistic term set and its use for designing a VIKOR method: The application in smart healthcare," *Journal of the Operational Research Society*, pp. 1–20, 2020.

[29] P. Li, J. Liu, Y. Yang, and C. Wei, "Evaluation of poverty-stricken families in rural areas using a novel case-based reasoning method for probabilistic linguistic term sets," *Computers & Industrial Engineering*, vol. 147, Article ID 106658, 2020.

[30] W. Ruan, Y. Li, S. Zhang, and C.-H. Liu, "Evaluation and drive mechanism of tourism ecological security based on the DPSIR-DEA model," *Tourism Management*, vol. 75, pp. 609–625, 2019.

[31] B. Han, H. Liu, and R. Wang, "Urban ecological security assessment for cities in the Beijing-Tianjin-Hebei metropolitan region based on fuzzy and entropy methods," *Ecological Modelling*, vol. 318, pp. 217–225, 2015.

[32] X. Jiang, "Urban ecological security evaluation and analysis based on fuzzy mathematics," *Procedia Engineering*, vol. 15, pp. 4451–4455, 2011.

[33] N. V. Solovjova, "Ecological risk modelling in developing resources of ecosystems characterized by varying vulnerability levels," *Ecological Modelling*, vol. 406, pp. 60–72, 2019.

[34] P. Li and C. Wei, "An emergency decision-making method based on D-S evidence theory for probabilistic linguistic term sets," *International Journal of Disaster Risk Reduction*, vol. 37, Article ID 101178, 2019.

[35] S. Zeng, Y. Hu, T. Balezentis, and D. Streimikiene, "A multi-criteria sustainable supplier selection framework based on neutrosophic fuzzy data and entropy weighting," *Sustainable Development*, vol. 28, pp. 1431–1440, 2020.

[36] Y. Du and X. Sun, "Influence paths of marine ranching ecological security in China based on probabilistic linguistic term sets and qualitative comparative analysis," *International Journal of Fuzzy Systems*, 2020.

[37] X. C. Zhang, C. Ma, S. F. Zhan, and W. P. Chen, "Evaluation and simulation for ecological risk based on emergy analysis and pressure-state-response model in a coastal city, China," *Procedia Environmental Sciences*, vol. 13, pp. 221–231, 2012.

[38] I. Mukuvari, S. K. Mafwila, and L. Chimuka, "Measuring the recovery of the northern benguela current large marine ecosystem (BCLME): an application of the DPSIR framework," *Ocean & Coastal Management*, vol. 119, pp. 227–233, 2016.

[39] R. L. Lewison, M. A. Rudd, W. Al-Hayek et al., "How the DPSIR framework can be used for structuring problems and facilitating empirical research in coastal systems," *Environmental Science & Policy*, vol. 56, pp. 110–119, 2016.

[40] Y. Du and X. Li, "Hierarchical DEMATEL method for complex systems," *Expert Systems with Applications*, vol. 167, Article ID 113871, 2020.

[41] P. K. Singh and P. Sarkar, "A framework based on fuzzy Delphi and DEMATEL for sustainable product development: a case of Indian automotive industry," *Journal of Cleaner Production*, vol. 246, Article ID 118991, 2020.

[42] M. Dalvi-Esfahani, A. Niknafs, D. J. Kuss, M. Nilashi, and S. Afrough, "Social media addiction: applying the DEMATEL approach," *Telematics and Informatics*, vol. 43, Article ID 101250, 2019.

[43] Y.-W. Du and W. Zhou, "New improved DEMATEL method based on both subjective experience and objective data," *Engineering Applications of Artificial Intelligence*, vol. 83, pp. 57–71, 2019.

[44] J. Cheng and J.-p. Tao, "Fuzzy comprehensive evaluation of drought vulnerability based on the analytic hierarchy process," *Agriculture and Agricultural Science Procedia*, vol. 1, pp. 126–135, 2010.

[45] Z.-H. Zou, Y. Yun, and J.-N. Sun, "Entropy method for determination of weight of evaluating indicators in fuzzy synthetic evaluation for water quality assessment," *Journal of Environmental Sciences*, vol. 18, no. 5, pp. 1020–1023, 2006.

[46] W. Li, W. Liang, L. Zhang, and Q. Tang, "Performance assessment system of health, safety and environment based on experts' weights and fuzzy comprehensive evaluation," *Journal of Loss Prevention in the Process Industries*, vol. 35, pp. 95–103, 2015.

[47] G. Zheng, K. Li, W. Bu, and Y. Wang, "Fuzzy comprehensive evaluation of human physiological state in indoor high temperature environments," *Building and Environment*, vol. 150, pp. 108–118, 2019.

[48] R. C. de Sousa-Felix, L. C. C. Pereira, W. N. Trindade, I. P. de Souza, R. M. Da Costa, and J. A. Jimenez, "Application of the DPSIR framework to the evaluation of the recreational and environmental conditions on estuarine beaches of the Amazon coast," *Ocean & Coastal Management*, vol. 149, pp. 96–106, 2017.

[49] M. Elliott, D. Burdon, J. P. Atkins et al., "And DPSIR begat DAPSI (W) R (M)!-a unifying framework for marine environmental management," *Marine Pollution Bulletin*, vol. 118, no. 1-2, pp. 27–40, 2017.

[50] X. Li, H. Li, D. Liu et al., "Connotation analysis and evaluation index system construction of regional agricultural soil and water resource composite system harmony," *Journal of Cleaner Production*, vol. 263, Article ID 121438, 2020.

[51] W. Lu, C. Xu, J. Wu, and S. Cheng, "Ecological effect assessment based on the DPSIR model of a polluted urban river during restoration: a case study of the Nanfei River, China," *Ecological Indicators*, vol. 96, pp. 146–152, 2019.

[52] P. Ma, G. Ye, X. Peng, J. Liu, J. Qi, and S. Jia, "Development of an index system for evaluation of ecological carrying capacity of marine ecosystems," *Ocean & Coastal Management*, vol. 144, pp. 23–30, 2017.

[53] Y. Du and J. Zhong, "Generalized combination rule for evidential reasoning approach and Dempster-shafer theory of evidence," *Information Sciences*, vol. 547, pp. 1201–1232, 2021.

[54] C. Bai, J. Sarkis, and Y. Dou, "Constructing a process model for low-carbon supply chain cooperation practices based on the DEMATEL and the NK model," *Supply Chain Management: An International Journal*, vol. 22, no. 3, 2017.

[55] J. Kaur, R. Sidhu, A. Awasthi, S. Chauhan, and S. Goyal, "A DEMATEL based approach for investigating barriers in green supply chain management in Canadian manufacturing firms," *International Journal of Production Research*, vol. 56, pp. 1-2, 2018.

Hindawi
Mathematical Problems in Engineering
Volume 2021, Article ID 6670196, 6 pages
https://doi.org/10.1155/2021/6670196

*Research Article*

# A Fuzzy-Decomposition Grey Modeling Procedure for Management Decision Analysis

**Jianhong Guo,[1,2] Che-Jung Chang,[1,2] Yingyi Huang,[1,2] and Kun-Peng Yu[1,2]**

[1]*TSL Business School, Quanzhou Normal University, No. 398, Donghai Street, Quanzhou, Fujian 362000, China*
[2]*Fujian University Engineering Research Center of Cloud Computing, Internet of Things and E-Commerce Intelligence, No. 398, Donghai Street, Quanzhou, Fujian 362000, China*

Correspondence should be addressed to Che-Jung Chang; r3795102@nckualumni.org.tw

Received 5 December 2020; Revised 19 January 2021; Accepted 27 January 2021; Published 10 February 2021

Academic Editor: Tomas Balezentis



To cope with the increasingly fierce market competition environment, enterprises need to quickly respond to business issues and maintain business advantages, which require timely and correct decisions. In this context, the general mathematical modeling method may cause overfitting phenomenon when using small data sets, so it is difficult to ensure good analysis performance. Therefore, it is significant for enterprises to use limited samples to analyze and forecast. Over the past few decades, the grey model and its extensions have been shown to be effective tools for processing small data sets. To further enforce the effectiveness of data uncertainty processing, a fuzzy-decomposition modeling procedure for grey models is developed. Specifically, Latent Information (LI) function is employed to decompose the initial series into three subseries; next, the three subseries are used to build three grey models and create the estimated values of the three subseries; finally, the weighted average method is applying to combine the estimated values of the three subseries into a single final predicted value. After the actual test on the monthly demand data of the thin-film transistor liquid crystal display panels, the proposed fuzzy-decomposition modeling procedure can result in good prediction outcomes and is thus an appropriate decision analysis tool for managers.

## 1. Introduction

Decision-making is the heart of administration and one of the most important tasks for managers [1, 2]. To cope with the increasingly fierce market competition environment, managers need to quickly respond to business problems and maintain business advantages, which require timely and correct decisions. However, uncertain events and uncontrollable factors often make decisions invalid and affect business performance. Forecasting analysis can help managers grasp the future development trends to reduce the impact of uncertainty [3] and then make meaningful decisions.

In various management situations, decisions that require an immediate response are the more difficult tasks for managers. Real-time grasp of the situation through a limited amount of data can enable managers to make appropriate treatments and obtain effective control and management [4]. Analyzing the occurrence of a new disease is another example. If the government can make a correct decision sooner, the possible harm and impact of the new disease on the country will be reduced. Decision-making should not wait until more people with infections appear; the more immediate decision-making can bring higher management value. Therefore, using a smaller amount of data to build a prediction model has more practical value.

Machine learning algorithms and statistical theory have been widely used in knowledge extraction for management, but these approaches are generally developed based on a large number of samples [5, 6]. These typical approaches may not provide satisfactory predictions when the sample size is small [7]. Multivariate analysis is another popular prediction method; its prediction accuracy depends on the choice of independent variables. If the established causal model cannot effectively explain the variation of the dependent variable, a prediction model with high variance will

be produced. In contrast, time series models only require historical data to predict their future trends [8]. However, it usually requires a large number of observations to get accurate prediction results. In all the above methods, the key factor that affects the prediction performance is the sample size, which limits their applicability in certain prediction situations. Therefore, how to use limited data to extract useful information to assist managers to make immediate decisions has great management significance and practical value [9].

Grey system theory was proposed by Deng [10], which mainly studies the uncertainty of insufficient information caused by limited samples. It provides strong technical support for solving small-data-set analysis and decision-making problems [11]. The typical first-order one-variable grey model, abbreviated as the typical GM (1, 1), is one of the most commonly used grey methods due to its convenience and easy calculation [12], which can use only four data points to predict the future trend of a time series with a favorable result [13]. The model and its extensions have succeeded applied to the fields of business [14], energy [15–18], environment [19–21], industry [22, 23], engineering [24], medicine [25], hydraulics [26], economics [27–30], and many other domains.

To further improve it, this study proposed a fuzzy-decomposition modeling procedure for enhancing the prediction quality of typical GM (1, 1). This new method can build a robust model based on Latent Information (LI) function [31] and bring a better forecasting accuracy than the typical GM (1, 1).

To confirm the validity of the proposed fuzzy-decomposition modeling procedure, one real case was selected to implement the experimental analysis for confirming the effectiveness and practical value investigating; the data is the total demand of thin-film transistor liquid crystal display (TFT-LCD) panels provided by a leading manufacturer in Taiwan. Experimental results showed that the fuzzy-decomposition modeling procedure can significantly improve the accuracy of the typical GM (1, 1) and is a useful decision analysis tool for managers.

The remaining parts of this paper are systematically presented as follows. Section 2 introduces the typical GM (1, 1), LI function, and the proposed modeling procedure. Section 3 addresses the data analysis and comparison which is applied to one real case. Finally, the conclusions are presented in Section 4.

## 2. Methodology

When the sample size is small, the main challenge is how to effectively extract useful information for modeling. Therefore, a fuzzy-decomposition modeling procedure was developed for this issue. The following subsections will introduce the two main components of the modeling procedure and the detailed steps of the proposed method.

*2.1. The Typical GM (1, 1).* Among the existing grey approaches, the typical GM (1, 1) is the simplest and widely used model. It can use four data to build a model and bring favorable performance. Its primary means is to use the generating operation to weaken the randomness of the data and then recognize the inherent regular pattern of the data to establish a fitting model [32]. At present, it is one of the important methods to solve the problem of small-sample analysis. The modeling process of typical GM (1, 1) is as follows:

(1) Suppose that the original nonnegative data series is $X^{(0)} = \{x^{(0)}(1), x^{(0)}(2), \ldots, x^{(0)}(n)\}$, where $n$ is the sample size, and $n \geq 4$. Let $x^{(0)}(k)$ represents the datum at the $k$th phase.

(2) Form an accumulated generating series $X^{(1)} = \{x^{(1)}(1), x^{(1)}(2), \ldots, x^{(1)}(n)\}$:

$$x^{(1)}(k) = \sum_{i=1}^{k} x^{(0)}(i), \quad k = 1, 2, \ldots, n. \tag{1}$$

(3) Calculate the background values $Z^{(1)} = \{z^{(1)}(1), z^{(1)}(2), \ldots, z^{(1)}(n)\}$:

$$z^{(1)}(k) = \frac{1}{2}\left[x^{(1)}(k-1) + x^{(1)}(k)\right], \quad k = 2, 3, \ldots, n. \tag{2}$$

(4) Formulate the grey differential equation:

$$x^{(0)}(k) + a z^{(1)}(k) = b. \tag{3}$$

(5) Expand equation (3) as a vector-matrix form such as equation (4); let $\mathbf{Y} = [x^{(0)}(2), x^{(0)}(3), \ldots, x^{(0)}(n)]^T$, $\hat{a} = \begin{bmatrix} a \\ b \end{bmatrix}$, and $\mathbf{B} = \begin{bmatrix} -z^{(1)}(2) & 1 \\ -z^{(1)}(3) & 1 \\ \vdots & \vdots \\ -z^{(1)}(n) & 1 \end{bmatrix}$.

$$\begin{bmatrix} x^{(0)}(2) \\ x^{(0)}(3) \\ \vdots \\ x^{(0)}(n) \end{bmatrix} = \begin{bmatrix} -z^{(1)}(2) & 1 \\ -z^{(1)}(3) & 1 \\ \vdots & \vdots \\ -z^{(1)}(n) & 1 \end{bmatrix} \times \begin{bmatrix} a \\ b \end{bmatrix}. \tag{4}$$

(5) Determine the pending coefficient vector $\hat{a}$:

$$\hat{a} = \left(\mathbf{B}^T\mathbf{B}\right)^{-1}\mathbf{B}^T\mathbf{Y}. \tag{5}$$

(6) Solve the ordinary differential equation $\mathrm{d}x^{(1)}/\mathrm{d}t + a x^{(1)} = b$ with the initial condition $x^{(0)}(1) = x^{(1)}(1)$ to build the grey forecasting model:

$$\begin{cases} \hat{x}^{(1)}(k+1) = \left[x^{(0)}(1) - \dfrac{b}{a}\right]e^{-ak} + \dfrac{b}{a}, \\ \hat{x}^{(0)}(k+1) = \hat{x}^{(1)}(k+1) - \hat{x}^{(1)}(k). \end{cases} \tag{6}$$

(7) Obtain the desired forecasting output with equation (6).

*2.2. LI Function.* When the sample size is small, the training process of a prediction model usually cannot be implemented effectively. This problem could be solved by increasing the amount of information used, while also enhancing the stability of the forecasting results. Therefore, in this study, we selected the LI function [31] to analyze the data behavior and then combined the obtained fuzzy data from the analysis with the typical GM (1, 1) to develop a robust modeling procedure.

The LI function was proposed by Chang et al. [31]; its development concept is to appropriately expand the possible margins of data by using four statistical indexes to fill the information gap. The degree of expansion of the data range in the LI function depends on the sample size. Specifically, when there are a large number of observations available, the overall outline of the data is clearer due to more information, so it is necessary to substantially expand the data range. On the contrary, if there is no adequate information, the data range must be greatly expanded. Therefore, in the LI function, the degree of expansion of the data range is calculated by the division operation between the range and the number of samples, and the ratio of expansion of the data range toward the left or right is determined by the skewness. The upper bound (UB) and the lower bound (LB) are the expanded boundaries, which are combined with the central tendency (CT) to form the LI function. The LI function is essentially a fuzzy membership function, whose value falls between 0 and 1, and the magnitude of the value represents the possibility of potential data occurrence. The complete process for forming the LI function is as follows:

(1) Suppose that the time series obtained is $X = \{x_1, x_2, \ldots, x_n\}$; in the set $X$, the element with the minimal value is $x_{\min}$, and the element with the maximal value is $x_{\max}$.

(2) Calculate the rang $R$:

$$\mathrm{R} = x_{\max} - x_{\min}. \tag{7}$$

(3) Determine the CT:

$$\mathrm{CT} = \frac{\sum_{i=1}^{n} i x_i}{\sum_{i=1}^{n} i}, \quad i = 1, 2, \ldots, n. \tag{8}$$

(4) Calculate the central location (CL) of the existing data:

$$\mathrm{CL} = \frac{x_{\min} + x_{\max}}{2}. \tag{9}$$

(5) Count the number of elements in the subset composed of data with values larger than CL, denoted by $N^+$; count the number of elements in the subset composed of data with values less than CL, denoted by $N^-$.

(6) Compute the increasing tendency (IT) and the decreasing tendency (DT):

$$\begin{cases} \mathrm{IT} = \dfrac{N^+}{N^+ + N^-}, \\ \mathrm{DT} = \dfrac{N^-}{N^+ + N^-}. \end{cases} \tag{10}$$

(7) Expand the domain range asymmetrically using IT and DT and then use equation (11) to obtain the extended UB and LB.

$$\begin{cases} \mathrm{UB} = x_{\max} + \mathrm{IT} \times \dfrac{R}{n}, \\ \mathrm{LB} = x_{\min} - \mathrm{DT} \times \dfrac{R}{n}. \end{cases} \tag{11}$$

(8) Combine CT, UB, and LB to generate a triangular LI function (Figure 1 is one of the possible shapes). Here, the LI value of CT is set to 1, and the LI value of the boundaries (UB and LB) is set to 0.

(9) Obtain the LI values of existing data using equation (12) (the properties of the generated triangle).

$$\text{LI value} = \begin{cases} \dfrac{x - \mathrm{LB}}{\mathrm{CT} - \mathrm{LB}}, & \text{if } x < \mathrm{CT}, \\ 1, & \text{if } x = \mathrm{CT}, \\ \dfrac{\mathrm{UB} - x}{\mathrm{UB} - \mathrm{CT}}, & \text{if } x > \mathrm{CT}. \end{cases} \tag{12}$$

*2.3. Fuzzy-Decomposition Modeling Procedure.* A time series usually contains many elements; decomposition can identify these elements and separate the time series data into basic components. The decomposed components are relatively easy to grasp, which will be helpful for model fitting. The concrete realization of this idea in the paper is to decompose the time series into three subseries through the LI function (data fuzzification); they are the upper series, central series, and lower series. These three subseries are, respectively, predicted using the grey model. Finally, simply set the weights of the three series according to the possibility to combine the estimated values of the three series into a single final predicted value (defuzzification). The computational process is detailed as follows:

(1) Denote $X^{(0)} = \{x^{(0)}(1), x^{(0)}(2), \ldots, x^{(0)}(n)\}$ as the initial series, where $n$ is the sample size, and $n \geq 4$. Let $x^{(0)}(k)$ represents the datum at the $k$th phase.

(2) Decompose the initial series into three subseries using LI function, referring to Section 2.2; these three series are $X_{\mathrm{UB}}^{(0)}$, $X_{\mathrm{CT}}^{(0)}$, and $X_{\mathrm{LB}}^{(0)}$.

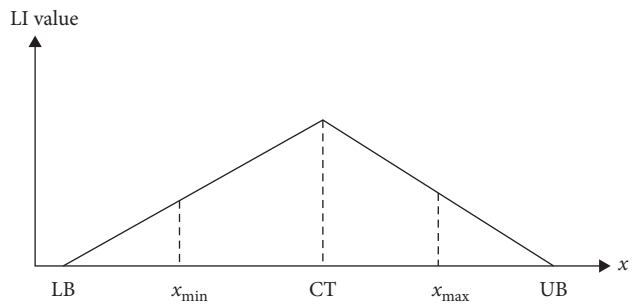



Figure 1: Shape of an asymmetric triangular LI function.

(3) Build three forecasting models, $GM_{UB}$, $GM_{CT}$, and $GM_{LB}$, based on $X_{UB}^{(0)}$, $X_{CT}^{(0)}$, and $X_{LB}^{(0)}$ using GM (1, 1), referring to Section 2.1.

(4) Employ the rank reciprocal weighting method to determine the weights of these three models. The probability of occurrence of $GM_{CT}$ is higher and the ranking is set to 1; the probability of occurrence of $GM_{UB}$ and $GM_{LB}$ is lower and it is impossible to distinguish which one is higher, so both rankings are set to 2. After calculating using equation (13), the weights of $GM_{CT}$, $GM_{UB}$, and $GM_{LB}$ are 0.5, 0.25, and 0.25, respectively.

$$w_i = \frac{1/R_i}{\sum_{i=1}^{n} 1/R_i}. \tag{13}$$

(5) Apply the weighted average method to obtain the final predicted value.

## 3. Empirical Analysis

To verify the applicability and effectiveness of the proposed fuzzy-decomposition modeling procedure, this study selected the total demand of TFT-LCD panels provided by a leading manufacturer in Taiwan between years 2010 to 2012 for empirical analysis. Detailed descriptions are provided in the following subsections.

*3.1. Total Demand Data of TFT-LCD Panels.* The development of TFT-LCD is already a mature manufacturing industry. Since the 20th century, China's active participation in manufacturing has caused fierce competition in this industry, and there is an imbalance in which supply exceeds the actual demand. Under this industrial circumstance, how to effectively reduce the inventory level and increase inventory turnover is an important issue facing TFT-LCD manufacturers.

TFT-LCD panel manufacturing is a highly capital-intensive industry. Economies of scale are an important way for this industry to maintain its price competitive advantage in the global business market. Although a moderate increase in production quantity can reduce the manufacturing costs, however, if the production quantity exceeds the critical value, the total cost may increase due to other expenses (such as management costs, storage costs, taxes, insurance, and interest). Therefore, the production quantity should not be increased without qualifications; instead, it should be maintained at a balance point in consideration of the total cost. To meet the above requirements in satisfying production-marketing coordination, a robust short-term demand prediction is requisite.

Because the demand for TFT-LCD panels is greatly affected by the business cycle, and the changes in the global business cycle have increased significantly in recent years, using a large number of long-term historical observations to forecast does not meet the real-time control requirements of production-marketing coordination. This study selected data set provided by a leading TFT-LCD panel manufacturer in Taiwan to confirm the practicability and effectiveness of the proposed method. The data set is the total monthly demand for TFT-LCD panels of the company, which contains 36-period observations ranging from January 2010 to December 2012. Due to the confidentiality requirement, this study had carried out data preprocessing, through min–max normalization to convert all data to fall into the [1, 2] interval; the normalized data is shown in Table 1. Equation (14) is the min–max normalization conversion formula used in this research; $x^{(0)}(k)$ is the prenormalized datum, $x_{min}$ is the minimum value in the prenormalized data set, $x_{max}$ is the maximum value in the prenormalized data set, and $n^{(0)}(k)$ is the postnormalized datum. In this study, four observations were used each time to build the model to predict the next output. A total of 32 models and 32 predicted outputs are obtained.

$$n^{(0)}(k) = \frac{x^{(0)}(k) - x_{min}}{x_{max} - x_{min}} + 1. \tag{14}$$

*3.2. A Modeling Example of the Proposed Modeling Procedure.* Here, the first four samples in the case of TFT-LCD are used as an example to illustrate the calculation details of the proposed procedure. That is, the actual demand from January 2010 to April 2010 is used to establish a model to predict the demand in May 2010, specifically, using series $X^{(0)} = \{1.135, 1.000, 1.231, 1.277\}$ as the inputs and then separating it into three derived series, namely, $X_{UB}^{(0)} = \{1.135, 1.169, 1.282, 1.312\}$, $X_{CT}^{(0)} = \{1.135, 1.045, 1.138, 1.194\}$, and $X_{LB}^{(0)} = \{1.135, 0.9663, 0.974, 0.965\}$. The typical GM (1, 1) is used to build models for these three derived series and obtain their estimated values in the fifth period; they are $x_{UB}^{(0)}(5) = 1.4019$, $x_{CT}^{(0)}(5) = 1.2811$, and $x_{LB}^{(0)}(5) = 0.9678$. The three estimated values are combined through the weighted average method to create the final predicted value $\hat{x}^{(0)}(5) = 1.2330$.

*3.3. Performance Evaluation and Comparison.* To verify the improvement effect of the proposed modeling procedure, the typical GM (1, 1) is used for comparison. The typical GM (1, 1) is one of the most popular models in grey system theory due to its ease of use. In the experiment, the typical GM (1, 1) is built from the same data, and the results are

Table 1: The monthly demand of TFT-LCD panel.

| Months | Demands | Months | Demands | Months | Demands |
|---|---|---|---|---|---|
| 2010/01 | 1.135 | 2011/01 | 1.587 | 2012/01 | 1.442 |
| 2010/02 | 1.000 | 2011/02 | 1.334 | 2012/02 | 1.446 |
| 2010/03 | 1.231 | 2011/03 | 1.768 | 2012/03 | 1.696 |
| 2010/04 | 1.277 | 2011/04 | 1.835 | 2012/04 | 1.718 |
| 2010/05 | 1.274 | 2011/05 | 1.851 | 2012/05 | 1.728 |
| 2010/06 | 1.176 | 2011/06 | 1.850 | 2012/06 | 1.714 |
| 2010/07 | 1.334 | 2011/07 | 2.000 | 2012/07 | 1.753 |
| 2010/08 | 1.133 | 2011/08 | 1.914 | 2012/08 | 1.742 |
| 2010/09 | 1.193 | 2011/09 | 1.658 | 2012/09 | 1.691 |
| 2010/10 | 1.241 | 2011/10 | 1.515 | 2012/10 | 1.932 |
| 2010/11 | 1.475 | 2011/11 | 1.542 | 2012/11 | 1.933 |
| 2010/12 | 1.749 | 2011/12 | 1.573 | 2012/12 | 1.848 |

Table 2: MAPE criteria.

| MAPE (%) | Forecasting power |
|---|---|
| <10 | Highly accurate forecasting |
| 10–20 | Good forecasting |
| 20–50 | Reasonable forecasting |
| >50 | Inaccurate forecasting |

compared with the proposed modeling procedure; it is also based on four observations to predict the next output.

Yokum and Armstrong [33] stated that the accuracy is an important index used to measure the forecasting ability of various methods. Liu and Lin [12] also considered that it is essential to confirm whether a grey model is suitable for forecast using an error-based measurement index. For the empirical analysis, the mean absolute percentage error (MAPE) is adopted to evaluate the forecasting performance; MAPE provides a performance benchmark and is a robust error-based measurement [34]. Table 2 shows the MAPE criteria [35] generally used to evaluate forecasting models. Equation (15) is the calculation formula of MAPE, where $\widehat{y}_i$ and $y_i$ are the forecasting and actual values of the $i$-th testing sample, and $m$ is the number of testing samples.

$$\text{MAPE} = \frac{1}{m}\sum_{i=1}^{m}\left|\frac{\widehat{y}_i - y_i}{y_i}\right| \times 100\%. \tag{15}$$

The empirical results showed that the proposed method resulted in a favorable prediction; its MAPE (8.14%) is slightly smaller than the MAPE (10.98%) of the typical GM (1, 1). Besides, its improvement in the MAPE reaches 25.87% compared with that of typical GM (1, 1). These indicated that the proposed method can improve the prediction ability of the grey model and bring better prediction results by the fuzzy-decomposition modeling procedure.

## 4. Conclusions and Discussion

Facing the increasingly fierce business competition environment, enterprises must quickly respond to business problems and maintain a competitive advantage, which depends on timely and correct decisions. Decision contexts usually encounter many uncontrollable events and factors. To overcome this uncertainty, enterprises must have appropriate forecasting techniques. Predictive analysis can help managers grasp the future development trends to ease the impact of uncertainty and then make meaningful decisions. In various management contexts, decisions that need immediate response are the more hard tasks for managers. Real-time grasp of the situation through a limited amount of data enables managers to carry out appropriate processing and obtain effective management and control. Therefore, it is of greater practical value to establish forecasting models under small data sets.

Grey system theory can build models with small data sets, which meets the needs of enterprise decision analysis. To further improve it, a fuzzy-decomposition modeling procedure based on LI function was developed in this study for enhancing the prediction performance of typical GM (1, 1). This method fits the model better by decomposing the series into three subseries and then obtains a more robust prediction output. Through the verification of the monthly demand for TFT-LCD panels, the experimental results showed that the proposed method has good predictive performance under small data sets. From the comparisons, it can be found that the proposed procedure outperforms the typical GM (1, 1) in the empirical case. Obviously, the proposed method has important practical value and is an effective tool to assist decision analysis with limited samples.

The proposed method maintains the features of the grey approach, which is a consistent model without stochastic processes; given the same input, only one result will be obtained, rather than a different result each time. If a model is built with small training samples in a stochastic process, managers would bear the risk of uncertainty when making decisions. In the future, the proposed method can be applied to solve the forecasting problems in other fields, such as medicine, energy, finance, engineering, and transportation, to further confirm its effectiveness. In addition, using more training samples to verify the predictive power of the proposed method is also a feasible direction for future work. Finally, researchers may use optimization algorithms to find a more appropriate defuzzification weight to acquire a better grey model.

## Data Availability

The data used in the experiment are listed in this article; anyone can use these data by citing this article.

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## Acknowledgments

This research was supported by the Fujian Provincial Social Science Planning Project of China under Grant no. FJ2019B099, Zhejiang Provincial Natural Science Foundation of China under Grant no. LY19G010002, Qianjiang Talent Program of Zhejiang Province (China), National Social Science Foundation of China under Grant no.

20BGL003, Natural Science Foundation of Fujian Province for Youths of China under Grant no. 2017J051165, and Humanistic and Social Science Youth Foundation of the Ministry of Education of China under Grant no. 17YJC630204.

## References

[1] N. Rekova, H. Telnova, O. Kachur, I. Golubkova, T. Baležentis, and D. Streimikiene, "Financial sustainability evaluation and forecasting using the Markov chain: the case of the wine business," *Sustainability*, vol. 12, no. 15, p. 6150, 2020.

[2] S. Zeng, Y. Hu, T. Balezentis, and D. Streimikiene, "A multi-criteria sustainable supplier selection framework based on neutrosophic fuzzy data and entropy weighting," *Sustainable Development*, vol. 28, no. 5, pp. 1431–1440, 2020.

[3] C. J. Chang, D. C. Li, W. L. Dai, and C. C. Chen, "Utilizing an adaptive grey model for short-term time series forecasting: a case study of wafer-level packaging," *Mathematical Problems in Engineering*, vol. 2013, Article ID 526806, , 2013.

[4] C.-J. Chang, D.-C. Li, C.-C. Chen, and W.-C. Chen, "Extrapolation-based grey model for small-data-set forecasting," *Economic Computation and Economic Cybernetics Studies and Research*, vol. 53, no. 1, pp. 171–182, 2019.

[5] D. A. Lind, W. G. Marchal, and S. A. Wathen, *Basic Statistics for Business and Economics*, McGraw-Hill Education, New York, NY, USA, 2013.

[6] I. H. Witten, E. Frank, and M. A. Hall, *Data Mining: Practical Machine Learning Tools and Techniques*, Elsevier Science, Amsterdam, Netherlands, 2011.

[7] S. F. Liu and Y. Lin, *Grey Information: Theory and Practical Applications*, Springer-Verlag, Berlin, Germany, 1st edition, 2006.

[8] R.-C. Tsaur, "Fuzzy exponential smoothing model by grey forecasting values," *Journal of the Chinese Institute of Industrial Engineers*, vol. 18, no. 6, pp. 95–103, 2001/01/01 2001.

[9] R.-C. Tsaur, "Forecasting analysis by fuzzy grey model GM (1, 1)," *Journal of the Chinese Institute of Industrial Engineers*, vol. 23, no. 5, pp. 415–422, 2006.

[10] J. L. Deng, "Control problems of grey systems," *Systems & Control Letters*, vol. 1, no. 5, pp. 288–294, 1982.

[11] W. Pan, L. Jian, and T. Liu, "Grey system theory trends from 1991 to 2018: a bibliometric analysis and visualization," *Scientometrics*, vol. 121, no. 3, pp. 1407–1434, 2019.

[12] S. F. Liu and Y. Lin, *Grey Systems: Theory and Applications*, Springer-Verlag, Berlin, Germany, 1st edition, 2010.

[13] N. Xie and R. Wang, "A historic review of grey forecasting models," *Journal of Grey System*, vol. 29, no. 4, pp. 1–29, 2017.

[14] I. E. Edem, S. A. Oke, and K. A. Adebiyi, "A novel grey-fuzzy-Markov and pattern recognition model for industrial accident forecasting," *Journal of Industrial Engineering International*, vol. 14, no. 3, pp. 455–489, 2018.

[15] D. Luo, M. Ambreen, A. Latif, and X. Wang, "Forecasting Pakistan's electricity based on improved discrete grey polynomial model," *Grey Systems: Theory and Application*, vol. 10, no. 2, pp. 215–230, 2020.

[16] Z.-X. Wang, Q. Li, and L.-L. Pei, "A seasonal GM (1, 1) model for forecasting the electricity consumption of the primary economic sectors," *Energy*, vol. 154, pp. 522–534, 2018.

[17] B. Zeng, H. Duan, Y. Bai, and W. Meng, "Forecasting the output of shale gas in China using an unbiased grey model and weakening buffer operator," *Energy*, vol. 151, pp. 238–249, 2018.

[18] A. Karaaslan and M. Gezen, "Forecasting of Turkey's sectoral energy demand by using fuzzy grey regression model," *International Journal of Energy Economics and Policy*, vol. 7, no. 1, pp. 67–77, 2017.

[19] Y.-J. Chiu, Y.-C. Hu, P. Jiang, J. Xie, and Y.-W. Ken, "A multivariate grey prediction model using neural networks with application to carbon dioxide emissions forecasting," *Mathematical Problems in Engineering*, vol. 2020, Article ID 8829948, , 2020.

[20] L. F. Wu, S. F. Liu, L. G. Yao, S. L. Yan, and D. L. Liu, "Grey system model with the fractional order accumulation," *Communications in Nonlinear Science and Numerical Simulation*, vol. 18, no. 7, pp. 1775–1785, 2013.

[21] P.-S. Yu, S.-T. Chen, C.-C. Wu, and S.-C. Lin, "Comparison of grey and phase-space rainfall forecasting models using a fuzzy decision method," *Hydrological Sciences Journal*, vol. 49, no. 4, 2004.

[22] R.-C. Tsaur, "Forecasting analysis by using fuzzy grey regression model for solving limited time series data," *Soft Computing*, vol. 12, no. 11, pp. 1105–1113, 2008.

[23] R.-C. Tsaur and Y.-C. Liao, "Forecasting lcd tv demand using the fuzzy grey model GM (1, 1)," *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, vol. 15, no. 6, pp. 753–767, 2007.

[24] R. C. Tsaur, "Seasonal forecasting of a decomposed fuzzy exponential smoothing model using grey estimated values," *Journal of the Chinese Institute of Engineers*, vol. 32, no. 1, pp. 17–31, 2009.

[25] C. Zor and F. Çebi, "Demand prediction in health sector using fuzzy grey forecasting," *Journal of Enterprise Information Management*, vol. 31, no. 6, pp. 937–949, 2018.

[26] Y.-H. Lin, C.-C. Chiu, P.-C. Lee, and Y.-J. Lin, "Applying fuzzy grey modification model on inflow forecasting," *Engineering Applications of Artificial Intelligence*, vol. 25, no. 4, pp. 734–743, 2012.

[27] K. Shi and M. Bai, "Using grey model to predict the governance indicators in China and India," *Journal of Grey System*, vol. 32, no. 1, pp. 16–28, 2020.

[28] X.-y. Zeng, L. Shu, G.-m. Huang, and J. Jiang, "Triangular fuzzy series forecasting based on grey model and neural network," *Applied Mathematical Modelling*, vol. 40, no. 3, pp. 1717–1727, 2016.

[29] X. Zeng, L. Shu, and J. Jiang, "Fuzzy time series forecasting based on grey model and Markov chain," *International Journal of Applied Mathematics*, vol. 46, no. 4, 2016.

[30] H.-L. Wong and J.-M. Shiu, "Comparisons of fuzzy time series and hybrid Grey model for non-stationary data forecasting," *Applied Mathematics & Information Sciences*, vol. 6, no. 2, pp. 409–416, 2012.

[31] C. J. Chang, D. C. Li, W. L. Dai, and C. C. Chen, "A latent information function to extend domain attributes to improve the accuracy of small-data-set forecasting," *Neurocomputing, A.; Proceedings Paper*, vol. 129, pp. 343–349, Apr 2014.

[32] J. L. Deng, "Introduction to grey system theory," *Journal of Grey System*, vol. 1, no. 1, pp. 1–24, 1989.

[33] J. T. Yokum and J. S. Armstrong, "Beyond accuracy: comparison of criteria used to select forecasting methods," (in English)," *International Journal of Forecasting*, vol. 11, no. 4, pp. 591–597, 1995.

[34] S. Makridakis, "Accuracy measures: theoretical and practical concerns," *International Journal of Forecasting*, vol. 9, no. 4, pp. 527–529, 12//1993.

[35] S. A. DeLurgio, *Forecasting Principles and Applications*, Irwin/McGraw-Hill, New York, NY, USA, 1998.

Hindawi
Mathematical Problems in Engineering
Volume 2021, Article ID 6684711, 8 pages
https://doi.org/10.1155/2021/6684711

*Research Article*

# Research and Development Investment Combination Forecasting Model of High-Tech Enterprises Based on Uncertain Information

**Qi Wei, Min Chen, and Chuan-yang Ruan**

*School of Business Administration, Guangdong University of Finance & Economics, 21 Luntou Road, Guangzhou 510320, China*

Correspondence should be addressed to Chuan-yang Ruan; ruancyang@163.com

Received 2 November 2020; Revised 12 January 2021; Accepted 22 January 2021; Published 31 January 2021

Academic Editor: Shouzhen Zeng



The high demand of the competitive market for innovation has brought the increase of research and development (R&D) investment. High-tech enterprises can reasonably control R&D cost and effectively manage R&D activities by accurately predicting R&D investment. Given the characteristics that high-tech enterprises have high uncertainty and frequently changing information in R&D investment, this paper uses the grey metabolic GM (1, 1) model and the exponential smoothing method in time series to establish a single prediction model of R&D investment in high-tech enterprises. With the analysis of the advantages and disadvantages of each single model, a combined forecast model of R&D investment in high-tech enterprises is thus established. The model was applied to the forecast of R&D investment of a high-tech enterprise in China from 2019 to 2023, and the results verified the higher accuracy and practicability of this model. The establishment of this model can provide effective support for high-tech enterprises in R&D cost management.

## 1. Introduction

In the era of knowledge economy, market competition for innovative products is becoming increasingly fierce. In order to meet the increasingly personalized needs of customers, the speed of product upgrading is continuously accelerated [1]. With the constant emergence of new technologies and new products, the investment in research and development (R&D) expenses required for enterprise innovation is increasing. With the increase in the intensity of R&D investment, however, most companies have not achieved a directly proportional increase in product output and revenue, and the problem of improper control of R&D costs is still common [2]. In addition, from the statistics on science and technology expenditures released by the Ministry of Science and Technology of China, it can be seen that the intensity of corporate R&D expenditure has been increasing in recent years as shown in Table 1. This phenomenon shows that, in the construction of an innovative country, the role of enterprises as the main body of technological innovation has become more prominent. Enterprises are the mainstay of technological innovation, and their R&D activities have an important impact on the promotion of national innovation-driven development [3]. As the main body of independent innovation, this is especially true of high-tech enterprises and corporate R&D investment is more effective in the high-tech field [4]. In order to improve the core competitiveness, high-tech enterprises must pay attention to R&D investment, effectively manage R&D investment, and achieve sustainable development based on a reasonable increase in input and output [3]. For this reason, effective management of R&D investment in high-tech enterprises has become an urgent problem to be addressed.

Cost forecasting is the prerequisite for all R&D activities, and the goal of forecasting is to provide important information on subsequent stages [5], so as to effectively control enterprise costs. At present, there have been many studies on the forecast of R&D investment in China, mainly focusing on the overall scientific research investment of the national and regional levels [6, 7]. The R&D activities of enterprises are uncertain and unknown, which are more prominent during the R&D activities of high-tech enterprises. This phenomenon will increase the degree of random fluctuations in R&D investment, making it very difficult to forecast

Table 1: Enterprise R&D expenditures in 2011–2018.

| Items | 2011 | 2012 | 2013 | 2014 | 2015 | 2016 | 2017 | 2018 |
|---|---|---|---|---|---|---|---|---|
| Enterprise R&D expenditures (100 million) | 6579.3 | 7842.2 | 9075.8 | 10060.6 | 10881.3 | 12144 | 13660.2 | 15233.7 |
| Growth rate (%) | 26.9 | 19.2 | 15.7 | 10.9 | 8.2 | 11.6 | 12.5 | 11.5 |
| Proportion of the whole society (%) | 75.7 | 76.2 | 76.6 | 77.3 | 76.8 | 77.5 | 77.6 | 77.4 |

Data source: "Annual Report on China's Science and Technology Expenditures Statistics."

accurately. Therefore, there are a few relevant studies on forecasting R&D investment of high-tech enterprises. Li and Hu (2017) used radial basis function neural network (RBF) and backpropagation neural network (BP) to construct an investment simulation system of R&D activity, based on the nonlinear relationship between GEM companies' R&D investment and its influencing factors. The study found that the RBF neural network model has better fitting and forecasting effects than BP neural network [8]. Chen (2013) carried out research on forecasting R&D costs of high-tech enterprises by applying the system dynamics model, effectively improving the ability and accuracy of forecasting enterprise R&D cost [9]. Some scholars recognize the importance of forecasting R&D costs of enterprises, and they believe that while disclosing historical data onto R&D expenses, attention should be paid to the provision of future forecast information [10].

Macroeconomic fluctuations and some sudden uncertain factors often have a major impact on R&D investment in a certain field. Most R&D investment focuses on the short-term changes in some economic indicators, and long-term forecasting of R&D investment and building more accurate prediction models have always been hot research issues in the field of prediction [11]. R&D investment forecasting is usually used to forecast single or a few closely related investment indicators. Therefore, the methods used for R&D investment forecasting are mainly time series methods [12, 13], semiparametric model methods (e.g., Rounaghi et al. [14] and Dordonnat et al. [15]), uncertain forecasting methods [16–19], and so on. R&D investment forecasting is very important to high-tech enterprises, but the uncontrollability in the forecasting process will increase the difficulty of forecasting and management [2]. It is difficult to accurately forecast the investment of R&D funds in enterprises, and it is rarely achieving through reasonable quantitative methods at present [20]. Scientific and reasonable forecasting of R&D funds investment in high-tech enterprises has crucial practical significance for promoting innovative enterprises and national construction.

The high demand of the competitive market for innovation has brought the increase of research and development (R&D) investment. High-tech enterprises can reasonably control R&D cost and effectively manage R&D activities by accurately predicting R&D investment. In view of this, given the characteristics that high-tech enterprises have high uncertainty and frequently changing information in R&D investment, this paper takes the investment of high-tech R&D funds as the research object, respectively using the grey metabolic GM (1, 1) model and the exponential smoothing method to establish a single forecasting model of high-tech enterprise R&D investment. However, the prediction accuracy of a single prediction model is poor, then according to the advantages and disadvantages of each model, a combined forecasting model for R&D funds investment of high-tech enterprises is thus established, which uses variance reciprocal method to assign weights. It is exemplified by the forecasting results that the combined forecasting model has higher forecasting precision and better fit with the actual situation. Thus, it is more suitable for the R&D investment forecasting of high-tech enterprises with a fast information update.

## 2. The Combined Forecasting Model of R&D Investment

The R&D activities of high-tech enterprises are characterized by high risks and high uncertainties. Their R&D investment is affected and restricted by the external environment and various factors, and the information on R&D innovation activities is updated fast and the product life cycle is short. There are many models to choose from for the prediction of R&D costs, but different methods have different scopes and characteristics. The selection of models and the different methods of solving lead to different accuracy of the results. Although many scholars have conducted extensive and in-depth research on R&D cost forecasting, a single forecasting method is mainly used in actual application, which makes the information contained in different forecasting methods ineffective. Individual studies have improved the prediction accuracy through combination forecasting, but the combined model is mainly limited to the same method. For example, Xiao and Zhou (2006) used the improved ant colony algorithm to solve the weighted average coefficient in the combination forecast. And they tested and verified the feasibility and effectiveness by applying this method to R&D funds in China [21].

The grey model can extract valuable information through the generation and development of sample data itself, which avoids discussing the relationship between other influencing factors [22]. The improved grey metabolic GM (1, 1) model can make full use of the new information generated by the time lapse, and the forecasting accuracy is higher. The exponential smoothing method based on time series can consider new information and historical data. Both of these two forecasting methods can better meet the requirements of high-tech enterprises' R&D investment forecasting. According to the characteristics of R&D investment of high-tech enterprises and the applicable scope of the forecasting model, the metabolic GM (1, 1) model optimized by the grey system and the exponential smoothing

method of the time series model is selected as the single model for forecasting.

*2.1. Grey Metabolic GM (1, 1) Model.* The grey system theory was proposed and developed by Deng in 1982. In theory, the GM (1, 1) model has been widely used as an effective forecasting tool [23–25], especially in areas with significant uncertainty and lack of data, such as green electronic materials [26], energy consumption [27], and electricity [28]. Grey prediction is to accumulate irregular historical data with randomness and uncertainty to generate a series of exponential growth laws, thereby establishing a prediction model of grey differential equations [29]. The R&D innovation system of high-tech enterprises is easily interfered by various uncertain factors of the development process. It contains both unknown information and known information, which is a typical grey system of "small sample, poor information" [30]. In addition, the rapid development of the system prompts the continuous update and increase of R&D information, and the forecasting of R&D investment of high-tech enterprises needs to consider information that is newer. However, the ordinary GM (1, 1) model uses the first component of the original sequence as the initialization, and the new information is not fully utilized in the forecasting [31]. According to the new information principle of grey system theory and the characteristics of high-tech enterprise R&D investment, the grey metabolic GM (1, 1) model is selected as the grey forecasting single model.

The ordinary modeling process of GM (1, 1) model is as follows [32].

Set the nonnegative original time sequence $X^{(0)}$:

$$X^{(0)} = \left(x^{(0)}(1), x^{(0)}(2), \ldots, x^{(0)}(n)\right). \tag{1}$$

Based on the initial sequence $X^{(0)}$, a new series $X^{(1)}$ can be obtained through a one-order accumulating generation operator (1-AGO):

$$X^{(1)} = \left(x^{(1)}(1), x^{(1)}(2), \ldots, x^{(1)}(n)\right), \tag{2}$$

where

$$x^{(1)}(k) = \sum_{i=1}^{n} x^{(0)}(i), \quad k = 1, 2, \ldots, n. \tag{3}$$

Then, generate a sequence $Z^{(1)}$ from the mean generation with consecutive neighbors of $X^{(1)}$:

$$Z^{(1)} = \left(z^{(1)}(2), z^{(1)}(3), \ldots, z^{(1)}(n)\right). \tag{4}$$

In the formula,

$$z^{(1)}(k) = 0.5x^{(1)}(k) + 0.5x^{(1)}(k+1), k = 2, 3, \ldots, n. \tag{5}$$

So, the basic form of the GM (1, 1) model is the first-order grey differential equation:

$$x^{(0)}(k) + az^{(1)}(k) = b, \tag{6}$$

where $a$ and $b$ are undefined constants, $a$ is development coefficient, and $b$ is system grey action quantity. Let $X^{(1)}$ satisfy the whitening equation of the GM (1, 1) model:

$$\frac{\mathrm{d}x^{(1)}}{\mathrm{d}t} + ax^{(1)} = b. \tag{7}$$

Moreover, set $\hat{a} = [a, b]^T$; then, the parameters can be estimated by the principle of least squares, that is, satisfying $[a, b]^T = (B^T B)^{-1} B^T Y$.

Let

$$Y = \begin{bmatrix} x^{(0)}(2) \\ x^{(0)}(3) \\ \vdots \\ x^{(0)}(n) \end{bmatrix}, \quad B = \begin{bmatrix} -z^{(1)}(2) & 1 \\ -z^{(1)}(3) & 1 \\ \vdots & \vdots \\ -z^{(1)}(n) & 1 \end{bmatrix}. \tag{8}$$

It can be solved by substituting $\hat{a}$ into the whitening equation, and the time response sequence becomes the following:

$$\hat{X}^{(1)}(k+1) = \left[x^{(0)}(1) - \frac{b}{a}\right]e^{-ak} + \frac{b}{a}, \quad k = 1, 2, \ldots, n. \tag{9}$$

Finally, the predicted value of the original sequence $X^{(0)}$ can be obtained through an inverse accumulated generating operation:

$$\begin{aligned} \hat{X}^{(0)}(k+1) &= \hat{X}^{(1)}(k+1) - \hat{X}^{(1)}(k) \\ &= \left(1 - e^{-a}\right)\left[x^{(0)}(1) - \frac{b}{a}\right]e^{-ak}, \quad k = 1, 2, \ldots, n. \end{aligned} \tag{10}$$

In the ordinary GM (1, 1) model, if only all historical data of the observation object are used, then it will be found that only little data closing to the real time have high forecast accuracy. Over time, random disturbance factors will continue to enter the system and have an impact in the future. This situation will lead to a gradual increase in forecast errors and gradually weaken the predictive significance of the model. The grey metabolic GM (1, 1) model is a grey forecasting model optimized by the ordinary GM (1, 1) model. It not only has the advantages of the ordinary GM (1, 1) model but also can consider random disturbance factors and make full use of the latest information carried by the original data sequence [22]. That makes the predicted value generated during the dynamic development process, and the predicted result is more accurate.

The grey metabolic GM (1, 1) model is not one-time forecasting, but one-by-one forecasting with successive replacement, continuously adding new updated information and removing useless old information to maintain the sequence of equal dimensions [33]. The new information is the forecasting data obtained through the model, which will directly affect the R&D system. The oldest information that is far away from the future time will have a small impact on the system and can be gradually discarded during the dynamic forecasting process. Keeping the system updated and developed can improve the forecasting accuracy of the model. Specifically, the forecasting process of the grey metabolic GM (1, 1) model is to build a GM (1, 1) model based on the known original data sequence $X^{(0)}$ to forecast the next data $x^{(0)}(n+1)$. Next, by putting the new information into the original sequence and removing the oldest data $x^{(0)}(1)$ with reduced information significance, a new prediction sequence $X^{(0)} = (x^{(0)}(2), x^{(0)}(3), \ldots, x^{(0)}(n+1))$ is generated [26]. Repeating the above steps to continuously build a new GM (1, 1) model to forecast new information one by one and substitute in turn until the predetermined forecasting target is completed [34], the required prediction data can be obtained.

*2.2. Exponential Smoothing Method.* The exponential smoothing method is a commonly used time series analysis and forecasting method developed from the moving average method. By calculating the smoothing value and working with the time series model, it can be used to predict the future development of the phenomenon or system [1]. This method assigns a greater weight to recent observations and a smaller weight to longer-term samples [35]. Then, they are calculated using the weighted average method in chronological order, making the model's predicted value able to reflect more recent information and include all historical data, so that the forecasting results are more in line with the actual development of the system [36]. The modeling logic and calculation method of the exponential smoothing method is easy to understand, and, more importantly, the model forecasting results are relatively stable. According to the number of times that the observation value is smoothed, it can be divided into single exponential smoothing, secondary exponential smoothing, and cubic exponential smoothing. Which exponential smoothing model is used for prediction is mainly based on the trend of the data value and the effect of the model and the actual value fitting. The basic formula of the exponential smoothing method is [37]

$$\widehat{x}_t = \alpha x_t^{(0)} + (1-\alpha)\widehat{x}_{t-1}, \quad t = 1, 2, \ldots, n. \tag{11}$$

In the formula, $\widehat{x}_t$ is the smoothing forecasting value in the time $t$, $x_t^{(0)}$ is the actual observation value in time $t$, $\widehat{x}_{t-1}$ is the smoothing forecasting value in the time t − 1, $\alpha$ is the smoothing coefficient, and its value range is [0, 1]. The value of $\alpha$ in the calculation of exponential smoothing is easily affected by subjective factors, so it is very important to determine a reasonable value of $\alpha$. Generally speaking, if the data fluctuates greatly, the value of $\alpha$ should be larger. The larger the value $\alpha$, the more attention the exponential smoothing prediction model pays to new information. Yet, in the actual forecasting process, the appropriate smoothing coefficient is mainly selected according to the forecasting accuracy. The basic formula can be extended as

$$\begin{aligned}\widehat{x}_{t+1} &= \alpha x_t^{(0)} + (1-\alpha)\alpha x_{t-1}^{(0)} \\ &\quad + (1-\alpha)^2 \alpha x_{t-2}^{(0)} + \cdots + (1-\alpha)^n \alpha x_{t-n}^{(0)} \\ &\quad + \cdots + (1-\alpha)^t x_1^{(0)}.\end{aligned} \tag{12}$$

It can be seen from the above formula that the forecasting value of the time t + 1 is actually a weighted sum of the time $t$ and the actual observation value of all previous periods in exponential form. After increasing the observation value of the latest period, the new data replaces the status of the old data, and the weight of the old data will gradually weaken so that the forecasting value can always reflect the latest data structure.

The grey metabolic GM (1, 1) model and exponential smoothing method are both suitable for R&D investment forecasting of high-tech enterprises. However, the single forecasting models have certain limitations. The characteristics of the two forecasting methods are different, and they can only reflect the future situation of the R&D system from an individual perspective, which cannot fully meet the forecasting requirements of the research object.

The advantage of the grey metabolism GM (1, 1) model is that there is less demand for the original data [38], and the more recent the data prediction, the more effective. The model that passes the posterior variance test is more suitable for medium and long-term forecasting [39]. However, its short-term forecasting effect is not as good as the time series model. The exponential smoothing method performs non-equal weight processing on data at different times [35], which means, as time goes on, it gradually reduces the degree of influence, which can offset or reduce the influence of abnormal factors, so that the forecasting model has higher stability. When the external factors change significantly, however, the results of the exponential smoothing method are likely to cause large deviations, which is more suitable for short-term forecasting.

*2.3. Combined Forecasting Model.* A single prediction model usually contains only part of the information of the prediction object, which makes it suffer from defects such as not extensive information sources and being affected by the setting of the model. According to the foregoing description, it can be found that the scope of application of the metabolic GM (1, 1) model and the exponential smoothing method can complement each other. If these two single models are properly combined through certain rules, effective system information can be extracted from the single predictive model. This makes the error caused by incomplete consideration in the prediction process smaller and the prediction accuracy higher. Hibon and Evgeniou (2005) also pointed out that the combination model may not be the optimal model based on large sample experiments, but the predicted risk of choosing the combination model is less

than choosing the single model [40]. Therefore, based on the limitations of the single model in practical application and the randomness of system influencing factors, this paper uses the combined forecasting model [41], which aims comprehensively to use the information provided by each single forecasting method to improve the forecasting accuracy [42]. Combined forecasting models assign weight to single models according to certain rules, and that improves the performance of single forecasting models by including more comprehensive information [42–44]. Suppose $X$ is the forecasting value of the combined model, $\hat{x}_i$ is the forecasting value of the $i$-th single model, the combined model weight is W, and $W = \sum_{i=1}^{n} w_i = 1$; then, the combined forecasting model is then given as

$$X = \sum_{i=1}^{m} w_i \hat{x}_i = w_1 \hat{x}_1 + w_2 \hat{x}_2 + \cdots + w_m \hat{x}_m, \quad i = 1, 2, \ldots, m. \tag{13}$$

The determination of the weighting coefficient in the combined forecasting model is very important. The distribution of weight directly affects the forecasting realization and the final effect, and reasonable weighting can effectively improve the forecasting accuracy of the model [24]. The variance reciprocal method is simple and effective in practical applications, so it is used frequently. In this paper, the variance reciprocal method is used to assign the combination weight. This method assigns different weights to the single model according to the size of the data's sum of squared errors. That is, the smaller the single models' sum of squared errors, the higher the forecasting accuracy and the greater the weight [20]. Assuming $e_i$ as the sum of squared errors of the $i$-th single model, then

$$w_i = \frac{e_i^{-1}}{\sum_{i=1}^{m} e_i^{-1}}, \quad i = 1, 2, \ldots, m. \tag{14}$$

In the formula,

$$e_i = \sum_{t=1}^{n} \left( x_t - \hat{x}_t(i) \right)^2, \tag{15}$$

where $x_t$ is the observed value of time $t$ and $\hat{x}(i)$ is the predicted value of the $i$-th method in time $t$.

## 3. Application and Result Analysis of the Combined Forecasting Model

Company A is a high-tech enterprise in the software and information technology service industry. Its investment in R&D from 2011 to 2019 is shown in Table 2. As the company attaches great importance to improving its independent R&D capabilities, it continuously increases its investment in new business R&D every year, and it constantly achieves new breakthroughs in conventional technologies and products, which enhances the company's product competitiveness. The relevant data of company A comes from the CSMAR database (http://www.csmar.com/Csmar.html).

The grey system theory requires that the sample data should not be less than 4, and this study selects the data from 2011 to 2015 as the model testing sample. In order to reflect the validity of the model forecasting, the root mean square error (RMSE) and mean absolute percentage error (MAPE) are selected as error evaluation indicators. The conventional GM (1, 1) model and the metabolic GM (1, 1) model were, respectively, predicted by MATLAB software. In addition, according to the accuracy test requirements of the grey model, a posterior variance test was performed on the GM (1, 1) model. The test result shows the ratios of mean square deviation were 0.0993 and 0.0861, both of which were less than 0.35. According to the above results, the test results are good and the model accuracy level is the first level. In view of the changing trend of the quadratic curve of the sample data, cubic exponential smoothing is selected as the exponential smoothing method. After repeated measurement and comparison, the smoothing coefficient selects $\alpha = 0.6$, which has higher simulation accuracy. The predicted value of the three individual models is compared with the actual value, and the forecasting error results are shown in Table 3.

Table 3 gives an overview that the mean absolute percentage error (MAPE) of the two grey models is less than 10%, which meets the requirements of forecasting accuracy and has good predictability. Comparing the fitting condition of forecasting each year, it is found that the forecasting error of the conventional GM (1, 1) model fluctuates to a large extent. The indicators of the grey metabolic GM (1, 1) model are all far less than the result of the conventional GM (1, 1) model, which shows higher forecasting reliability. The prediction results in Table 3 show that the metabolic GM (1, 1) model is more suitable for the prediction of R&D investment of high-tech enterprises. The MAPE of the triple exponential smoothing method is 3.73, and the fitting result of the model is good, which can meet the actual forecasting requirements of high-tech enterprise R&D investment, and it can also reflect their rapidly changing information characteristics.

In order to make up for the shortcomings of a single model, a combined forecasting model is introduced. The variance reciprocal method is used to calculate the weights of each single model, which are $w_1 = 0.27$, $w_2 = 0.217473$, and the combined prediction results are obtained. The forecasting results and errors are calculated according to the combined forecasting model. The calculation results are shown in Table 4.

Table 4 shows that the forecasting errors for all other years are below 3% except for 2016. Compared with the grey metabolic GM (1, 1) model and the cubic exponential smoothing method, all evaluation indicators of the combined forecasting model are greatly reduced, which is much lower than the results of the two single forecasting models. This shows that this model has a very high degree of fitting and forecasting accuracy to the original data, and it can overcome the limitations of the single model, which is more suitable for the actual situation of R&D investment forecasting in high-tech enterprise.

Therefore, according to the data from 2011 to 2019, the company's R&D investment for the next five years is

Table 2: Company A's R&D investment in 2011–2019.

| Year | 2011 | 2012 | 2013 | 2014 | 2015 | 2016 | 2017 | 2018 | 2019 |
|---|---|---|---|---|---|---|---|---|---|
| R&D investment (10 K) | 9925.50 | 10613.17 | 11626.77 | 13382.29 | 16908.87 | 21368.59 | 24582.22 | 28864.64 | 34579.33 |

Table 3: Forecast results and errors of a single model.

| Year | Actual value (10 K) | Predictive value | | | Relative error (%) | | |
|---|---|---|---|---|---|---|---|
| | | GM (1, 1) model | Grey metabolic GM (1, 1) model | The cubic exponential smoothing method | GM (1, 1) model | The grey metabolic GM (1, 1) model | The cubic exponential smoothing method |
| 2016 | 21368.59 | 19312.77 | 19312.79 | 20539.64 | 9.62 | 9.62 | 3.88 |
| 2017 | 24582.22 | 22722.01 | 23130.88 | 25011.67 | 7.57 | 5.90 | 1.75 |
| 2018 | 28864.64 | 26733.07 | 27462.58 | 30221.20 | 7.38 | 4.86 | 4.70 |
| 2019 | 34579.33 | 31452.19 | 32159.04 | 36168.23 | 9.04 | 7.00 | 4.59 |
| SDE | | | | | 2434.68 | 1819.34 | 1321.17 |
| RMSE | | | | | 2434.68 | 1819.34 | 1144.16 |
| MAPE | | | | | 8.40 | 6.85 | 3.73 |

Table 4: Forecast results and errors of the combined model.

| Year | Predictive value (10 K) | Prediction error (10 K) | Relative error (%) |
|---|---|---|---|
| 2016 | 20208.39 | 1160.20 | 5.43 |
| 2017 | 24503.85 | 78.37 | 0.32 |
| 2018 | 29476.37 | −611.73 | 2.12 |
| 2019 | 20208.39 | −506.42 | 1.46 |
| SDE | 813.00 | | |
| RMSE | 704.07 | | |
| MAPE | 2.33 | | |

Table 5: Forecast value of company A's R&D investment from 2020 to 2024.

| Year | Predictive value (10 K) ) | | |
|---|---|---|---|
| | Grey metabolic GM (1, 1) model | The cubic exponential smoothing method | Combination model |
| 2020 | 41194.72 | 41139.51 | 41183.84 |
| 2021 | 49283.67 | 48680.73 | 49164.83 |
| 2022 | 58578.37 | 57193.49 | 58305.41 |
| 2023 | 69190.17 | 66677.78 | 68694.98 |
| 2024 | 81769.00 | 77133.60 | 80855.36 |

forecasted. The weights are calculated based on the predicted values of the two single models, and the combined prediction model is obtained as follows:

$$X = 0.8029\hat{x}_1 + 0.1971\hat{x}_2. \tag{16}$$

The results are shown in Table 5. The company will continue to increase its R&D investment in the next few years, and the amount of R&D investment will continue to rise steadily.

## 4. Conclusion

R&D innovation is an important indicator of the current market competitiveness of enterprises, and it is regarded as the engine room of high-tech enterprises. While increasing R&D investment, however, effective forecasting of R&D funds is a problem that must be solved. In this paper, based on the information update characteristics of high-tech enterprises' R&D investment and the limitations of the single forecasting model, a combined forecasting model based on grey metabolic GM (1, 1) model and exponential smoothing method is established to study its R&D investment forecasting in high-tech enterprises. The application in the forecasting proves the practicability and effectiveness of the combined forecasting model for improving forecasting accuracy.

High-tech enterprises can make more accurate and reasonable forecasting of their own R&D capital demand, and they can rationally allocate their R&D funds according to the forecasting value, so as to effectively manage R&D costs and improve their economic and social benefits. In addition, high-tech enterprises are susceptible to external factors. While committed to the management of their own R&D fund, they must always pay attention to the social and even global technology investment. In the future, we will use neural networks [8, 45], bilateral matching [46, 47], fuzzy information integration [48], distance measures [49], and other methods to compare and analyze our own research

and development innovation and closely follow the development trend of domestic and foreign competitive markets.

Of course, the prediction results of any model cannot be completely accurate. For a complex system such as enterprise R&D investment that is simultaneously affected by internal and external environments, if a longer-term forecast is to be made, the external economic factors of the enterprise should be considered. In addition, there are many forecasting methods that can be selected, and further research on the selection of single models and the combination of multiple rules on the basis of the characteristics of R&D activities of high-tech enterprises is worthy of research.

## Data Availability

All data included in this study are available from the corresponding author upon request. Most of the data are already in the article.

## Conflicts of Interest

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

## Acknowledgments

This research was supported by the Philosophy and Social Science Foundation of Guangdong Province (no. GD16XYJ26), the National Statistical Science Foundation of China (no. 2019LY29), and the Characteristic and Innovative Foundation for Humanities and Social Sciences of Education Department of Guangdong Province (no. 2018WTSCX041).

## References

[1] J. Yang and L. Yu, "Using a system dynamics approach for comparing forecast methods of short life- cycle products' demand," *Chinese Journal of Management Science*, vol. 20, no. S1, pp. 55–60, 2012.

[2] X. Song and D. Zhang, "Discussion on the control strategy of enterprise R&D cost," *Communication in Finance and Accounting*, vol. 32, pp. 89-90, 2010.

[3] K. Yao, J. Tang, and Y. Jiang, "R&D input/R&D project progress and fluctuation of stock price," *Chinese Journal of Management Science*, vol. 21, no. S1, pp. 205–213, 2013.

[4] R. Ortega-Argilés, M. Piva, L. Potters, and M. Vivarelli, "Is corporate R&D investment in high-tech sectors more effective?" *Contemporary Economic Policy*, vol. 28, no. 3, pp. 353–365, 2010.

[5] Innovation Research Interchange, "2019 R&D trends forecast: results from the innovation research interchange's annual survey," *Research-Technology Management*, vol. 62, no. 2, pp. 21–30, 2019.

[6] Y. Li and L. Zhang, "The current situation and development trend of R&D investment in the pearl river delta region," *Science & Technology and Economy*, vol. 31, no. 5, pp. 66–70, 2018.

[7] D. Wu, "Analysis and forecast on the evolution trend of R&D input intensity in China from global perspective," *Science and Technology Management Research*, vol. 377, no. 3, pp. 9–14, 2017.

[8] J. Li and Z. Hu, "Neural network prediction model for R&D of GEM: BP model or RBF model?" *Science and Technology Management Research*, vol. 37, no. 5, pp. 183–190, 2017.

[9] H. Chen, *Research on R&D Cost Forecast of High Technology Corporate Based on System Dynamic Theory*, Harbin University of Commerce, Harbin, China, 2013.

[10] L. Liang and Y. Xiong, "The status quo of research and development expenses disclosure of listed companies in my country and suggestions," *Finance and Accounting*, vol. 10, pp. 22-23, 2005.

[11] H. Ahumada and M. Cornejo, "Forecasting food prices: the case of corn, soybeans and wheat," *International Journal of Forecasting*, vol. 32, no. 3, pp. 838–848, 2016.

[12] R. Liu, J. Yang, and C. Ruan, "Expected stock return and mixed frequency variance risk premium data," *Journal of Ambient Intelligence and Humanized Computing*, vol. 11, no. 9, pp. 3585–3596, 2020.

[13] M. L. Diop and W. Kengne, "Piecewise autoregression for general integer-valued time series," *Journal of Statistical Planning and Inference*, vol. 211, pp. 271–286, 2021.

[14] M. M. Rounaghi, M. R. Abbaszadeh, and M. Arashi, "Stock price forecasting for companies listed on Tehran stock exchange using multivariate adaptive regression splines model and semi-parametric splines technique," *Physica A: Statistical Mechanics and Its Applications*, vol. 438, pp. 625–633, 2015.

[15] V. Dordonnat, A. Pichavant, and A. Pierrot, "GEFCom2014 probabilistic electric load forecasting using time series and semi-parametric regression models," *International Journal of Forecasting*, vol. 32, no. 3, pp. 1005–1011, 2016.

[16] J. Wang, T. Niu, P. Du, and W. Yang, "Ensemble probabilistic prediction approach for modeling uncertainty in crude oil price," *Applied Soft Computing*, vol. 95, Article ID 106509, 2020.

[17] S. Zeng, X. Peng, T. Baležentis, and D. Streimikiene, "Prioritization of low-carbon suppliers based on Pythagorean fuzzy group decision making with self-confidence level," *Economic Research-Ekonomska Istraživanja*, vol. 32, no. 1, pp. 1073–1087, 2019.

[18] H. Cai, J. Feng, Q. Yang, W. Li, X. Li, and J. Lee, "A virtual metrology method with prediction uncertainty based on Gaussian process for chemical mechanical planarization," *Computers in Industry*, vol. 119, p. 103228, 2020.

[19] D. Luo, S. Zeng, and J. Chen, "A probabilistic linguistic multiple attribute decision making based on a new correlation coefficient method and its application in hospital assessment," *Mathematics*, vol. 8, no. 1–16, p. 340, 2020.

[20] H. Chen, "Review of research documents on enterprise R&D cost forecast," *Economic Research Guide*, vol. 5, pp. 107-108, 2013.

[21] Z. Xiao and G. Zhou, "The application of combining forecasting based on ant colony algorithm to R&D funds in China," *Science of Science and Management of S. & T.*vol. 27, no. 9, pp. 19–21, 2006.

[22] W. Cheng and T. Yao, "Application of an information renewal GM (1, 1) model predicting in coal consumption forecast of China," *Mathematics in Practice and Theory*, vol. 45, no. 16, pp. 93–98, 2015.

[23] Z. Zhou and C. Yin, "Application of gray metabolic model in the prediction of the cotton output in China," *Asian Agricultural Research*, vol. 3, no. 1, pp. 1–6, 2011.

[24] Z. Ceylan, S. Bulkan, and S. Elevli, "Prediction of medical waste generation using SVR, GM (1, 1) and ARIMA models: a

case study for megacity Istanbul," *Journal of Environmental Health Science and Engineering*, vol. 18, pp. 687–697, 2020, doi: 10.1007/s40201-020-00495-8.

[25] S. Huang, W. Wang, C. Zeng et al., "Application of gray metabolic GM (1, 1) model in prediction of annual total yields of Chinese aquatic products," *Asian Agricultural Research*, vol. 5, no. 3, p. 21, 2013.

[26] Y.-C. Lee, C.-H. Wu, and S.-B. Tsai, "Grey system theory and fuzzy time series forecasting for the growth of green electronic materials," *International Journal of Production Research*, vol. 52, no. 10, pp. 2931–2945, 2014.

[27] L. Hua and N. Xie, "Forecasting analysis of Chinese energy consumption and control strategy under policy impact," *Chinese Journal of Management Science*, vol. 22, no. 7, pp. 18–25, 2014.

[28] S. Ding, K. W. Hipel, and Y.-G. Dang, "Forecasting China's electricity consumption using a new grey prediction model," *Energy*, vol. 149, pp. 314–328, 2018.

[29] L. De and X. Xiao, "The application of the gm (1, 1) metabolism model to error data processing of NC machine tools," *International Journal of Plant Engineering & Management*, vol. 5, no. 2, pp. 41–45, 2003.

[30] S. Dai, D. Niu, and Y. Han, "Forecasting of energy-related CO2 emissions in China based on GM (1, 1) and least squares support vector machine optimized by modified shuffled frog leaping algorithm for sustainability," *Sustainability*, vol. 10, no. 4, p. 958, 2018.

[31] Y. X. Luo, "Non-equidistant step by step optimum new information GM (1, 1) and its application," *Systems Engineering—Theory & Practice*, vol. 30, no. 12, pp. 2254–2258, 2010.

[32] J. Deng, "Introduction to grey system theory," *The Journal of Grey System*, vol. 1, no. 1, pp. 1–24, 1989.

[33] H. Xiong and P. Li, "Application of dynamic GM (1, 1 model in predicting the collected amount of municipal solid waste in hubei Province," *Mathematics in Practice and Theory*, vol. 49, no. 1, pp. 9–14, 2019.

[34] D. Akay and M. Atak, "Grey prediction with rolling mechanism for electricity demand forecasting of Turkey," *Energy*, vol. 32, no. 9, pp. 1670–1675, 2007.

[35] E. Ostertagova and O. Ostertag, "Forecasting using simple exponential smoothing method," *Acta Electrotechnica et Informatica*, vol. 12, no. 3, p. 62, 2012.

[36] Q. Xue, F. Mou, and Z. Tu, "Application of combination forecast method to chongqing's GDP prediction," *Journal of Chongqing Technology Business University (Natural Science Education)*, vol. 34, no. 1, pp. 56–63, 2017.

[37] E. S. Gardner and S. Everette, "Exponential smoothing: the state of the art," *Journal of Forecasting*, vol. 4, no. 1, pp. 1–28, 1985.

[38] M. Mao and E. C. Chirwa, "Application of grey model GM (1, 1) to vehicle fatality risk estimation," *Technological Forecasting and Social Change*, vol. 73, no. 5, pp. 588–605, 2006.

[39] S. J. Feng, Y. D. Ma, Z. L. Song, and J. Ying, "Forecasting the energy consumption of China by the grey prediction model," *Energy Sources, Part B: Economics, Planning, and Policy*, vol. 7, no. 4, pp. 376–389, 2012.

[40] M. Hibon and T. Evgeniou, "To combine or not to combine: selecting among forecasts and their combinations," *International Journal of Forecasting*, vol. 21, no. 1, pp. 15–24, 2005.

[41] J. M. Bates and C. W. J. Granger, "The combination of forecasts," *Journal of the Operational Research Society*, vol. 20, no. 4, pp. 451–468, 1969.

[42] D. E. Rapach and J. K. Strauss, "Forecasting US employment growth using forecast combining methods," *Journal of Forecasting*, vol. 27, no. 1, pp. 75–93, 2008.

[43] S. G. Hall and J. Mitchell, "Combining density forecasts," *International Journal of Forecasting*, vol. 23, no. 1, pp. 1–13, 2007.

[44] W. Zheng, D.-H. Lee, and Q. Shi, "Short-term freeway traffic flow prediction: bayesian combined neural network approach," *Journal of Transportation Engineering*, vol. 132, no. 2, pp. 114–121, 2006.

[45] S. Varela-Santos and P. Melin, "A new approach for classifying coronavirus COVID-19 based on its manifestation on chest X-rays using texture features and neural networks," *Information Sciences*, vol. 545, pp. 403–414, 2021.

[46] D. Yu and Z. Xu, "Intuitionistic fuzzy two-sided matching model and its application to personnel-position matching problems," *Journal of the Operational Research Society*, vol. 71, no. 2, pp. 312–321, 2020.

[47] J. W. Hatfield, R. Jagadeesan, and S. D. Kominers, "Matching in networks with bilateral contracts: corrigendum," *American Economic Journal: Microeconomics*, vol. 12, no. 3, pp. 277–285, 2020.

[48] Z. Mu, S. Zeng, and P. Wang, "Novel approach to multi-attribute group decision-making based on interval-valued Pythagorean fuzzy power Maclaurin symmetric mean operator," *Computers & Industrial Engineering*, 2021, In Press, Article ID 107049.

[49] J. Wang, S. Zeng, and C. Zhang, "Single-valued neutrosophic linguistic logarithmic weighted distance measures and their application to supplier selection of fresh aquatic products," *Mathematics*, vol. 8, no. 3, p. 439, 2020.

Hindawi
Mathematical Problems in Engineering
Volume 2021, Article ID 6687229, 12 pages
https://doi.org/10.1155/2021/6687229

*Research Article*

# Study on Spatial Imbalance and Determinants of E-Commerce Development in Zhejiang, China

**Haidong Zhong**,[1] **Jinhui Zhang**,[2] **Shaozhong Zhang**,[3] **and Wen Zheng**[1]

[1]*Digital Industry Research Institute, Zhejiang Wanli University, Ningbo 315100, Zhejiang, China*
[2]*Logistics and E-Commerce School, Zhejiang Wanli University, Ningbo 315100, China*
[3]*College of Information and Intelligence Engineering, Zhejiang Wanli University, Ningbo 315100, China*

Correspondence should be addressed to Wen Zheng; zhengwen0722@163.com

Received 16 November 2020; Revised 23 December 2020; Accepted 6 January 2021; Published 18 January 2021

Academic Editor: Chonghui Zhang



As a world-famous and well-developed e-commerce region, the development of e-commerce in Zhejiang province has always attracted people's wide attention. Based on publicly available e-commerce transaction-related data, basic geographic data, and regional economic and social development data, we use the Gini coefficient to measure the imbalance of e-commerce development in Zhejiang province during 2017–2019. With the help of spatial analyst tools in ArcGIS desktop, the cluster and outlier analysis method is used to study the spatial pattern of e-commerce development in the province at the district or county-level city scale. To explore the causes of spatial aggregation and imbalance of e-commerce in Zhejiang province quantitatively, the paper proposes a geographical weighted regression (GWR) model with 15 economic and social development-related indicators. GWR and ordinary least squares (OLS) analysis indicate that 5 of the 15 selected indicators are highly related to the development of regional e-commerce development in Zhejiang, China.

## 1. Introduction

As the largest e-commerce market in the world, China's e-commerce has maintained rapid development for many years. According to the report on the development of e-commerce in China, China's e-commerce volume reached 34.81 trillion yuan, of which the online retail sales reached 10.63 trillion yuan, which is year-on-year growth of 16.5%. The online retail sales of physical goods reached 8.52 trillion yuan, accounting for 20.7% of the total retail sales of consumer goods. The number of e-commerce employees reached 51,256,500 (http://www.chinanews.com/cj/2020/06-30/9225677.shtml). With the continuous expansion of the application fields and application depth of e-commerce, various regions have introduced policies to promote the development of local e-commerce and improved the service systems such as credit certification, payment guarantee, and brand promotion, training mechanism for scientific and technological talents, and development of logistics distribution. The development of e-commerce has created a lot of job opportunities for urban and rural residents and become one of the important drivers for the sustained economic growth. Therefore, the spatial distribution characteristics of e-commerce can reflect the development of China's economy and information technology to a large extent.

Due to the accumulation of many economic and social development related factors, the spatial distribution of the development level of e-commerce in China is obviously different. From four aspects, the development level, growth potential, application penetration, and support environment, China's E-Commerce Development Index Report (2018) (http://www.cac.gov.cn/2019-05/29/c_1124554997.htm) evaluated the development level of e-commerce in each province in China comprehensively. According to the report, the integration of e-commerce and traditional economy is deepening, and the characteristics of horizontal spatial agglomeration of e-commerce development are

obvious. In 2018, Guangdong province, Zhejiang province, Beijing city, Shanghai city, and Jiangsu province are the top five online retailers at the province (municipality directly under the central government) scale, accounting for 72.3% of China's online retail sales [1]. The overall development of e-commerce at the province scale is basically "strong in the east and weak in the west," and the cluster effect is emerging. The spatial distribution characteristics of e-commerce development also reflect the imbalance of regional economic and social development to a large extent.

Zhejiang province ranks the second in online retail sales of China. Meanwhile, Hangzhou city (the provincial capital city of Zhejiang province) is the home to the world's super e-commerce giant Alibaba Group (https://www.alibaba.com/). Currently, among Chinese e-commerce platforms, TMALL, owned by Alibaba Group, takes the first place in the online retail market with a share of 57.7%, which has an obvious driving effect on the traditional economy. Zhejiang province supports the development of new forms of business, such as Alibaba's new retail store Super Species, NetEase Koala, and NetEase experience stores, and helps Hangzhou build a new retail benchmark city. By promoting county-level e-commerce development vigorously, Zhejiang province has built 1,253 e-commerce villages, over 4,500 county-level e-commerce public service centres and village-level e-commerce service stations. Analysing the spatial distribution pattern and its influencing factors of the level of e-commerce development from the perspective of county region is meaningful for promoting the development of e-commerce in county regions.

Under the background of the in-depth development of the Internet, regional development is gradually integrated, and the energy radiation of the regional development of the central city is gradually transferred to the space of the district and county level, and the county region has become an important node of the network development. Analysing the spatial connection of county economic and social development factors can better adapt to the demand change of modern service industry development in the new era.

Therefore, this paper explores the imbalance of e-commerce development at provincial and municipal scales. Meanwhile, it studies the relationship between economic and social development and the spatial pattern of e-commerce development at the district or county-level city scale. It is hoped that the study can provide theoretical support and scientific basis for the collaborative development of e-commerce and other related industries in the exploration at the county level. The main contributions of this paper lie in the following three aspects:

(1) We measure the imbalance of regional e-commerce development in Zhejiang, China quantitatively. Referring to the internationally widely used measures of income gap between residents of countries or regions, we calculate the Gini coefficient based on online retail sales at both city scale and district or county-level city scale.

(2) We employ a geographical perspective to study the spatial pattern of e-commerce development in Zhejiang, China. Based on the publicly available e-commerce development data and basic geographic information data, we use GIS technology and spatial statistical analysis tools in ArcGIS software to investigate the spatial aggregation pattern of online retail sales at the district or county-level city scale.

(3) We attempt to explore the reasons for the spatial distribution of online retail sales in Zhejiang, China. Referring to the existing research, we establish a geographical weighted regression (GWR) model to explain the relationship between regional e-commerce development and economic and social development. To verify the fitness of the proposed model, we use spatial autocorrelation analysis, geographical weighted regression analysis, etc., to analyse quantitatively with publicly available economic development related data, such as GDP, disposable fiscal income, and income of urban residents.

## 2. Literature Review

In recent years, booming e-commerce has attracted considerable scholars' attention to the impact of this kind of online transaction pattern. They find that the need for retail stores and facilities in commercial cores is somewhat challenged in the e-commerce scenario: the need for trips to bustling downtown shopping locations decreases rapidly, and supermarkets, shops, etc. are increasingly replaced by warehousing, retail outlets, or other types of storage facilities far from downtown cities [2]. To accomplish transactions based on virtual cyberspace, it requires a reconstruction of upstream, midstream, and downstream industrial supply chains and logistics systems from consumers' point of view [3, 4]. Accompanied by the rapidly growing global e-commerce market, cities all over the world are experiencing an operational change from physicality to virtuality, and the pattern of urban planning, enterprise production, and sales and economic and social development modes are also changing accordingly [5].

As an important part of information economy, e-commerce is of special significance in China and has played a very important role in improving people's lives. In addition to the progress made in the use of e-commerce among individuals, it is also important and meaningful to understand the extent and importance of differences between geographical regions [6, 7]. There are many indicators published to measure the level of access to and use of information and communication technologies (ICTs), such as the ICT Development Index released by the International Telecommunication Union [8], the Digital Economy and Society Index [9] announced by the European Commission in 2017, and the Networked Readiness Index [10] reported by the World Economic Forum in 2017. However, all these indicators were developed at the country scale.

As early as 2002, the spatial econometric method has been used to explore the spread of e-commerce in various countries and the impact of the policy environment on the development path of e-commerce [11]. Many studies focus

on regional e-commerce competitiveness and the impact of e-commerce on local economic development based on indices irregularly released by many well-known e-commerce enterprises such as Alibaba Group (https://www.alibabagroup.com/en/global/home) and Jingdong Group (https://corporate.jd.com/home). AliResearch (http://www.aliresearch.com/) was the first research institute to publish a nationwide Alibaba e-commerce Development Index (eEDI) using the number of sellers, the number of buyers, and e-commerce transaction volume on Taobao (http://www.taobao.com/) and TMall (http://www.tmall.com/) platforms in 2015. eEDI is the weighted average of the Internet business index and the online shopping index. The value of eEDI ranges from 0 to 100, and the higher the value, the higher the development level of regional e-commerce. In addition, there are some well-known universities and research institutions such as the National Engineering Laboratory for E-commerce Transaction Technology of Tsinghua University and the Chinese Academy of Social Sciences, who jointly publish the China E-Commerce Development Index Report (http://necc.nufe.edu.cn/) annually. There are a great many existing studies that use eEDI or the China e-commerce Development Index (including size index, growth index, and penetration index) and support index to evaluate the development level of regional e-commerce [12–14]. However, only the calculation methods of eEDI and the China E-Commerce Development Index are disclosed, and the raw data that generate them are completely unavailable. The authority of these regional e-commerce development inequality or imbalance evaluation indicators also needs to be further improved.

There is a line of research on the factors affecting the development of regional e-commerce at the provincial level and city level. Studies show there is an obvious spatial aggregation in the development of e-commerce in China: the level of e-commerce in the eastern coastal areas is higher than that in the western inland areas; significant high-high spatial aggregation areas mainly exist in Zhejiang province, Jiangsu province, Fujian province, etc.; significant low-low spatial aggregation areas are mainly located in Tibet Autonomous Region, Gansu province, Qinghai province, and so on [15–17]. The spatial difference of the e-commerce development level between provinces and regions does not change obviously in China, and it shows a development trend of gradient reduction and staggered distribution from the east coast to the interior [17]. In addition, the least square method, spatial lag model, spatial error model, and geographic regression analysis based on eEDI show that determinants of e-commerce development at the county level include the balance of savings deposits of urban and rural residents, urbanisation rate, per capita GDP, proportion of nonagricultural industries, disposable income of urban residents, mobile phone users, Internet users, and fixed-line phone users [16, 18–20].

Overall, the research on the spatial distribution pattern of e-commerce is mostly conducted at the national or provincial scales, but it lacks the scale of city and county or county-level city studies in China. Although many studies select the number of consumer-to-consumer (C2C) stores [21] or the number of Taobao towns and Taobao villages [22] to establish a regional e-commerce development level evaluation system, they use only C2C data from Alibaba and much of the data is not publicly available or not regularly released.

In this paper, we attempt to further analyse the spatial distribution pattern of the regional e-commerce development level in Zhejiang, China, from the perspective of county or district region scale based on publicly available and regularly release statistical yearbook data. Furthermore, we focus on applying spatial statistical analysis methods to explore the relationship between the regional e-commerce development pattern and some economic and social development factors. We hope the research results can provide some meaningful references to the regional policy making process to promote e-commerce and social and economic development.

## 3. Study Area and Data Sources

*3.1. Study Area.* Zhejiang province is located in the core area of the Yangtze River Delta Economic Zone in China, and it is bounded by east longitude 118°01′–123°10′ and north latitude 27°02′–31°11′, with an area of 105,500 square kilometres (as shown in Figure 1). By the end of 2019, the province had 11 cities under its jurisdiction (including two subprovincial cities), 20 county-level cities, 32 counties, one autonomous county, and 37 municipal districts. Zhejiang province slopes from southwest to northeast with complicated terrain, it has a humid monsoon climate, and natural conditions are superior. The total resident population of the province was 58.5 million at the end of 2019 and 1.13 million more than the previous year.

Zhejiang province is one of the provinces with the smallest economic disparities in China, with Hangzhou city, Ningbo city, Shaoxing city, and Wenzhou city as its four economic pillars. The world's largest port, the Ningbo-Zhoushan Port, is located in the province. In 2019, gross domestic product (GDP) of Zhejiang province reached 6,235.2 billion yuan ($903.9 billion), 6.8 percent higher than the previous year at comparable prices, and the per capita GDP was 107,624 yuan ($15,601). In October 2019, it was selected as the National Digital Economy Innovation and Development Pilot Zone.

*3.2. Data Source.* All the e-commerce transaction-related data used in this paper come from the Department of Commerce of Zhejiang province (http://www.zcom.gov.cn/). The data contain transaction volumes of both e-commerce and cross-border e-commerce in each municipal district (county or county-level city) in Zhejiang province from 2017 to 2019. In addition, the data are published monthly and publicly available. After simple operations such as splitting and merging, quarterly data and yearly data can be obtained correspondingly.

Indicators used to explore e-commerce and regional economic and social development, such as GDP, per capita GDP, local fiscal revenue, and transportation expenditure,


Figure 1: The geographical location of Zhejiang province.

are collected from the 2019 statistical yearbook of 11 cities in Zhejiang province. Basic geographic data of the province, including area, city, and county level or county boundaries, are gathered from national basic data and public data-sharing websites.

## 4. Research Methods

*4.1. Spatial Imbalance Analysis.* The paper employs the Gini coefficient [23], which is a common index used to measure the income gap of residents in a country or region, to analyse the spatial imbalance of e-commerce development at the county level of Zhejiang province. The specific method is shown in the following formula:

$$\text{Gin} = \sum p_k^2 \lambda_k G_k + \frac{1}{2} \sum p_k p_h \left|\lambda_k - \lambda_h\right| + R, \tag{1}$$

where Gin is the overall Gini coefficient that is adopted to measure the spatial imbalance degree of e-commerce development at the county scale, $p_k$ represents the share of group $k$, $\lambda_k$ represents the ratio of the average regional e-commerce transaction volume of group $k$ to the entire sample value, and $G_k$ is the Gini coefficient of group $k$. The first term on the right of the formula reflects the intragroup gap or disequilibrium; the second term reflects the intergroup disequilibrium; and $R$ is the remaining term, reflecting the interaction caused by the overlap between different groups.

*4.2. LISA Analysis.* Local indicators of spatial association (LISA) analysis is commonly used to explore the local index of spatial relations. In a LISA graph, clustering is divided into four cases, high-high, low-high, low-low, and high-low, each of which identifies an area and its relationship to its neighbours. High-high indicates that a high-level area is surrounded by other high levels; low-high indicates that a low-level area is surrounded by other high-level areas; low-low indicates that a low-level area is surrounded by other low-level areas; and high-low indicates that a high-level area is surrounded by other low-level areas [24].

*4.3. Spatial Autocorrelation Analysis.* Spatial autocorrelation analysis explores spatial dependency effects between different variables with a set of statistical methods in conventional linear statistical models. It originates from biometrics, and it has become one of the basic methods of theoretical geography [25]. Moran's I is one of the most popular methods to measure the overall pattern across a geographic landscape and degree of interdependence between different locations [26, 27]. Moran's I measures spatial autocorrelation by both factor location and factor value, which can be applied to evaluate whether the expressed patterns are clustering, discrete, or random. It can be calculated by the following formula:

$$I = \frac{n}{S_0} \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} w_{i,j} z_i z_j}{\sum_{i=1}^{n} z_i^2}, \tag{2}$$

where $z_i$ is the dispersion value between the observed value and the value of expectation at location $i$, $w_{i,j}$ is the proximity relation value of locations $i$ and $j$, $n$ is the sum of all observation locations, $S_0$ is the sum of all proximity relation values, and $z_i$ can be calculated with the following formula:

$$z_i = \frac{I - E[I]}{\sqrt{[V]}}, \tag{3}$$

where $E[I] = (1/(n-1))$ and $[V] = E[I^2] - E[I]^2$.

Moran's I > 0 represents positive spatial correlation, and the greater the value, the more obvious the spatial correlation. Moran's I < 0 represents negative spatial correlation, and the smaller the value, the greater the spatial difference. Otherwise, Moran's I = 0, and the space is random. With the help of the tool kit for spatial statistical analysis in ArcGIS software, spatial pattern of a set of factors and related

attributes can be measured by calculating the value of Moran's I index, the $z$-score, and the $p$ value at the same time. The $z$-score is used to identify the calculated spatial pattern, and the $p$ value is an area approximation based on a curve of a known distribution. Very high or very low (negative) $z$-scores appear at both ends of the normal distribution, which are associated with very small $p$ values. If the $z$-score is smaller than −1.96 and or greater than 1.96, the calculated statistical result is basically reliable at the confidence level of 0.05. If the $z$-score is smaller than −1.96, the calculated distribution is discrete. Otherwise, the calculated distribution can be viewed as random [28].

*4.4. GWR Analysis.* GWR is a local form of linear regression used to model spatial variation relationships. It extends the general linear regression model. In GWR, the regression coefficient of a specific location is not assumed constant as estimated by using all the information, but it is a variable obtained by using the subsample information of the observed values of the neighbourhood region. By adapting the regression equation to each element in the dataset, GWR provides a local model for the variables or processes you are trying to understand/predict. The GWR model can be represented as follows [29]:

$$y_i = \beta_0(u_i, v_i) + \sum_{i=1}^{k} \beta_i(u_i, v_i) x_{ik} + \varepsilon_i, \tag{4}$$

where $\beta_i$ represents a parameter vector that needs to be estimated, $(u_i, v_i)$ is geographical coordinate value at location $i$, $\beta_i$ is a $k+1$ function of $(u_i, v_i)$, and $\varepsilon_i$ is the random error at location $i$ and satisfies the assumptions of zero mean, homoscedasticity, and independence. In GWR, each region has a corresponding estimation function, whose logarithmic likelihood function can be expressed as [30, 31]

$$\begin{aligned} \operatorname{Log} L &= L[\beta_0(u,v), \beta_1(u,v), \ldots, \beta_k(u,v) | M] \\ &= -\frac{1}{2\sigma^2} \sum_{i=1}^{n} [y_i - \beta_0(u_i, v_i)] - \sum_{j=1}^{k} \beta_k(u_i, v_i) x_i + \alpha, \end{aligned} \tag{5}$$

where $\alpha$ is a constant, $M = [y_i, x_{ij}, (ui, vi), i = 1, 2, 3, \ldots, n, j = 1, 2, 3, \ldots, k]$, and $y_i$ is an approximate value of $\beta_k(u_i, v_i)$.

## 5. Imbalance and Spatial Pattern of E-Commerce Development in Zhejiang Province

*5.1. Analysis on Unbalanced Development of E-Commerce.* Statistics show that the development level of e-commerce varies greatly among cities in Zhejiang province. As shown in Figure 2, the development of Hangzhou city is far ahead compared to other cities, whether measured in terms of online sales or online spending from 2017 to 2019. In terms of online retail sales, the gap between Hangzhou city and other cities is widening during 2017–2019. In 2017, Hangzhou city accounts for 32.3% of online sales in Zhejiang province, and the proportion rises to 36.8% in 2019. Also, Figure 2 indicates that the development level of e-commerce in Quzhou city, Zhoushan city, and Lishui city is relatively low.

To draw a clear profile of the imbalance of e-commerce development in recent years, we use the Gini coefficient and its decomposition method to measure the degree of imbalance of online retail sales in cities of Zhejiang province. Due to data statistics and public data access restrictions, only 3 years' online retail sales of cities in the province are collected. However, these data reveal a very unbalanced state of e-commerce development at the provincial scale from 2017 to 2019 (as shown in Figure 3). Even at the lowest point in 2018, the Gini coefficient was 0.491, which indicates a large e-commerce development gap, according to international practice. In 2019, this value reached 0.525, which represented a wide gap of e-commerce development in Zhejiang province.

To further study the balanced development status of e-commerce at the city level, we collect online retail sales data of all districts and county-level cities in Zhejiang province from 2017 to 2019. The Gini coefficient of online retail sales in cities of the province (as shown in Figure 4) reveals the following. (1) The unbalanced development of e-commerce in each city is quite obvious. Seven out of 11 cities have Gini coefficient of online retail sales greater than 0.4, which indicates a large e-commerce development gap between districts and county-level cities in the seven cities. Only two cities (Quzhou city and Shaoxing city) have Gini coefficient of online retail sales between 0.2 and 0.3, which shows an acceptable average level of e-commerce development in the two cities. (2) The imbalance of e-commerce development in cities with high online retail sales is evident. Taking Hangzhou city and Jinhua city as examples, online retail sales in these two cities ranked first and second, respectively, in Zhejiang province nearly every year from 2017 to 2019 (as shown in Figure 2). However, in Hangzhou city, the 3-year Gini coefficients of online retail sales are 0.469, 0.474, and 0.481, respectively, during 2017–2019. In Jinhua city, the 3-year Gini coefficients of online retail sales are 0.644, 0.633, and 0.616, respectively, during the same period. According to the international convention on the Gini coefficient, when the Gini coefficient of online retail sales reaches more than 0.5, it represents that the level of e-commerce development varies greatly in these cities.

*5.2. Analysis on Spatial Pattern of E-Commerce Development.* Online retail sales data of 89 districts and county-level cities during 2017–2019 in Zhejiang province are imported into ArcGIS Desktop to analyse the spatial aggregation pattern of the e-commerce development level at a more specific spatial scale. The results of 2017, 2018, and 2019 are basically the same, and thus, we only display the result of 2019 in 5-bit colour ramp display mode (as shown in Figure 5).

As we can find from Figure 5, spatial aggregation characteristics of e-commerce development are highly significant. However, high-high clusters indicate that well-

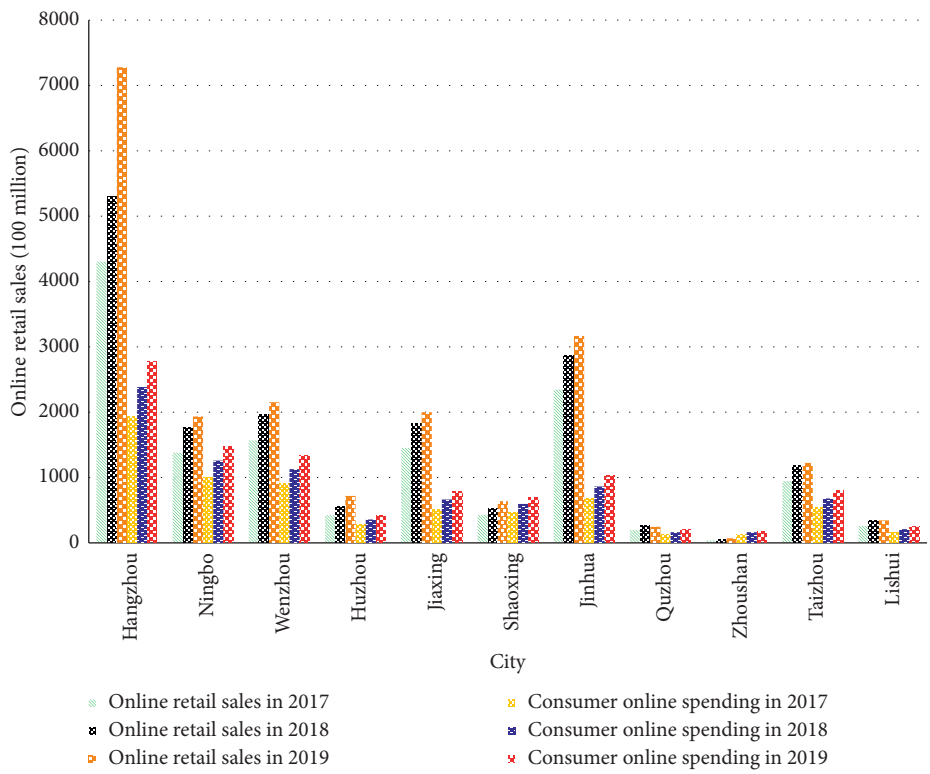



Figure 2: The overall e-commerce trade profile of Zhejiang province.

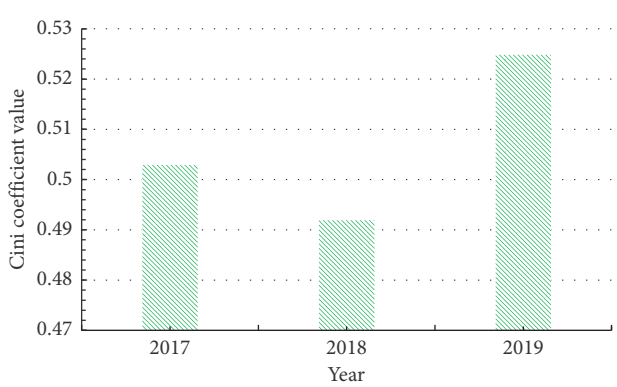



Figure 3: The Gini coefficient of online retail sales in Zhejiang province during 2017–2019.

developed e-commerce districts or county-level cities exist in the middle-north area of Zhejiang province. Meanwhile, there are some low-low clusters which indicate that underdeveloped e-commerce districts or county-level cities exist in the middle-north area of the province. High-high clusters are mainly located in some districts and county-level cities of Hangzhou city, Jinhua city, Huzhou city, and Jiaxing city. Low-low clusters are mainly located in some districts and county-level cities of Quzhou city, Lishui city, and Wenzhou city.

To validate the spatial aggregation of e-commerce development in Zhejiang province quantitatively, we adapt the cluster and outlier analysis (Anselin Local Moran's I) with the help of spatial analyst tools in ArcGIS Desktop. The online retail sales data of 89 districts and county-level cities (with INVERSE_DISTANCE as the conceptualisation and EUCLIDEAN as the distance method) and the spatial autocorrelation analysis report are shown in Figure 6. According to Anselin Local Moran's I analysis result interpretation specification, the $z$-score of 11.3606 implies less than 1% likelihood that this clustered pattern could be the result of random chance.

## 6. Influence Factors of Spatially Imbalanced E-Commerce Development in Zhejiang Province

It is generally believed that e-commerce development is influenced to a large extent by many regional economic and social development-related factors. To explore the specific influencing factors and spatial differences of the development of e-commerce in Zhejiang province, we use the GWR and the OLS methods to examine the relationship between the level of economic, social, and technological development and online retail sales at the district or county-level city scale.

*6.1. Influence Factors and Model Building.* We select online retail sales (OnlineRetailSale) as the dependent variable, and gross domestic product (GDP), per capita gross domestic product (PerGDP), population of permanent residents (Population), administrative area (Area), local financial revenue (LocalRevenue), local transportation expenditure (TransExpenditure), postal revenue (PostalRevenue),

number of Internet users (InternetUsers), retail sales of consumer goods (RetailSales), urban per capita disposable income (UrbanPerCapitaIncome), rural per capita disposable income (RuralPerCapitaIncome), gross product of transportation, warehousing, and postal services (TransportationWarehousingPostalGross), electricity use of transportation, warehousing, and postal services (TransportationWarehousingPostalElectricity), gross product of wholesale and retail (WholesaleRetail), and gross product of the tertiary industry (TertiaryIndustry) as explanatory variables to build the following GWR model:

$$\begin{aligned} y_i = {} & \beta_0(u_i, v_i) + \sum_{j=1}^{k} \beta_1(u_i, v_i) x_{ij}(\text{GDP}) + \sum_{j=1}^{k} \beta_2(u_i, v_i) x_{ij}(\text{PerGDP}) \\ & + \sum_{j=1}^{k} \beta_3(u_i, v_i) x_{ij}(\text{Population}) + \sum_{j=1}^{k} \beta_4(u_i, v_i) x_{ij}(\text{Area}) \\ & + \sum_{j=1}^{k} \beta_5(u_i, v_i) x_{ij}(\text{Local Revenue}) + \sum_{j=1}^{k} \beta_6(u_i, v_i) x_{ij}(\text{TransExpenditure}) \\ & + \sum_{j=1}^{k} \beta_7(u_i, v_i) x_{ij}(\text{PostalRevenue}) + \sum_{j=1}^{k} \beta_8(u_i, v_i) x_{ij}(\text{InternetUsers}) \\ & + \sum_{j=1}^{k} \beta_9(u_i, v_i) x_{ij}(\text{RetailSales}) + \sum_{j=1}^{k} \beta_{10}(u_i, v_i) x_{ij}(\text{UrbanPerCapitaIncome}) \\ & + \sum_{j=1}^{k} \beta_{11}(u_i, v_i) x_{ij}(\text{RuralPerCapitaIncome}) + \sum_{j=1}^{k} \beta_{12}(u_i, v_i) x_{ij}(\text{TransportationWarehousingPostalGross}) \\ & + \sum_{j=1}^{k} \beta_{13}(u_i, v_i) x_{ij}(\text{TransportationWarehousingPostalElectricity}) + \sum_{j=1}^{k} \beta_{14}(u_i, v_i) x_{ij}(\text{WholesaleRetail}) \\ & + \sum_{j=1}^{k} \beta_{15}(u_i, v_i) x_{ij}(\text{TertiaryIndustry}) + \varepsilon_i, \quad i = 1, 2, 3, \ldots, n. \end{aligned} \tag{6}$$

*6.2. Regression Result Analysis.* Table 1 reports the results of the GWR analysis in ArcGIS Desktop 10.7. In the table, the adjusted $R^2$ value is 0.789332616, which indicates that the established GWR model can explain about 78.9% of the spatial distribution of e-commerce development in Zhejiang province. Also, the adjusted $R^2$ value indicates a strong spatial autocorrelation of e-commerce development with the selected economic and social development related factors (OnlineRetailSale, GDP, PerGDP, Population, Area, LocalRevenue, TransExpenditure, PostalRevenue, InternetUsers, RetailSales, UrbanPerCapitaIncome, RuralPerCapitaIncome, TransportationWarehousingPostalGross TransportationWarehousing, PostalElectricity, WholesaleRetail, and TertiaryIndustry) at the district or county-level city scale.

In addition, the OLS model is used for multivariate stepwise regression analysis to explore a strong correlation between some variables and eliminate the influence of multicollinearity. The optimal model is selected according to the fitting effect, and the results are shown in Tables 2 and 3. In the two tables, an asterisk beside a number is used to label a statistically significant $p$ value ($p < 0.01$).

In Table 2, we find that six variables, LocalRevenue, PostalRevenue, InternetUsers, RetailSales, TransportationWarehousingPostalGross, and TertiaryIndustry, are closely related to the dependent variable. Among these variables, TertiaryIndustry is negatively related to the dependent variable OnlineRetailSale. This may be because TertiaryIndustry (gross product of the tertiary industry) includes the gross product from many industries, such as transportation, warehousing, and postal services, information transmission, computer services and software industry, wholesale and retail trade, accommodation and catering industry, finance, real estate, leasing, and business services, scientific research, technical services and geological survey, water conservancy, environment, and public facilities management, residents' services and other services, education, health, social security and social welfare, culture, sports and entertainment, public administration, and social organisations, international organisations, and other industries.

Internet users, logistics, express delivery, and warehousing are important parts in the development of e-commerce, and therefore, it is consistent with our common sense that postal revenue (PostalRevenue), number of Internet users (InternetUsers), gross product of transportation, warehousing, and postal services

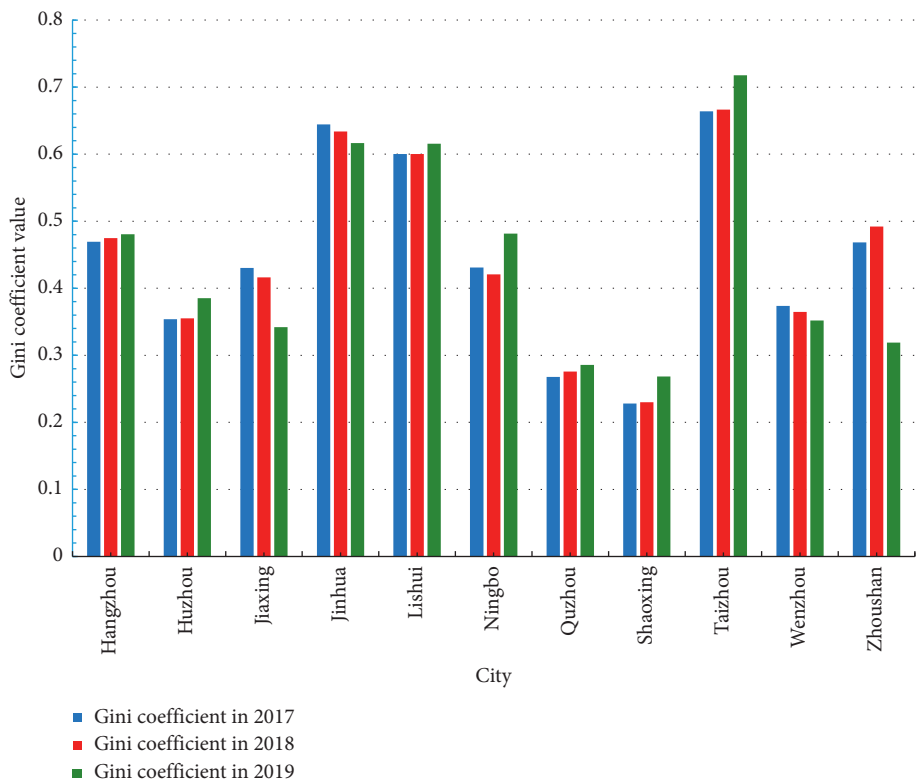



Figure 4: The Gini coefficient of online retail sales in cities of Zhejiang province.

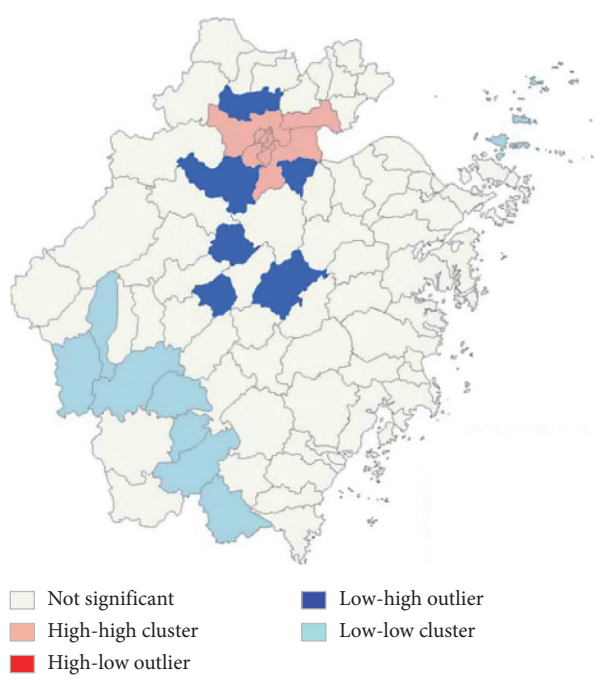



Figure 5: Spatial aggregation analysis of online retail sales at the district or county-level city scale.

(TransportationWarehousingPostalGross) are positively related to OnlineRetailSale.

E-commerce-related enterprises account for a large proportion of all enterprises in Zhejiang province, and they pay a large amount of taxes and fees to the government. This may be the reason why local financial revenue (LocalRevenue) is positively related to OnlineRetailSale. E-commerce in Zhejiang is well developed, and most residents have formed the habit of online shopping which accounts for a large proportion in their daily consumption.

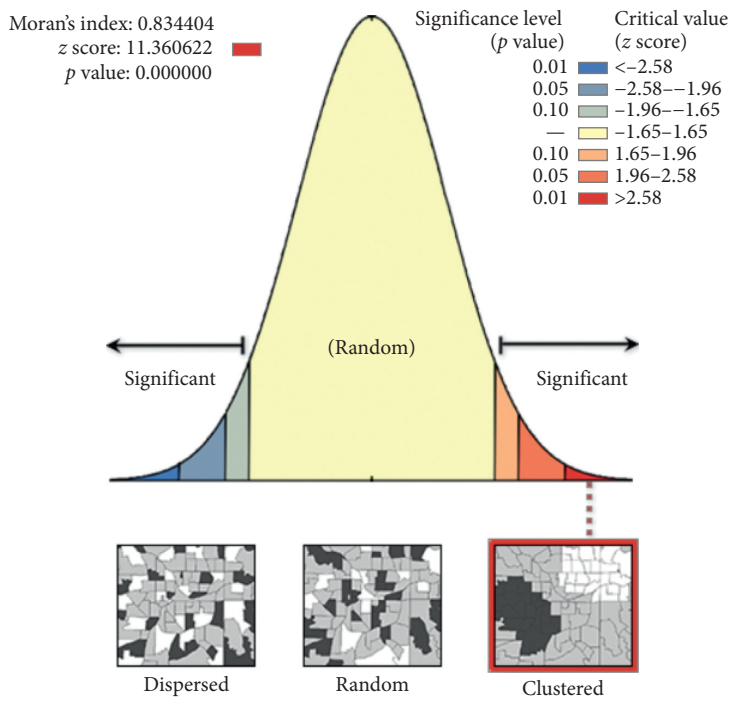



Figure 6: Spatial autocorrelation analysis report of online retail sales at the district or county-level city scale.

Table 1: GWR analysis result with ArcGIS spatial relationships tool.

| Object ID | Varname | Variable | Definition |
|---|---|---|---|
| 1 | Bandwidth | 1.652645326 | |
| 2 | Residual squares | $2.64019E+14$ | |
| 3 | Effective number | 20.43252159 | |
| 4 | Sigma | 1962270.415 | |
| 5 | AICc | 2851.544042 | |
| 6 | $R^2$ | 0.789332616 | |
| 7 | $R^2$ adjusted | 0.761287222 | |
| 8 | Dependent field | 0 | OnlineRetailSale |
| 9 | Explanatory field | 1 | GDP |
| 10 | Explanatory field | 2 | PerGDP |
| 11 | Explanatory field | 3 | Population |
| 12 | Explanatory field | 4 | Area |
| 13 | Explanatory field | 5 | LocalRevenue |
| 14 | Explanatory field | 6 | TransExpenditure |
| 15 | Explanatory field | 7 | PostalRevenue |
| 16 | Explanatory field | 8 | InternetUsers |
| 17 | Explanatory field | 9 | RetailSales |
| 18 | Explanatory field | 10 | UrbanPerCapitalIncome |
| 19 | Explanatory field | 11 | RuralPerCapitalIncome |
| 20 | Explanatory field | 12 | TransportationWarehousingPostalGross |
| 21 | Explanatory field | 13 | TransportationWarehousingPostalElectricity |
| 22 | Explanatory field | 14 | WholesaleRetail |
| 23 | Explanatory field | 15 | TertiaryIndustry |

This may explain why retail sales of consumer goods (RetailSales) is positively related to OnlineRetailSale.

Table 3 indicates the overall reliability of the OLS analysis result. We do not have multiple models to check the smallest AICc information criterion model. However, the rather high adjusted $R^2$ value of 0.7501227 shows a good performance of our GWR model. In addition, the model established in the paper has good fitness according to the evaluation criteria of the joint $F$-statistic, joint Wald statistic, Koenker (BP) statistic, and Jarque–Bera statistic. The residual vs. predicted scatterplot has little structure, and it looks random (as shown in Figure 7). According to the rules for interpretation of OLS analysis results (https://desktop.arcgis.com/zh-cn/arcmap/10.3/tools/spatial-statistics-toolbox/interpreting-ols-results.htm), it indicates a valuable fitness of the proposed e-commerce development spatial

Table 2: Summary of OLS results.

| Variable | Coefficient | StdError | t-statistic | Probability | Robust_SE | Robust_t | Robust_Pr | VIF |
|---|---|---|---|---|---|---|---|---|
| Intercept | 1634599.5 | 2381296 | 0.68643 | 0.494613 | 1764610.88 | 0.92632 | 0.357325 | — |
| GDP | 0.280109 | 0.24274 | 1.15393 | 0.252291 | 0.313456 | 0.89361 | 0.37446 | 23.62808 |
| PerGDP | 3.203353 | 10.797 | 0.29669 | 0.76755 | 5.963631 | 0.53715 | 0.592802 | 6.151398 |
| Population | −760.6725 | 19370 | −0.03927 | 0.968779 | 22906.3123 | −0.03321 | 0.973596 | 9.036138 |
| Area | −652.6249 | 464.224 | −1.40584 | 0.164018 | 512.886693 | −1.27245 | 0.207252 | 2.505199 |
| LocalRevenue | 0.085829 | 0.59704 | 0.14376 | 0.036086* | 0.889384 | 0.0965 | 0.023381* | 6.103148 |
| TransExpenditure | 7.124505 | 14.3365 | 0.49695 | 0.620723 | 18.430816 | 0.38655 | 0.700216 | 2.497873 |
| PostalRevenue | 6.475981 | 3.27634 | 1.97659 | 0.031866* | 5.105642 | 1.2684 | 0.0208688* | 2.057148 |
| InternetUsers | 41532.433 | 20297.3 | 2.0462 | 0.044338* | 12422.2993 | 3.34338 | 0.001313* | 3.072969 |
| RetailSales | 0.559669 | 0.28557 | 1.95987 | 0.02383* | 0.259745 | 2.15469 | 0.034483* | 6.76502 |
| UrbanPerCapitalIncome | 47.106312 | 70.0908 | 0.67208 | 0.503656 | 54.03746 | 0.87173 | 0.0386206* | 7.373611 |
| RuralPerCapitalIncome | −187.5783 | 98.2379 | −1.90943 | 0.070138 | 75.432815 | −2.48669 | 0.015177* | 6.453628 |
| TransportationWarehousingPostalGross | 0.30451 | 0.5212 | −0.58424 | 0.016086* | 0.2266 | 1.34382 | 0.0183172* | 1.127603 |
| TransportationWarehousingPostalElectricity | 21.390363 | 27.1108 | 0.789 | 0.432664 | 31.238967 | 0.68473 | 0.495678 | 1.178223 |
| WholesaleRetail | 0.04045 | 0.21772 | 0.18579 | 0.853121 | 0.243053 | 0.16643 | 0.868281 | 1.460752 |
| TertiaryIndustry | −0.579123 | 0.23344 | −2.48085 | 0.015408* | 0.252954 | −2.28944 | 0.024943* | 5.212943 |

Table 3: OLS diagnostics report.

| Input features | | Dependent variable | OnlineRetailSale |
|---|---|---|---|
| Number of observations | 89 | Akaike's information criterion (AICc) | 2876.63345 |
| Multiple $R$-squared | 0.586245 | Adjusted $R$-squared | 0.7501227 |
| Joint $F$-statistic | 6.89554 | Prob (>$F$), (15, 73) degrees of freedom | 0.000000* |
| Joint Wald statistic | 207.832259 | Prob (>chi-squared), (15) degrees of freedom | 0.000000* |
| Koenker (BP) statistic | 34.123774 | Prob (>chi-squared), (15) degrees of freedom | 0.003272* |
| Jarque–Bera statistic | 204.382134 | Prob (>chi-squared), (2) degrees of freedom | 0.000000* |

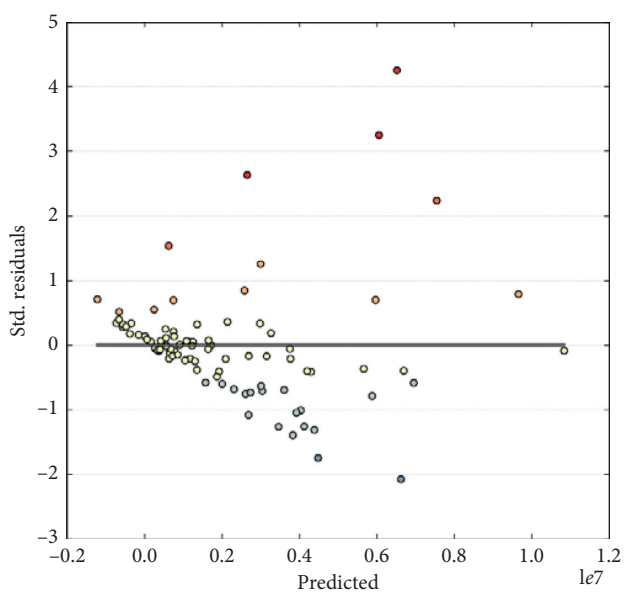



Figure 7: Residual vs. predicted plot.

distribution influencing the GWR model in the previous section.

## 7. Conclusions

A number of studies have shown that the rapid and widespread application of e-commerce is shaping the new business model and order in China, which has a profound influence on the industrial spatial organisation. Also, the unbalanced development of e-commerce in China is very prominent. As one of the earliest and fastest provinces to develop e-commerce in China, Zhejiang has maintained the momentum of rapid growth in recent years. Undoubtedly, the rapid development of e-commerce has promoted the regional economic transformation and development in the province to a large extent. In this paper, we focus on the empirical study of spatial imbalance and the impact factors of regional e-commerce development in Zhejiang, China. Based on the existing theoretical analysis framework, we collect three types of publicly available datasets to analyse quantitatively. The main results coming from our study lie in the following three aspects:

(1) The Gini coefficient based analysis shows that the unbalanced development of e-commerce in Zhejiang province is very serious during 2017–2019. Generally, the lowest Gini coefficient was 0.491 in 2018, and it reached 0.525 in 2019, which indicates a large e-commerce development gap, according to international practice. At the city level, online retail sales of the provincial capital city of Zhejiang province, Hangzhou city, and Jinhua city rank first and second, respectively. The online retail sales of Zhoushan city and Lishui city are relatively low. However, city-level analysis shows that the unbalanced development of e-commerce in cities with large online retail sales is quite serious. Taking Jinhua city as an example, Gini coefficients of online

retail sales were 0.644, 0.633, and 0.616, respectively, from 2017 to 2019.

(2) Quantitative analysis shows that the spatial aggregation of e-commerce development is highly significant. With the help of spatial analyst tools in ArcGIS Desktop, the spatial pattern of e-commerce development at the district or county-level city scale is explored. And, we find high-high clusters which indicate that well-developed e-commerce districts or county-level cities exist in the middle-north area of Zhejiang province. Specifically, these clusters are mainly located in some districts and county-level cities of Hangzhou city, Jinhua city, Huzhou city, and Jiaxing city. Meanwhile, low-low clusters indicate that underdeveloped e-commerce districts or county-level cities exist in the middle-north area of the province. These clusters are mainly located in some districts and county-level cities of Quzhou city, Lishui city, and Wenzhou city. The spatial quantitative analysis validated by Anselin Local Moran's I shows more than 99% likelihood that the clustered pattern of Zhejiang province is dependable.

(3) Geostatistical analysis shows that some economic and social development-related factors can explain the spatial pattern of e-commerce in Zhejiang province. We collect 15 economic and social development indicators of each district or county-level city as explanatory variables and online retail sales as the dependent variable to establish a GWR model. GWR analysis and OLS analysis indicate that the established model can explain about 78.9% of the spatial pattern of e-commerce development. We find that local financial revenue, postal revenue, number of Internet users, retail sales of consumer goods, gross product of transportation, warehousing, and postal services, and gross product of the tertiary industry are highly correlated with online retail sales. Specifically, local financial revenue, postal revenue, number of Internet users, retail sales of consumer goods, and gross product of transportation, warehousing, and postal services are highly positively correlated with online retail sales. Gross product of the tertiary industry and online retail sales are highly negatively correlated.

China is the largest e-commerce market in the world and has maintained rapid development in recent years. The development of e-commerce in China has its own special reasons that are closely related to population, geographical environment, traffic conditions, and economic and social development factors. Our findings come from the study of Zhejiang province that give birth to the world's super e-commerce giant Alibaba Group. The statistical index system of regional e-commerce and economic and social development in other countries may vary from China. However, the development of e-commerce is inevitably affected by regional economic development and geographical pattern. Therefore, we believe the findings of the paper can also give a good reference for other countries or regions to improve the development of e-commerce.

Due to data collection and availability reasons, we just collect 15 social and economic related indictors to explore the relationship between them and the spatial aggregation of e-commerce development in Zhejiang province. There may be many other factors that impact the spatial distribution of e-commerce development in the province. This is one of the points that we need to further study in the future. In addition, we also intend to look for data from other countries besides China for comparative study.

## Data Availability

The data used to support the findings of the study are available at http://www.zcom.gov.cn/.

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## Acknowledgments

The study was partly supported by the Zhejiang Province Philosophy and Social Science Planning "Zhijiang Youth Project" (Grant no. 19ZJQN20YB), Zhijiang Youth Action Project: Study on Mobile E-Commerce Recommendation (Grant no. G306), Ningbo Natural Science Foundation Key Project (Grant no. 2019A610046), Ningbo Philosophy and Social Science Planning Project (project title: the Role and Development Strategy of Cross-border e-commerce in Ningbo's Transition from a Big Foreign Trade City to a Strong Foreign Trade City), Public Technology Research Project of Zhejiang Province (Grant no. LGF20F020004), Humanities and Social Sciences Research Project of the Ministry of Education of China (Grant no. 20YJAZH130), and Achievements of scientific research and innovation team project of Zhejiang Wanli University in 2021.

## References

[1] Mo Fangli and S. Jia, *China's E-Commerce Report (2018)*, Department of E-Commerce and Information Technology, Ministry of Commerce of China, Beijing, China, 2019.

[2] K. M. Nahiduzzaman, A. S. Aldosary, and I. Mohammed, "Framework analysis of E-commerce induced shift in the spatial structure of a city," *Journal of Urban Planning and Development*, vol. 145, no. 3, Article ID 04019006, 2019.

[3] F. K. Andoh-Baidoo, K.-M. Osei-Bryson, K. Amoako-Gyampah, and K. Amoako-Gyampah, "Effects of firm and IT characteristics on the value of e-commerce initiatives: an inductive theoretical framework," *Information Systems Frontiers*, vol. 14, no. 2, pp. 237–259, 2012.

[4] I. Cárdenas, J. Beckers, and T. Vanelslander, "E-commerce last-mile in Belgium: developing an external cost delivery index," *Research in Transportation Business & Management*, vol. 24, pp. 123–129, 2017.

[5] D. R. Agrawal and D. E. Wildasin, "Sales taxation, spatial agglomeration, and the internet," 2019, https://ssrn.com/abstract=3009785.

[6] I. Novo-Corti and M. Barreiro-Gen, "Public policies based on social networks for the introduction of technology at home: demographic and socioeconomic profiles of households," *Computers in Human Behavior*, vol. 51, pp. 1216–1228, 2015.

[7] B. R. Schlichter and L. Danylchenko, "Measuring ICT usage quality for information society building," *Government Information Quarterly*, vol. 31, no. 1, pp. 170–184, 2014.

[8] ITU-International Telecommunication Union, "ICT development index 2017," 2018, http://www.itu.int/net4/ITU-D/idi/2017/index.html.

[9] European Commission, "Digital economy and society index," 2017, https://digital-agenda-data.eu/datasets/desi/indicators.

[10] World Economic Forum, "Shaping the future of digital economy and society," 2017, https://www.weforum.org/communities/the-future-of-the-digital-economy-and-society.

[11] J. Gibbs, K. L. Kraemer, and J. Dedrick, "Environment and policy factors shaping global E-commerce diffusion: a cross-country comparison," *The Information Society*, vol. 19, no. 1, pp. 5–18, 2003.

[12] Y. Si, X. Li, Z. Wei, and T. Chen, "Spatial transition and restructuring of Taobao villages in Internet+ era," *Planners*, vol. 32, no. 5, pp. 117–123, 2016.

[13] Y. Tang, "A spatial difference study on China's E-commerce development level," *Economic Geography*, vol. 35, no. 5, pp. 9–14, 2015.

[14] Y. D. Wei, J. Lin, and L. Zhang, "E-commerce, Taobao villages and regional development in China," *Geographical Review*, vol. 110, no. 3, pp. 380–405, 2019.

[15] C. Deng, B. Xie, X. Li, Li Yang, D. Zhu, and F. Zhang, "Evaluation of agricultural land intensive use in hunan based on principal component analysis method," *Tropical Geography*, vol. 31, no. 1, pp. 65–71, 2011.

[16] H. Feilong, B. Wang, and S. Wang, "Influencing factors and spatial difference of E-commerce development level at county scale in northeast China," *Area Research Development*, vol. 35, no. 4, pp. 16–31, 2016.

[17] X. Wang, J. Ding, and J. Zhao, "The temporal and spatial differentiation of E-commerce development and its driving effect on economic growth—an empirical analysis based on China's inter-provincial panel data," *Jiangsu Agricultural Science*, vol. 45, no. 20, pp. 309–314, 2017.

[18] X. Liu, Z. Ding, X. Huang, M. Wang, and F. Wang, "Spatial distribution characteristics and influence factors of E-commerce development level in China: based on EDI of 1915 counties," *Economic Geography*, vol. 38, no. 11, pp. 11–21, 2018.

[19] J. Yu, L. Wang, and N. Li, "E-shops spatial distribution rule: a case study of Taobao net," *Economic Geography*, vol. 30, no. 8, pp. 1248–1253, 2010.

[20] C. Zhang, C. Chen, D. Streimikiene, and T. Balezentis, "Intuitionistic fuzzy multimoora approach for multi-criteria assessment of the energy storage technologies," *Applied Soft Computing*, vol. 79, pp. 410–423, 2019.

[21] H. Zhong, S. Zhang, L. Hua, and Y. Nie, "Spatial pattern of C2C E-commerce online shops in China," *Economic Geography*, vol. 34, no. 4, pp. 91–96, 2014.

[22] B. Zhu, Y. Song, G. Li, and T. Yu, "Spatial aggregation pattern and influencing factors of Taobao village in China under the C2C E-commerce mode," *Economic Geography*, vol. 36, no. 6, pp. 92–98, 2016.

[23] P. J. Lambert and J. R. Aronson, "Inequality decomposition analysis and the Gini coefficient revisited," *The Economic Journal*, vol. 103, no. 420, pp. 1221–1227, 1993.

[24] L. Anselin, "Spatial econometrics: methods and models," *Economic Geography*, vol. 65, no. 2, pp. 160–162, 1989.

[25] C. Yan-guang, "Reconstructing the mathematical process of spatial autocorrelation based on Moran's statistics," *Geographical Research*, vol. 28, no. 6, pp. 1449–1463, 2009.

[26] M. Tiefelsdorf, "The saddlepoint approximation of Moran's I's and local Moran's II's reference distributions and their numerical evaluation," *Geographical Analysis*, vol. 34, no. 3, pp. 187–206, 2002.

[27] Y. Tillé, M. M. Dickson, G. Espa, and D. Giuliani, "Measuring the spatial balance of a sample: a new measure based on Moran's I index," *Spatial Statistics*, vol. 23, pp. 182–192, 2018.

[28] T. B. Carrijo and A. R. da Silva, "Modified Moran's I for small samples," *Geographical Analysis*, vol. 49, no. 4, pp. 451–467, 2017.

[29] C.-H. Lin and T.-H. Wen, "Using geographically weighted regression (GWR) to explore spatial varying relationships of immature mosquitoes and human densities with the incidence of dengue," *International Journal of Environmental Research and Public Health*, vol. 8, no. 7, pp. 2798–2815, 2011.

[30] S. Farber and A. Páez, "A systematic investigation of cross-validation in GWR model estimation: empirical analysis and Monte Carlo simulations," *Journal of Geographical Systems*, vol. 9, no. 4, pp. 371–396, 2007.

[31] W. Mijia, "Analysis of endogenous growth factors in China's macro economy: an empirical analysis based on the geographical weighted regression (GWR) model," *China Economic Studies*, vol. 3, pp. 24–30, 2009.

Hindawi
Mathematical Problems in Engineering
Volume 2021, Article ID 6622064, 9 pages
https://doi.org/10.1155/2021/6622064

*Research Article*

# Production-Use Water Pricing and Corporate Water Use in China: An Evolutionary Game Theory Model

**Yao Xiao,[1,2] Qiao Peng,[2,3] Wanting Xu,[1,2] and Hongye Xiao[4,5]**

[1]*Center for Innovation and Development studies, Beijing Normal University, Zhuhai 519000, China*
[2]*Economics and Resource Management, Beijing Normal University, Beijing 100875, China*
[3]*Beijing Key Lab of Study on Sci-tech Strategy for Urban Green Development, Beijing Normal University, Beijing 100875, China*
[4]*School of economics, Henan University, Kaifeng 475001, China*
[5]*School of Statistics, Tianjin University of Finance and Economics, Tianjin 300204, China*

Correspondence should be addressed to Qiao Peng; 201731410009@mail.bnu.edu.cn and Wanting Xu; 202031410004@mail.bnu.edu.cn

Received 20 November 2020; Revised 22 December 2020; Accepted 2 January 2021; Published 12 January 2021

Academic Editor: Chonghui Zhang



Decisions related to pricing production-use water are a critical issue that local governments in China are facing. Its significance has increased in recent years, as a serious corporate water-supply shortage has surfaced with rapid economic development and urbanization. Different from developed countries, the pricing of production-use water is a complex issue in China that involves the distribution of benefits among local governments, water-supply companies, and water-consuming companies, where the overall balance is affected by every slight adjustment. Based on the evolutionary game theory, this study constructs an evolutionary game model involving water-supply companies and water-consuming companies with a systematic analysis of the interaction process between the policy formulation related to water pricing by water-supply companies and the decision making related to water consumption by water-consuming companies. The research finds that the difficulty of balancing corporate financial benefits and public water conservation benefits has led to the complexity of water pricing. Moreover, raising water prices will not necessarily cause companies to save water, but it will increase the production cost of the entire economy. This is the direct cause of low water prices, implemented by water-supply companies, in many regions of China.

## 1. Introduction

China is a water-scarce country, with their per capita water resources being only one-fourth of the world average. With the acceleration of economic growth and urbanization, the gap between water supply and demand is further widening [1, 2]. In the face of increasingly severe pressure for water, the Chinese government is promoting green development strategies to reduce the intensity of water use and promote water recycling [3, 4], and water conservation projects are being constructed widely to facilitate efficient water supply. Nevertheless, the problem remains only partially solved, as the abovementioned policies mostly act on the macrolevel and cannot address inefficiency in corporate water use. At present, every 10,000 USD increase in the Chinese economy costs as much as 1456 $m^3$ of water [1]. Therefore, it is urgently important to maximize the role of water pricing in resource distribution and use price tools to boost the efficient use of water.

A number of studies have shown that low water prices are the key reason for China's current low-efficient water use [5, 6]. However, many local governments and water-supply companies persistently implement low water prices [7]. This can be explained by the particularity of water resources, as water pricing for corporate use is not only related to enterprise production but also fundamental to social stability and national security. In developed countries such as Europe and the United States, the price of water for corporate use is mainly decided by market mechanisms [8]. For example, the United States does not have a unified regulatory body for

water prices, and adjustments are performed by the market. Different water-supply organizations adopt different pricing models according to market demand, and one water-supply organization can adopt a variety of price strategies for different types of users. In the United Kingdom, pricing of corporate-use water is also decided by market supply and demand under the input-output model with a price ceiling set by the government. France and other countries have adopted similar pricing methods. Water pricing in China, however, is approved by the government after being determined collaboratively by local water-supply companies, administrative bodies of water resources at all levels, and pricing and finance departments [9]. Water-supply companies are incapable of setting prices by market demand alone. Therefore, mutual influence exists between the decision making of water-supply companies and the water-saving strategies of water-consuming companies, as shown in Figure 1.

When setting water prices, water-supply companies must consider both their own financial benefits and public welfare. That is, they must simultaneously play both corporate and government roles. Meanwhile, their water price policy directly affects the water costs of water-consuming companies, which are, in turn, reflected in water-saving strategies. When making water-saving decisions, local water-consuming companies primarily consider their own benefits and give less consideration to public welfare. Saving water not only reduces water costs for the companies but also serves the public interest, providing both social and environmental benefits [10, 11]. As the latter is also a goal for water-supply companies, their water conservation strategies impact the water pricing decisions of water-supply companies. Therefore, we can observe the mutual influence between the water pricing strategies of water-supply companies and the conservation strategies of water-consuming company. Moreover, due to information asymmetry, either the water-supply side or the water-consumer side can predict the actions of the other. As a result, either side can only make profit-maximizing decisions based on their predictions of the other's actions, and they can continue to improve such predictions by learning from the actual results, which then become references for future decision making. Such information asymmetry and the mutual influence of behavioral decision making has undoubtedly increased the complexity and uncertainty of water pricing and corporate water-saving decision making. Therefore, the interests of all stakeholders must be fully considered, and in-depth research be conducted. As an important tool for studying the decision-making behavior of multiple stakeholders in a complex and uncertain environment, evolutionary game theory provides an effective analysis framework for studying the pricing of water resources and corporate water-saving strategies.

In recent years, natural resources and environmental issues have become further complicated, as the decision-making process has begun to involve multiple subjects [12–14]; this has led to the wide application of evolutionary game theory in the field. Cai et al.[15], Zhang et al.[16], Troeva et al.[17], Jiang et al.[18], and others adopted evolutionary game methods to study the effects of pollution control by polluter companies under government supervision. Maezuru [19], Li et al.[20], Zhao [21], and Niu et al.[22] used evolutionary game tools in their research of the third-party governance model of environmental pollution. Lu et al.[23], Lu, S. B et al. [24], Fang et al.[25], and Lu et al.[26] applied evolutionary game theory to study cross-regional water pollution control issues. Filho et al.[27] and Xin et al.[28] used game theory methods to investigate the impact of water rights prices on the water-saving strategies of water-consuming companies. Some scholars have also studied how to improve the efficiency of the green supply of upstream corporations under environmental supervision [29–32]. Also applying evolutionary game theory, Chen [33] studied the issue of ecological and environmental protection in tourist destinations. Xin and Li [34] adopted the traditional duopoly game model in their study of the water price game in coastal irrigation areas in free market environments, as well as the impact of water price on the efficiency of agricultural irrigation.

Unlike the market-based pricing mechanisms of developed countries, the pricing of corporate water use in China is mainly negotiated between water-supply companies and the relevant government departments. The addition of the government as a decision-making agent has diversified the roles and goals of water-supply companies, further complicating the pricing process of corporate water use in China at the regional level. Through a review of the relevant literature, we find that a few scholars have employed evolutionary game methods to investigate the issue of regional pricing in corporate water use and companies' water-saving strategies in China. This study is innovative in its attempt to adopt evolutionary game theory to build an evolutionary game model of the water pricing and water-saving strategies adopted by water-supply companies and water-consuming companies, respectively. It analyzes the relationship between water pricing, set by local Chinese water-supply companies and the water conservation strategies of water-consuming companies; here, evolutionary game theory will demonstrate its the advantages in solving complex decision-making problems with multiple agents. In the end, the study explains the recent persistence of low pricing strategies in some regions of China in order to provide references for researching the pricing and consumption decisions of other resources, as well as China's green development.

## 2. Methods

Evolutionary game is a recent development of game theory. Originating in the field of biology, it was first proposed by Smith et al.[35] when studying symmetric population games; they claim that the basic concept of evolutionary game is the evolutionarily stable strategy (ESS). Evolution game theory abandons the hypothesis of hyperrationality found in classical game theory and replaces it with the assumption of bounded rationality. By combining the equilibrium analysis in classical game theory with the dynamic evolution analysis in biology, the evolutionary game theory is established. Evolutionary game is a formal model that places the

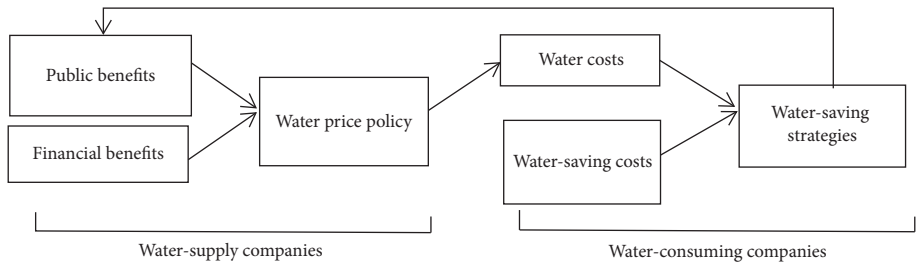



Figure 1: Diagram of the relationship between the pricing strategies of water-supply companies and the conservation strategies of water-consuming companies.

interaction between individuals under the context of evolution [36]. The behavior of an individual is determined by the type of gene, and individuals with different gene types show different behaviors. For example, individuals with combative "eagle" genes and those with mild "doves" genes make different choices when competing for food (that is, fight or concession), and these genotypes can be passed on to the next generation. Individuals with a certain gene type behave accordingly and reap certain benefits that represent their level of adaptation in the natural environment. The greater the benefits, the greater their likelihood of survival in the environment (i.e., through natural selection) and the higher their probability of passing the genotype on to the next generation. That is, the gene fittest for survival will be passed on, while the unfit genes are gradually eliminated, finally resulting in a stable state where biological evolution is achieved. The retained genotypes and their behavioral strategies are, thus, called evolutionarily stable strategies.

In its subsequent development, evolutionary game theory has received extensive attention from the scientific community and is applied in the research of other fields, especially in the social sciences [37]. In the 1990s, economists and social scientists applied biological evolutionary game theory to the research of social problems, and they made particular breakthroughs in the study of the transformation of social systems, the formation of social habits, and social governance. Economists use the ideas mentioned above to study the decision making of multiple interacting microagents in an uncertain environment. The gene type is replaced with the type of strategy that the participant can choose; the level of adaptation is replaced by the amount of benefit of a strategy chosen by the participant; and natural selection is replaced by the principle of bounded rational individuals chasing maximum benefits as the determinant of the probability of a certain strategy being chosen by an individual. The greater the benefit of a strategy is, the higher its probability of being rechosen; this concept corresponds to "heredity" in biology. Participants constantly adjust their strategies according to the size of benefits offered, to the point where all participants finish with the adjustment process with only one strategy retained: the ESS. This strategy can resist the disturbance of any strategy adopted by the participants and restore itself to the original evolutionarily stable state. Evolutionary game theory differs from classical game theory in that it focuses on the dynamic adjustment process of the strategies by individuals with bounded rationality in uncertain environments. It assumes that the participant is not a hyperrational individual with infinite reasoning ability, and unlike classical game, it does not see the participant as capable of achieving the Nash equilibrium strategy through calculation and making choices accordingly. Rather, it is through the process of incessantly repeating the game that the participant learns from the results, adjusts their strategies, and gradually narrows their choices to a stable Nash equilibrium strategy.

In evolutionary game theory, the mode of learning or the strategy adjustment of the participant mainly follows the replicator dynamic equation proposed by Taylor et al.[38]. That is, game participants with bounded rationality adopt the replicator dynamic equation to adjust their strategies. The basic assumption is that the growth rate of the probability of a strategy being adopted is proportional to its level of adaptation. The ESS is obtained by solving a differential equation, and the dynamic equation can be expressed as follows [39, 40].

$$F(x_v) = \frac{\mathrm{d}x_v}{\mathrm{d}t} = x_v\left[E(x_v) - \overline{E}\right], \tag{1}$$

$x_v$ is the probability of a strategy being adopted by the game participant, $E(x_v)$ is the expected benefit of adopting the $v$ strategy, and $\overline{E}$ is the average benefit the participant gains when adopting all possible strategies.

## 3. Model Construction and Solution

*3.1. Hypotheses and Construction.* This paper adopts reduction in water consumption per unit of output of water-consuming companies to indicate water-saving behavior and uses $\alpha$ as the water consumption coefficient. The water-consuming company decides whether to adopt water-saving technologies based upon water prices and water-saving costs, which correspond to the two water consumption coefficients of $\alpha_1$ and $\alpha_2$, respectively, representing a non-water-saving strategy and a water-saving strategy, with $\alpha_1 > \alpha_2$. $P_1$ and $P_2$, respectively, represent the low-price and high-price strategies of the water-supply companies, with $P_1 < P_2$. Table 1 shows all the combinations of the four strategies adopted by water-consuming companies and water-supply companies. To build a concise and realistic game model, we make the following assumptions, based on actual situations.

*Hypothesis 1.* Water-consuming companies pursue maximum corporate benefit. As water pricing is subject to the

Table 1: Game model of water-consuming companies and water-supplying companies.

| Water-consuming companies | Water-supply companies | |
|---|---|---|
| | Low water price $(y)$ $P_1$ | High water price $(1-y)$ $P_2$ |
| Non-water-saving $(x)$ $(\alpha_1)$ | $U_1^u = Y_1 - C - P_1W_1$<br>$U_1^s = P_1W_1$ | $U_3^u = Y_2 - C - P_2W_3$<br>$U_3^s = P_2W_3 - L(\Delta P) + g(W_1 - W_3)$ |
| Water saving<br>$(1-x)$ $(\alpha_2)$ | $U_2^u = Y_1 - C - P_1W_2 - f(\Delta\alpha)$<br>$U_2^s = P_1W_2 + g(W_1 - W_2)$ | $U_4^u = Y_2 - C - P_2W_4 - f(\Delta\alpha)$<br>$U_4^s = P_2W_4 + g(W_1 - W_4) - L(\Delta P)$ |

regulations of local government, water-supply companies assume certain social responsibilities and, therefore, pursue the maximum sum of corporate profits and social benefits.

*Hypothesis 2.* Taking the non-water-saving, low water price combination strategy as the baseline, the reduction $(\Delta W)$ in water consumption can bring social benefits $g(\Delta W)$. We assume that the revenue function $g(\Delta W)$ is a linear increasing function, and $g(\Delta W = 0) = 0$. Water-consuming companies need an investment $f(\Delta\alpha)$ in water-saving technologies, where $\Delta\alpha = \alpha_1 - \alpha_2$ represents the advancement of water-saving technologies. $f\prime > 0$ indicates that the more advanced the water-saving technologies are, the greater the cost; $f\prime\prime > 0$ indicates the increasing marginal cost of water-saving technologies; and $f(\Delta\alpha = 0) = 0$ indicates that the cost of investment is 0 when a water-consuming company does not adopt any advanced water-saving technologies.

*Hypothesis 3.* When water-supply companies raise water prices, the increase in water costs will have a certain impact on the local economy, resulting in a decline in total economic output. Therefore, this study assumes that increasing water prices will cause a loss $L(\Delta P)$ to the entire economy, $\Delta P = P_2 - P_1$. $L\prime > 0$ indicates that the greater the increase in water prices is, the stronger the impact on the economy; $L\prime\prime < 0$ indicates that diminishing marginal effect of the changes in water price changes.

*Hypothesis 4.* A water-consuming company's decision on adopting water-saving technologies is not only dependent on the costs of water and water-saving technologies but also on the costs of other production factors. Since the latter is not the major focus of this study, we assume that the costs $(C)$ of other factors of production remain unchanged.

*Hypothesis 5.* Water-consuming companies adopt non-water-saving strategies with a probability of $x$ and water-saving strategies with a probability of $(1-x), 0 < = x < = 1$. Water-supply companies adopt low-price strategies with a probability of $y$ and high-price strategies with a probability of $(1-y)$, $0 < = y < = 1$.

Suppose $W_1 = \alpha_1 Y_1$, $W_2 = \alpha_2 Y_1$, $W_3 = \alpha_1 Y_2$, and $W_4 = \alpha_2 Y_2$, respectively, represent the water consumption under four strategy combinations (i.e., [non-water-saving, low water price]; [water saving, low water price]; [non-water-saving, high water price]; and [water saving, high water price]). Based on the abovementioned assumptions, we can calculate the corresponding benefits under different combinations of strategies adopted by water-consuming companies and water-supply companies, as shown in Table 1. As $\alpha_1 > \alpha_2$ and $Y_1 > Y_2$, we have $W_1 > W_2 > W_4$ and $W_1 > W_3 > W_4$.

*3.2. Model Solution and Analysis.* According to Table 1, we can obtain the dynamic replicator equations of water-consuming companies and water-supply companies:

(1) Let $U_{\alpha_1}^u$ denote the expected benefit when the water-consuming company adopts

non-water-saving strategies; let $U_{\alpha_2}^u$ denote the expected benefit when the water-consuming company adopts water-saving strategies; let $U^u$ denote the average expected benefit of the water-consuming company.

Expected benefits of water-consuming companies when adopting non-water-saving strategies:

$$U_{\alpha_1}^u = yU_1^u + (1-y)U_3^u. \tag{2}$$

Expected benefits of water-consuming companies when adopting water-saving strategies:

$$U_{\alpha_2}^u = yU_2^u + (1-y)U_4^u. \tag{3}$$

Average expected benefits of water-consuming companies:

$$U^u = xU_{\alpha_1}^u + (1-x)U_{\alpha_2}^u. \tag{4}$$

The dynamic replicator equation of water-consuming companies:

$$\begin{aligned} \frac{\mathrm{d}x}{\mathrm{d}t} &= x\left(U_{\alpha_1}^u - U^u\right) = x(1-x)\left(U_{\alpha_1}^u - U_{\alpha_2}^u\right) \\ &= x(1-x)\{f(\Delta\alpha) - P_2(W_3 - W_4) \\ &\quad + y[P_2(W_3 - W_4) - P_1(W_1 - W_2)]\}. \end{aligned} \tag{5}$$

As the dynamic replicator equation of water-consuming companies becomes stable, that is, when $(\mathrm{d}x/\mathrm{d}t) = 0$,

$$\begin{aligned} x_1^* &= 0, \\ x_2^* &= 1, \\ y^* &= b = \frac{P_2(W_3 - W_4) - f(\Delta\alpha)}{P_2(W_3 - W_4) - P_1(W_1 - W_2)}. \end{aligned} \tag{6}$$

(2) Let $U^s_{P_1}$ denote the expected benefit when the water-supply company adopts low-price strategies; let $U^s_{P_2}$ denote the expected benefit when the water-supply company adopts high-price strategies; and let $U^s$ denote the average expected benefit of the water-supply company. Expected benefits of water-supply companies when adopting low-price strategies are

$$U^g_{P_1} = xU^g_1 + (1-x)U^g_2. \tag{7}$$

Expected benefits of water-supply companies when adopting high-price strategies:

$$U^g_{P_2} = xU^g_3 + (1-x)U^g_4. \tag{8}$$

Average expected benefits of water-supply companies:

$$U^g = yU^g_{P_1} + (1-y)U^g_{P_2}. \tag{9}$$

The dynamic replicator equation of water-supply companies:

$$\begin{aligned}\frac{\mathrm{d}y}{\mathrm{d}t} = y\left(U^g_{P_1} - U^g\right) = y\,(1-y)\{&L(\Delta P) - g(W_2 - W_4) \\ &+ (P_1W_2 - P_2W_4) + x\,[g(W_2 - W_4) \\ &+ (P_2W_4 - P_1W_2) - (P_2W_3 - P_1W_1) \\ &- g(W_1 - W_3)]\}.\end{aligned} \tag{10}$$

As the dynamic replicator equation of water-supply companies becomes stable, that is, when $(\mathrm{d}y/\mathrm{d}t) = 0, y^*_1 = 0$, and $y^*_2 = 1$,

$$x^* = a = \frac{-L(\Delta P) + g(W_2 - W_4) + (P_2W_4 - P_1W_2)}{g(W_2 - W_4) + (P_2W_4 - P_1W_2) - g(W_1 - W_3) - (P_2W_3 - P_1W_1)}. \tag{11}$$

From the analysis mentioned above, we can observe that when $0 < a < 1$ and $0 < b < 1$, the equilibrium points of the two-dimensional dynamic system composed of the water price strategies of water-supply companies and the water-saving strategies of water-supply company are, respectively, $(0,0)$, $(0,1)$, $(1,0)$, $(1,1)$, and $(a,b)$.

*3.3. Analysis of Evolutionary Equilibrium Stability.* Let $F(x) = (\mathrm{d}x/\mathrm{d}t)$ and $G(y) = (\mathrm{d}y/\mathrm{d}t)$. We use the Jacobian matrix to analyze the stability of each local equilibrium point of the two-dimensional dynamic system:

$$\begin{aligned}
J &= \begin{bmatrix} \dfrac{\partial F(x)}{\partial x} & \dfrac{\partial F(x)}{\partial y} \\ \dfrac{\partial G(y)}{\partial x} & \dfrac{\partial G(y)}{\partial y} \end{bmatrix} = \begin{bmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{bmatrix}, \\
c_{11} &= (1-2x)\{f(\Delta\alpha) - P_2(W_3 - W_4) + y\,[P_2(W_3 - W_4) - P_1(W_1 - W_2)]\}, \\
c_{12} &= x\,(1-x)\,[P_2(W_3 - W_4) - P_1(W_1 - W_2)], \\
c_{21} &= y\,(1-y)\,[g(W_2 - W_4) + (P_2W_4 - P_1W_2) - (P_2W_3 - P_1W_1) - g\,(W_1 - W_3)], \\
c_{22} &= (1-2y)\{L\,(\Delta P) - g\,(W_2 - W_4) + (P_1W_2 - P_2W_4) + x\,[g\,(W_2 - W_4) \\
&\quad + (P_2W_4 - P_1W_2) - (P_2W_3 - P_1W_1) - g\,(W_1 - W_3)]\}.
\end{aligned} \tag{12}$$

The stable equilibrium point of the dynamic replicator equation must satisfy the following two conditions:

$$\begin{aligned}\mathrm{tr}\,(J) &= c_{11} + c_{22} < 0, \\ \det\,(J) &= c_{11}c_{22} - c_{12}c_{21} > 0.\end{aligned} \tag{13}$$

To simplify the expression and make the economic implication more obvious, we carry out the following transformation:

(1) Suppose $\Delta\pi_1 = P_1(W_1 - W_2)$ and $\Delta\pi_2 = P_2(W_3 - W_4)$, respectively, denote the situations of low water price $P_1$ and high water price $P_2$, where water costs

drop for water-consuming companies owing to the adoption of water-saving technologies that reduces water consumption

(2) Suppose $\Delta g_1 = g(W_1 - W_3)$ and $\Delta g_2 = g(W_2 - W_4)$, respectively, denote the situations of water-consuming companies not adopting water-saving technologies $\alpha_1$ and adopting water-saving technologies $\alpha_2$, where water consumption drops due to an increase in water price by water-supply companies, which, in turn, increases public benefit

(3) Suppose $\Delta z_1 = P_2W_3 - P_1W_1$ and $\Delta z_2 = P_2W_4 - P_1W_2$, respectively, denote the situations of water-consuming companies not adopting water-saving technologies $\alpha_1$ and adopting water-saving technologies $\alpha_2$, where water-supply companies gain more corporate benefit

We can obtain Table 2 from the Jacobian matrix of the dynamic replicator equation.

We can know from Table 2 and the stability condition of the dynamic replicator's equation that first, when $\Delta\pi_1 < f(\Delta\alpha) < \Delta\pi_2$ and $\Delta g_1 + \Delta z_1 < L(\Delta P) < \Delta g_2 + \Delta z_2$, the binary differential dynamic system composed of water-supply companies and water-consuming companies has two stable equilibrium points, as shown in Figure 2, which are O (0, 0) and B (1, 1), and a saddle point exists as H (a, b), of which

$$a = \frac{-L(\Delta P) + \Delta g_2 + \Delta z_2}{\Delta g_2 + \Delta z_2 - \Delta g_1 - \Delta z_1},$$
$$b = \frac{\Delta\pi_2 - f(\Delta\alpha)}{\Delta\pi_2 - \Delta\pi_1}. \quad (14)$$

The game between water-supply companies and water-consuming companies on water price and water-saving may eventually converge to the equilibrium point of O (0, 0) or B (1, 1). The respective probabilities are as follows:

$$P_o = S_{\mathrm{OAHC}} = \frac{1}{2}(a+b)$$
$$= \frac{1}{2}\left(\frac{-L(\Delta P) + \Delta g_2 + \Delta z_2}{\Delta g_2 + \Delta z_2 - \Delta g_1 - \Delta z_1} + \frac{\Delta\pi_2 - f(\Delta\alpha)}{\Delta\pi_2 - \Delta\pi_1}\right), \quad (15)$$
$$P_B = 1 - S_{\mathrm{OAHC}}.$$

Conclusion 1: The probability $P_B$ that the decision-making adjustments of water-supply companies and water-consuming companies converge to the equilibrium point of B (1, 1) is inversely proportional to $\Delta g_2$, $\Delta z_2$, $\Delta g_1$, and $\Delta z_1$ and directly proportional to $L(\Delta P)$.

The economic implication is that when water-supply companies raise the price of water, the smaller the increase in the corporate benefit of revenue and the public benefit of water saving, the more likely that the evolutionary game converges to the strategy combination of [no water saving, low water price]. Meanwhile, the greater the impact that increasing the price of water has on the economy, the more likely that the evolutionary game converges to the strategy combination of [no water-saving, low water price], and *vice versa*.

Table 2: The Jacobian matrix of the equilibrium points.

| Equilibrium points | $c_{11}$ | $c_{12}$ | $c_{21}$ | $c_{22}$ |
|---|---|---|---|---|
| (0,0) | $f(\Delta\alpha) - \Delta\pi_2$ | 0 | 0 | $L(\Delta P) - \Delta g_2 - \Delta z_2$ |
| (0,1) | $f(\Delta\alpha) - \Delta\pi_1$ | 0 | 0 | $-[L(\Delta P) - \Delta g_2 - \Delta z_2]$ |
| (1,0) | $-[f(\Delta\alpha) - \Delta\pi_2]$ | 0 | 0 | $L(\Delta P) - \Delta g_1 - \Delta z_1$ |
| (1,1) | $-[f(\Delta\alpha) - \Delta\pi_1]$ | 0 | 0 | $-[L(\Delta P) - \Delta g_1 - \Delta z_1]$ |
| (a, b) | 0 | $c_{12}^*$ | $c_{21}^*$ | 0 |

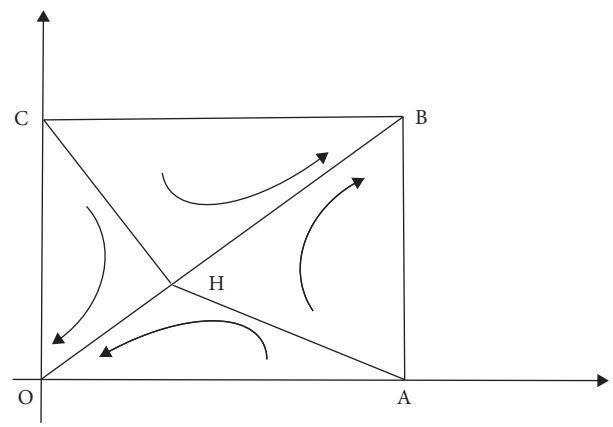



Figure 2: Phase diagram of the evolution of decision making between water-consuming and water-supplying companies.

Conclusion 2: The probability $P_B$ that the decision-making adjustments of water-supply companies and water-consuming companies converge to the equilibrium point of B (1, 1) is inversely proportional to $\Delta\pi_2$ and $\Delta\pi_1$ and directly proportional to $f(\Delta\alpha)$.

The economic implication is that when water-consuming companies adopt water-saving technologies, the smaller the decrease in water costs and the higher the cost of water-saving technologies, the more likely that the evolutionary game converges to the strategy combination of [no water-saving, low water price], and *vice versa*.

Second, when $\Delta\pi_2 < f(\Delta\alpha) < \Delta\pi_1$ and $\Delta g_2 + \Delta z_2 < L(\Delta P) < \Delta g_1 + \Delta z_1$, the binary differential dynamic system composed of water-supplying companies and water-consuming companies has two stable equilibrium points, as shown in Figure 3, which are A (1, 0) and C (0, 1), and a saddle point exists as H (a, b). Due to the rarity of the strategy combinations of [water-saving, low water price] or [non-water-saving, high water price] in real life, this article does not go further into their discussions.

In general, the analysis explains why the state of equilibrium (i.e., [non-water-saving, low water prices]) exists in certain regions. This is because increases in water prices made by water-supply companies have a strong impact $L(\Delta P)$ on the economy; however, they have insufficient corporate benefits and public benefits of water conservation. This leads to the persistence of low pricing strategies by some local water-supply companies, for fear they will lose benefits by raising water prices.

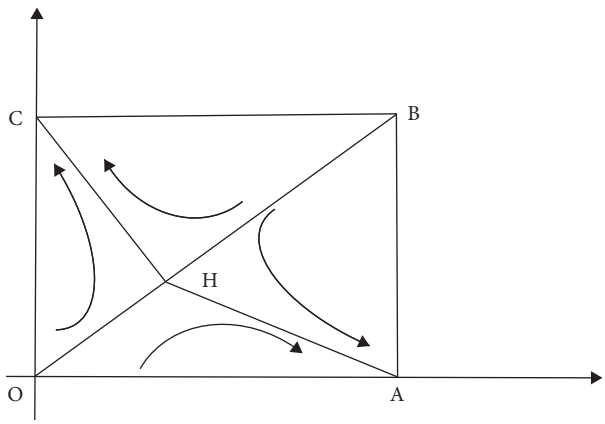



Figure 3: Phase diagram of decision-making evolution between water-consuming and water-supplying enterprises.

According to the water price data of water-supply companies in various regions of China [2], it is found that China's Inner Mongolia, Shanxi, and Hebei provinces implement the low water price policy.Water resources of these regions are relatively scarce. Although high water price can improve the efficiency of water use and promote enterprises to save water and, thus, improve the overall social public benefits, since the cost of water-saving technologies $f(\Delta\alpha)$ is relatively high in these areas and they are not adopted by water-supply companies, which high water price cannot greatly promote enterprises to save water, it can only slightly increase the public benefit of water saving. Moreover, higher water prices increase the cost of production, so local governments and the water sectors continue to keep water prices low based on the tradeoff between cost and benefit.

## 4. Conclusions and Suggestion

This study adopts the evolutionary game methodology of biology to create an evolutionary game model of water-supply companies and water-consuming companies. It presents a systematic analysis of the decision-making process regarding water pricing and water consumption and explains the low water price strategies in some regions in China. Based on these, the article reaches the following conclusions:

(1) Through an analysis of the persistence of low water price strategies in many regions in China using our evolutionary game model, we find the model is an efficient tool in describing the decision-making process of multiple interacting microagents, especially for decision making in uncertain environments by individuals with bounded rationality. This research shows that water-supply companies and water-consuming companies are related stakeholders and entities with bounded rationality. Moreover, they are unsure about each other's decisions and must constantly adjust their own strategies accordingly. Evolutionary games capture the complexity, uncertainty, and dynamics of the decision-making process of water-supply companies and water-consuming companies by employing dynamic replicator equations. Its unique advantages are proven.

(2) The greater the corporate benefit and public benefit are when water-supply companies raise water prices, and the smaller its impact on the economy and the more likely that the evolutionary game converges to the strategy combination of [water-saving, high water price]. Otherwise, it is more likely for it to converge to the strategy combination of [non-water-saving, low water price].

(3) The greater the decrease in water costs is when water-consuming companies adopt water-saving technologies and the lower their cost, the more likely that the evolutionary game converges to the strategy combination of [water-saving, high water price]. Otherwise, it is more likely for it converge to the strategy combination of [non-water-saving, low water price].

Conclusions (2) and (3) show that even when water-supply companies and the government raise water prices, water-consuming companies will still shrink from adopting water-saving technologies due to their huge costs. As a result, high water price will neither improve water conservation nor boost water efficiency; instead, it will adversely reduce the public benefit of water-supply companies brought by water-supply companies and further inhibit them from raising the price of water. However, when the cost of water-saving technologies is low, raising water prices will have an immediate effect; that is, water-consuming companies will start saving water. Nevertheless, only when their impact on the economy is limited will water-supply companies willingly choose high water prices, thus improving water conservation and boosting water efficiency.

Therefore, this article suggests that when water-supply companies and the government are considering promoting water conservation by raising water prices, they can provide relevant policy support, such as preferential subsidies or tax reductions, to water-consuming companies to soothe the induced economic impact. The government can also provide targeted subsidies to water-consuming companies that are investing in water-saving technologies, thus encouraging innovations in water-saving technologies and, thereby, enhancing water conservation through pricing.

## 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 paper was supported by the Institute of National Accounts (Zhuhai), Beijing Normal University. This research was funded by the National Natural Science Foundation of China (Grant no. 71873019) and the National Social Science Fund of China (Grant no. 18ZDA123).

## References

[1] W. Fang, S. Huang, G. Huang et al., "Copulas-based risk analysis for inter-seasonal combinations of wet and dry conditions under a changing climate," *International Journal of Climatology*, vol. 39, no. 4, pp. 2005–2021, 2018.

[2] S. Liu, S. Huang, Y. Xie et al., "Identification of the non-stationarity of floods: changing patterns, causes, and implications," *Water Resources Management*, vol. 33, no. 3, pp. 939–953, 2018.

[3] J. Chang, A. Guo, Y. Wang et al., "Reservoir operations to mitigate drought effects with a hedging policy triggered by the drought prevention limiting water level," *Water Resources Research*, vol. 55, no. 2, pp. 904–922, 2019.

[4] Y. Shang, P. Hei, S. Lu et al., "China's energy-water nexus: assessing water conservation synergies of the total coal consumption cap strategy until 2050," *Applied Energy*, vol. 210, pp. 643–660, 2018.

[5] J. Zhao, H. Ni, X. Peng, J. Li, G. Chen, and J. Liu, "Impact of water price reform on water conservation and economic growth in China," *Economic Analysis and Policy*, vol. 51, pp. 90–103, 2016.

[6] J. Li and X. C. Ma, "Econometric analysis of industrial water use efficiency in China," *Environment Development and Sustainability*, vol. 17, no. 5, pp. 1–18, 2015.

[7] L. Mu, C. Wang, B. Xue, H. Wang, and S. Li, "Assessing the impact of water price reform on farmers' willingness to pay for agricultural water in northwest China," *Journal of Cleaner Production*, vol. 234, pp. 1072–1081, 2019.

[8] G.. G.. Rojas and R. Hilda, "Water price policy and its institutional role as an economic instrument for water management," *Water Policy in Mexico*, vol. 20, pp. 137–152, 2019.

[9] Z. Chen, H. Wang, and X. Qi, "Pricing and water resource allocation scheme for the south-to-north water diversion project in China," *Water Resources Management*, vol. 27, no. 5, pp. 1457–1472, 2013.

[10] W. Wang, H. Xie, N. Zhang, and D. Xiang, "Sustainable water use and water shadow price in China's urban industry," *Resources Conservation & Recycling*, vol. 128, pp. 489–498.

[11] World Bank, "China country water resources partnership strategy (2013-2020)," *World Bank Other Operational Studies*, vol. 50, no. 3, pp. 133–139, 2013.

[12] S. Zeng, Y. Hu, T. Balezentis, and D. Streimikiene, "A multi-criteria sustainable supplier selection framework based on neutrosophic fuzzy data and entropy weighting," *Sustainable Development*, vol. 28, no. 5, pp. 1431–1440, 2020.

[13] S. Zeng, X. Peng, T. Baležentis, and D. Streimikiene, "Prioritization of low-carbon suppliers based on Pythagorean fuzzy group decision making with self-confidence level," *Economic Research-Ekonomska Istraživanja*, vol. 32, no. 1, pp. 1073–1087, 2019.

[14] C. Zhang, C. Chen, D. Streimikiene, and T. Balezentis, "Intuitionistic fuzzy MULTIMOORA approach for multi-criteria assessment of the energy storage technologies," *Applied Soft Computing*, vol. 79, pp. 410–423, 2019.

[15] L.-ru. Cai, H.-wei. Wang, and W.. Zeng, "Mixed-strategy game simulation on environmental pollution based on system dynamics," in *Proceedings of the Control & Decision Conference*, pp. 2199–2204, IEEE, Yantai, Shandong, China, July 2008.

[16] X.. Zhang, Q.. Li, and J.. Yao, *Study on Secondary Lead Enterprises Pollution Monitoring with an Asymmetric Evolutionary Game Model*, pp. 1–4, World Automation Congress, Taipei, Taiwan, 2012.

[17] M. Troeva and V. Lukin, "On a game-theoretic model of environmental pollution problem," *Advances in Dynamic Games*, pp. 223–236, 2013.

[18] K. Jiang, D. You, R. Merrill, and Z. Li, "Implementation of a multi-agent environmental regulation strategy under Chinese fiscal decentralization: an evolutionary game theoretical approach," *Journal of Cleaner Production*, vol. 214, pp. 902–915, 2019.

[19] M. Maezuru, "Stochastic differential game of transboundary pollution and international environmental policy," *Studies in Regional Science*, vol. 40, no. 2, pp. 413–425, 2010.

[20] H.. Li, "Genlong. Guo, "A differential game analysis of multipollutant transboundary pollution in river basin," *Physica A: Statistical Mechanics and Its Applications*, vol. 535, p. 12248, 2019.

[21] Y. Zhao, "Government's control countermeasures against environmental pollution by introducing third-party constraints," *Nature Environment and Pollution Technology*, vol. 19, no. 02, pp. 571–576, 2020.

[22] X. Niu, X. Wang, J. Gao, and X. Wang, "Has third-party monitoring improved environmental data quality? An analysis of air pollution data in China," *Journal of Environmental Management*, vol. 253, Article ID 109698, 12 pages, 2020.

[23] S. Lu, H. Bao, and H. Pan, "Urban water security evaluation based on similarity measure model of vague sets," *International Journal of Hydrogen Energy*, vol. 41, no. 35, pp. 15944–15950, 2016.

[24] S. Lu, Y. Shang, and Y. Li, "A research on the application of fuzzy iteration clustering in the water conservancy project," *Journal of Cleaner Production*, vol. 151, pp. 356–360, 2017.

[25] W. Fang, S. Z. Huang, K Ren et al., "Examining the applicability of different sampling techniques in the development of decomposition-based streamflow forecasting models," *Journal of Hydrology*, vol. 568, pp. 534–550, 2019.

[26] S. Lu, X. Wu, H. Sun, W. Li, and Y. Tang, "The multi-user evolutionary game simulation in water quality-based water source system," *Environmental Geochemistry and Health*, vol. 42, no. 3, pp. 863–879, 2020.

[27] F. A. S. Filho, U. Lall, and R. L. L. Porto, "Role of price and enforcement in water allocation: insights from Game Theory," *Water Resources Research*, vol. 44, no. 12, pp. 12424–12431, 2008.

[28] B. Xin and M. Sun, "A differential oligopoly game for optimal production planning and water savings," *European Journal of Operational Research*, vol. 269, no. 1, pp. 206–217, 2017.

[29] Q.-h. ZHU and Y.-j. DOU, "Evolutionary game model between governments and core enterprises in greening supply chains," *Systems Engineering - Theory & Practice*, vol. 27, no. 12, pp. 85–89, 2007.

[30] S. Barari, G. Agarwal, W. J. Zhang, B.. Mahanty, and M. . K. Tiwari, "A decision framework for the analysis of green supply chain contracts: an evolutionary game approach," *Expert Systems with Applications*, vol. 39, no. 3, pp. 2965–2976, 2012.

[31] P. Ji, X. Ma, and G. Li, "Developing green purchasing relationships for the manufacturing industry: an evolutionary game theory perspective," *International Journal of Production Economics*, vol. 166, pp. 155–162, 2015.

[32] Y. Tian, K. Govindan, and Q. Zhu, "A system dynamics model based on evolutionary game theory for green supply chain management diffusion among Chinese manufacturers," *Journal of Cleaner Production*, vol. 80, pp. 96–105, 2014.

[33] Si-hua. Chen, "An evolutionary game study of an ecological industry chain based on multi-agent simulation:A case study of the poyang lake eco-economic zone," *Sustainability*, vol. 9, no. 7, p. 1165, 2017.

[34] B. Xin and Y. Li, "Bifurcation and chaos in a price game of irrigation water in a coastal irrigation district," *Discrete Dynamics in Nature and Society*, vol. 201310 pages, 2013.

[35] J. M. Smith and G. R. Price, "The logic of animal conflict," *Nature*, vol. 246, no. 5427, pp. 15–18, 1973.

[36] J. Smith, *Evolution and the Theory of Games*, Cambridge University Press, Cambridge, Englang, 1982.

[37] J. Hofoauer and K. Siglnund, "Evolutionary game dynamies," *Bulletin-Ameriean Mathematieal Soeiety*, vol. 40, no. 4, pp. 479–520, 2003.

[38] P. D. Taylor and L. B. Jonker, "Evolutionary stable strategies and game dynamics," *Mathematical Biosciences*, vol. 40, no. 1-2, pp. 145–156, 1978.

[39] Y. Yang and W. Yang, "Does whistleblowing work for air pollution control in China? A study based on three-party evolutionary game model under incomplete information," *Sustainability*, vol. 11, no. 2, p. 324, 2019.

[40] Q. Peng and Y. Xiao, "Will third-party treatment effectively solve issues related to industrial pollution in China?" *Sustainability*, vol. 12, no. 18, p. 7685, 2020.

Hindawi
Mathematical Problems in Engineering
Volume 2023, Article ID 9759528, 1 page
https://doi.org/10.1155/2023/9759528

*Retraction*

# Retracted: Probabilistic Hesitant Fuzzy Methods for Prioritizing Distributed Stream Processing Frameworks for IoT Applications

**Mathematical Problems in Engineering**

Received 17 October 2023; Accepted 17 October 2023; Published 18 October 2023



This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

(1) Discrepancies in scope
(2) Discrepancies in the description of the research reported
(3) Discrepancies between the availability of data and the research described
(4) Inappropriate citations
(5) Incoherent, meaningless and/or irrelevant content included in the article
(6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

## References

[1] Z. Lin, C. Huang, and M. Lin, "Probabilistic Hesitant Fuzzy Methods for Prioritizing Distributed Stream Processing Frameworks for IoT Applications," *Mathematical Problems in Engineering*, vol. 2021, Article ID 6655477, 12 pages, 2021.

Hindawi
Mathematical Problems in Engineering
Volume 2021, Article ID 6655477, 12 pages
https://doi.org/10.1155/2021/6655477

*Research Article*

# Probabilistic Hesitant Fuzzy Methods for Prioritizing Distributed Stream Processing Frameworks for IoT Applications

**Zhimin Lin**,[1] **Chao Huang**,[2] **and Mingwei Lin**[2]

[1]*College of Economics, Fujian Agriculture and Forestry University, Fuzhou 350002, China*
[2]*College of Mathematics and Informatics, Fujian Normal University, Fuzhou 350117, Fujian, China*

Correspondence should be addressed to Mingwei Lin; linmwcs@163.com

Received 23 October 2020; Revised 15 December 2020; Accepted 21 December 2020; Published 11 January 2021

Academic Editor: Shouzhen Zeng



Distributed stream processing frameworks (DSPFs) are the vital engine, which can handle real-time data processing and analytics for IoT applications. How to prioritize DSPFs and select the most suitable one for special IoT applications is an open issue. To help developers of IoT applications to solve this complex issue, a novel probabilistic hesitant fuzzy multicriteria decision making (MCDM) model is put forward in this paper. To characterize the requirements for large-scale IoT data stream processing, a novel evaluation criteria system including qualitative and quantitative criteria is established. To accurately model the collective opinions from skilled developers and consider their psychological distance, the definition of probabilistic hesitant fuzzy sets (PHFSs) is used. To derive the importance degrees of criteria, a novel probabilistic hesitant fuzzy best-worst (PHFBW) method is proposed based on the score value. To prioritize the DSPFs and choose the most suitable one, a novel probabilistic hesitant fuzzy MULTIMOORA method is put forward. Finally, a practical case composed of four Apache stream processing frameworks, namely, Storm, Flink, Spark, and Samza, is studied. The obtained results indicate that throughput, latency, and reliability are considered to be the three most important criteria, and Flink is the most suitable stream framework.



## 1. Introduction

Internet of things (IoT) technology [1] is a new computing paradigm, which uses a large number of physical things for continuously monitoring and collecting data from surrounding objects, transmitting the collected data over the network, and feeding the collected data into backend servers. These physical things may be smartphones, wearable devices, tablets, sensors, and cameras. It has been widely used in various domains, such as transportation, health care, logistics, and agriculture [2]. In the IoT applications, millions of IoT devices are deployed and they continuously output large amounts of data [3], which are valuable for the enterprises to make reasonable business decisions in realtime [4]. However, how to process and analyze the IoT stream data are a big challenge for enterprises since traditional batch processing architecture cannot process large amounts of data in realtime. Even worse, data are produced continuously at a high speed [5].

The distributed stream processing frameworks (DSPFs) [6] are the practicable technique solution, which can be used to fulfil such large-scale data processing and analytics for IoT applications in realtime [7, 8]. The DSPFs have become a vital component of each IoT solution stack [9]. There are so many kinds of DSPFs that it is difficult for enterprises to choose the most suitable one since the DSPFs have different features [10] and enterprises have conflicting requirements for creating their IoT applications. The wrong choice may lead to failures in developing IoT applications. Thus, how to evaluate the DSPFs and choose the most suitable one is a critical step for creating IoT applications [11]. Up to now, there are no research studies focusing on how to evaluate DSPFs and select the most suitable one to support the requirements for large-scale IoT data stream processing.

In this paper, we plan to formulate the process of evaluating DSPFs and choosing the most suitable one to be a multicriteria decision making (MCDM) problem since some DSPFs should be evaluated with respect to their criteria. To

the best of our knowledge, it is the first study that focuses on addressing this problem. The contributions of our study are summarized as follows:

(1) To characterize multiple requirements for large-scale IoT data stream processing, a hybrid evaluation criteria system composed of qualitative and quantitative criteria is established for DSPFs.

(2) To accurately model collective opinions from a group of experienced professionals in the technical committee and also consider the psychological distance among linguistic terms, the concept of probabilistic hesitant fuzzy sets (PHFSs) is introduced.

(3) A novel probabilistic hesitant fuzzy best-worst (PHFBW) method is put forward for computing the weights of criteria. Afterward, the importance of degrees of criteria are analyzed.

(4) To prioritize the DSPFs, we put forward a novel probabilistic hesitant fuzzy MULTIMOORA method to derive the ranking values and ranking orders of the DSPFs by using three subsystems and then propose an extended Borda method to fuse the ranking values and ranking orders.

This study can help the enterprise to make correct decisions according to the requirements for large-scale IoT data stream processing. It is easy to extend this study for solving the other decision-making problems in the organization management. In this paper, the following contents are organized as follows:

In Section 2, the research results of DSPFs and information representation in the MCDM problem are briefly given. Four DSPFs and some basic knowledge about probabilistic hesitant fuzzy sets are provided in Section 3. In Section 4, a hybrid evaluation criteria system composed of qualitative criteria and quantitative criteria is established. Then, we propose a novel probabilistic hesitant fuzzy best-worst method for deriving the importance degrees of criteria and a new probabilistic hesitant fuzzy MULTIMOORA method to determine the ranking order of four DSPFs. The numerical analysis is used to show the implementation processes of the probabilistic hesitant fuzzy MCDM model in Section 5. Finally, Section 6 presents some valuable conclusions.

## 2. Literature Review

In this section, the research studies focusing on DSPFs and information representation in the decision-making process are briefly reviewed.

*2.1. Review on Streaming Frameworks.* There are many research results on DSPFs. Various DSPFs have been proposed for special purposes, such as multimedia streaming framework [12], P2P live framework [13], and fraud detection framework [14]. To process genomics data in a fast and efficient way, a novel sequence aligner was implemented on Apache Spark [15]. The multiquery component of Apache Flink was optimized for big data [16]. An efficient tool was put forward by Espinosa et al. [17] for testing the functions of Apache Flink. Researchers also used the streaming frameworks for the health status predictions [18], congestion prediction [19], and precise medicine [20]. To the best of our knowledge, there are no research studies focusing on evaluating DSPFs for large-scale IoT data stream processing.

*2.2. Information Representation.* In the early stage of the decision-making evolution, crisp values are usually adopted by human beings to express their opinions [21]. Due to the uncertainty in human beings' complicated activities, fuzzy sets [22] were proposed for describing uncertain information or vague information. To further highlight human beings' hesitant attitudes, the concept of hesitant fuzzy sets (HFSs) [23] was proposed so that several possible fuzzy values from the interval [0 and 1] can be used to express the quantitative hesitant information or group preference information. Nevertheless, the HFSs may distort the original opinions when they are used to model the group preference information since they do not have the ability to contain the probability information of each fuzzy value. To solve this defect, the probabilistic HFSs (PHFSs) [24, 25] were developed to accurately model the group preference information without losing probability information [26–48].

In some cases, human beings prefer to use the qualitative tools for expressing their opinions. For example, human beings may use the linguistic terms "high" or "low" when evaluating the maturity of streaming frameworks. The fuzzy linguistic method was put forward in [49] to portray these linguistic terms. Although there are some extensions of the fuzzy linguistic method, such as linguistic 2-tuple concepts [50] and virtual linguistic term model [51], they still have the limitation that they cannot contain several linguistic terms simultaneously. Motivated by HFSs, two qualitative tools: hesitant fuzzy linguistic term sets (HFLTSs) [52] and extended HFLTSs [53] were proposed for expressing the qualitative hesitant information of individuals [54] or the group preference information from a group of skilled experts. Similar to the HFSs, HFLTSs and extended HFLTSs also cannot contain the probability information of each linguistic term. Hence, the idea of probabilistic linguistic term sets (PLTSs) [55] was implemented to associate each linguistic term with probability information. Because of the strong capability of expressing the group preference information in the qualitative context, PLTSs have been applied into various fields, such as edge computing [56] and evaluation of hospitals [57].

## 3. Preliminaries

In this section, the introductions of the DSPFs are given, and then, the knowledge on probabilistic hesitant fuzzy sets is given.

*3.1. Introductions of Four DSPFs.* There are many well-known DSPFs that have the ability to perform the IoT data stream processing. After screening DSPFs, the enterprise chooses to evaluate four DSPFs of Apache for the large-scale

IoT data stream processing according to its requirements. The four Apache streaming frameworks are introduced as follows:

*3.1.1. Apache Storm.* Storm is a well-known streaming framework [58], which is equipped with various queueing and database technologies and can be also compatible with any programming language. It can handle streaming events at a high speed. The benchmarking results show that Storm has the ability to process the streaming events at more than 1,000,000 events per second per node. It also has a flexible topology that allows streaming events to be processed in any way and repartitioned from node to node in any way.

*3.1.2. Apache Flink.* Flink [59] can not only process the collected data in batches but also provide the way of event streaming processing. It can be deployed on all the mainstream cluster platforms, and it also has the ability to process streaming events at in-memory speed and at an arbitrary scale. When it is configured for the purpose of high availability, Flink has the ability to scale to thousands of cores and trillions of events per day, while still keeping low latency and high throughput.

*3.1.3. Apache Spark.* Spark [60] is a scalable streaming framework that supports the functions of high-throughput and fault-tolerant processing. The processed streaming data in Spark can be collected from various sources, processed, and fed into file systems, databases, and live dashboards. Different from other frameworks, it processes data in microbatches, not the event streaming way. Since it can process data in extremely small batches, these extremely small batches can be solved in rapid succession, closely approximate to the real-time requirement of event streaming. Moreover, it is broadly applied in the industrial environments. Hence, in this paper, it is compared with the native streaming frameworks.

*3.1.4. Apache Samza.* Samza [61] is equipped with a scalable and high-performance storage scheme, which allows organizations to execute stateful streaming applications. Hence, stateful streaming processing is a core function of Samza. This excellent feature makes Samza smoothly execute extremely complicated streaming jobs. It can migrate jobs from one node to another without influencing the overall performance.

*3.2. Knowledge on PLTSs and PHFSs.* The linguistic term set [62], abbreviated to LTS, is the data source of PLTSs. It consists of several ordered linguistic terms that mathematically represent the natural language such as "high" and "good". It is defined as $S = \{s_\rho | \rho = -l, \ldots, 0, \ldots, l\}$. When the maturity of a streaming framework is evaluated, we can use the following LTS: $S = \{s_{-2} = \text{very low}, s_{-1} = \text{low}, s_0 = \text{Neutral}, s_1 = \text{high}, \text{and } s_2 = \text{very high}\}$.

*Definition 1* (See [55]). Let $S$ be an LTS, then the PLTS can be mathematically defined as

$$H = \left\{ (L_\alpha, p_\alpha) | L_\alpha \in S; \sum_{\alpha=1}^{|H|} p_\alpha \le 1; \quad 0 \le p_\alpha \le 1 \right\}, \tag{1}$$

where $L_\alpha$ is a linguistic term from $S$ and $p_\alpha$ is its probability information, $|H|$ denotes the number of elements within the set of $H$.

*Definition 2* (See [24]). The PHFS is mathematically defined as

$$F = \left\{ (f_\alpha, p_\alpha) | f_\alpha \in [0,1]; \sum_{\alpha=1}^{|F|} p_\alpha \le 1; \quad 0 \le p_\alpha \le 1 \right\}, \tag{2}$$

where $f_\alpha$ denotes the $\alpha$th fuzzy value from the unit interval and $|F|$ is the number of elements within the set of $F$.

In the qualitative linguistic context, there exist two methods for calculating linguistic terms: (1) the semantic method mapping linguistic terms into fuzzy values by considering psychological distances between linguistic terms; (2) the symbolic method using the subscripts of linguistic terms directly [54]. Therefore, using the semantic method, PLTSs can be transformed into PHFSs.

If the psychological distances between any two consecutive linguistic terms are equal, then the PLTSs can be transformed using the following definition:

*Definition 3.* Let $S = \{s_\rho | \rho = -l, \ldots, 0, \ldots, l\}$ be an LTS and $H = \{(L_\alpha, p_\alpha) | L_\alpha \in S\}$ denote any PLTS, then the PLTS is transformed into the PHFS $F = \{(g(L_\alpha), p_\alpha) | \alpha = 1, 2, \ldots, |H|\}$ using the following function:

$$\begin{gathered} g\colon [-l, l] \longrightarrow [0,1], \\ g(L_\alpha) = \frac{\operatorname{Sub}(L_\alpha)}{2l} + \frac{1}{2}, \end{gathered} \tag{3}$$

where $|H|$ denotes the number of elements within the set of $H$ and $\operatorname{Sub}(L_\alpha)$ is the subscript of linguistic term $L_\alpha$.

*Definition 4* (See [24]). Let $F = \{(f_\alpha, p_\alpha) | \alpha = 1, 2, \ldots, |F|\}$ denote a PHFS, then $S(F) = \sum_{\alpha=1}^{|F|} f_\alpha p_\alpha$ is defined to be the score value of this PHFS.

It is difficult to compute various measures between PHFSs when they have different numbers of elements. Therefore, they should be normalized using the following definition:

*Definition 5.* Suppose that $F_1 = \{(f_\alpha^1, p_\alpha^1) | \alpha = 1, 2, \ldots, |F_1|\}$ and $F_2 = \{(f_\alpha^2, p_\alpha^2) | \alpha = 1, 2, \ldots, |F_2|\}$ are PHFSs with $|F_1| > |F_2|$, then $|F_1| - |F_2|$ elements should be added into the PHFS $F_2$ and the added elements are the minimum fuzzy value in the PHFS $F_2$ and associated with the probability of zero. At the same time, the elements within the PHFSs are rearranged according to the descending order of the values of $f_\alpha^1 p_\alpha^1$ and $f_\alpha^2 p_\alpha^2$.

*Definition 6.* Suppose that $F_1 = \{(f_\alpha^1, p_\alpha^1) | \alpha = 1, 2, \ldots, |F_1|\}$ and $F_2 = \{(f_\alpha^2, p_\alpha^2) | \alpha = 1, 2, \ldots, |F_2|\}$ are PHFSs with $|F_1| = |F_2|$, then the distance between these two PHFSs is computed as

$$d(F_1, F_2) = \sqrt{\frac{\sum_{\alpha=1}^{|F_1|} \left(f_\alpha^1 p_\alpha^1 - f_\alpha^2 p_\alpha^2\right)^2}{|F_1|}}, \tag{4}$$

where $|F_1|$ and $|F_2|$ are the numbers of elements in $F_1$ and $F_2$.

## 4. Methodology

In this section, an evaluation criteria system is put forward according to the requirements for supporting large-scale IoT data stream processing, and a novel probabilistic hesitant fuzzy best-worst method is proposed to determine the importance degrees of criteria. Finally, to select the most suitable one from four DSPFs, we put forward a novel probabilistic hesitant fuzzy MULTIMOORA method.

*4.1. Evaluation Criteria Set.* To comprehensively characterize the requirements for large-scale IoT data stream processing, we need to establish a hybrid evaluation criteria system as shown in Figure 1.

It can be seen that this evaluation criteria system consists of four qualitative criteria and three quantitative criteria. We give a description of these seven criteria as follows:

*4.1.1. Maintainability.* The criterion maintainability measures the ease with which the DSPFs can be changed so that they can be compatible with the existing IT systems of enterprises and adapt to the change of the existing IT systems.

*4.1.2. Developer Friendliness.* The developer friendliness measures the ease for developers to deploy the DSPFs and program so as to perform the large-scale IoT data stream processing. It is measured from the following four aspects: (1) ease of understanding this model, documentation, and code; (2) number of parameters, which should be tuned; (3) job history and debuggability; (4) APIs.

*4.1.3. Framework Complexity.* The criterion complexity measures the ease of operations of DSPFs and their compatibilities. It can be measured from four aspects: (1) ease of setup and monitoring; (2) the complexity of dependencies; (3) version limitations; (4) multitenancy support.

*4.1.4. Framework Maturity.* This criterion can measure the maturity of an organization's streaming framework development process. It can be measured from the following factors: (1) community support; (2) number of contributions to the open-source communities; (3) how fast a bug is fixed; (4) release frequency; (5) notable companies using the streaming framework.

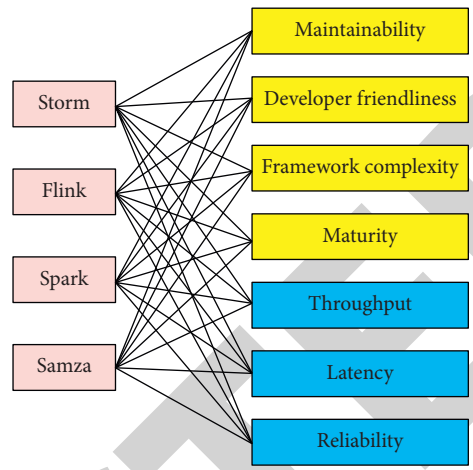



Figure 1: Evaluating four streaming frameworks according to the evaluation criteria system.

*4.1.5. Throughput.* It is an important metric for evaluating the performances of streaming frameworks. It measures the rate at which streaming events are processed per second.

*4.1.6. Latency.* It is an important metric for measuring the real-time feature of streaming frameworks. It actually measures the time that is taken to complete one event.

*4.1.7. Reliability.* The criterion is a metric used to measure the probability that streaming frameworks experience crashes or failures during a given amount of time.

As shown in Figure 1, it can be seen that evaluating these four DSPFs with respect to the evaluation criteria system should be formulated to be an MCDM problem, in which four DSPFs are denoted as

$$A = \{a_1 = \text{Storm}, a_2 = \text{Flink}, a_3 = \text{Spark}, a_4 = \text{Samza}\}, \tag{5}$$

and seven criteria are denoted as

$$\begin{aligned} C = \{&c_1 = \text{maintainability}, c_2 = \text{developer friendliness}, c_3 \\ &= \text{framework complexity}, c_4 = \text{maturity}, c_5 \\ &= \text{throughput}, c_6 = \text{latency}, \text{ and } c_7 = \text{reliability}\}. \end{aligned} \tag{6}$$

Therefore, evaluating these four DSPFs with respect to this evaluation criteria system can be transformed into solving the above MCDM problem. To evaluate these four DSPFs, the enterprise establishes the technical committee, which is composed of ten experts denoted as $\{D_1, D_2, \ldots, D_{10}\}$. Each expert chooses one linguistic term from the following LTS $S = \{s_{-2} = \text{very low}, s_{-1} = \text{low}, s_0 = \text{Neutral}, s_1 = \text{high}, \text{and } s_2 = \text{very high}\}$ to express his/her preference information over each DSPF with respect to each criterion. We can derive the group preference information of each DSPF with respect to each criterion using the following definition:

*Definition 7.* Let $S = \{s_\rho | \rho = -2, -1, 0, 1, 2\}$ be an LTS and $E_e = \{\ell^e\}$ $(e = 1, 2, \ldots, 10)$ be the preference information of the expert $D_e$, then the group preference information over each DSPF with respect to each criterion can be derived as

$$H = \left\{ (L_\alpha, p_\alpha) | L_\alpha \in E_1 \cup E_2, \ldots, \cup E_{10}, p_\alpha = \frac{\sum_{e=1}^{10} n_e}{10} \right\}, \tag{7}$$

with

$$n_e = \begin{cases} 1, & \text{if } L_\alpha \in E_e, \\ 0, & \text{if } L_\alpha \notin E_e, \end{cases} \tag{8}$$

where the group preference information $H$ is actually a PLTS.

All the obtained PLTSs are used to construct a probabilistic linguistic decision matrix (PLDM) $H_{4\times7}$ as

$$H_{4\times7} = \begin{array}{c} \\ a_1 \\ a_2 \\ a_3 \\ a_4 \end{array} \begin{bmatrix} c_1 & c_2 & \cdots & c_7 \\ H_{11} & H_{12} & \cdots & H_{17} \\ H_{21} & H_{22} & \cdots & H_{27} \\ H_{31} & H_{32} & \cdots & \vdots \\ H_{41} & H_{42} & \cdots & H_{47} \end{bmatrix}, \tag{9}$$

where the element $H_{ij}$ is a PLTS and it is the group preference information of the DSPF $a_i$ with respect to criterion $c_j$.

In order to consider the psychological distances among two consecutive linguistic terms, Definition 3 is used to transform the PLDM $H_{4\times7}$ to a probabilistic hesitant fuzzy decision matrix (PHFDM) $F_{4\times7}$.

*4.2. Probabilistic Hesitant Fuzzy Best-Worst Method.* The best-worst method [63] is a subjective method, which is used to determine the importance of degrees of criteria according to the preference information from the organization. Compared with the AHP (analytic hierarchy process), the best-worst method requires less times for pairwise comparisons among the streaming frameworks. Moreover, it is easier to be understood. Because of these advantages, the best-worst method is extended to develop a subjective probabilistic hesitant fuzzy best-worst (PHFBW) method, whose steps are summarized as follows:

(i) *Step 1.* The most important criterion $c_b$ and least important criterion $c_w$ should be determined by the technical committee from the evaluation criteria set as follows:

$$C = \{c_1 = \text{maintainability}, c_2 = \text{developer friendliness}, c_3 = \text{framework complexity}, c_4 = \text{maturity}, c_5 = \text{throughput}, c_6 = \text{latency, and } c_7 = \text{reliability}\}. \tag{10}$$

(ii) *Step 2.* Each expert $D_e$ from the technical committee (TC) evaluates the intensity of the most important criterion $c_b$ over other criteria using the following LTS:

$$S = \{s_{-2} = \text{very low}, s_{-1} = \text{low}, s_0 = \text{Neutral}, s_1 = \text{high}, s_2 = \text{very high}\}, \tag{11}$$

and then obtain the most-to-all (MtA) vector as

$$\text{MtA}_e = \left(\ell^e_{b1}, \ell^e_{b2}, \ldots, \ell^e_{bj}, \ldots, \ell^e_{b7}\right), \tag{12}$$

where $\ell^e_{bj}$, a linguistic term from $S$, is the intensity of the most important criterion $c_b$ over the criterion $c_j$.

(iii) *Step 3.* Each expert $D_e$ from the technical committee need to assess the intensity of each criterion over the least important criterion $c_w$ using the LTS $S$ and obtain the all-to-least (AtL) vector as

$$\text{AtL}_e = \left(\ell^e_{1w}, \ell^e_{2w}, \ldots, \ell^e_{jw}, \ldots, \ell^e_{7w}\right), \tag{13}$$

where $\ell^e_{jw}$, a linguistic term from $S$, represents the intensity of each criterion $c_j$ over the least important criterion $c_w$.

(iv) *Step 4.* Definition 7 is used to aggregate the preference information of ten experts and obtain the following probabilistic linguistic MtA (PLMtA) vector as follows:

$$\text{PLMtA} = \left(H_{b1}, H_{b2}, \ldots, H_{bj}, \ldots, H_{b7}\right), \tag{14}$$

where $H_{bj}$ denotes a PLTS and it means the group preference information about the intensity of the most important criterion $c_b$ over the criterion $c_j$.

(v) *Step 5.* Definition 7 is used to aggregate the preference information of ten experts and obtain the following probabilistic linguistic AtL (PLAtL) vector as follows:

$$\text{PLAtL} = \left(H_{1w}, H_{2w}, \ldots, H_{jw}, \ldots, H_{7w}\right), \tag{15}$$

where $H_{jw}$ denotes a PLTS and it means the group preference information on the intensity of the criterion $c_j$ over the least important criterion $c_w$.

(vi) *Step 6.* Definition 3 is used to transform the PLMtA vector into probabilistic hesitant fuzzy MtA (PHFMtA) vector as

$$\text{PHFMtA} = \left(F_{b1}, F_{b2}, \ldots, F_{bj}, \ldots, F_{b7}\right), \tag{16}$$

where $F_{bj}$ denotes a PHFS.

(vii) *Step 7*. Definition 3 is used to transform the PLAtL vector into the probabilistic hesitant fuzzy AtL (PHFAtL) vector as

$$\text{PHFAtL} = \left(F_{1w}, F_{2w}, \ldots, F_{jw}, \ldots, F_{7w}\right), \tag{17}$$

where $F_{jw}$ denotes a PHFS.

(viii) *Step 8*. The elements in the PHFMtA and PHFAtL vectors are the PHFSs that are complex information structures. To facilitate the computation process, we use the score values of PHFSs to approximately represent PHFSs in the PHFMtA and PHFAtL vectors. The above two vectors can be transformed into

$$\begin{aligned} S(\text{PHFMtA}) &= \left(S(F_{b1}), S(F_{b2}), \ldots, S\left(F_{bj}\right), \ldots, S(F_{b7})\right), \\ S(\text{PHFAtL}) &= \left(S(F_{1w}), S(F_{2w}), \ldots, S\left(F_{jw}\right), \ldots, S(F_{7w})\right), \end{aligned} \tag{18}$$

where $S(F_{bj})$ and $S(F_{jw})$ are the score values of PHFSs $F_{bj}$ and $F_{jw}$.

(ix) *Step 9*. If the PHFMtA and PHFAtL vectors are completely consistent, the weights of criteria should satisfy the following formulas: $\omega_b/(\omega_b + \omega_j) = S(F_{bj})$ and $\omega_j/(\omega_j + \omega_w) = S(F_{jw})$.

In fact, the PHFMtA and PHFAtL vectors cannot satisfy the condition of completely consistent. Thus, the optimal weights of criteria should satisfy Model 1.

$$\min \max_j \left\{ \left| \frac{\omega_b}{\left(\omega_b + \omega_j\right)} - S\left(F_{bj}\right) \right|, \left| \frac{\omega_j}{\left(\omega_j + \omega_w\right)} - S\left(F_{jw}\right) \right| \right\}. \tag{19}$$

To obtain the solutions from Model 1, a slack variable $\xi$ is introduced. Then, Model 1 is equivalently transformed into Model 2

$$\begin{aligned} &\min \ \xi, \\ &\sum_{j=1}^{n} \omega_j = 1, \quad \omega_j \geq 0, j = 1, 2, \ldots, n, \\ &s.t. \quad \left| \frac{\omega_b}{\left(\omega_b + \omega_j\right)} - S\left(F_{bj}\right) \right| \leq \xi, \\ &\left| \frac{\omega_j}{\left(\omega_j + \omega_w\right)} - S\left(F_{jw}\right) \right| \leq \xi. \end{aligned} \tag{20}$$

Then, the weights of the above seven criteria can be derived by solving Model 2.

The advantage of this subjective method for determining the weights is that the technical committee can determine the most and least important criteria according to their requirements for large-scale IoT data stream processing, and that can reflect the intensities of the most important criterion over others, and the intensities of the criteria over the least important criterion. Therefore, this subjective method can integrate with the group preference information from experts to prioritize criteria reasonably according to their special requirements.

*4.3. Probabilistic Hesitant Fuzzy MULTIMOORA Method.* The MULTIMOORA method [64] uses the ratio subsystem (RS), reference point subsystem (RPS), and full-multiplicative form subsystem (FMFS) to obtain ranking values and ranking results. For determining the final ranking result, the dominancy theory is used to aggregate the ranking values and ranking results of three subsystems. The experimental results in [65] showed that the MULTIMOORA method obtains better decision performance than some well-known decision-making methods. However, it has not been extended to process the PHFS information. In this subsection, we put forward a novel probabilistic hesitant fuzzy MULTIMOORA (PHF-MULTIMOORA) method to rank four DSPFs with respect to their criteria. The steps are listed as

(i) *Step 1*. The RS model is used to compute the ranking values of four DSPFs as

$$R_1(a_i) = \sum_{j=1}^{n_b} \omega_j S\left(F_{ij}\right) - \sum_{j=n_b+1}^{7} \omega_j S\left(F_{ij}\right), \tag{21}$$

where $R_1(a_i)$ is the ranking value of the DSPF $a_i$ by using the RS model, $n_b$ is the number of benefit-type criteria that have positive impacts on the ranking value and $7 - n_b$ is the number of cost-type criteria that show negative impacts on the ranking value. The DSPF having a higher ranking value is better, hence these DSPFs are prioritized according to the descending order of the ranking values, and then, the ranking order of these four DSPFs is determined as $O_1 = \{o_1(a_1), o_1(a_2), o_1(a_3), o_1(a_4)\}$, where $o_1(a_i)$ is the position of the DSPF $a_i$.

(ii) *Step 2*. The RPS model is used to derive the ranking values of four DSPFs as

$$R_2(a_i) = \max_{j \in \{1,2,\ldots,7\}} \omega_j \frac{d\left(F_j^+, F_{ij}\right)}{d\left(F_j^+, F_j^-\right)}, \tag{22}$$

where $F_j^+$ and $F_j^-$ denote the best and worst values of DSPFs with respect to the criterion $c_j$. They can be computed as follows:the best value of DSPFs with respect to criterion $c_j$ can be determined as

$$F_j^+ = \begin{cases} \max\limits_{i\in\{1,2,3,4\}} \{F_{ij}\}, & \text{if criterion} c_j \text{ is benefit} - \text{type}, \\ \min\limits_{i\in\{1,2,3,4\}} \{F_{ij}\}, & \text{if criterion} c_j \text{ is cost} - \text{type}, \end{cases} \tag{23}$$

and the worst value of DSPFs with respect to criterion $c_j$ can be determined as

$$F_j^- = \begin{cases} \max\limits_{i\in\{1,2,3,4\}} \{F_{ij}\}, & \text{if criterion} c_j \text{ is cost} - \text{type}, \\ \min\limits_{i\in\{1,2,3,4\}} \{F_{ij}\}, & \text{if criterion} c_j \text{ is benefit} - \text{type}, \end{cases} \tag{24}$$

where $\max\limits_{i\in\{1,2,3,4\}} \{F_{ij}\}$ denotes $F_{ij}$ with the largest score value and $\min\limits_{i\in\{1,2,3,4\}} \{F_{ij}\}$ is $F_{ij}$ having the smallest score value.

The DSPF having the smaller ranking value is better. Therefore, these four DSPFs can be ranked according to the ascending order of their ranking values and then the ranking order of these four DSPFs is determined as $O_2 = \{o_2(a_1), o_2(a_2), o_2(a_3), o_2(a_4)\}$, where $o_2(a_i)$ is the position of the DSPF $a_i$.

(iii) *Step 3.* The FMFS model is used to compute the ranking values of four DSPFs as

$$R_3(a_i) = \frac{\sqrt[n_b]{\prod_{j=1}^{n_b}\left(2-\left(1-S\left(F_{ij}\right)\right)^{\omega_j}\right)}}{\sqrt[7-n_b]{\prod_{j=n_b+1}^{7}\left(2-\left(1-S\left(F_{ij}\right)\right)^{\omega_j}\right)}}. \tag{25}$$

The DSPF owning a larger ranking value is better, thus these four DSPFs should be prioritized according to the descending order of their ranking values. The ranking order of these four DSPFs can be determined as $O_3 = \{o_3(a_1), o_3(a_2), o_3(a_3), \text{ and } o_3(a_4)\}$.

(iv) *Step 4.* Aggregate the ranking values and ranking orders of three subsystems into the final ranking values.

In the original MULTIMOORA method [64], the dominance theory was applied to aggregate the ranking orders of subsystems. However, it does not consider their ranking values [66]. In this paper, a novel Borda is extended to aggregate the ranking values and ranking orders from three subsystems. Therefore, RS ($Q_1$), RPS ($Q_2$), and FMFS ($Q_3$) are considered as three criteria of DSPFs, and these four DSPFs are associated with the ranking values $R_k(a_i)$ and ranking orders $o_k(a_i)$ with respect to three criteria $Q_k (k = 1, 2, 3)$. The fusion of these ranking values and ranking orders from three subsystems can be transformed into the problem that how to fuse two matrices: ranking value matrix $R = (R_k(a_i))_{4\times3}$ and ranking order matrix $O = (o_k(a_i))_{4\times3}$.

$$R = (R_k(a_i))_{4\times3} = \begin{array}{c} \\ a_1 \\ a_2 \\ a_3 \\ a_4 \end{array} \begin{bmatrix} Q_1 & Q_2 & Q_3 \\ R_1(a_1) & R_2(a_1) & R_3(a_1) \\ R_1(a_2) & R_2(a_2) & R_3(a_2) \\ R_1(a_3) & R_2(a_3) & R_3(a_3) \\ R_1(a_4) & R_2(a_4) & R_3(a_4) \end{bmatrix},$$
$$O = (o_k(a_i))_{4\times3} = \begin{array}{c} \\ a_1 \\ a_2 \\ a_3 \\ a_4 \end{array} \begin{bmatrix} Q_1 & Q_2 & Q_3 \\ o_1(a_1) & o_2(a_1) & o_3(a_1) \\ o_1(a_2) & o_2(a_2) & o_3(a_2) \\ o_1(a_3) & o_2(a_3) & o_3(a_3) \\ o_1(a_4) & o_2(a_4) & o_3(a_4) \end{bmatrix}. \tag{26}$$

Before computing the final ranking values of DSPFs, the ranking value matrix should be normalized to be $\widetilde{R} = (\widetilde{R}_k(a_i))_{4\times3}$, where the element $\widetilde{R}_k(a_i) = R_k(a_i)/\sqrt{\sum_{i=1}^{4}(R_k(a_i))^2}$.

According to the Borda rule [66], the DSPF with a larger value is better. However, in the RPS, the DSPF with a smaller value is better. It is in conflict with the Borda rule. Therefore, the final ranking value $f(a_i)$ of the DSPF $a_i$ is calculated as

$$f(a_i) = \widetilde{R}_1(a_i) * \frac{4 - o_1(a_i) + 1}{4(4+1)/2} - \widetilde{R}_2(a_i) * \frac{o_2(a_i)}{4(4+1)/2} + \widetilde{R}_3(a_i) * \frac{4 - o_3(a_i) + 1}{4(4+1)/2}. \tag{27}$$

From the above equation, it can be noted that the DSPF with a higher ranking value is better. Thus, the final ranking order of DSPFs is derived according to the descending order of the final ranking values $f(a_i)$.

## 5. Numerical Analysis

In this section, the numerical analysis is presented to show the implementation process of the proposed PHF-BW method and PHF-MULTIMOORA method.

*5.1. How to Determine the Importance Degrees of Criteria.* According to the steps of the PHF-BW method, the process for determining importance degrees of criteria is implemented as

(i) *Step 1.* According to the requirements for large-scale IoT data stream processing, the technical committee selects the criterion throughput ($c_5$) as the most important criterion and then selects the criterion framework complexity ($c_3$) as the least important one from the evaluation criteria system.

(ii) *Steps 2–5.* Ten experts in the technical committee provide their preference information on the intensity of the criterion $c_5$ over the other criteria and

intensity of all the criteria over the criterion $c_3$ and construct the PLMtA and PLAtL vectors as

$$\begin{aligned}\text{PLMtA} &= (\{s_0(0.2), s_2(0.8)\}, \{s_1(0.3), s_2(0.7)\}, \{s_2(1)\}, \{s_1(0.8), s_2(0.2)\}, \{s_0(1)\}, \{s_0(0.6), s_1(0.4)\}, \{s_0(0.5), s_1(0.5)\}),\\ \text{PLAtL} &= (\{s_0(0.3), s_1(0.7)\}, \{s_1(0.9), s_2(0.1)\}, \{s_0(1)\}, \{s_0(0.6), s_2(0.4)\}, \{s_2(1)\}, \{s_1(0.2), s_2(0.8)\}, \{s_1(0.5), s_2(0.5)\}).\end{aligned} \tag{28}$$

(iii) *Steps 6–7*. Definition 3 is used to transform the PLMtA vector into PHFMtA vector and PLAtL vector into PHFAtL vector as

$$\begin{aligned}\text{PHFMtA} &= (\{(0.5, 0.2), (1, 0.8)\}, \{(0.75, 0.3), (1, 0.7)\}, \{(1, 1)\}, \{(0.75, 0.8), (1, 0.2)\}, \{(0.5, 1)\}, \{(0.5, 0.6), (0.75, 0.4)\},\\ &\quad \{(0.5, 0.5), (0.75, 0.5)\}),\\ \text{PHFAtL} &= (\{(0.5, 0.3), (0.75, 0.7)\}, \{(0.75, 0.9), (1, 0.1)\}, \{(0.5, 1)\}, \{(0.5, 0.6), (1, 0.4)\}, \{(1, 1)\}, \{(0.75, 0.2), (1, 0.8)\},\\ &\quad \{(0.75, 0.5), (1, 0.5)\}).\end{aligned} \tag{29}$$

(iv) *Step 8*. Definition 4 is used to compute the score values of the elements in the PHFMtA vector and PHFAtL vector as

$$\begin{aligned}S(\text{PHFMtA}) &= (0.9, 0.925, 1, 0.8, 0.5, 0.6, 0.625),\\ S(\text{PHFAtL}) &= (0.675, 0.775, 0.5, 0.7, 1, 0.95, 0.875).\end{aligned} \tag{30}$$

(v) *Step 9*. The above score values is brought into Model 2.

By solving Model 2, the weights of the seven criteria are derived as shown in Figure 2.

From Figure 2, it can be noted that the most important criterion is throughput ($c_5$) followed by latency ($c_6$) and reliability ($c_7$). The least important criterion is framework complexity ($c_3$).

*5.2. How to Rank DSPFs.* Ten experts in the technical committee are called to evaluate four DSPFs with respect to four qualitative criteria by using the LTS $S$ and Definition 7 is applied to aggregate the individual preference information for constructing their group preference information. The preference information of four DSPFs with respect to three quantitative criteria are derived from the benchmarking results that are presented in Ref. [67]. To make the information representation be not different, the information of throughput, latency, and reliability are expressed by using the PLTSs. Finally, all the group preference information for qualitative and quantitative criteria are applied to construct the PLDM $H_{4\times7}$ as shown in Table 1.

For the evaluation criteria system, the framework complexity and latency are cost-type criteria and others are benefit-type.

(i) *Step 1*. The RS model is used to compute the ranking values of four DSPFs as

$$\begin{aligned}&R_1(a_1) = 0.6294,\\ &R_1(a_2) = 0.7126,\\ &R_1(a_3) = 0.2391,\\ &R_1(a_4) = 0.3935,\\ &O_1 = \{o_1(a_2), o_1(a_1), o_1(a_4), o_1(a_3)\}, \quad a_2 > a_1 > a_4 > a_3.\end{aligned} \tag{31}$$

(ii) *Step 2*. The RPS model is used to compute the ranking values of four DSPFs as

$$\begin{aligned}&R_2(a_1) = 0.3600,\\ &R_2(a_2) = 0.0673,\\ &R_2(a_3) = 0.2610,\\ &R_2(a_4) = 0.3600,\\ &O_2 = \{o_2(a_2), o_2(a_3), o_2(a_1) \sim o_2(a_4)\}, \quad a_2 > a_3 > a_1 \sim a_4.\end{aligned} \tag{32}$$

(iii) *Step 3*. The FMFS model is used to compute the ranking values of four DSPFs as

$$\begin{aligned}&R_3(a_1) = 1.6615,\\ &R_3(a_2) = 1.4345,\\ &R_3(a_3) = 0.8650,\\ &R_3(a_4) = 1.1665,\\ &O_3 = \{o_3(a_1), o_3(a_2), o_3(a_4), o_3(a_3)\}, \quad a_1 > a_2 > a_4 > a_3.\end{aligned} \tag{33}$$

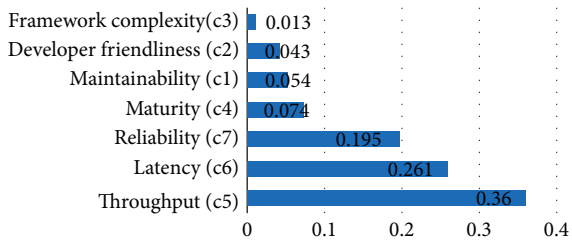



Figure 2: Importance degrees of criteria.

Table 1: The group preference information.

| | Storm | Flink | Spark | Samza |
|---|---|---|---|---|
| Maintainability | $\left\{\begin{matrix}(s_2, 0.8),\\(s_1, 0.2)\end{matrix}\right\}$ | $\left\{\begin{matrix}(s_2, 0.9),\\(s_1, 0.1)\end{matrix}\right\}$ | $\left\{\begin{matrix}(s_2, 0.9),\\(s_1, 0.1)\end{matrix}\right\}$ | $\left\{\begin{matrix}(s_2, 0.8),\\(s_1, 0.2)\end{matrix}\right\}$ |
| Developer friendliness | $\{(s_2, 1.0)\}$ | $\left\{\begin{matrix}(s_2, 0.7),\\(s_1, 0.3)\end{matrix}\right\}$ | $\{(s_1, 1.0)\}$ | $\left\{\begin{matrix}(s_1, 0.8),\\(s_0, 0.2)\end{matrix}\right\}$ |
| Framework complexity | $\left\{\begin{matrix}(s_{-1}, 0.8),\\(s_0, 0.2)\end{matrix}\right\}$ | $\{(s_{-1}, 1.0)\}$ | $\left\{\begin{matrix}(s_{-2}, 0.9),\\(s_{-1}, 0.1)\end{matrix}\right\}$ | $\{(s_1, 1.0)\}$ |
| Maturity | $\{(s_2, 1.0)\}$ | $\left\{\begin{matrix}(s_2, 0.7),\\(s_1, 0.3)\end{matrix}\right\}$ | $\{(s_1, 1.0)\}$ | $\left\{\begin{matrix}(s_1, 0.8),\\(s_0, 0.2)\end{matrix}\right\}$ |
| Throughput | $\{(s_1, 1.0)\}$ | $\{(s_2, 1.0)\}$ | $\{(s_2, 1.0)\}$ | $\{(s_1, 1.0)\}$ |
| Latency | $\{(s_{-2}, 1.0)\}$ | $\{(s_{-2}, 1.0)\}$ | $\{(s_2, 1.0)\}$ | $\{(s_{-2}, 1.0)\}$ |
| Reliability | $\{(s_2, 1.0)\}$ | $\{(s_2, 1.0)\}$ | $\{(s_{-2}, 1.0)\}$ | $\{(s_1, 1.0)\}$ |

Table 2: The ranking orders of these decision-making methods.

| Methods | Rank orders |
|---|---|
| Our proposed method | $a_2 > a_1 > a_4 > a_3$ |
| VIKOR | $a_2 > a_1 \sim a_4 > a_3$ ($a_1, a_2$, and $a_4$ are the compromise solutions) |
| TOPSIS | $a_2 > a_4 > a_3 > a_1$ |

(iv) *Step 4.* The ranking values and ranking orders of three subsystems are aggregated into the final ranking values as

$$\begin{aligned} f(a_1) &= 0.2438, \\ f(a_2) &= 0.4217, \\ f(a_3) &= -0.0351, \\ f(a_4) &= -0.0243. \end{aligned} \tag{34}$$

Hence, the final ranking order of DSPFs is $a_2 > a_1 > a_4 > a_3$ and the most suitable DSPF is Apache Flink. The Flink shows equal or better performances than the other three DSPFs in terms of throughput, latency, and reliability. As for the benchmarking testing results, Flink did not experience any crashes or failures. Moreover, Flink has enriched community support that can make subsequent development, deployment, and maintenance well. Thus, it can be seen that the result of our proposed PHF-MULTIMOORA method is reasonable.

From the implementation process, it can be noted that the three models achieve different ranking orders. Our proposed PHF-MULTIMOORA method can fuse these different ranking orders into the final one. Therefore, the final ranking order is more reliable and robust.

*5.3. Comparative Analysis.* To show the superiority of the proposed PHF-MULTIMOORA method, we compare the proposed PHF-MULTIMOORA method with the existing TOPSIS and VIKOR methods [68]. We use the existing TOPSIS and VIKOR methods in Ref. [68] to handle the PLDM $H_{4\times7}$ in Table 1. The ranking orders of these two methods are listed in Table 2.

As shown in Table 2, the best DSPF obtained from our proposed method is $a_2$, which is the same as that of the existing TOPSIS method. However, their ranking orders are different. The existing VIKOR method has three compromise solutions $a_1, a_2$, and $a_4$. It does not have a unique solution.

## 6. Conclusions

In this study, an evaluation criteria system consisting of seven criteria is proposed for characterizing the requirements of ranking the DSPFs, and the process of ranking the DSPFs with respect to the evaluation criteria system is

formulated as an MCDM problem. A novel PHF-BW method is proposed to derive the weight values of seven criteria, and a novel PHF-MULTIMOORA method is proposed to rank these four DSPFs. The results from the numerical analysis show that the most important criterion is throughput followed by low latency and high reliability. Flink is selected as the most suitable DSPF. It is easy to extend this study for evaluating other IT systems according to the special requirements of enterprises.

In future research, we plan to combine the subjective method and objective method for determining the weight values of criteria and use picture fuzzy sets for accurately modelling the collective opinions.

## Data Availability

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

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## References

[1] M. K. Pandit, R. N. Mir, and M. A. Chishti, "Adaptive task scheduling in IoT using reinforcement learning," *International Journal of Intelligent Computing and Cybernetics*, vol. 13, no. 3, pp. 261–282, 2020.

[2] O. Elijah, T. A. Rahman, I. Orikumhi, C. Y. Leow, and M. H. D. N. Hindia, "An overview of Internet of Things (IoT) and data analytics in agriculture: benefits and challenges," *IEEE Internet of Things Journal*, vol. 5, no. 5, pp. 3758–3773, 2018.

[3] M. Lin, C. Huang, Z. Xu, and R. Chen, "Evaluating IoT platforms using integrated probabilistic linguistic MCDM method," *IEEE Internet of Things Journal*, vol. 7, no. 11, pp. 11195–11208, 2020.

[4] A. Nazir, R. N. Mir, and S. Qureshi, "Exploring compression and parallelization techniques for distribution of deep neural networks over edge-fog continuum-a review," *International Journal of Intelligent Computing and Cybernetics*, vol. 13, no. 3, pp. 331–364, 2020.

[5] M. Chernyshev, Z. Baig, O. Bello, and S. Zeadally, "Internet of things (IoT): research, simulators, and testbeds," *IEEE Internet of Things Journal*, vol. 5, no. 3, pp. 1637–1647, 2018.

[6] H. Isah, T. Abughofa, S. Mahfuz, D. Ajerla, F. Zulkernine, and S. Khan, "A survey of distributed data stream processing frameworks," *IEEE Access*, vol. 7, pp. 154300–154316, 2019.

[7] J.-H. Choi, J. Park, H. D. Park, and O.-G. Min, "DART: fast and efficient distributed stream processing framework for internet of things," *ETRI Journal*, vol. 39, no. 2, pp. 202–212, 2017.

[8] F. E. Ayo, O. Folorunso, F. T. Ibharalu, and I. A. Osinuga, "Hate speech detection in Twitter using hybrid embeddings and improved cuckoo search-based neural networks," *International Journal of Intelligent Computing and Cybernetics*, vol. 13, no. 4, pp. 485–525, 2020.

[9] A. I. Maarala, X. Su, and J. Riekki, "Semantic reasoning for context-aware internet of things applications," *IEEE Internet of Things Journal*, vol. 4, no. 2, pp. 461–473, 2017.

[10] T. Li, J. Tang, and J. Xu, "Performance modeling and predictive scheduling for distributed stream data processing," *IEEE Transactions on Big Data*, vol. 2, no. 4, pp. 353–364, 2016.

[11] S. Dolev, P. Florissi, E. Gudes, S. Sharma, and I. Singer, "A survey on geographically distributed big-data processing using MapReduce," *IEEE Transactions on Big Data*, vol. 5, no. 1, pp. 60–80, 2019.

[12] M. A. Jan, M. Usman, X. He, and A. Ur Rehman, "SAMS: a seamless and authorized multimedia streaming framework for WMSN-based IoMT," *IEEE Internet of Things Journal*, vol. 6, no. 2, pp. 1576–1583, 2019.

[13] F. Liu, B. Li, L. Zhong, B. Li, H. Jin, and X. Liao, "Flash crowd in P2P live streaming systems: fundamental characteristics and design implications," *IEEE Transactions on Parallel and Distributed Systems*, vol. 23, no. 7, pp. 1227–1239, 2012.

[14] F. Carcillo, A. Dal Pozzolo, Y.-A. Le Borgne, O. Caelen, Y. Mazzer, and G. Bontempi, "SCARFF: a scalable framework for streaming credit card fraud detection with spark," *Information Fusion*, vol. 41, pp. 182–194, 2018.

[15] S. Rathee and A. Kashyap, "StreamAligner: a streaming based sequence aligner on Apache Spark," *Journal of Big Data*, vol. 5, no. 1, 8 pages, 2018.

[16] R. Sahal, M. H. Khafagy, and F. A. Omara, "Big data multi-query optimisation with Apache Flink," *International Journal of Web Engineering and Technology*, vol. 13, no. 1, pp. 78–97, 2018.

[17] C. V. Espinosa, E. Martin-Martin, A. Riesco, and J. Rodriguez-Hortala, "FlinkCheck: property-based testing for Apache flink," *IEEE Access*, vol. 7, pp. 150369–150382, 2019.

[18] L. R. Nair, S. D. Shetty, and S. D. Shetty, "Applying spark based machine learning model on streaming big data for health status prediction," *Computers & Electrical Engineering*, vol. 65, pp. 393–399, 2018.

[19] F.-H. Tseng, J.-H. Hsueh, C.-W. Tseng, Y.-T. Yang, H.-C. Chao, and L.-D. Chou, "Congestion prediction with big data for real-time highway traffic," *IEEE Access*, vol. 6, pp. 57311–57323, 2018.

[20] J. R. Sutton, R. Mahajan, O. Akbilgic, and R. Kamaleswaran, "PhysOnline: an open source machine learning pipeline for real-time analysis of streaming physiological waveform," *IEEE Journal of Biomedical and Health Informatics*, vol. 23, no. 1, pp. 59–65, 2019.

[21] M. Lin, X. Li, and L. Chen, "Linguistic q-rung orthopair fuzzy sets and their interactional partitioned Heronian mean aggregation operators," *International Journal of Intelligent Systems*, vol. 35, no. 2, pp. 217–249, 2020.

[22] L. A. Zadeh, "Fuzzy sets," *Information and Control*, vol. 8, no. 3, pp. 338–353, 1965.

[23] V. Torra, "Hesitant fuzzy sets," *International Journal of Intelligent Systems*, vol. 25, no. 6, pp. 529–539, 2010.

[24] Z. Xu and W. Zhou, "Consensus building with a group of decision makers under the hesitant probabilistic fuzzy environment," *Fuzzy Optimization and Decision Making*, vol. 16, no. 4, pp. 481–503, 2017.

[25] B. Zhu and Z. Xu, "Probability-hesitant fuzzy sets and the representation of preference relations," *Technological and Economic Development of Economy*, vol. 24, no. 3, pp. 1029–1040, 2018.

[26] W. Zhou and Z. Xu, "Probability calculation and element optimization of probabilistic hesitant fuzzy preference relations based on expected consistency," *IEEE Transactions on Fuzzy Systems*, vol. 26, no. 3, pp. 1367–1378, 2018.

[27] M. Lin, Q. Zhan, and Z. Xu, "Decision making with probabilistic hesitant fuzzy information based on multiplicative

consistency," *International Journal of Intelligent Systems*, vol. 35, no. 8, pp. 1233–1261, 2020.

[28] M. Lin, Q. Zhan, Z. Xu, and R. Chen, "Group decision-making model with hesitant multiplicative preference relations based on regression method and feedback mechanism," *IEEE Access*, vol. 6, pp. 61130–61150, 2018.

[29] M. Lin, Q. Zhan, Z. Xu, and R. Chen, "Group decision making with probabilistic hesitant multiplicative preference relations based on consistency and consensus," *IEEE Access*, vol. 6, pp. 63329–63344, 2018.

[30] L. Wang and N. Li, "Pythagorean fuzzy interaction power Bonferroni mean aggregation operators in multiple attribute decision making," *International Journal of Intelligent Systems*, vol. 35, no. 1, pp. 150–183, 2020.

[31] L. Wang, H. Garg, and N. Li, "Pythagorean fuzzy interactive Hamacher power aggregation operators for assessment of express service quality with entropy weight," *Soft Computing*, 2020.

[32] D. D. Luo, S. Z. Zeng, and J. Chen, "A probabilistic linguistic multiple attribute decision making based on a new correlation coefficient method and its application in hospital assessment," *Mathematics*, vol. 8, no. 3, 340 pages, 2020.

[33] S. Zeng, D. Luo, C. Zhang, and X. Li, "A correlation-based TOPSIS method for multiple attribute decision making with single-valued neutrosophic information," *International Journal of Information Technology & Decision Making*, vol. 19, no. 1, pp. 343–358, 2020.

[34] S. Zeng, S.-M. Chen, and K.-Y. Fan, "Interval-valued intuitionistic fuzzy multiple attribute decision making based on nonlinear programming methodology and TOPSIS method," *Information Sciences*, vol. 506, pp. 424–442, 2020.

[35] C. H. Zhang, C. Chen, D. Streimikiene, and T. Balezentis, "Intuitionistic fuzzy multimoora approach for multi-criteria assessment of the energy storage technologies," *Applied Soft Computing*, vol. 79, pp. 410–423, 2020.

[36] M. Lin, Z. Xu, Y. Zhai, and Z. Yao, "Multi-attribute group decision-making under probabilistic uncertain linguistic environment," *Journal of the Operational Research Society*, vol. 69, no. 2, pp. 157–170, 2018.

[37] M. W. Lin, C. Huang, and Z. S. Xu, "Multimoora based MCDM model for site selection of car sharing station under picture fuzzy environment," *Sustainable Cities and Society*, vol. 53, Article ID 101873, 2020.

[38] C. Huang, M. Lin, and Z. Xu, "Pythagorean fuzzy MULTIMOORA method based on distance measure and score function: its application in multicriteria decision making process," *Knowledge and Information Systems*, vol. 62, no. 11, pp. 4373–4406, 2020.

[39] M. Lin, W. Xu, Z. Lin, and R. Chen, "Determine OWA operator weights using kernel density estimation," *Economic Research-Ekonomska Istraživanja*, vol. 33, no. 1, pp. 1441–1464, 2020.

[40] M. Lin, H. Wang, and Z. Xu, "TODIM-based multi-criteria decision-making method with hesitant fuzzy linguistic term sets," *Artificial Intelligence Review*, vol. 53, no. 5, pp. 3647–3671, 2020.

[41] M. Lin, H. Wang, Z. Xu, Z. Yao, and J. Huang, "Clustering algorithms based on correlation coefficients for probabilistic linguistic term sets," *International Journal of Intelligent Systems*, vol. 33, no. 12, pp. 2402–2424, 2018.

[42] M. W. Lin, Z. Y. Chen, Z. S. Xu, X. J. Gou, and F. Herrera, "Score function based on concentration degree for probabilistic linguistic term sets: an application to TOPSIS and VIKOR," *Information Sciences*, 2020.

[43] Z. Ayag and F. Samanlioglu, "A hesitant fuzzy linguistic terms set-based AHP-TOPSIS approach to evaluate ERP software packages," *International Journal of Intelligent Computing and Cybernetics*, 2020.

[44] M. Qiyas, M. A. Khan, S. Khan, and S. Abdullah, "Concept of Yager operators with the picture fuzzy set environment and its application to emergency program selection," *International Journal of Intelligent Computing and Cybernetics*, vol. 13, no. 4, pp. 455–483, 2020.

[45] H. Li, L. Lv, F. Li, L. Wang, and Q. Xia, "A novel approach to emergency risk assessment using FMEA with extended multimoora method under interval-valued Pythagorean fuzzy environment," *International Journal of Intelligent Computing and Cybernetics*, vol. 13, no. 1, pp. 41–65, 2020.

[46] M. Lin, Y. Chen, and R. Chen, "Bibliometric analysis on Pythagorean fuzzy sets during 2013-2020," *International Journal of Intelligent Computing and Cybernetics*, 2020.

[47] G. F. Can and S. Demirok, "Universal usability evaluation by using an integrated fuzzy multi criteria decision making approach," *International Journal of Intelligent Computing and Cybernetics*, vol. 12, no. 2, pp. 194–223, 2019.

[48] K. Chen, P. Chen, L. Yang, and L. Jin, "Grey clustering evaluation based on AHP and interval grey number," *International Journal of Intelligent Computing and Cybernetics*, vol. 12, no. 1, pp. 127–137, 2019.

[49] L. A. Zadeh, "The concept of a linguistic variable and its application to approximate reasoning-I," *Information Sciences*, vol. 8, no. 3, pp. 199–249, 1975.

[50] F. Herrera and L. Martinez, "A 2-tuple fuzzy linguistic representation model for computing with words," *IEEE Transactions on Fuzzy Systems*, vol. 8, no. 6, pp. 746–752, 2000.

[51] Z. Xu and H. Wang, "On the syntax and semantics of virtual linguistic terms for information fusion in decision making," *Information Fusion*, vol. 34, pp. 43–48, 2017.

[52] R. M. Rodriguez, L. Martinez, and F. Herrera, "Hesitant fuzzy linguistic term sets for decision making," *IEEE Transactions on Fuzzy Systems*, vol. 20, no. 1, pp. 109–119, 2012.

[53] H. Wang, "Extended hesitant fuzzy linguistic term sets and their aggregation in group decision making," *International Journal of Computational Intelligence Systems*, vol. 8, no. 1, pp. 14–33, 2015.

[54] H. Liao, X. Wu, X. Liang, J. Xu, and F. Herrera, "A new hesitant fuzzy linguistic oreste method for hybrid multi-criteria decision making," *IEEE Transactions on Fuzzy Systems*, vol. 26, no. 6, pp. 3793–3807, 2018.

[55] Q. Pang, H. Wang, and Z. Xu, "Probabilistic linguistic term sets in multi-attribute group decision making," *Information Sciences*, vol. 369, pp. 128–143, 2016.

[56] M. Lin, Z. Chen, H. Liao, and Z. Xu, "Electre II method to deal with probabilistic linguistic term sets and its application to edge computing," *Nonlinear Dynamics*, vol. 96, no. 3, pp. 2125–2143, 2019.

[57] X. Wu, H. Liao, Z. Xu, A. Hafezalkotob, and F. Herrera, "Probabilistic linguistic multimoora: a multicriteria decision making method based on the probabilistic linguistic expectation function and the improved Borda rule," *IEEE Transactions on Fuzzy Systems*, vol. 26, no. 6, pp. 3688–3702, 2018.

[58] Apache Storm-a distributed stream processing computation framework. http://storm.apache.org.

[59] Apache Flink–Stateful Computations over Data Streams. https://flink.apache.org/.

[60] Apache Spark Is a Unified Analytics Engine for Large-Scale Data Processing. https://spark.apache.org/.

Hindawi
Mathematical Problems in Engineering
Volume 2020, Article ID 6671001, 21 pages
https://doi.org/10.1155/2020/6671001

*Research Article*

# Decision-Making Based on q-Rung Orthopair Fuzzy Soft Rough Sets

**Yinyu Wang**,[1] **Azmat Hussain**,[2] **Tahir Mahmood**,[2] **Muhammad Irfan Ali**,[3] **Hecheng Wu**,[1] **and Yun Jin**[4]

[1]*College of Economic and Management, Nanjing University of Aeronautics and Astronautics, Nanjing 211106, China*
[2]*Department of Mathematics and Statistics, Faculty of Basic and Applied Sciences, International Islamic University Islamabad, Islamabad, Pakistan*
[3]*Department of Mathematics, Islamabad Model College for Boys, G-11/1, Islamabad, Pakistan*
[4]*Wuxi Vocational College of Science and Technology, Wuxi 214028, China*

Correspondence should be addressed to Tahir Mahmood; tahirbakhat@iiu.edu.pk

Received 17 October 2020; Revised 11 November 2020; Accepted 23 November 2020; Published 22 December 2020

Academic Editor: Chonghui Zhang



Recently, Yager presented the new concept of q-rung orthopair fuzzy (q-ROF) set (q-ROFS) which emerged as the most significant generalization of Pythagorean fuzzy set (PFS). From the analysis of q-ROFS, it is clear that the rung q is the most significant feature of this notion. When the rung q increases, the orthopair adjusts in the boundary range which is needed. Thus, the input range of q-ROFS is more flexible, resilient, and suitable than the intuitionistic fuzzy set (IFS) and PFS. The aim of this manuscript is to investigate the hybrid concept of soft set ($S_t$S) and rough set with the notion of q-ROFS to obtain the new notion of q-ROF soft rough (q-ROF$S_t$R) set (q-ROF$S_t$RS). In addition, some averaging aggregation operators such as q-ROF$S_t$R weighted averaging (q-ROF$S_t$RWA), q-ROF$S_t$R ordered weighted averaging (q-ROF$S_t$ROWA), and q-ROF$S_t$R hybrid averaging (q-ROF$S_t$RHA) operators are presented. Then, the basic desirable properties of these investigated averaging operators are discussed in detail. Moreover, we investigated the geometric aggregation operators, such as q-ROF$S_t$R weighted geometric (q-ROF$S_t$RWG), q-ROF$S_t$R ordered weighted geometric (q-ROF$S_t$ROWG), and q-ROF$S_t$R hybrid geometric (q-ROF$S_t$RHG) operators, and proposed the basic desirable characteristics of the investigated geometric operators. The technique for multicriteria decision-making (MCDM) and the stepwise algorithm for decision-making by utilizing the proposed approaches are demonstrated clearly. Finally, a numerical example for the developed approach is presented and a comparative study of the investigated models with some existing methods is brought to light in detail which shows that the initiated models are more effective and useful than the existing methodologies.

## 1. Introduction

Decision-making has always been a hot topic under consideration by the researchers. Multicriteria decision-making (MCDM) has a high prospective and discipline manner to improve and evaluate multiple conflicting criteria in all areas of decision-making. In this competitive environment, an enterprise needs the most accurate and rapid response to change the customer needs. So, MCDM has the ability to handle successfully the evaluation process of multiple contradictory criteria. For an intelligent decision, the experts analyze each and every character of an alternative and then they take the decision. For an intelligent and successful decision, the experts require careful preparation and analysis of each and every character for an alternative and then they can take a good decision if they are armed with all the data and information that they need. To handle this complexity, Zadeh [1] originated the dominant and pioneer concept of the fuzzy set. For each domain in the fuzzy set, a value is assigned from the unit interval and is called membership grade (MemG). From the inception of the fuzzy set, it has been generalized in different directions in which one of the

most significant concepts is intuitionistic fuzzy (IF) set (IFS). Atanassov [2] initiated this dominant concept of IFS which is characterized by two mappings called MemG and nonmembership grade (NMemG). IFS is defined on the basis of restriction that the sum of MemG and NMemG must not exceed the unit interval $[0, 1]$. The notion of IFS appears as a hot research area after its origination. In literature, researchers used different techniques to handle the ranking with score or accuracy functions but all these techniques had some drawbacks. So, Ali et al. [3] initiated a graphical technique for ranking the IF values. Xu [4] investigated the series of aggregation operators such as IF weighted averaging (IFWA), IF ordered weighted averaging (IFOWA), and IF hybrid averaging (IFHA) operators under IF environment. The series of geometric operators, namely, IF weighted geometric (IFWG), IF ordered weighted geometric (IFOWG), and IF hybrid geometric (IFHG) operators, were presented by Xu and Yager [5]. Zhao et al. [6] initiated the idea of generalized IFWA, generalized IFOWA, and generalized IFHA operators by utilizing the IF information. Wang and Liu [7, 8] originated the concepts of geometric and averaging aggregation operators by utilizing Einstein operations. Zeng et al. [9] investigated a novel score function of intuitionistic fuzzy and then presented its application in decision-making. In a different scenario, the professional experts are restricted to provide their choices in the range of IFS. To cover this shortcoming, Yager [10] investigated the powerful paradigm of Pythagorean fuzzy (PF) set (PFS) in which the square sum of MemG and NMemG must lie between the real numbers 0 and 1. PFS relaxes and widens the boundary range by providing additional space to the decision-maker. Yager [11] originated the geometric and averaging aggregation operation by using PF information. Peng and Yang [12] initiated the concept of subtraction and division operators and proved some of its basic properties. Peng and Yang [13] investigated the notion of PF Choquet integral average and PF Choquet integral geometric operators. Garg [14, 15] proposed some PF Einstein averaging and PF Einstein geometric operators and presented their basic characteristics. Garg [16] investigated confidence PF weighted and ordered weighted averaging operators with their basic properties. Wei and Lu [17] originated the notions of PF power averaging and geometric operators and presented their desirable characteristics of these investigated operators. Wei [18] presented some interaction averaging and geometric operators by suing PF information. The concept of Hamacher operations for PF averaging and geometric operators was presented by Wu and Wei [19]. In the PF environment, the decision-makers are restricted to their boundary limitation and they cannot provide their preferred values freely. Due to these restrictions, some decisive information cannot be effectively handled by PFS.

Recently, Yager [20] presented the new concept of q-rung orthopair fuzzy (q-ROF) set (q-ROFS) from which the most significant generalization of PFS emerged. In q-ROFS, the sum of $q$th power of MemG and $q$th power of NMemG must be confined to the unit interval $[0, 1]$ and, furthermore, when the rung $q$ increases, then the range of orthopair satisfies the boundary restriction which is needed. Thus, the concept of q-ROFS is more useful and powerful than IFS and PFS because these are the special cases of q-ROFS. The basic properties of q-ROFS are proposed by Yager and Alajlan [21] and have been utilized in knowledge representation. Ali [22] proposed another view of q-ROFS by using the idea of orbits. Liu and Wang [23] proposed the concepts of q-ROF weighted averaging (q-ROFWA) and q-ROF weighted geometric (q-ROFWG). Liu and Liu [24] presented the combined study of Bonferroni mean (BM) operators with q-ROFS to investigate the q-ROF BM operators and also study q-ROF geometric BM operators with their desirable properties. Jana et al. [25] initiated the q-ROF Dombi averaging and geometric aggregation operators with their fundamental desirable characteristics. Wang et al. [26] investigated the combined concept of Muirhead means (MM) operators with q-ROFS to get the new aggregation operators that are q-ROF MM operators. Joshi and Gegov [27] initiated the concept of the confidence level of experts to the original information under q-ROF environment to propose some aggregation operators such as confidence q-ROFWA (Cq-ROFWA) and confidence q-ROFWG (Cq-ROFWG), Cq-ROFOWA, and Cq-ROFOWG operators. Yang et al. [28] presented the idea of q-RO normal fuzzy sets and defined the operational laws and score function for it. They also initiated some aggregation operators for the same concept that is q-RONFWA and q-RONFOWG. Furthermore, Hussain et al. [29] proposed hesitant q-ROFWA and hesitant q-ROFWG operators and discussed their desirable properties. Hussain et al. [30] proposed the generalized and group generalized averaging operation by using q-ROF information.

The dominant theory of rough set was first proposed by Pawlak [31] which generalized the classical set theory to cope with the imprecise, vague, and uncertain information. By the definition of Pawlak rough set, a universal set is characterized by two approximation sets known as upper and lower approximations. The lower approximation consists of those alternatives which contain a subset and the upper approximation consists of those alternatives having nonempty intersection with a subset. Further equivalence relation plays a key role in Pawlak rough set for approximations but this condition too restricts the practical and theoretical aspects of rough set. So, researchers used the generalized structure by using the nonequivalence structure; for details, see [32–38]. From the inception, researchers used the hybrid study of rough set theory with different concepts. The hybrid study of rough set and IFS was proposed by Chakrabarty et al. [39] to obtain the notion of IF rough set (IFRS) and IFRS became the hot and progressive research area for the researchers; for details, see [40–43]. Zhou and Wu [44] proposed the combined study of rough set and IFS by using crisp and fuzzy approximation. Zhou and Wu [45] initiated the constructive and axiomatic approach under the IF rough environment. Hussain et al. [46] investigated the idea of rough PF ideals by using the algebraic structure of semigroups. Zeng [47] proposed a new MCDM technique based on probabilistic information by using the PF environment. Hussain et al. [48] proposed the notion of q-ROF rough set by utilizing fuzzy $\beta$-covering and fuzzy $\beta$-covering

neighborhood. Molodtsov [49] originated the prominent and pioneer concept of soft set ($S_t$S) which generalized the classical set theory and is free from inherent complexity which the contemporary theories faced. It is observed that $S_t$S has a very close relation with fuzzy set and rough set. The $S_t$S theory is an essential concept and powerful mathematical tool for coping the uncertain, ambiguous, and imprecise data. Maji et al. [50, 51] proposed the hybrid notion of $S_t$S with fuzzy set and IFS to obtain fuzzy $S_t$S and IF$S_t$S which play a key role among these theories. Ali et al. [52] improved some existing definitions and operations in $S_t$S theory. The concept of generalized IF$S_t$S was proposed by Agarwal et al. [53]. Arora and Garg [54] presented the concept of IF$S_t$ weighted averaging (IF$S_t$WA) and IF$S_t$ weighted geometric (IF$S_t$WG) operators. Garg and Arora [55] proposed the notion of some power averaging and geometric aggregation operators by utilizing generalized IF$S_t$S. Arora [56] initiated the notion of IF$S_t$WA and IF$S_t$WG by using the Einstein operations. Feng et al. [57] improved some existing literature related to generalized IF $S_t$S and proposed some new operations for the developed concept. The combined study $S_t$S, rough set, and PFS were presented by Hussain et al. [58] to achieve the new concept of soft rough PFS and PF soft rough set. Riaz and Hashmi [59] presented the hybrid study of $S_t$S, rough set, PFS, and m-polar fuzzy set to get the new notion of Pythagorean m-polar fuzzy soft rough set. Hussain et al. [60] originated the hybrid structure of $S_t$S with q-ROFS to get the prominent notion of q-ROF soft (q-ROF$S_t$) set (q-ROF$S_t$S) and proposed some aggregation operators such as q-ROF$S_t$ weighted averaging (q-ROF$S_t$WA), q-ROF$S_t$ ordered weighted averaging (q-ROF$S_t$OWA), and q-ROF$S_t$ hybrid averaging (q-ROF$S_t$HA).

q-Rung orthopair fuzzy soft rough sets, a hybrid intelligent structure of soft sets, rough sets, and q-rung orthopair fuzzy sets are a powerful mathematical tool to deal with indeterminate, inconsistent, and incomplete information, which has caught the attention of the researchers. From the analysis, it is observed that aggregation operators have great importance in decision-making to aggregate the collective evaluated information of different sources into a single value. According to the best of our knowledge up till now, no application of the aggregation operators with the hybridization of q-ROFS with soft set and rough set is reported in q-ROF environment. Therefore, this motivates the current work of q-ROF$S_t$ rough study, and, further, we will investigate aggregation operators based on soft rough information that are q-ROF$S_t$RWA, q-ROF$S_t$ROWA, q-ROF$S_t$RHA, q-ROF$S_t$RWG, q-ROF$S_t$ROWG, and q-ROF$S_t$RHG operators.

The design of the remaining sections of the manuscript is summarized as follows: Section 2 consists of a brief study of the basic notions connecting the link with the coming sections. Section 3 is devoted to investigating the hybrid concept of $S_t$S and rough set with the notion of q-ROFS to obtain the new concept of q-ROF$S_t$RS. In Section 4, we presented the averaging aggregation operators such as q-ROF$S_t$RWA, q-ROF$S_t$ROWA, and q-ROF$S_t$RHA. Furthermore, the basic desirable properties of investigated averaging operators that are Idempotency, Boundedness, Monotonicity, shift invariance, and Homogeneity are investigated in detail. Section 5 is devoted to the geometric aggregation operators such as q-ROF$S_t$RWG, q-ROF$S_t$ROWG, and q-ROF$S_t$RHG. Moreover, the basic desirable characteristics of these investigated geometric operators that are Idempotency, Boundedness, Monotonicity, shift invariance, and Homogeneity are investigated in detail. In Section 6, the technique for MCDM and the stepwise algorithm for decision-making are demonstrated by utilizing the proposed approach. In Section 7, a numerical example for the developed approach is presented and a comparative study of the investigated models with some existing methods is given in detail which shows that the investigated models are more effective and useful than existing approaches. The final Section 8 consists of a conclusion of the manuscript.

## 2. Preliminaries

This section consists of some basic notions including IFS, PFS, q-ROFS, and q-ROF$S_t$S which will be helpful in on word sections.

*Definition 1* (see [2]). Consider a universe $Y$ and an IFS$\mathcal{T}$ on set $Y$ denoted and defined as

$$\mathcal{T} = \{\langle s, \beta_{\mathcal{T}}(s), \vartheta_{\mathcal{T}}(s)\rangle | s \in Y\}, \tag{1}$$

in which $\beta_{\mathcal{T}}, \vartheta_{\mathcal{T}}: Y \longrightarrow [0,1]$ denotes the MemG and NMemG of an alternative $s \in Y$ to the set $\mathcal{T}$ having the condition that $0 \le \beta_{\mathcal{T}}(s) + \vartheta_{\mathcal{T}}(s) \le 1$. $\pi_{\mathcal{T}} = 1 - (\beta_{\mathcal{T}}(s) + \vartheta_{\mathcal{T}}(s))$ is called hesitancy degree of $s \in Y$.

*Definition 2* (see [10]). Consider a universe $Y$ and a PFS$\mathcal{T}$ on set $Y$ defined and denoted as

$$\mathcal{T} = \{\langle s, \beta_{\mathcal{T}}(s), \vartheta_{\mathcal{T}}(s)\rangle | s \in Y\}, \tag{2}$$

in which $\beta_{\mathcal{T}}, \vartheta_{\mathcal{T}}: Y \longrightarrow [0,1]$ denotes the MemG and NMemG of an alternative $s \in Y$ to the set $\mathcal{T}$ having the condition that $0 \le (\beta_{\mathcal{T}}(s))^2 + (\vartheta_{\mathcal{T}}(s))^2 \le 1$. $\pi_{\mathcal{T}} = \sqrt{1 - (\beta_{\mathcal{T}}(s))^2 - (\vartheta_{\mathcal{T}}(s))^2}$ is called hesitancy degree of $s \in Y$.

*Definition 3* (see [20]). Consider $Y$ as a universe of discourse and a q-ROFS$\mathcal{T}$ on set $Y$ is an object of the form

$$\mathcal{T} = \left\{\langle s, \beta_{\mathcal{T}}(s), \vartheta_{\mathcal{T}}(s)\rangle_q | s \in Y \text{ and } q \ge 1\right\}, \tag{3}$$

in which $\beta_{\mathcal{T}}, \vartheta_{\mathcal{T}}: Y \longrightarrow [0,1]$ shows the MemG and NMemG of an alternative $s \in Y$ to the set $\mathcal{T}$ having the condition that $0 \le (\beta_{\mathcal{T}}(s))^q + (\vartheta_{\mathcal{T}}(s))^q \le 1$. The hesitancy degree is shown as $\pi_{\mathcal{T}} = \sqrt[q]{1 - (\beta_{\mathcal{T}}(s))^q - (\vartheta_{\mathcal{T}}(s))^q}$ for each $s \in Y$.

Molodtsov [49] originated the prominent and pioneer concept of soft set ($S_t$S) which generalized the classical set theory and is free from inherent complexity which the contemporary theories faced. It is observed that $S_t$S has very close relation with fuzzy set and rough set. The $S_t$S theory is an essential concept and powerful mathematical tool for

coping the uncertain, ambiguous and imprecise data which is defined as:

*Definition 4* (see [49]). Let $Y$ be a fixed set and $\mathscr{E}$ be set of parameters with $\mathscr{B} \subseteq \mathscr{E}$. Then the pair $(\mathscr{T}, \mathscr{B})$ is said to be $S_t$S, where $\mathscr{T}$ is function given as $\mathscr{T}: \mathscr{B} \longrightarrow P(Y)$. $P(Y)$ denotes the collection of all subsets of $Y$.

*Definition 5* (see [50]). Let $(\mathscr{T}, \mathscr{B})$ a $S_t$S over $Y$ with $\mathscr{B} \subseteq \mathscr{E}$. Then $(\mathscr{T}^*, \mathscr{B})$ is known as fuzzy $S_t$S over $Y$, where $\mathscr{T}^*$ is a function given as $\mathscr{T}^*: \mathscr{B} \longrightarrow F(Y)$. $F(Y)$ denotes the collection of all fuzzy subsets of $Y$ and mathematically it is gives as

$$\mathscr{T}^*_{c_j} = \left\{\langle s_i, \beta_j(s_i)\rangle | s_i \in Y\right\}. \tag{4}$$

*Definition 6* (see [60]). Consider a universal set $Y$. Let $\mathscr{E}$ be set of parameters and $\mathscr{B} \subseteq \mathscr{E}$. Then a q-ROFS$_t$S is a pair $(\mathscr{T}, \mathscr{B})$ over set $Y$ and $\mathscr{T}$ is a mapping given as $\mathscr{T}: \mathscr{B} \longrightarrow \mathrm{q-ROFS}_t\mathrm{S}(Y)$ in which $\mathrm{q-ROFS}_t\mathrm{S}(Y)$ contains the collection of all q-ROFSs. Then, the q-ROFS$_t$S is denoted and defined as

$$\mathscr{T}_{c_j}(s_i) = \left\{\langle s_i, \beta_j(s_i), \vartheta_j(s_i)\rangle_q | s_i \in Y \text{ and } q \geq 1\right\}, \tag{5}$$

in which $\beta_j(s_i), \vartheta_j(s_i)$ represents the MemG and NMemG of an alternative $s_i \in Y$ to the set $\mathscr{T}_{c_j}$ satisfying that $0 \leq (\beta_j(s_i))^q + (\vartheta_j(s_i))^q \leq 1$.

$\pi_{\mathscr{T}} = \sqrt[q]{1 - (\beta_j(s_i))^q - (\vartheta_j(s_i))^q}$ is known to be hesitancy degree of $s_i \in Y$. For simplicity $\mathscr{T}_{c_j}(s_i) = (s_i, \beta_j(s_i), \vartheta_j(s_i)_q)$ is denoted as $\mathscr{T}_{c_j} = (\beta_{ij}, \vartheta_{ij})$ if there is no confusion and is called q-ROFS$_t$ number (q-ROFS$_t$N).

Considering two q-ROFS$_t$Ns $\mathscr{T}_{c_j} = (\beta_{1j}, \vartheta_{1j})$ for $(j = 1, 2)$, reference [20] defined the following operation are defined on them:

(i) $\mathscr{T}_{c_1} \cup \mathscr{T}_{c_2} = (\max(\beta_{11}, \beta_{12}), \min(\vartheta_{11}, \vartheta_{12}))$
(ii) $\mathscr{T}_{c_1} \cap \mathscr{T}_{c_2} = (\min(\beta_{11}, \beta_{12}), \max(\vartheta_{11}, \vartheta_{12}))$
(iii) $\mathscr{T}_{c_1} \oplus \mathscr{T}_{c_2} = (\sqrt[q]{\beta_{11}^q + \beta_{12}^q - \beta_{11}^q\beta_{12}^q}, \vartheta_{11}\vartheta_{12})$
(iv) $\mathscr{T}_{c_1} \oplus \mathscr{T}_{c_2} = (\beta_{11}\beta_{12}, \sqrt[q]{\vartheta_{11}^q + \vartheta_{12}^q - \vartheta_{11}^q\vartheta_{12}^q})$
(v) $\mathscr{T}_{c_1} \leq \mathscr{T}_{c_2}$ if $(\beta_{11} \leq \beta_{12}, \vartheta_{11} \geq \vartheta_{12})$
(vi) $\mathscr{T}^c_{c_1} = (\vartheta_{11}, \beta_{11})$ where $\mathscr{T}^c_{c_1}$ represents the complement of $\mathscr{T}_{c_1}$
(vii) $\alpha\mathscr{T}_{c_1} = (\sqrt[q]{1 - (1 - \beta_{11}^q)^\alpha}, \vartheta_{11}^\alpha)$ for $\alpha \geq 1$
(viii) $\mathscr{T}^\alpha_{c_1} = (\beta_{11}^\alpha, \sqrt[q]{1 - (1 - \vartheta_{11}^q)^\alpha})$ for $\alpha \geq 1$

## 3. q-ROF Soft Rough Set (q-ROF$S_t$RS)

The concept of $S_t$S theory, is the generalization of classical set theory and is free from inherent complexity which the contemporary theories faced. The $S_t$S theory and rough set theory are essential concepts and powerful mathematical tools for coping the uncertain, ambiguous, and imprecise data. Motivated from the combining study of soft rough set, this section is devoted to the hybridized study of q-ROFS with $S_t$S and rough set to obtain the new concept of q-ROFS$_t$RS. Some basic operations, a new score function, and some desirable characteristics of the proposed concept are investigated in detail.

*Definition 7.* Let $(\mathscr{T}, \mathscr{E})$ be a q-ROFS$_t$S over $Y$. Any subset $\mathscr{L}$ of $Y \times \mathscr{E}$ is set to a q-ROFS$_t$ relation from $Y$ to $\mathscr{E}$ and is defined as

$$\mathscr{L} = \left\{\langle (s_i, c_j), \beta(s_i, c_j), \vartheta(s_i, c_j)\rangle | (s_i, c_j) \in Y \times \mathscr{E}\right\}, \tag{6}$$

where $\beta: Y \times \mathscr{E} \longrightarrow [0, 1]$ and $\vartheta: Y \times \mathscr{E} \longrightarrow [0, 1]$ denote the MemG and NMemG with $0 \leq [\beta(s_i, c_j)]^q + [\vartheta(s_i, c_j)]^q \leq 1$ for all $(s_i, c_j) \in Y \times \mathscr{E}$.

If $Y = \{s_1, s_2, \ldots, s_m\}$ and $\mathscr{E} = \{c_1, c_2, \ldots, c_n\}$, then, q-ROFS$_t$ relation $\mathscr{L}$ from $Y$ to $\mathscr{E}$ can be presented in Table 1.

*Definition 8.* Consider a universal $Y$, $\mathscr{E}$ as the set of parameters and $(\mathscr{T}, \mathscr{E})$ as a q-ROFS$_t$S. Let $\mathscr{L}$ be an arbitrary q-ROFS$_t$ relation from set $Y$ to $\mathscr{E}$. The pair $(\mathscr{T}, \mathscr{E}, \mathscr{L})$ is said to be q-ROFS$_t$ approximation space. For any optimum decision normal object $\mathscr{M} \in \mathrm{q-ROFS}(\mathscr{E})$, then the lower and upper approximation of $\mathscr{M}$ with respect to approximation space $(\mathscr{T}, \mathscr{E}, \mathscr{L})$ are denoted and defined as

$$\underline{\mathscr{L}}(\mathscr{M}) = \left\{\left(s_i, \underline{\beta_j}(s_i), \underline{\vartheta_j}(s_i)\right) | s_i \in Y\right\}, \tag{7}$$

$$\overline{\mathscr{L}}(\mathscr{M}) = \left\{\left(s_i, \overline{\beta_j}(s_i), \overline{\vartheta_j}(s_i)\right) | s_i \in Y\right\}, \tag{8}$$

where $\underline{\beta_j}(s_i) = \wedge_{c_j \in \mathscr{E}}[\beta_{\mathscr{L}}(s_i, c_j) \wedge \beta_{\mathscr{M}}(c_j)]$, $\underline{\vartheta_j}(s_i) = \vee_{c_j \in \mathscr{E}}[\vartheta_{\mathscr{L}}(s_i, c_j) \vee \vartheta_{\mathscr{M}}(c_j)]$ and $\overline{\beta_j}(s_i) = \vee_{c_j \in \mathscr{E}}[\beta_{\mathscr{L}}(s_i, c_j) \vee \beta_{\mathscr{M}}(c_j)]$, $\overline{\vartheta_j}(s_i) = \wedge_{c_j \in \mathscr{E}}[\vartheta_{\mathscr{L}}(s_i, c_j) \wedge \vartheta_{\mathscr{M}}(c_j)]$ such that $0 \leq [\underline{\beta_j}(s_i)]^q + [\underline{\vartheta_j}(s_i)]^q \leq 1$ and $0 \leq [\overline{\beta_j}(s_i)]^q + [\overline{\vartheta_j}(s_i)]^q \leq 1$.

It is observed that $\underline{\mathscr{L}}(\mathscr{M})$ and $\overline{\mathscr{L}}(\mathscr{M})$ are two q-ROFSs in $Y$. Thus, the operators $\underline{\mathscr{L}}(\mathscr{M}), \overline{\mathscr{L}}(\mathscr{M}): \mathrm{q-ROFS}_t(\mathscr{E}) \longrightarrow \mathrm{q-ROFS}_t(Y)$ are, respectively, known as lower and upper q-ROFS$_t$R approximation operators. Therefore, q-ROFS$_t$RS is a pair $\mathscr{L}(\mathscr{M}) = (\underline{\mathscr{L}}(\mathscr{M}), \overline{\mathscr{L}}(\mathscr{M})) = (s_i, (\underline{\beta_j}(s_i), \underline{\vartheta_j}(s_i)), (\overline{\beta_j}(s_i), \overline{\vartheta_j}(s_i)))$.

For simplicity, we can write $\mathscr{L}(\mathscr{M}) = (\underline{\mathscr{L}}(\mathscr{M}), \overline{\mathscr{L}}(\mathscr{M})) = (s_i, (\underline{\beta_j}(s_i), \underline{\vartheta_j}(s_i)), (\overline{\beta_j}(s_i), \overline{\vartheta_j}(s_i)))$ as $\mathscr{L}_{c_j}(\mathscr{M}_i) = (\underline{\mathscr{L}_{c_j}}(\mathscr{M}_i), \overline{\mathscr{L}_{c_j}}(\mathscr{M}_i)) = ((\underline{\beta_{ij}}, \underline{\vartheta_{ij}})), (\overline{\beta_{ij}}, \overline{\vartheta_{ij}})$ and call q-ROFS$_t$R number (q-ROFS$_t$RN), if there is no confusion.

*Remark 1*

(a) If $q = 1$ is fixed, then the developed q-ROFS$_t$R approximation operators reduce to IFS$_t$R approximation operators
(b) If $q = 2$ is fixed, then the developed q-ROFS$_t$R approximation operators reduce to PFS$_t$R approximation operators

Consider the following example to better understand the concept of q-ROFS$_t$R approximation operators.

*Example 1.* Suppose a decision-maker $Z$ buys a house, as given in set $Y = \{s_1, s_2, s_3, s_4, s_5\}$ under consideration. Let the

Table 1: q-ROFS$_t$ relation $\mathscr{L}$ from $Y$ to $\mathscr{E}$.

| $\mathscr{L}$ | $c_1$ | $c_2$ | $\cdots$ | $c_n$ |
|---|---|---|---|---|
| $s_1$ | $(\beta(s_1,c_1),\vartheta(s_1,c_1))$ | $(\beta(s_1,c_2),\vartheta(s_1,c_2))$ | $\cdots$ | $(\beta(s_1,c_n),\vartheta(s_1,c_n))$ |
| $s_2$ | $(\beta(s_2,c_1),\vartheta(s_2,c_1))$ | $(\beta(s_2,c_2),\vartheta(s_2,c_2))$ | $\cdots$ | $(\beta(s_2,c_n),\vartheta(s_2,c_n))$ |
| $\vdots$ | $\vdots$ | $\vdots$ | $\ddots$ | $\vdots$ |
| $s_m$ | $(\beta(s_m,c_1),\vartheta(s_m,c_1))$ | $(\beta(s_m,c_2),\vartheta(s_m,c_2))$ | $\cdots$ | $(\beta(s_m,c_n),\vartheta(s_m,c_n))$ |

parameter set $\mathscr{E}=\{c_1,c_2,c_3,c_4\}$ where $c_1$ = beautiful, $c_2$ = size, $c_3$ = expensive, and $c_4$ = location. A decision-maker $Z$ wants to purchase a house from the available houses which fulfill the utmost extent of given parameters. Consider the decision-maker $Z$ presents the gorgeous of houses in form of q-ROFS$_t$ relation $\mathscr{L}$ from set $Y$ to $\mathscr{E}$ and is given in Table 2.

Consider a decision-maker $Z$ presents the optimum normal decision object $\mathscr{M}$ which is a q-ROF subset over parameter set $\mathscr{E}$; that is,

$$\mathscr{M}=\{(c_1,0.9,0.2),(c_2,0.4,0.6),(c_3,0.8,0.4),(c_4,0.5,0.1)\}. \tag{9}$$

Now, by using equations (7) and (8), we have

$$\begin{aligned}
&\underline{\beta_1}(s_1)=0.4,\\
&\underline{\vartheta_1}(s_1)=0.6,\\
&\underline{\beta_2}(s_2)=0.3,\\
&\underline{\vartheta_2}(s_2)=0.7,\\
&\underline{\beta_3}(s_3)=0.2,\\
&\underline{\vartheta_3}(s_3)=0.9,\\
&\underline{\beta_4}(s_4)=0.4,\\
&\underline{\vartheta_4}(s_4)=0.6,\\
&\underline{\beta_5}(s_5)=0.3,\\
&\underline{\vartheta_5}(s_5)=0.7,\\
&\overline{\beta_1}(s_1)=0.9,\\
&\overline{\vartheta_1}(s_1)=0.1,\\
&\overline{\beta_2}(s_2)=0.9,\\
&\overline{\vartheta_2}(s_2)=0.1,\\
&\overline{\beta_3}(s_3)=0.95,\\
&\overline{\vartheta_3}(s_3)=0.1\\
&\overline{\beta_4}(s_4)=0.93,\\
&\overline{\vartheta_4}(s_4)=0.1,\\
&\overline{\beta_5}(s_5)=0.9,\\
&\overline{\vartheta_5}(s_5)=0.1.
\end{aligned} \tag{10}$$

Now, to get the lower and upper q-ROFS$_t$R approximation operators,

Table 2: q-ROFS$_t$ relation from set $Y$ to $\mathscr{E}$ for $q\geq 3$.

| $\mathscr{L}$ | $c_1$ | $c_2$ | $c_3$ | $c_4$ |
|---|---|---|---|---|
| $s_1$ | $(0.9,0.4)$ | $(0.8,0.2)$ | $(0.7,0.3)$ | $(0.65,0.2)$ |
| $s_2$ | $(0.8,0.5)$ | $(0.5,0.1)$ | $(0.85,0.2)$ | $(0.3,0.7)$ |
| $s_3$ | $(0.6,0.9)$ | $(0.2,0.6)$ | $(0.6,0.1)$ | $(0.95,0.3)$ |
| $s_4$ | $(0.7,0.4)$ | $(0.93,0.4)$ | $(0.4,0.2)$ | $(0.5,0.1)$ |
| $s_5$ | $(0.3,0.7)$ | $(0.78,0.25)$ | $(0.8,0.15)$ | $(0.7,0.4)$ |

$$\begin{aligned}
\underline{\mathscr{L}}(\mathscr{M})&=\{(s_1,0.7,\ 0.6),(s_2,0.3,\ 0.7),(s_3,0.2,\ 0.9),\\
&\quad(s_4,0.4,\ 0.6),(s_5,0.3,\ 0.7)\},\\
\overline{\mathscr{L}}(\mathscr{M})&=\{(s_1,0.9,\ 0.1),(s_2,0.9,\ 0.1),(s_3,0.95,\ 0.1),\\
&\quad(s_4,0.93,\ 0.1),(s_5,0.9,\ 0.1)\}.
\end{aligned} \tag{11}$$

Therefore,

$$\begin{aligned}
\mathscr{L}(\mathscr{M})&=\left(\underline{\mathscr{L}}(\mathscr{M}),\overline{\mathscr{L}}(\mathscr{M})\right)=\{(s_1,\ (0.7,\ 0.6),\ (0.9,\ 0.1)),\\
&\quad(s_2,\ (0.3,\ 0.7),\ (0.9,\ 0.1)),(s_3,\ (0.2,\ 0.9),\\
&\quad(0.95,\ 0.1)),(s_4,\ (0.4,\ 0.6),\ (0.93,\ 0.1)),\\
&\quad(s_5,\ (0.3,\ 0.7),\ (0.9,\ 0.1))\}.
\end{aligned} \tag{12}$$

*Definition 9.* Consider $\mathscr{L}_{c_j}(\mathscr{M}_1)=(\underline{\mathscr{L}_{c_j}}(\mathscr{M}_1),\overline{\mathscr{L}_{c_j}}(\mathscr{M}_1))$ for $(j=1,2)$ are the two q-ROFS$_t$RNs. Then, the following operators are defined on them:

(i) $\mathscr{L}_{c_1}(\mathscr{M}_1)\cup\mathscr{L}_{c_2}(\mathscr{M}_1)=\{(\underline{\mathscr{L}_{c_1}}(\mathscr{M}_1)\cup\underline{\mathscr{L}_{c_2}}(\mathscr{M}_1)),(\overline{\mathscr{L}_{c_1}}(\mathscr{M}_1)\cup\overline{\mathscr{L}_{c_2}}(\mathscr{M}_1))\}$

(ii) $\mathscr{L}_{c_1}(\mathscr{M}_1)\cap\mathscr{L}_{c_2}(\mathscr{M}_1)=\{(\underline{\mathscr{L}_{c_1}}(\mathscr{M}_1)\cap\underline{\mathscr{L}_{c_2}}(\mathscr{M}_1)),(\overline{\mathscr{L}_{c_1}}(\mathscr{M}_1)\cap\overline{\mathscr{L}_{c_2}}(\mathscr{M}_1))\}$

(iii) $\mathscr{L}_{c_1}(\mathscr{M}_1)\oplus\mathscr{L}_{c_2}(\mathscr{M}_1)=\{(\underline{\mathscr{L}_{c_1}}(\mathscr{M}_1)\oplus\underline{\mathscr{L}_{c_2}}(\mathscr{M}_1)),(\overline{\mathscr{L}_{c_1}}(\mathscr{M}_1)\oplus\overline{\mathscr{L}_{c_2}}(\mathscr{M}_1))\}$

(iv) $\mathscr{L}_{c_1}(\mathscr{M}_1)\otimes\mathscr{L}_{c_2}(\mathscr{M}_1)=\{(\underline{\mathscr{L}_{c_1}}(\mathscr{M}_1)\otimes\underline{\mathscr{L}_{c_2}}(\mathscr{M}_1)),(\overline{\mathscr{L}_{c_1}}(\mathscr{M}_1)\otimes\overline{\mathscr{L}_{c_2}}(\mathscr{M}_1))\}$

(v) $\mathscr{L}_{c_1}(\mathscr{M}_1)\subseteq\mathscr{L}_{c_2}(\mathscr{M}_1)=(\underline{\mathscr{L}_{c_1}}(\mathscr{M}_1)\subseteq\underline{\mathscr{L}_{c_2}}(\mathscr{M}_1))$ and $(\overline{\mathscr{L}_{c_1}}(\mathscr{M}_1)\subseteq\overline{\mathscr{L}_{c_2}}(\mathscr{M}_1))$

(vi) $\alpha\mathscr{L}_{c_1}(\mathscr{M}_1)=(\alpha\underline{\mathscr{L}_{c_1}}(\mathscr{M}_1),\alpha\overline{\mathscr{L}_{c_1}}(\mathscr{M}_1))$ for $\alpha\geq 1$

(vii) $(\mathscr{L}_{c_1}(\mathscr{M}_1))^{\alpha}=((\underline{\mathscr{L}_{c_1}}(\mathscr{M}_1))^{\alpha},(\overline{\mathscr{L}_{c_1}}(\mathscr{M}_1))^{\alpha})$ for $\alpha\geq 1$

(viii) $\mathscr{L}_{c_1}(\mathscr{M}_1)^c=(\underline{\mathscr{L}_{c_j}}(\mathscr{M}_1)^c,\overline{\mathscr{L}_{c_j}}(\mathscr{M}_1)^c)$, where $\underline{\mathscr{L}_{c_j}}(\mathscr{M}_1)^c$ are $\overline{\mathscr{L}_{c_j}}(\mathscr{M}_1)^c$, the complements of q-ROFS$_t$R approximation operators $\underline{\mathscr{L}_{c_j}}(\mathscr{M}_1)$ and $\overline{\mathscr{L}_{c_j}}(\mathscr{M}_1)$, that is, $\underline{\mathscr{L}_{c_j}}(\mathscr{M}_1)^c=(\underline{\vartheta_{ij}},\underline{\beta_{ij}})$

(ix) $\mathscr{L}(\mathscr{M}_1) = \mathscr{L}(\mathscr{M}_2)$ iff $\underline{\mathscr{L}}(\mathscr{M}_1) = \underline{\mathscr{L}}(\mathscr{M}_2)$ and $\overline{\mathscr{L}}(\mathscr{M}_1) = \overline{\mathscr{L}}(\mathscr{M}_2)$

*Definition 10.* Let $\mathscr{L}_{c_1}(\mathscr{M}_1) = (\underline{\mathscr{L}_{c_1}}(\mathscr{M}_1), \overline{\mathscr{L}_{c_1}}(\mathscr{M}_1)) = ((\underline{\beta_{11}}, \underline{\vartheta_{11}}), (\overline{\beta_{11}}, \overline{\vartheta_{11}})$ be a q-ROFS$_t$RN. Then, the score function is given as

$$\mathrm{SC}\left(\mathscr{L}_{c_1}(\mathscr{M}_1)\right) = \frac{1}{2}\left(\underline{\beta_{11}}^q + \overline{\beta_{11}}^q - \underline{\vartheta_{11}}^q - \overline{\vartheta_{11}}^q\right), \quad \mathrm{SC}\left(\mathscr{L}_{c_1}(\mathscr{M}_1)\right) \in [-1, 1] \text{ and } q \geq 1. \tag{13}$$

The greater the score value, the greater the q-ROFS$_t$RN.

**Proposition 1.** *Let $(\mathscr{T}, \mathscr{E}, \mathscr{L})$ be q-ROFS$_t$ approximation space. For any two $\mathscr{L}(\mathscr{M}_1) = (\underline{\mathscr{L}}(\mathscr{M}_1), \overline{\mathscr{L}}(\mathscr{M}_1))$ and $\mathscr{L}(\mathscr{M}_2) = (\underline{\mathscr{L}}(\mathscr{M}_2), \overline{\mathscr{L}}(\mathscr{M}_2))$ q-ROFS$_t$RSs over a common universe set $Y$, then the following properties hold:*

*(i)* $\sim(\sim\mathscr{L}(\mathscr{M}_1)) = \mathscr{M}_1$, *where* $\sim\mathscr{L}(\mathscr{M}_1)$ *is the complement of* $\mathscr{L}(\mathscr{M}_1)$, $\mathscr{L}(\mathscr{M}_1) \cup \mathscr{L}(\mathscr{M}_2) = \mathscr{L}(\mathscr{M}_2) \cup \mathscr{L}(\mathscr{M}_1)$, $\mathscr{L}(\mathscr{M}_1) \cap \mathscr{L}(\mathscr{M}_2) = \mathscr{L}(\mathscr{M}_2) \cap \mathscr{L}(\mathscr{M}_1)$

*(ii)* $\sim(\mathscr{L}(\mathscr{M}_1) \cup \mathscr{L}(\mathscr{M}_2)) = (\sim\mathscr{L}(\mathscr{M}_1)) \cap (\sim\mathscr{L}(\mathscr{M}_2))$

*(iii)* $\sim(\mathscr{L}(\mathscr{M}_1) \cap \mathscr{L}(\mathscr{M}_2)) = (\sim\mathscr{L}(\mathscr{M}_1)) \cup (\sim\mathscr{L}(\mathscr{M}_2))$

*(iv) If* $\mathscr{M}_1 \subseteq \mathscr{M}_2$, *then* $\mathscr{L}(\mathscr{M}_1) \subseteq \mathscr{L}(\mathscr{M}_2)$

*(v)* $\mathscr{L}(\mathscr{M}_1 \cup \mathscr{M}_2) \supseteq \mathscr{L}(\mathscr{M}_1) \cup \mathscr{L}(\mathscr{M}_2)$

*(vi)* $\mathscr{L}(\mathscr{M}_1 \cap \mathscr{M}_2) \subseteq \mathscr{L}(\mathscr{M}_1) \cap \mathscr{L}(\mathscr{M}_2)$

## 4. q-Rung Orthopair Fuzzy Soft Rough Averaging (q-ROFS$_t$RA) Aggregation Operator

This section is devoted to the analysis of q-ROFS$_t$R averaging aggregation operators such as q-ROFS$_t$RWA, q-ROFS$_t$ROWA, and q-ROFS$_t$RHA operators. We will present the fundamental properties of these operators in detail.

### *4.1. q-Rung Orthopair Fuzzy Soft Rough Weighted Averaging (q-ROFS$_t$RWA) Operator*

*Definition 11.* Let $\mathscr{L}_{c_j}(\mathscr{M}_i) = (\underline{\mathscr{L}_{c_j}}(\mathscr{M}_i), \overline{\mathscr{L}_{c_j}}(\mathscr{M}_i))$ $(i = 1, 2, \ldots, m,\ j = 1, 2, \ldots, n)$ be the collection of q-ROFS$_t$RNs. Let the weight vectors $t = (t_1, t_2, \ldots, t_m)^T$ and $v = (v_1, v_2, \ldots, v_n)^T$ for experts $s_i$ and parameters $c_j$ with $\sum_{i=1}^m t_i = 1$, $\sum_{j=1}^n v_j = 1$, and $0 \leq t_i,\ v_j \leq 1$, respectively. The q-ROFS$_t$RWA operator is defined as

$$\begin{aligned} &\mathrm{q-ROFS}_t\mathrm{RWA}\left(\mathscr{L}_{c_1}(\mathscr{M}_1), \ldots, \mathscr{L}_{c_n}(\mathscr{M}_m)\right) \\ &= \left(\oplus_{j=1}^n v_j\left(\oplus_{i=1}^m t_i \underline{\mathscr{L}_{c_j}}(\mathscr{M}_i)\right), \oplus_{j=1}^n v_j\left(\oplus_{i=1}^m t_i \overline{\mathscr{L}_{c_j}}(\mathscr{M}_i)\right)\right). \end{aligned} \tag{14}$$

In view of the above definition, Theorem 1 illustrates the aggregated result for q-ROFS$_t$RWA.

**Theorem 1.** *Let $\mathscr{L}_{c_j}(\mathscr{M}_i) = (\underline{\mathscr{L}_{c_j}}(\mathscr{M}_i), \overline{\mathscr{L}_{c_j}}(\mathscr{M}_i))$ $(i = 1, 2, \ldots, m,\ j = 1, 2, \ldots, n)$ be the collection of q-ROFS$_t$RNs. Let the weight vectors $t = (t_1, t_2, \ldots, t_m)^T$ and $v = (v_1, v_2, \ldots, v_n)^T$ for experts $s_i$ and parameters $c_j$ with $\sum_{i=1}^m t_i = 1$, $\sum_{j=1}^n v_j = 1$ and $0 \leq t_i,\ v_j \leq 1$, respectively. Then, q-ROFS$_t$RWA operator is given as*

$$\begin{aligned} &\mathrm{q-ROFS}_t\mathrm{RWA}\left(\mathscr{L}_{c_1}(\mathscr{M}_1), \ldots, \mathscr{L}_{c_n}(\mathscr{M}_m)\right) = \left[\oplus_{j=1}^n v_j\left(\oplus_{i=1}^m t_i \underline{\mathscr{L}_{c_j}}(\mathscr{M}_i)\right), \oplus_{j=1}^n v_j\left(\oplus_{i=1}^m t_i \overline{\mathscr{L}_{c_j}}(\mathscr{M}_i)\right)\right] \\ &= \left[\left\{\sqrt[q]{1 - \prod_{j=1}^n\left(\prod_{i=1}^m\left(1 - \underline{\beta_{ij}}^q\right)^{t_i}\right)^{v_j}}, \prod_{j=1}^n\left(\prod_{i=1}^m \underline{\vartheta_{ij}}^{t_i}\right)^{v_j}\right\}, \left\{\sqrt[q]{1 - \prod_{j=1}^n\left(\prod_{i=1}^m\left(1 - \overline{\beta_{ij}}^q\right)^{t_i}\right)^{v_j}}, \prod_{j=1}^n\left(\prod_{i=1}^m \overline{\vartheta_{ij}}^{t_i}\right)^{v_j}\right\}\right]. \end{aligned} \tag{15}$$

*Proof.* To get the required proof, we will use the method of mathematical induction.

As by operational law,

$$\begin{aligned} \mathscr{L}_{c_1}(\mathscr{M}_1) \oplus \mathscr{L}_{c_1}(\mathscr{M}_2) &= \left(\left(\underline{\beta_{11}}, \underline{\vartheta_{11}}\right) \oplus \left(\underline{\beta_{12}}, \underline{\vartheta_{12}}\right), \left(\overline{\beta_{11}}, \overline{\vartheta_{11}}\right) \oplus \left(\overline{\beta_{12}}, \overline{\vartheta_{12}}\right)\right) \\ &= \left[\left(\sqrt[q]{\underline{\beta_{11}}^q + \underline{\beta_{12}}^q + \underline{\beta_{11}}^q \underline{\beta_{12}}^q}, \underline{\vartheta_{11}}\,\underline{\vartheta_{12}}\right), \left(\sqrt[q]{\overline{\beta_{11}}^q + \overline{\beta_{12}}^q + \overline{\beta_{11}}^q \overline{\beta_{12}}^q}, \overline{\vartheta_{11}}\,\overline{\vartheta_{12}}\right)\right], \\ \alpha\mathscr{L}_{c_1}(\mathscr{M}_1) &= \left[\left(\sqrt[q]{1 - \left(1 - \underline{\beta_{11}}^q\right)^\alpha}, \underline{\vartheta_{11}}^\alpha\right), \left(\sqrt[q]{1 - \left(1 - \overline{\beta_{11}}^q\right)^\alpha}, \overline{\vartheta_{11}}^\alpha\right)\right]. \end{aligned} \tag{16}$$

Suppose the result is true for $m = 2$ and $n = 2$; that is,

$$\begin{aligned}&\mathrm{q-ROFS}_t\mathrm{RWA}\left(\mathscr{L}_{c_j}(\mathscr{M}_i), \mathscr{L}_{c_j}(\mathscr{M}_i)\right)\\&=\left(\oplus_{j=1}^{2} v_j\left(\oplus_{i=1}^{2} t_i \underline{\mathscr{L}_{c_j}}(\mathscr{M}_i)\right), \oplus_{j=1}^{2} v_j\left(\oplus_{i=1}^{2} t_i \overline{\mathscr{L}_{c_j}}(\mathscr{M}_i)\right)\right).\end{aligned} \tag{17}$$

Now, consider

$$\begin{aligned}\mathrm{q-ROFS}_t\mathrm{RWA}\left(\mathscr{L}_{c_j}(\mathscr{M}_i), \mathscr{L}_{c_j}(\mathscr{M}_i)\right) &= \left[\oplus_{j=1}^{2} v_j\left(\oplus_{i=1}^{2} t_i \underline{\mathscr{L}_{c_j}}(\mathscr{M}_i)\right), \oplus_{j=1}^{2} v_j\left(\oplus_{i=1}^{2} t_i \overline{\mathscr{L}_{c_j}}(\mathscr{M}_i)\right)\right]\\&=\left[\left\{v_1\left(t_1 \underline{\mathscr{L}_{c_1}}(\mathscr{M}_1) \oplus t_2 \underline{\mathscr{L}_{c_1}}(\mathscr{M}_2)\right) \oplus v_2\left(t_1 \underline{\mathscr{L}_{c_2}}(\mathscr{M}_1) \oplus t_2 \underline{\mathscr{L}_{c_2}}(\mathscr{M}_2)\right)\right\},\right.\\&\quad\left.\left\{v_1\left(t_1 \overline{\mathscr{L}_{c_j}}(\mathscr{M}_1) \oplus t_2 \overline{\mathscr{L}_{c_j}}(\mathscr{M}_2)\right) \oplus v_2\left(t_1 \overline{\mathscr{L}_{c_j}}(\mathscr{M}_1) \oplus t_2 \overline{\mathscr{L}_{c_j}}(\mathscr{M}_2)\right)\right\}\right],\\\mathrm{q-ROFS}_t\mathrm{RWA}\left(\mathscr{L}_{c_j}(\mathscr{M}_i), \mathscr{L}_{c_j}(\mathscr{M}_i)\right) &= \left[\left\{\sqrt[q]{1-\prod_{j=1}^{2}\left(\prod_{i=1}^{2}\left(1-\underline{\beta_{ij}}^{q}\right)^{t_i}\right)^{v_j}}, \prod_{j=1}^{2}\left(\prod_{i=1}^{2}\underline{\vartheta_{ij}}^{t_i}\right)^{v_j}\right\},\right.\\&\quad\left.\left\{\left(\sqrt[q]{1-\prod_{j=1}^{2}\left(\prod_{i=1}^{2}\left(1-\overline{\beta_{ij}}^{q}\right)^{t_i}\right)^{v_j}}, \prod_{j=1}^{2}\left(\prod_{i=1}^{2}\overline{\vartheta_{ij}}^{t_i}\right)^{v_j}\right)\right\}\right].\end{aligned} \tag{18}$$

The result is true for $m = 2$ and $n = 2$.

Now, consider the result is for $n = k_1$ and $m = k_2$:

$$\begin{aligned}&\mathrm{q-ROFS}_t\mathrm{RWA}\left(\mathscr{L}_{c_j}(\mathscr{M}_i), \mathscr{L}_{c_j}(\mathscr{M}_i), \ldots, \mathscr{L}_{c_{k_1}}(\mathscr{M}_{k_2})\right)\\&=\left[\left\{\sqrt[q]{1-\prod_{j=1}^{k_1}\left(\prod_{i=1}^{k_2}\left(1-\underline{\beta_{ij}}^{q}\right)^{t_i}\right)^{v_j}}, \prod_{j=1}^{k_1}\left(\prod_{i=1}^{k_2}\underline{\vartheta_{ij}}^{t_i}\right)^{v_j}\right\}, \left\{\left(\sqrt[q]{1-\prod_{j=1}^{k_1}\left(\prod_{i=1}^{k_2}\left(1-\overline{\beta_{ij}}^{q}\right)^{t_i}\right)^{v_j}}, \prod_{j=1}^{k_1}\left(\prod_{i=1}^{k_2}\overline{\vartheta_{ij}}^{t_i}\right)^{v_j}\right)\right\}\right].\end{aligned} \tag{19}$$

Suppose the result holds for $n = k_1 + 1$ and $m = k_2 + 1$, so we have

$$\begin{aligned}&\mathrm{q-ROFS}_t\mathrm{RWA}\left[\left(\mathscr{L}_{c_j}(\mathscr{M}_i), \mathscr{L}_{c_j}(\mathscr{M}_i), \ldots, \mathscr{L}_{c_{k_1}}(\mathscr{M}_{k_2})\right), \mathscr{L}_{c_{k_1+1}}(\mathscr{M}_{k_2+1})\right]\\&=\left[\oplus_{j=1}^{k_1} v_j\left(\oplus_{i=1}^{k_2} t_i \underline{\mathscr{L}_{c_j}}(\mathscr{M}_i)\right) \oplus v_{k_1+1}\left(t_{k_2+1} \underline{\mathscr{L}_{c_{k_1+1}}}(\mathscr{M}_{k_2+1})\right), \oplus_{j=1}^{k_1} v_j\left(\oplus_{i=1}^{k_2} t_i \overline{\mathscr{L}_{c_j}}(\mathscr{M}_i)\right) \oplus v_{k_1+1}\left(t_{k_2+1} \overline{\mathscr{L}_{c_{k_1+1}}}(\mathscr{M}_{k_2+1})\right)\right]\\&=\left[\left\{\sqrt[q]{1-\prod_{j=1}^{k_1+1}\left(\prod_{i=1}^{k_2+1}\left(1-\underline{\beta_{ij}}^{q}\right)^{t_i}\right)^{v_j}}, \prod_{j=1}^{k_1+1}\left(\prod_{i=1}^{k_2+1}\underline{\vartheta_{ij}}^{t_i}\right)^{v_j}\right\}, \left\{\left(\sqrt[q]{1-\prod_{j=1}^{k_1+1}\left(\prod_{i=1}^{k_2+1}\left(1-\overline{\beta_{ij}}^{q}\right)^{t_i}\right)^{v_j}}, \prod_{j=1}^{k_1+1}\left(\prod_{i=1}^{k_2+1}\overline{\vartheta_{ij}}^{t_i}\right)^{v_j}\right)\right\}\right].\end{aligned} \tag{20}$$

This implies the result is true for $n = k_1 + 1$ and $m = k_2 + 1$. Therefore, the result holds for all $m, n \geq 1$.

Since it is clear that $\underline{\mathscr{L}_{c_j}}(\mathscr{M}_i)$ and $\overline{\mathscr{L}_{c_j}}(\mathscr{M}_i)$ are q-ROFNs, by Definition 7, we have that $\oplus_{j=1}^{n} v_j(\oplus_{i=1}^{m} t_i \underline{\mathscr{L}_{c_j}}(\mathscr{M}_i))$ and $\oplus_{j=1}^{n} v_j(\oplus_{i=1}^{m} t_i \overline{\mathscr{L}_{c_j}}(\mathscr{M}_i))$ are also q-ROFNs. Therefore, $\mathrm{q-ROFS}_t\mathrm{RWA}$ $(\mathscr{L}_{c_1}(\mathscr{M}_1), \ldots, \mathscr{L}_{c_n}(\mathscr{M}_m))$ is also a q-ROFS$_t$RN in approximation space $(\mathscr{T}, \mathscr{E}, \mathscr{L})$. □

*Example 2.* Let $Y = \{s_1, s_2, s_3\}$ be a set and $\mathscr{M} = \{c_1, c_2\} \subseteq \mathscr{E}$ a set of parameters with weight vector $t = (0.25, 0.3, 0.45)^T$ for

$s_i$ $(i = 1, 2, 3)$ and $v = (0.55, 0.45)$ for $c_j$ $(j = 1, 2)$. Then, q-ROFS$_t$RNs is given in Table 3:

$$\begin{aligned}
&\text{q}-\text{ROFS}_t\text{RWA}\left(\mathscr{L}_{c_1}(\mathscr{M}_1), \ldots, \mathscr{L}_{c_n}(\mathscr{M}_m)\right) = \left[\oplus_{j=1}^{2} v_j\left(\oplus_{i=1}^{3} t_i \underline{\mathscr{L}_{c_j}}(\mathscr{M}_i)\right), \oplus_{j=1}^{2} v_j\left(\oplus_{i=1}^{3} t_i \overline{\mathscr{L}_{c_j}}(\mathscr{M}_i)\right)\right] \\
&= \left[\left\{\sqrt[3]{1-\left(\left[\left(1-0.9^3\right)^{0.25}\left(1-0.7^3\right)^{0.25}\left(1-0.29^3\right)^{0.45}\right]^{0.55}\left[\left(1-0.55^3\right)^{0.25}\left(1-0.92^3\right)^{0.25}\left(1-0.4^3\right)^{0.45}\right]^{0.45}\right)},\right.\right. \\
&\left.\left(0.3^{0.25}0.1^{0.3}0.25^{0.45}\right)^{0.55}\left(0.2^{0.25}0.3^{0.3}0.85^{0.45}\right)^{0.45}\right\}, \\
&\left\{\sqrt[3]{1-\left(\left[\left(1-0.4^3\right)^{0.25}\left(1-0.2^3\right)^{0.25}\left(1-0.65^3\right)^{0.45}\right]^{0.55}\left[\left(1-0.76^3\right)^{0.25}\left(1-0.6^3\right)^{0.25}\left(1-0.88^3\right)^{0.45}\right]^{0.45}\right)},\right. \\
&\left.\left.\left(0.4^{0.25}0.75^{0.3}0.15^{0.45}\right)^{0.55}\left(0.14^{0.25}0.3^{0.3}0.12^{0.45}\right)^{0.45}\right\}\right] \\
&= [(0.831432, 0.255487), (0.72581, 0.26258)].
\end{aligned} \tag{21}$$

From the analysis of Theorem 1, some related characteristics of q-ROFS$_t$RWA operator are given as follows:

**Theorem 2.** *Let* $\mathscr{L}_{c_j}(\mathscr{M}_i) = (\underline{\mathscr{L}_{c_j}}(\mathscr{M}_i), \overline{\mathscr{L}_{c_j}}(\mathscr{M}_i))$ $(i = 1, 2, \ldots, m,\ j = 1, 2, \ldots, n)$ *be the collection of q-ROFS$_t$RNs.* Let the weight vectors $t = (t_1, t_2, \ldots, t_m)^T$ and $v = (v_1, v_2, \ldots, v_n)^T$ for experts $s_i$ and parameters $c_j$ with $\sum_{i=1}^{m} t_i = 1$, $\sum_{j=1}^{n} v_j = 1$ and $0 \le t_i, v_j \le 1$, respectively. Then, the following properties hold for q-ROFS$_t$RWA operator:

*(i) (Idempotency): if* $\mathscr{L}_{c_j}(\mathscr{M}_i) = \mathscr{P}_c(\mathscr{N})$ *(for all* $i = 1, 2, \ldots, m$ *and* $j = 1, 2, \ldots n$*), where* $\mathscr{P}_c(\mathscr{M}) = (\underline{\mathscr{P}_c}(\mathscr{M}), \overline{\mathscr{P}_c}(\mathscr{M})) = ((\underline{b}, \underline{d}), (\overline{b}, \overline{d}))$*, then*

$$\begin{aligned}
&\text{q}-\text{ROFS}_t\text{RWA}\left(\mathscr{L}_{c_1}(\mathscr{M}_1), \mathscr{L}_{c_2}(\mathscr{M}_2), \ldots, \mathscr{L}_{c_n}(\mathscr{M}_m)\right) \\
&= \mathscr{P}_c(\mathscr{M}).
\end{aligned} \tag{22}$$

*(ii) (Boundedness): let* $(\mathscr{L}_{c_j}(\mathscr{M}_i))^- = (\min_j \min_i \underline{\mathscr{L}_{c_j}}(\mathscr{M}_i), \max_j \max_i \overline{\mathscr{L}_{c_j}}(\mathscr{M}_i))$ *and* $(\mathscr{L}_{c_j}(\mathscr{M}_i))^+ = (\max_j \max_i \underline{\mathscr{L}_{c_j}}(\mathscr{M}_i), \min_j \min_i \overline{\mathscr{L}_{c_j}}(\mathscr{M}_i))$*. Then,*

$$\begin{aligned}
&\left(\mathscr{L}_{c_j}(\mathscr{M}_i)\right)^- \le \text{q}-\text{ROFS}_t\text{RWA}\left(\mathscr{L}_{c_1}(\mathscr{M}_1), \mathscr{L}_{c_2}\right. \\
&\left.(\mathscr{M}_2), \ldots, \mathscr{L}_{c_n}(\mathscr{M}_m)\right) \le \left(\mathscr{L}_{c_j}(\mathscr{M}_i)\right)^+.
\end{aligned} \tag{23}$$

*(iii) (Monotonicity): let* $\mathscr{P}_{c_j}(\mathscr{N}_i) = (\underline{\mathscr{P}_{c_j}}(\mathscr{N}_i), \overline{\mathscr{P}_{c_j}}(\mathscr{N}_i))$ $(i = 1, 2, \ldots, m,\ j = 1, 2, \ldots, n)$ *be another collection of q-ROFS$_t$RNs such that* $\underline{\mathscr{P}_{c_j}}(\mathscr{N}_i) \le \underline{\mathscr{L}_{c_j}}(\mathscr{M}_i)$ *and* $\overline{\mathscr{P}_{c_j}}(\mathscr{N}_i) \le \overline{\mathscr{L}_{c_j}}(\mathscr{M}_i)$*. Then,*

$$\begin{aligned}
&\text{q}-\text{ROFS}_t\text{RWA}\left(\mathscr{P}_{c_1}(\mathscr{N}_1), \mathscr{P}_{c_2}(\mathscr{N}_2), \ldots, \mathscr{P}_{c_n}(\mathscr{N}_m)\right) \\
&\le \text{q}-\text{ROFS}_t\text{RWA}\left(\mathscr{L}_{c_1}(\mathscr{M}_1), \mathscr{L}_{c_2}(\mathscr{M}_2), \ldots, \mathscr{L}_{c_n}(\mathscr{M}_m)\right).
\end{aligned} \tag{24}$$

*(iv) (Shift invariance): let* $\mathscr{P}_c(\mathscr{N}) = (\underline{\mathscr{P}_c}(\mathscr{N}), \overline{\mathscr{P}_c}(\mathscr{N})) = ((\underline{b}, \underline{d}), (\overline{b}, \overline{d}))$ *be any other q − ROFS$_t$RN. Then,*

$$\begin{aligned}
&\text{q}-\text{ROFS}_t\text{RWA}\left(\mathscr{L}_{c_1}(\mathscr{M}_1) \oplus \mathscr{P}_c(\mathscr{N}), \mathscr{L}_{c_2}(\mathscr{M}_2)\right. \\
&\left.\oplus \mathscr{P}_c(\mathscr{N}), \ldots, \mathscr{L}_{c_n}(\mathscr{M}_m) \oplus \mathscr{P}_c(\mathscr{N})\right) \\
&= \text{q}-\text{ROFS}_t\text{RWA}\left(\mathscr{L}_{c_1}(\mathscr{M}_1), \mathscr{L}_{c_2}(\mathscr{M}_2), \ldots, \mathscr{L}_{c_n}(\mathscr{M}_m)\right) \\
&\oplus \mathscr{P}_c(\mathscr{N}).
\end{aligned} \tag{25}$$

*(v) (Homogeneity): for any real number* $\lambda > 0$*,*

$$\begin{aligned}
&\text{q}-\text{ROFS}_t\text{RWA}\left(\lambda\mathscr{L}_{c_1}(\mathscr{M}_1), \lambda\mathscr{L}_{c_2}(\mathscr{M}_2), \ldots, \lambda\mathscr{L}_{c_n}(\mathscr{M}_m)\right) \\
&= \lambda\text{q}-\text{ROFS}_t\text{RWA}\left(\mathscr{L}_{c_1}(\mathscr{M}_1), \mathscr{L}_{c_2}(\mathscr{M}_2), \ldots, \mathscr{L}_{c_n}(\mathscr{M}_m)\right).
\end{aligned} \tag{26}$$

*Remark 2*

(a) If $q = 1$, so, in this case, the developed q-ROFS$_t$RWA operator reduces to IFS$_t$RWA operator

(b) If $q = 2$, so, in this case, the developed q-ROFS$_t$RWA operator reduces to PFS$_t$RWA operator

Table 3: Tabular representation of $\mathscr{L}_{c_j}(\mathscr{M}_i) = (\underline{\mathscr{L}_{c_j}}(\mathscr{M}_i), \overline{\mathscr{L}_{c_j}}(\mathscr{M}_i))$.

| $\mathscr{L}$ | $c_1$ | $c_2$ |
|---|---|---|
| $s_1$ | ((0.9, 0.3)), (0.8, 0.4) | ((0.55, 0.2)), (0.76, 0.14) |
| $s_2$ | ((0.7, 0.1), (0.2, 0.75)) | ((0.92, 0.3), (0.6, 0.3)) |
| $s_3$ | ((0.92, 0.25), (0.65, 0.15)) | ((0.4, 0.85), (0.88, 0.12)) |

Table 4: Tabular representation of $\mathscr{L}_{\delta c_j}(\mathscr{M}_i) = (\underline{\mathscr{L}_{\delta c_j}}(\mathscr{M}_i), \overline{\mathscr{L}_{\delta c_j}}(\mathscr{M}_i))$.

| $\mathscr{L}$ | $c_1$ | $c_2$ |
|---|---|---|
| $s_1$ | ((0.9, 0.2), (0.8, 0.4)) | ((0.92, 0.3), (0.6, 0.3)) |
| $s_2$ | ((0.92, 0.25), (0.65, 0.15)) | ((0.55, 0.2), (0.76, 0.14)) |
| $s_3$ | ((0.7, 0.1), (0.2, 0.75)) | ((0.4, 0.85), (0.88, 0.12)) |

(c) If the soft parameter $c_1$ is only one (means $n = 1$), then, in this case, the developed q-ROFS$_t$RWA operator reduces to q-ROFRWA operator

*4.2. q-Rung Orthopair Fuzzy Soft Rough Ordered Weighted Averaging (q-ROFS$_t$ROWA) Operator.* This subsection presents the detailed study of q-ROFS$_t$ROWA operator and some of its desirable characteristics such as Idempotency, Boundedness, and Monotonicity. The basic advantage of q-ROFS$_t$ROWG operator is to weight the ordered position of the q-ROFVs instead of weighting the values themselves.

*Definition 12.* Let $\mathscr{L}_{c_j}(\mathscr{M}_i) = (\underline{\mathscr{L}_{c_j}}(\mathscr{M}_i), \overline{\mathscr{L}_{c_j}}(\mathscr{M}_i))$ $(i = 1, 2, \ldots, m,\ j = 1, 2, \ldots, n)$ be the collection of q-ROFS$_t$RNs. Let the weight vectors $t = (t_1, t_2, \ldots, t_m)^T$ and $v = (v_1, v_2, \ldots, v_n)^T$ for experts $s_i$ and parameters $c_j$ with $\sum_{i=1}^m t_i = 1$, $\sum_{j=1}^n v_j = 1$, and $0 \le t_i, v_j \le 1$, respectively. The q-ROFS$_t$ROWA operator is defined as

$$\begin{aligned}&\text{q}-\text{ROFS}_t\text{ROWA}\left(\mathscr{L}_{c_1}(\mathscr{M}_1), \ldots, \mathscr{L}_{c_n}(\mathscr{M}_m)\right)\\&= \left(\oplus_{j=1}^n v_j\left(\oplus_{i=1}^m t_i \underline{\mathscr{L}_{\delta c_j}}(\mathscr{M}_i)\right), \oplus_{j=1}^n v_j\left(\oplus_{i=1}^m t_i \overline{\mathscr{L}_{\delta c_j}}(\mathscr{M}_i)\right)\right).\end{aligned} \tag{27}$$

In view of the above definition, the aggregated result for q-ROFS$_t$ROWA is described in Theorem 3.

**Theorem 3.** *Let $\mathscr{L}_{c_j}(\mathscr{M}_i) = (\underline{\mathscr{L}_{c_j}}(\mathscr{M}_i), \overline{\mathscr{L}_{c_j}}(\mathscr{M}_i))$ $(i = 1, 2, \ldots, m,\ j = 1, 2, \ldots, n)$ be the collection of q-ROFS$_t$RNs. Let the weight vectors $t = (t_1, t_2, \ldots, t_m)^T$ and $v = (v_1, v_2, \ldots, v_n)^T$ for experts $s_i$ and parameters $c_j$ with $\sum_{i=1}^m t_i = 1$, $\sum_{j=1}^n v_j = 1$, and $0 \le t_i, v_j \le 1$, respectively. Then, q-ROFS$_t$ROWA operator is given as*

$$\begin{aligned}&\text{q}-\text{ROFS}_t\text{ROWA}\left(\mathscr{L}_{c_1}(\mathscr{M}_1), \ldots, \mathscr{L}_{c_n}(\mathscr{M}_m)\right) = \left(\oplus_{j=1}^n v_j\left(\oplus_{i=1}^m t_i \underline{\mathscr{L}_{\delta c_j}}(\mathscr{M}_i)\right), \oplus_{j=1}^n v_j\left(\oplus_{i=1}^m t_i \overline{\mathscr{L}_{\delta c_j}}(\mathscr{M}_i)\right)\right)\\&= \left[\left(\sqrt[q]{1 - \prod_{j=1}^n \left(\prod_{i=1}^m \left(1 - \underline{\beta_{\delta ij}}^q\right)^{t_i}\right)^{v_j}}, \prod_{j=1}^n \left(\prod_{i=1}^m \underline{\vartheta_{\delta ij}}^{t_i}\right)^{v_j}\right), \left(\sqrt[q]{1 - \prod_{j=1}^n \left(\prod_{i=1}^m \left(1 - \overline{\beta_{\delta ij}}^q\right)^{t_i}\right)^{v_j}}, \prod_{j=1}^n \left(\prod_{i=1}^m \overline{\vartheta_{\delta ij}}^{t_i}\right)^{v_j}\right)\right],\end{aligned} \tag{28}$$

*where $\mathscr{L}_{\delta c_j}(\mathscr{M}_i) = (\underline{\mathscr{L}_{\delta c_j}}(\mathscr{M}_i), \overline{\mathscr{L}_{\delta c_j}}(\mathscr{M}_i))$ denotes the largest value of the permutation from ith row and jth column of the collection $i \times j$ q-ROFS$_t$RNs $\mathscr{L}_{c_j}(\mathscr{M}_i) = (\underline{\mathscr{L}_{c_j}}(\mathscr{M}_i), \overline{\mathscr{L}_{c_j}}(\mathscr{M}_i))$.*

*Example 3.* Consider Table 3 of Example 2, for the collection q-ROFS$_t$RNs $\mathscr{L}_{c_j}(\mathscr{M}_i) = (\underline{\mathscr{L}_{c_j}}(\mathscr{M}_i), \overline{\mathscr{L}_{c_j}}(\mathscr{M}_i))$ and the new ordered of tabular representation of $\mathscr{L}_{c_j}(\mathscr{M}_i)$ through score function is given in Table 4.

Now, $\text{q}-\text{ROFS}_t\text{ROWA}(\mathscr{L}_{c_1}(\mathscr{M}_1), \ldots, \mathscr{L}_{c_n}(\mathscr{M}_m)) = [\oplus_{j=1}^2 v_j(\oplus_{i=1}^3 t_i \underline{\mathscr{L}_{\delta c_j}}(\mathscr{M}_i)), \oplus_{j=1}^2 v_j(\oplus_{i=1}^3 t_i \overline{\mathscr{L}_{\delta c_j}}(\mathscr{M}_i))]$

$$\begin{aligned}&\text{q}-\text{ROFS}_t\text{ROWA}\left(\mathscr{L}_{c_1}(\mathscr{M}_1), \ldots, \mathscr{L}_{c_n}(\mathscr{M}_m)\right)\\&= ((0.838595, 0.261642), (0.727318, 0.255189)).\end{aligned} \tag{29}$$

From the analysis of Theorem 3, the following desirable properties hold for q-ROFS$_t$ROWA operator.

**Theorem 4.** *Let $\mathscr{L}_{c_j}(\mathscr{M}_i) = (\underline{\mathscr{L}_{c_j}}(\mathscr{M}_i), \overline{\mathscr{L}_{c_j}}(\mathscr{M}_i))$ $(i = 1, 2, \ldots, m,\ j = 1, 2, \ldots, n)$ be the collection of q-ROFS$_t$RNs. Let the weight vectors $t = (t_1, t_2, \ldots, t_m)^T$ and $v = (v_1, v_2, \ldots, v_n)^T$ for experts $s_i$ and parameters $c_j$ with $\sum_{i=1}^m t_i = 1$, $\sum_{j=1}^n v_j = 1$, and $0 \le t_i, v_j \le 1$, respectively. Then, the following properties hold for q-ROFS$_t$ROWA operator:*

*(i) (Idempotency): if $\mathscr{L}_{\delta c_j}(\mathscr{M}_i) = \mathscr{P}_c(\mathscr{N})$ (for all $i = 1, 2, \ldots, m$ and $j = 1, 2, \ldots, n$), where $\mathscr{P}_c(\mathscr{M}) = (\underline{\mathscr{P}_c}(\mathscr{M}), \overline{\mathscr{P}_c}(\mathscr{M})) = ((\underline{b}, \underline{d}), (\overline{b}, \overline{d}))$, then*

$$\text{q}-\text{ROFS}_t\text{ROWA}\left(\mathscr{L}_{c_1}(\mathscr{M}_1), \mathscr{L}_{c_2}(\mathscr{M}_2), \ldots, \mathscr{L}_{c_n}(\mathscr{M}_m)\right) = \mathscr{P}_c(\mathscr{M}). \tag{30}$$

*(ii) (Boundedness): let $(\mathscr{L}_{\delta c_j}(\mathscr{M}_i))^- = (\min_j \min_i \underline{\mathscr{L}_{\delta c_j}}(\mathscr{M}_i), \max_j \max_i \overline{\mathscr{L}_{\delta c_j}}(\mathscr{M}_i))$ and $(\mathscr{L}_{\delta c_j}(\mathscr{M}_i))^+ = (\max_j \max_i \underline{\mathscr{L}_{\delta c_j}}(\mathscr{M}_i), \min_j \min_i \overline{\mathscr{L}_{\delta c_j}}(\mathscr{M}_i))$. Then,*

$$\begin{aligned}&\left(\mathscr{L}_{\delta c_j}(\mathscr{M}_i)\right)^- \le \text{q}-\text{ROFS}_t\text{ROWA}\left(\mathscr{L}_{c_1}(\mathscr{M}_1),\right.\\&\left.\mathscr{L}_{c_2}(\mathscr{M}_2), \ldots, \mathscr{L}_{c_n}(\mathscr{M}_m)\right) \le \left(\mathscr{L}_{\delta c_j}(\mathscr{M}_i)\right)^+.\end{aligned} \tag{31}$$

*(iii) (Monotonicity): let $\mathscr{P}_{c_j}(\mathscr{N}_i) = (\underline{\mathscr{P}_{c_j}}(\mathscr{N}_i), \overline{\mathscr{P}_{c_j}}(\mathscr{N}_i))$ $(i = 1, 2, \ldots, m,\ j = 1, 2, \ldots, n)$ be another*

*collection of q-ROFS$_t$RNs such that* $\underline{\mathscr{P}_{c_j}}(\mathscr{N}_i) \le \underline{\mathscr{L}_{c_j}}(\mathscr{M}_i)$ *and* $\overline{\mathscr{P}_{c_j}}(\mathscr{N}_i) \le \overline{\mathscr{L}_{c_j}}(\mathscr{M}_i)$. *Then,*

$$\begin{aligned}&\text{q}-\text{ROFS}_t\text{ROWA}\left(\mathscr{P}_{c_1}(\mathscr{M}_1), \mathscr{P}_{c_2}(\mathscr{M}_2), \ldots, \mathscr{P}_{c_n}(\mathscr{M}_m)\right)\\ &\quad \le \text{q}-\text{ROFS}_t\text{ROWA}\left(\mathscr{L}_{c_1}(\mathscr{M}_1), \mathscr{L}_{c_2}(\mathscr{M}_2), \ldots, \mathscr{L}_{c_n}(\mathscr{M}_m)\right).\end{aligned} \tag{32}$$

*(iv) (Shift invariance): let* $\mathscr{P}_c(\mathscr{N}) = (\underline{\mathscr{P}_c}(\mathscr{N}), \overline{\mathscr{P}_c}(\mathscr{N})) = ((\underline{b}, \underline{d}), (\overline{b}, \overline{d}))$ *be any other q − ROFS$_t$RN. Then,*

$$\begin{aligned}&\text{q}-\text{ROFS}_t\text{ROWA}\left(\mathscr{L}_{c_1}(\mathscr{M}_1) \oplus \mathscr{P}_c(\mathscr{N}), \mathscr{L}_{c_2}(\mathscr{M}_2)\right.\\ &\quad \left.\oplus \mathscr{P}_c(\mathscr{N}), \ldots, \mathscr{L}_{c_n}(\mathscr{M}_m) \oplus \mathscr{P}_c(\mathscr{N})\right)\\ &= \text{q}-\text{ROFS}_t\text{ROWA}\left(\mathscr{L}_{c_1}(\mathscr{M}_1), \mathscr{L}_{c_2}(\mathscr{M}_2), \ldots, \mathscr{L}_{c_n}(\mathscr{M}_m)\right)\\ &\quad \oplus \mathscr{P}_c(\mathscr{N}).\end{aligned} \tag{33}$$

*(v) (Homogeneity): for any real number* $\lambda > 0$,

$$\begin{aligned}&\text{q}-\text{ROFS}_t\text{ROWA}\left(\lambda\mathscr{L}_{c_1}(\mathscr{M}_1), \lambda\mathscr{L}_{c_2}(\mathscr{M}_2), \ldots,\right.\\ &\quad \left.\lambda\mathscr{L}_{c_n}(\mathscr{M}_m)\right) = \lambda\text{q}-\text{ROFS}_t\text{ROWA}\\ &\quad \left(\mathscr{L}_{c_1}(\mathscr{M}_1), \mathscr{L}_{c_2}(\mathscr{M}_2), \ldots, \mathscr{L}_{c_n}(\mathscr{M}_m)\right).\end{aligned} \tag{34}$$

*Remark 3*

(a) If $q = 1$, then the developed q-ROFS$_t$ROWA operator reduces to IFS$_t$ROWA operator

(b) If $q = 2$, then the developed q-ROFS$_t$ROWA operator reduces to PFS$_t$ROWA operator

(c) If the soft parameter $c_1$ is one (means $n = 1$), then the developed q-ROFS$_t$ROWA operator reduces to q-ROFROWA operator

*4.3. q-Rung Orthopair Fuzzy Soft Rough Hybrid Averaging (q-ROFS$_t$RHA) Operator.* From the analysis of Definitions 11 and 12, it is clear that the q-ROFS$_t$RWA operator weights only the q-ROFVs, while q-ROFS$_t$ROWA operator weights the ordered position of the q-ROFVs instead of weighting the values themselves. To overcome this limitation and motivated by the idea of combining weighted averaging and the ordered weighted averaging by using the combined notion of soft rough set, we present q-ROFS$_t$RHA operator, which weights both the given q-ROFV and its ordered position. The basic desirable properties of the developed operator are presented in detail.

*Definition 13.* Let $\mathscr{L}_{c_j}(\mathscr{M}_i) = (\underline{\mathscr{L}_{c_j}}(\mathscr{M}_i), \overline{\mathscr{L}_{c_j}}(\mathscr{M}_i))$ $(i = 1, 2, \ldots, m,\ j = 1, 2, \ldots, n)$ be the collection of q-ROFS$_t$RNs. *Let the weight vectors* $k = (k_1, k_2, \ldots, k_m)^T$ *and* $l = (l_1, l_2, \ldots, l_n)^T$ *for experts* $s_i$ *and parameters* $c_j$ *with* $\sum_{i=1}^m t_i = 1$, $\sum_{j=1}^n v_j = 1$, *and* $0 \le k_i, l_j \le 1$. *Consider* $t = (t_1, t_2, \ldots, t_m)^T$ *and* $v = (v_1, v_2, \ldots, v_n)^T$ *as the associated weight vectors of experts* $s_i$ *and parameters* $c_j$ *with* $\sum_{i=1}^m t_i = 1$, $\sum_{j=1}^n v_j = 1$, *and* $0 \le t_i, v_j \le 1$, *respectively. The q-ROFS$_t$RHA operator is defined as*

$$\text{q}-\text{ROFS}_t\text{RHA}\left(\mathscr{L}_{c_1}(\mathscr{M}_1), \ldots, \mathscr{L}_{c_n}(\mathscr{M}_m)\right) = \left(\oplus_{j=1}^n v_j\left(\oplus_{i=1}^m t_i \underline{\mathscr{L}^*_{\delta c_j}}(\mathscr{M}_i)\right), \oplus_{j=1}^n v_j\left(\oplus_{i=1}^m t_i \overline{\mathscr{L}^*_{\delta c_j}}(\mathscr{M}_i)\right)\right). \tag{35}$$

From the above definition, the aggregated result for q-ROFS$_t$RHA operator is described in Theorem 5.

**Theorem 5.** *Let* $\mathscr{L}_{c_j}(\mathscr{M}_i) = (\underline{\mathscr{L}_{c_j}}(\mathscr{M}_i), \overline{\mathscr{L}_{c_j}}(\mathscr{M}_i))$ $(i = 1, 2, \ldots, m,\ j = 1, 2, \ldots, n)$ *be the collection of q-ROFS$_t$RNs. Let the weight vectors* $k = (k_1, k_2, \ldots, k_m)^T$ *and* $l = (l_1, l_2, \ldots, l_n)^T$ *of experts* $s_i$ *and parameters* $c_j$ *with* $\sum_{i=1}^m t_i = 1$, $\sum_{j=1}^n v_j = 1$, *and* $0 \le k_i, l_j \le 1$. *Consider the associated weight vectors* $t = (t_1, t_2, \ldots, t_m)^T$ *and* $v = (v_1, v_2, \ldots, v_n)^T$ *of experts* $s_i$ *and parameters* $c_j$ *with* $\sum_{i=1}^m t_i = 1$, $\sum_{j=1}^n v_j = 1$, *and* $0 \le t_i, v_j \le 1$, *respectively. Then, q-ROFS$_t$RHA operator is given as*

$$\begin{aligned}&\text{q}-\text{ROFS}_t\text{RHA}\left(\mathscr{L}_{c_1}(\mathscr{M}_1), \ldots, \mathscr{L}_{c_n}(\mathscr{M}_m)\right) = \left(\oplus_{j=1}^n v_j\left(\oplus_{i=1}^m t_i \underline{\mathscr{L}^*_{\delta c_j}}(\mathscr{M}_i)\right), \oplus_{j=1}^n v_j\left(\oplus_{i=1}^m t_i \overline{\mathscr{L}^*_{\delta c_j}}(\mathscr{M}_i)\right)\right)\\ &= \left[\left(\sqrt[q]{1 - \prod_{j=1}^n\left(\prod_{i=1}^m\left(1 - \underline{\beta^*_{\delta ij}}^q\right)^{t_i}\right)^{v_j}}, \prod_{j=1}^n\left(\prod_{i=1}^m \underline{\vartheta^*_{\delta ij}}^{t_i}\right)^{v_j}\right), \left(\sqrt[q]{1 - \prod_{j=1}^n\left(\prod_{i=1}^m\left(1 - \overline{\beta^*_{\delta ij}}^q\right)^{t_i}\right)^{v_j}}, \prod_{j=1}^n\left(\prod_{i=1}^m \overline{\vartheta^*_{\delta ij}}^{t_i}\right)^{v_j}\right)\right],\end{aligned} \tag{36}$$

*where* $\mathscr{L}^*_{\delta c_j}(\mathscr{M}_i) = nkl\mathscr{L}_{c_j} = (nkl\,\underline{\mathscr{L}_{c_j}}(\mathscr{M}_i), nkl\overline{\mathscr{L}_{c_j}}(\mathscr{M}_i))$ *denotes the largest value of the permutation from ith row and jth column of the collection* $i \times j$ *q-ROFS$_t$RNs* $\mathscr{L}_{c_j}(\mathscr{M}_i) = (\underline{\mathscr{L}_{c_j}}(\mathscr{M}_i), \overline{\mathscr{L}_{c_j}}(\mathscr{M}_i))$ *and n represents the balancing coefficient.*

*Example 4.* Consider Table 3 of Example 2, for the collection q-ROFS$_t$RNs $\mathscr{L}_{c_j}(\mathscr{M}_i) = (\underline{\mathscr{L}_{c_j}}(\mathscr{M}_i), \overline{\mathscr{L}_{c_j}}(\mathscr{M}_i))$ with $k = (0.33, 0.37, 0.3)^T$ and $l = (0.42, 0.58)^T$ as the weight vectors of experts $s_i$ and parameters $c_j$. Consider $t = (0.36, 0.34, 0.3)^T$ and $v = (0.55, 0.45)^T$ as the associated weight vectors of experts $s_i$ and parameters $c_j$. The tabular representation of $\mathscr{L}^*_{\delta c_j}(\mathscr{M}_i)$ through operation law and score function is given in Tables 5 and 6. Now,

$$\begin{aligned}
&\text{q}-\text{ROFS}_t\text{RHA}\left(\mathscr{L}_{c_1}(\mathscr{M}_1), \ldots, \mathscr{L}_{c_n}(\mathscr{M}_m)\right) = \left(\oplus_{j=1}^{2} v_j\left(\oplus_{i=1}^{3} t_i \underline{\mathscr{L}^*_{\delta c_j}}(\mathscr{M}_i)\right), \oplus_{j=1}^{2} v_j\left(\oplus_{i=1}^{3} t_i \overline{\mathscr{L}^*_{\delta c_j}}(\mathscr{M}_i)\right)\right),\\
&\text{q}-\text{ROFS}_t\text{RHA}\left(\mathscr{L}_{c_1}(\mathscr{M}_1), \ldots, \mathscr{L}_{c_n}(\mathscr{M}_m)\right) = [(0.701609, 0.129765), (0.600425, 0.122969)].
\end{aligned} \tag{37}$$

From the analysis of Theorem 5, the following characteristics hold in q-ROFS$_t$RHA operator.

**Theorem 6.** *Let $\mathscr{L}_{c_j}(\mathscr{M}_i) = (\underline{\mathscr{L}_{c_j}}(\mathscr{M}_i), \overline{\mathscr{L}_{c_j}}(\mathscr{M}_i))$ $(i = 1, 2, \ldots, m,\ j = 1, 2, \ldots, n)$ be the collection of q-ROFS$_t$RNs. Let the weight vectors $k = (k_1, k_2, \ldots, k_m)^T$ and $l = (l_1, l_2, \ldots, l_n)^T$ of experts $s_i$ and parameters $c_j$ with $\sum_{i=1}^{m} t_i = 1$, $\sum_{j=1}^{n} v_j = 1$, and $0 \le k_i, l_j \le 1$. Consider the associated weight vectors $t = (t_1, t_2, \ldots, t_m)^T$ and $v = (v_1, v_2, \ldots, v_n)^T$ of experts $s_i$ and parameters $c_j$ with $\sum_{i=1}^{m} t_i = 1$, $\sum_{j=1}^{n} v_j = 1$, and $0 \le t_i, v_j \le 1$, respectively. Then, the following properties hold for q-ROFS$_t$RHA operator:*

*(i) (Idempotency): if $\mathscr{L}^*_{\delta c_j}(\mathscr{M}_i) = \mathscr{P}_c(\mathscr{N})$ (for all $i = 1, 2, \ldots, m$ and $j = 1, 2, \ldots, n$) where $\mathscr{P}_c(\mathscr{M}) = (\underline{\mathscr{P}_c}(\mathscr{M}), \overline{\mathscr{P}_c}(\mathscr{M})) = ((\underline{b}, \underline{d}), (\overline{b}, \overline{d}))$, then*

$$\text{q}-\text{ROFS}_t\text{RHA}\left(\mathscr{L}_{c_1}(\mathscr{M}_1), \mathscr{L}_{c_2}(\mathscr{M}_2), \ldots, \mathscr{L}_{c_n}(\mathscr{M}_m)\right) = \mathscr{P}_c(\mathscr{M}). \tag{38}$$

*(ii) (Boundedness): let $(\mathscr{L}^*_{\delta c_j}(\mathscr{M}_i))^- = (\min_j \min_i \underline{\mathscr{L}^*_{\delta c_j}}(\mathscr{M}_i), \max_j \max_i \overline{\mathscr{L}^*_{\delta c_j}}(\mathscr{M}_i))$ and $(\mathscr{L}^*_{\delta c_j}(\mathscr{M}_i))^+ = (\max_j \max_i \underline{\mathscr{L}^*_{\delta c_j}}(\mathscr{M}_i), \min_j \min_i \overline{\mathscr{L}^*_{\delta c_j}}(\mathscr{M}_i))$. Then,*

$$\left(\mathscr{L}^*_{\delta c_j}(\mathscr{M}_i)\right)^- \le \text{q}-\text{ROFS}_t\text{RHA}\left(\mathscr{L}_{c_1}(\mathscr{M}_1), \mathscr{L}_{c_2}(\mathscr{M}_2), \ldots, \mathscr{L}_{c_n}(\mathscr{M}_m)\right) \le \left(\mathscr{L}^*_{\delta c_j}(\mathscr{M}_i)\right)^+. \tag{39}$$

*(iii) (Monotonicity): let $\mathscr{P}_{c_j}(\mathscr{N}_i) = (\underline{\mathscr{P}_{c_j}}(\mathscr{N}_i), \overline{\mathscr{P}_{c_j}}(\mathscr{N}_i))$ $(i = 1, 2, \ldots, m,\ j = 1, 2, \ldots, n)$ be another collection of q-ROFS$_t$RNs such that $\underline{\mathscr{P}_{c_j}}(\mathscr{N}_i) \le \underline{\mathscr{L}_{c_j}}(\mathscr{M}_i)$ and $\overline{\mathscr{P}_{c_j}}(\mathscr{N}_i) \le \overline{\mathscr{L}_{c_j}}(\mathscr{M}_i)$. Then,*

$$\begin{aligned}
&\text{q}-\text{ROFS}_t\text{RHA}\left(\mathscr{P}_{c_1}(\mathscr{M}_1), \mathscr{P}_{c_2}(\mathscr{M}_2), \ldots, \mathscr{P}_{c_n}(\mathscr{M}_m)\right)\\
&\quad \le \text{q}-\text{ROFS}_t\text{RHA}\left(\mathscr{L}_{c_1}(\mathscr{M}_1), \mathscr{L}_{c_2}(\mathscr{M}_2), \ldots, \mathscr{L}_{c_n}(\mathscr{M}_m)\right).
\end{aligned} \tag{40}$$

*(iv) (Shift invariance): let $\mathscr{P}_c(\mathscr{N}) = (\underline{\mathscr{P}_c}(\mathscr{N}), \overline{\mathscr{P}_c}(\mathscr{N})) = ((\underline{b}, \underline{d}), (\overline{b}, \overline{d}))$ be any other q − ROFS$_t$RN. Then,*

$$\begin{aligned}
&\text{q}-\text{ROFS}_t\text{RHA}\left(\mathscr{L}_{c_1}(\mathscr{M}_1) \oplus \mathscr{P}_c(\mathscr{N}), \mathscr{L}_{c_2}(\mathscr{M}_2)\right.\\
&\quad \left.\oplus \mathscr{P}_c(\mathscr{N}), \ldots, \mathscr{L}_{c_n}(\mathscr{M}_m) \oplus \mathscr{P}_c(\mathscr{N})\right) = \text{q}-\text{ROFS}_t\text{RHA}\\
&\quad \left(\mathscr{L}_{c_1}(\mathscr{M}_1), \mathscr{L}_{c_2}(\mathscr{M}_2), \ldots, \mathscr{L}_{c_n}(\mathscr{M}_m)\right) \oplus \mathscr{P}_c(\mathscr{N}).
\end{aligned} \tag{41}$$

*(v) (Homogeneity): for any real number $\lambda > 0$,*

$$\begin{aligned}
&\text{q}-\text{ROFS}_t\text{RHA}\left(\lambda\mathscr{L}_{c_1}(\mathscr{M}_1), \lambda\mathscr{L}_{c_2}(\mathscr{M}_2), \ldots, \lambda\mathscr{L}_{c_n}(\mathscr{M}_m)\right)\\
&\quad = \lambda\text{q}-\text{ROFS}_t\text{RHA}\left(\mathscr{L}_{c_1}(\mathscr{M}_1), \mathscr{L}_{c_2}(\mathscr{M}_2), \ldots, \mathscr{L}_{c_n}(\mathscr{M}_m)\right).
\end{aligned} \tag{42}$$

*Remark 4*

(a) If $q = 1$, then q-ROFS$_t$RHA operator reduces to IFS$_t$RHA operator
(b) If $q = 2$, then q-ROFS$_t$RHA operator reduces to PFS$_t$RHA operator
(c) If soft parameter $c_1$ is one (means $n = 1$), then q-ROFS$_t$RHA operator reduces to q-ROFRHA operator
(d) If $kl = ((1/n), (1/n), \ldots, (1/n))^T$, then the proposed q-ROFS$_t$RHA operator reduces to q-ROFS$_t$RWA operator
(e) If $tv = ((1/n), (1/n), \ldots, (1/n))^T$, then the proposed q-ROFS$_t$RHA operator reduces to q-ROFS$_t$ROWA operator

## 5. q-Rung Orthopair Fuzzy Soft Rough Geometric (q-ROFS$_t$RG) Aggregation Operator

This section is devoted to the study of q-ROFS$_t$R geometric aggregation operators such as q-ROFS$_t$RWG, q-ROFS$_t$ROWG, and q-ROFS$_t$RHG operators. We will present the fundamental properties of these operators in detail.

Table 5: Tabular representation by using operation law for $\mathscr{L}^*_{\delta c_j}(\mathscr{M}_i) = nkl\mathscr{L}_{c_j} = (nkl\,\underline{\mathscr{L}_{c_j}}(\mathscr{M}_i), nkl\overline{\mathscr{L}_{c_j}}(\mathscr{M}_i))$.

| $\mathscr{L}$ | $c_1$ | $c_2$ |
|---|---|---|
| $s_1$ | ((0.7483, 0.1247), (0.6366, 0.1663)) | ((0.4629, 0.1148), (0.6561, 0.0804)) |
| $s_2$ | ((0.5624, 0.0466), (0.1551, 0.3497)) | ((0.8533, 0.1931), (0.5254, 0.1931)) |
| $s_3$ | ((0.7574, 0.0945), (0.4853, 0.0567)) | ((0.3238, 0.4437), (0.7661, 0.0626)) |

Table 6: Tabular representation after using score function $\mathscr{L}^*_{\delta c_j}(\mathscr{M}_i) = (\underline{\mathscr{L}^*_{\delta c_j}}(\mathscr{M}_i), \overline{\mathscr{L}^*_{\delta c_j}}(\mathscr{M}_i))$.

| $\mathscr{L}$ | $c_1$ | $c_2$ |
|---|---|---|
| $s_1$ | ((0.7483, 0.1247), (0.6366, 0.1663)) | ((0.8533, 0.1931), (0.5254, 0.1931)) |
| $s_2$ | ((0.7574, 0.0945), (0.4853, 0.0567)) | ((0.3238, 0.4437), (0.7661, 0.0626)) |
| $s_3$ | ((0.5624, 0.0466), (0.1551, 0.3497)) | ((0.4629, 0.1148), (0.6561, 0.0804)) |

*5.1. q-Rung Orthopair Fuzzy Soft Rough Weighted Geometric (q-ROFS$_t$RWG) Operator*

*Definition 14.* Let $\mathscr{L}_{c_j}(\mathscr{M}_i) = (\underline{\mathscr{L}_{c_j}}(\mathscr{M}_i), \overline{\mathscr{L}_{c_j}}(\mathscr{M}_i))\,(i = 1, 2, \ldots, m,\ j = 1, 2, \ldots, n)$ be the collection of q-ROFS$_t$RNs. Let the weight vectors $t = (t_1, t_2, \ldots, t_m)^T$ and $v = (v_1, v_2, \ldots, v_n)^T$ of experts $s_i$ and parameters $c_j$ with $\sum_{i=1}^m t_i = 1$, $\sum_{j=1}^n v_j = 1$, and $0 \le t_i, v_j \le 1$, respectively. The q-ROFS$_t$RWG operator is defined as

$$\text{q}-\text{ROFS}_t\text{RWG}\left(\mathscr{L}_{c_1}(\mathscr{M}_1), \ldots, \mathscr{L}_{c_n}(\mathscr{M}_m)\right) = \left[\oplus_{j=1}^n \left\{\oplus_{i=1}^m \left(\underline{\mathscr{L}_{c_j}}(\mathscr{M}_i)\right)^{t_i}\right\}^{v_j}, \oplus_{j=1}^n \left\{\oplus_{i=1}^m \left(\overline{\mathscr{L}_{c_j}}(\mathscr{M}_i)\right)^{t_i}\right\}^{v_j}\right]. \tag{43}$$

Based on the above definition the aggregated result for q-ROFS$_t$RWG operator is given in Theorem 7.

**Theorem 7.** *Let $\mathscr{L}_{c_j}(\mathscr{M}_i) = (\underline{\mathscr{L}_{c_j}}(\mathscr{M}_i), \overline{\mathscr{L}_{c_j}}(\mathscr{M}_i))$ $(i = 1, 2, \ldots, m,\ j = 1, 2, \ldots, n)$ be the collection of q-ROFS$_t$RNs. Let the weight vectors $t = (t_1, t_2, \ldots, t_m)^T$ and $v = (v_1, v_2, \ldots, v_n)^T$ of experts $s_i$ and parameters $c_j$ with $\sum_{i=1}^m t_i = 1$, $\sum_{j=1}^n v_j = 1$, and $0 \le t_i, v_j \le 1$, respectively. Then, q-ROFS$_t$RWG operator is given as*

$$\begin{aligned}\text{q}-\text{ROFS}_t\text{RWG}\left(\mathscr{L}_{c_1}(\mathscr{M}_1), \ldots, \mathscr{L}_{c_n}(\mathscr{M}_m)\right) &= \left[\oplus_{j=1}^n \left\{\oplus_{i=1}^m \left(\underline{\mathscr{L}_{c_j}}(\mathscr{M}_i)\right)^{t_i}\right\}^{v_j}, \oplus_{j=1}^n \left\{\oplus_{i=1}^m \left(\overline{\mathscr{L}_{c_j}}(\mathscr{M}_i)\right)^{t_i}\right\}^{v_j}\right] \\ &= \left[\left\{\prod_{j=1}^n \left(\prod_{i=1}^m \underline{\beta_{ij}}^{t_i}\right)^{v_j}, \sqrt[q]{1 - \prod_{j=1}^n \left(\prod_{i=1}^m \left(1 - \underline{\vartheta_{ij}}^q\right)^{t_i}\right)^{v_j}}\right\}, \left\{\prod_{j=1}^n \left(\prod_{i=1}^m \overline{\beta_{ij}}^{t_i}\right)^{v_j}, \sqrt[q]{1 - \prod_{j=1}^n \left(\prod_{i=1}^m \left(1 - \overline{\vartheta_{ij}}^q\right)^{t_i}\right)^{v_j}}\right\}\right].\end{aligned} \tag{44}$$

Since it is clear that $\underline{\mathscr{L}_{c_j}}(\mathscr{M}_i)$ and $\overline{\mathscr{L}_{c_j}}(\mathscr{M}_i)$ are q-ROFNs, by Definition 7, we have that $\oplus_{j=1}^n \left\{\oplus_{i=1}^m (\underline{\mathscr{L}_{c_j}}(\mathscr{M}_i))^{t_i}\right\}^{v_j}$ and $\oplus_{j=1}^n \left\{\oplus_{i=1}^m (\overline{\mathscr{L}_{c_j}}(\mathscr{M}_i))^{t_i}\right\}^{v_j}$ are also q-ROFNs. Therefore,

$\mathrm{q-ROFS}_t\mathrm{RWG}(\mathscr{L}_{c_1}(\mathscr{M}_1),\ldots,\mathscr{L}_{c_n}(\mathscr{M}_m))$ is also a q-ROFS$_t$RN in approximation space $(\mathscr{T},\mathscr{E},\mathscr{L})$.

*Example 5.* Consider Table 3 of Example 2, for the collection q-ROFS$_t$RNs $\mathscr{L}_{c_j}(\mathscr{M}_i)=(\underline{\mathscr{L}_{c_j}}(\mathscr{M}_i),\overline{\mathscr{L}_{c_j}}(\mathscr{M}_i))$. Then, the aggregated result for $\mathrm{q-ROFS}_t\mathrm{RWG}$ is given as

$$\begin{aligned}
&\mathrm{q-ROFS}_t\mathrm{RWG}\left(\mathscr{L}_{c_1}(\mathscr{M}_1),\ldots,\mathscr{L}_{c_n}(\mathscr{M}_m)\right)=\left[\oplus_{j=1}^{2}\left\{\oplus_{i=1}^{3}\left(\underline{\mathscr{L}_{c_j}}(\mathscr{M}_i)\right)^{t_i}\right\}^{v_j},\oplus_{j=1}^{2}\left\{\oplus_{i=1}^{3}\left(\overline{\mathscr{L}_{c_j}}(\mathscr{M}_i)\right)^{t_i}\right\}^{v_j}\right]\\
&=\Bigg[\Bigg\{\left(0.9^{0.25}0.7^{0.3}0.92^{0.45}\right)^{0.55}\left(0.55^{0.25}0.92^{0.3}0.4^{0.45}\right)^{0.45},\\
&\sqrt[3]{1-\left(\left[\left(1-0.3^3\right)^{0.25}\left(1-0.1^3\right)^{0.25}\left(1-0.25^3\right)^{0.45}\right]^{0.55}\left[\left(1-0.2^3\right)^{0.25}\left(1-0.3^3\right)^{0.25}\left(1-0.85^3\right)^{0.45}\right]^{0.45}\right)}\Bigg\},\\
&\Bigg\{\left(0.8^{0.25}0.2^{0.3}0.65^{0.45}\right)^{0.55}\left(0.76^{0.25}0.6^{0.3}0.88^{0.45}\right)^{0.45},\\
&\sqrt[3]{1-\left(\left[\left(1-0.4^3\right)^{0.25}\left(1-0.75^3\right)^{0.25}\left(1-0.15^3\right)^{0.45}\right]^{0.55}\left[\left(1-0.14^3\right)^{0.25}\left(1-0.3^3\right)^{0.25}\left(1-0.12^3\right)^{0.45}\right]^{0.45}\right)}\Bigg\}\Bigg]\\
&=[(0.715607,0.509925),(0.573442,0.484819)].
\end{aligned}\tag{45}$$

From the analysis of Theorem 7, the following hold for q-ROFS$_t$RWG operator.

**Theorem 8.** *Let $\mathscr{L}_{c_j}(\mathscr{M}_i)=(\underline{\mathscr{L}_{c_j}}(\mathscr{M}_i),\overline{\mathscr{L}_{c_j}}(\mathscr{M}_i))$ $(i=1,2,\ldots,m,\ j=1,2,\ldots,n)$ be the collection of q-ROFS$_t$RNs. Let the weight vectors $t=(t_1,t_2,\ldots,t_m)^T$ and $v=(v_1,v_2,\ldots,v_n)^T$ of experts $s_i$ and parameters $c_j$ with $\sum_{i=1}^{m}t_i=1$, $\sum_{j=1}^{n}v_j=1$, and $0\le t_i,\ v_j\le 1$, respectively. Then, the following properties hold for q-ROFS$_t$RWG operator:*

*(i) (Idempotency): if $\mathscr{L}_{c_j}(\mathscr{M}_i)=\mathscr{P}_c(\mathscr{N})$ (for all $i=1,2,\ldots,m$ and $j=1,2,\ldots,n$) where $\mathscr{P}_c(\mathscr{M})=(\underline{\mathscr{P}_c}(\mathscr{M}),\overline{\mathscr{P}_c}(\mathscr{M}))=((\underline{b},\underline{d}),(\overline{b},\overline{d}))$, then*

$$\begin{aligned}
&\mathrm{q-ROFS}_t\mathrm{RWG}\left(\mathscr{L}_{c_1}(\mathscr{M}_1),\mathscr{L}_{c_2}(\mathscr{M}_2),\ldots,\mathscr{L}_{c_n}(\mathscr{M}_m)\right)\\
&=\mathscr{P}_c(\mathscr{M}).
\end{aligned}\tag{46}$$

*(ii) (Boundedness): let $(\mathscr{L}_{c_j}(\mathscr{M}_i))^-=(\min_j\min_i\underline{\mathscr{L}_{c_j}}(\mathscr{M}_i),\max_j\ \max_i\overline{\mathscr{L}_{c_j}}(\mathscr{M}_i))$ and $(\mathscr{L}_{c_j}(\mathscr{M}_i))^+=(\max_j\ \max_i\underline{\mathscr{L}_{c_j}}(\mathscr{M}_i),\min_j\min_i\overline{\mathscr{L}_{c_j}}(\mathscr{M}_i))$. Then,*

$$\begin{aligned}
&\left(\mathscr{L}_{c_j}(\mathscr{M}_i)\right)^-\le\mathrm{q-ROFS}_t\mathrm{RWG}\left(\mathscr{L}_{c_1}(\mathscr{M}_1),\mathscr{L}_{c_2}(\mathscr{M}_2),\ldots,\right.\\
&\left.\mathscr{L}_{c_n}(\mathscr{M}_m)\right)\le\left(\mathscr{L}_{c_j}(\mathscr{M}_i)\right)^+.
\end{aligned}\tag{47}$$

*(iii) (Monotonicity): let $\mathscr{P}_{c_j}(\mathscr{N}_i)=(\underline{\mathscr{P}_{c_j}}(\mathscr{N}_i),\overline{\mathscr{P}_{c_j}}(\mathscr{N}_i))$ $(i=1,2,\ldots,m,\ j=1,2,\ldots,n)$ be another collection of q-ROFS$_t$RNs such that $\underline{\mathscr{P}_{c_j}}(\mathscr{N}_i)\le\underline{\mathscr{L}_{c_j}}(\mathscr{M}_i)$ and $\overline{\mathscr{P}_{c_j}}(\mathscr{N}_i)\le\overline{\mathscr{L}_{c_j}}(\mathscr{M}_i)$. Then,*

$$\begin{aligned}
&\mathrm{q-ROFS}_t\mathrm{RWG}\left(\mathscr{P}_{c_1}(\mathscr{M}_1),\mathscr{P}_{c_2}(\mathscr{M}_2),\ldots,\mathscr{P}_{c_n}(\mathscr{M}_m)\right)\\
&\le\mathrm{q-ROFS}_t\mathrm{RWG}\left(\mathscr{L}_{c_1}(\mathscr{M}_1),\mathscr{L}_{c_2}(\mathscr{M}_2),\ldots,\mathscr{L}_{c_n}(\mathscr{M}_m)\right).
\end{aligned}\tag{48}$$

*(iv) (Shift invariance): let $\mathscr{P}_c(\mathscr{N})=(\underline{\mathscr{P}_c}(\mathscr{N}),\overline{\mathscr{P}_c}(\mathscr{N}))=((\underline{b},\underline{d}),(\overline{b},\overline{d}))$ be any other q-ROFS$_t$RN. Then,*

$$\begin{aligned}
&\mathrm{q-ROFS}_t\mathrm{RWG}\left(\mathscr{L}_{c_1}(\mathscr{M}_1)\oplus\mathscr{P}_c(\mathscr{N}),\mathscr{L}_{c_2}(\mathscr{M}_2)\right.\\
&\left.\oplus\mathscr{P}_c(\mathscr{N}),\ldots,\mathscr{L}_{c_n}(\mathscr{M}_m)\oplus\mathscr{P}_c(\mathscr{N})\right)\\
&=\mathrm{q-ROFS}_t\mathrm{RWG}\left(\mathscr{L}_{c_1}(\mathscr{M}_1),\mathscr{L}_{c_2}(\mathscr{M}_2),\ldots,\mathscr{L}_{c_n}(\mathscr{M}_m)\right)\\
&\oplus\mathscr{P}_c(\mathscr{N}).
\end{aligned}\tag{49}$$

*(v) (Homogeneity): for any real number $\lambda>0$,*

$$\begin{aligned}
&\mathrm{q-ROFS}_t\mathrm{RWG}\left(\lambda\mathscr{L}_{c_1}(\mathscr{M}_1),\lambda\mathscr{L}_{c_2}(\mathscr{M}_2),\ldots,\lambda\mathscr{L}_{c_n}(\mathscr{M}_m)\right)\\
&=\lambda\mathrm{q-ROFS}_t\mathrm{RWG}\left(\mathscr{L}_{c_1}(\mathscr{M}_1),\mathscr{L}_{c_2}(\mathscr{M}_2),\ldots,\mathscr{L}_{c_n}(\mathscr{M}_m)\right).
\end{aligned}\tag{50}$$

*Remark 5*

(a) If $q=1$, then q-ROFS$_t$RWG operator reduces to IFS$_t$RWG operator

(b) If $q=2$, then q-ROFS$_t$RWG operator reduces to PFS$_t$RWG operator

(c) If soft parameter $c_1$ is one (means $n=1$), then q-ROFS$_t$RWG operator reduces to q-ROFRWG operator

*5.2. q-Rung Orthopair Fuzzy Soft Rough Ordered Weighted Geometric (q-ROFS$_t$ROWG) Operator.* Here, we will put forward the detailed study of q-ROFS$_t$ROWG operator and some of its fundamental properties. The q-ROFS$_t$ROWG operator weights the ordered position of the q-ROFVs instead of weighting the values themselves.

*Definition 15.* Let $\mathscr{L}_{c_j}(\mathscr{M}_i) = (\underline{\mathscr{L}_{c_j}}(\mathscr{M}_i), \overline{\mathscr{L}_{c_j}}(\mathscr{M}_i))$ $(i = 1, 2, \ldots, m,\ j = 1, 2, \ldots, n)$ be ~~the~~ collection of q-ROFS$_t$RNs. Let the weight vectors $t = (t_1, t_2, \ldots, t_m)^T$ and $v = (v_1, v_2, \ldots, v_n)^T$ of experts $s_i$ and parameters $c_j$ with $\sum_{i=1}^{m} t_i = 1$, $\sum_{j=1}^{n} v_j = 1$, and $0 \le t_i,\ v_j \le 1$, respectively. The q-ROFS$_t$ROWG operator is defined as

$$\text{q}-\text{ROFS}_t\text{ROWG}\left(\mathscr{L}_{c_1}(\mathscr{M}_1), \ldots, \mathscr{L}_{c_n}(\mathscr{M}_m)\right) = \left[\oplus_{j=1}^{n}\left\{\oplus_{i=1}^{m}\left(\underline{\mathscr{L}_{\delta c_j}}(\mathscr{M}_i)\right)^{t_i}\right\}^{v_j}, \oplus_{j=1}^{n}\left\{\oplus_{i=1}^{m}\left(\overline{\mathscr{L}_{\delta c_j}}(\mathscr{M}_i)\right)^{t_i}\right\}^{v_j}\right]. \tag{51}$$

In view of the above definition, the aggregated result for q-ROFS$_t$ROWG operator is described in Theorem 9.

**Theorem 9.** *Let $\mathscr{L}_{c_j}(\mathscr{M}_i) = (\underline{\mathscr{L}_{c_j}}(\mathscr{M}_i), \overline{\mathscr{L}_{c_j}}(\mathscr{M}_i))$ $(i = 1, 2, \ldots, m,\ j = 1, 2, \ldots, n)$ be ~~the~~ collection of q-ROFS$_t$RNs. Let the weight vectors $t = (t_1, t_2, \ldots, t_m)^T$ and $v = (v_1, v_2, \ldots, v_n)^T$ of experts $s_i$ and parameters $c_j$ with $\sum_{i=1}^{m} t_i = 1$, $\sum_{j=1}^{n} v_j = 1$, and $0 \le t_i,\ v_j \le 1$, respectively. Then, q-ROFS$_t$ROWG operator is given as*

$$\begin{aligned}\text{q}-\text{ROFS}_t\text{ROWG}\left(\mathscr{L}_{c_1}(\mathscr{M}_1), \ldots, \mathscr{L}_{c_n}(\mathscr{M}_m)\right) &= \left[\oplus_{j=1}^{n}\left\{\oplus_{i=1}^{m}\left(\underline{\mathscr{L}_{\delta c_j}}(\mathscr{M}_i)\right)^{t_i}\right\}^{v_j}, \oplus_{j=1}^{n}\left\{\oplus_{i=1}^{m}\left(\overline{\mathscr{L}_{\delta c_j}}(\mathscr{M}_i)\right)^{t_i}\right\}^{v_j}\right] \\ &= \left[\left\{\prod_{j=1}^{n}\left(\prod_{i=1}^{m}\underline{\beta_{\delta ij}}^{t_i}\right)^{v_j}, \sqrt[q]{1-\prod_{j=1}^{n}\left(\prod_{i=1}^{m}\left(1-\underline{\vartheta_{\delta ij}}^{q}\right)^{t_i}\right)^{v_j}}\right\}, \left\{\prod_{j=1}^{n}\left(\prod_{i=1}^{m}\overline{\beta_{\delta ij}}^{t_i}\right)^{v_j}, \sqrt[q]{1-\prod_{j=1}^{n}\left(\prod_{i=1}^{m}\left(1-\overline{\vartheta_{\delta ij}}^{q}\right)^{t_i}\right)^{v_j}}\right\}\right],\end{aligned} \tag{52}$$

*where $\mathscr{L}_{\delta c_j}(\mathscr{M}_i) = (\underline{\mathscr{L}_{\delta c_j}}(\mathscr{M}_i), \overline{\mathscr{L}_{\delta c_j}}(\mathscr{M}_i))$ denotes the largest value of the permutation from ith row and jth column of the collection $i \times j$ q-ROFS$_t$RNs, $\mathscr{L}_{c_j}(\mathscr{M}_i) = (\underline{\mathscr{L}_{c_j}}(\mathscr{M}_i), \overline{\mathscr{L}_{c_j}}(\mathscr{M}_i))$.*

*Remark 6*

(a) If $q = 1$, then q-ROFS$_t$ROWG operator reduces to IFS$_t$ROWG operator

(b) If $q = 2$, then q-ROFS$_t$ROWG operator reduces to PFS$_t$ROWG operator

(c) If the soft parameter $c_1$ is one (means $n = 1$), then q-ROFS$_t$ROWG operator reduces to q-ROFROWG operator

*5.3. q-Rung Orthopair Fuzzy Soft Rough Hybrid Geometric (q-ROFS$_t$RHG) Operator.* From the analysis of Definitions 14 and 15, it is clear that the q-ROFS$_t$RWG operator weights only the q-ROFVs, while q-ROFS$_t$ROWG operator weights the ordered position of the q-ROFVs instead of weighting themselves. Motivated by the idea of combining weighted geometric and the ordered weighted geometric by using the combined notion of soft rough set, we present q-ROFS$_t$RHG operator, which weights both the given q-ROFV and its ordered position. The basic desirable properties of the developed operator are presented in detail.

*Definition 16.* Let $\mathscr{L}_{c_j}(\mathscr{M}_i) = (\underline{\mathscr{L}_{c_j}}(\mathscr{M}_i), \overline{\mathscr{L}_{c_j}}(\mathscr{M}_i))$ $(i = 1, 2, \ldots, m,\ j = 1, 2, \ldots, n)$ be the collection of q-ROFS$_t$RNs. Let the weight vectors $k = (k_1, k_2, \ldots, k_m)^T$ and $l = (l_1, l_2, \ldots, l_n)^T$ of experts $s_i$ and parameters $c_j$ with $\sum_{i=1}^{m} t_i = 1$, $\sum_{j=1}^{n} v_j = 1$, and $0 \le k_i,\ l_j \le 1$. Consider the associated weight vectors $t = (t_1, t_2, \ldots, t_m)^T$ and $v = (v_1, v_2, \ldots, v_n)^T$ of experts $s_i$ and parameters $c_j$ with $\sum_{i=1}^{m} t_i = 1$, $\sum_{j=1}^{n} v_j = 1$, and $0 \le t_i,\ v_j \le 1$, respectively. The q-ROFS$_t$RHG operator is defined as

$$\text{q}-\text{ROFS}_t\text{RHG}\left(\mathscr{L}_{c_1}(\mathscr{M}_1), \ldots, \mathscr{L}_{c_n}(\mathscr{M}_m)\right) = \left[\oplus_{j=1}^{n}\left\{\oplus_{i=1}^{m}\left(\underline{\mathscr{L}^*_{\delta c_j}}(\mathscr{M}_i)\right)^{t_i}\right\}^{v_j}, \oplus_{j=1}^{n}\left\{\oplus_{i=1}^{m}\left(\overline{\mathscr{L}^*_{\delta c_j}}(\mathscr{M}_i)\right)^{t_i}\right\}^{v_j}\right]. \tag{53}$$

From the above definition, the aggregated value for q-ROFS$_t$RHG operator is described in Theorem 11.

**Theorem 10.** *Let* $\mathscr{L}_{c_j}(\mathscr{M}_i) = (\underline{\mathscr{L}_{c_j}}(\mathscr{M}_i), \overline{\mathscr{L}_{c_j}}(\mathscr{M}_i))$ $(i = 1, 2, \ldots, m,\ j = 1, 2, \ldots, n)$ *be the collection of q-ROFS$_t$RNs. Let the weight vectors* $k = (k_1, k_2, \ldots, k_m)^T$ *and* $l = (l_1, l_2, \ldots, l_n)^T$ *of experts* $s_i$ *and parameters* $c_j$ *with* $\sum_{i=1}^m t_i = 1$, $\sum_{j=1}^n v_j = 1$, *and* $0 \le k_i, l_j \le 1$. *Consider the associated weight vectors* $t = (t_1, t_2, \ldots, t_m)^T$ *and* $v = (v_1, v_2, \ldots, v_n)^T$ *of experts* $s_i$ *and parameters* $c_j$ *with* $\sum_{i=1}^m t_i = 1$, $\sum_{j=1}^n v_j = 1$, *and* $0 \le t_i, v_j \le 1$, *respectively. Then, q-ROFS$_t$RHG operator is given as*

$$\begin{aligned}
\text{q}-\text{ROFS}_t\text{RHG}\left(\mathscr{L}_{c_1}(\mathscr{M}_1), \ldots, \mathscr{L}_{c_n}(\mathscr{M}_m)\right) &= \left[\oplus_{j=1}^n \left\{\oplus_{i=1}^m \left(\underline{\mathscr{L}^*_{\delta c_j}}(\mathscr{M}_i)\right)^{t_i}\right\}^{v_j}, \oplus_{j=1}^n \left\{\oplus_{i=1}^m \left(\overline{\mathscr{L}^*_{\delta c_j}}(\mathscr{M}_i)\right)^{t_i}\right\}^{v_j}\right] \\
&= \left[\left\{\prod_{j=1}^n \left(\prod_{i=1}^m \underline{\beta^*_{\delta ij}}^{t_i}\right)^{v_j}, \sqrt[q]{1 - \prod_{j=1}^n \left(\prod_{i=1}^m \left(1 - \underline{\vartheta^*_{\delta ij}}^{q}\right)^{t_i}\right)^{v_j}}\right\}, \left\{\prod_{j=1}^n \left(\prod_{i=1}^m \overline{\beta^*_{\delta ij}}^{t_i}\right)^{v_j}, \sqrt[q]{1 - \prod_{j=1}^n \left(\prod_{i=1}^m \left(1 - \overline{\vartheta^*_{\delta ij}}^{q}\right)^{t_i}\right)^{v_j}}\right\}\right],
\end{aligned} \tag{54}$$

*where* $\mathscr{L}^*_{\delta c_j}(\mathscr{M}_i) = (\mathscr{L}_{c_j})^{nkl} = (\underline{(\mathscr{L}_{c_j}}(\mathscr{M}_i))^{nkl}, (\overline{\mathscr{L}_{c_j}}(\mathscr{M}_i))^{nkl})$ *denotes the largest value of the permutation from ith row and jth column of the collection* $i \times j$ *q-ROFS$_t$RNs* $\mathscr{L}_{c_j}(\mathscr{M}_i) = (\underline{\mathscr{L}_{c_j}}(\mathscr{M}_i), \overline{\mathscr{L}_{c_j}}(\mathscr{M}_i))$ *and n represents the balancing coefficient.*

*Example 6.* Consider Tables 2, 4, and 5 of Examples 2 and 4, for the collection q-ROFS$_t$RNs $\mathscr{L}_{c_j}(\mathscr{M}_i) = (\underline{\mathscr{L}_{c_j}}(\mathscr{M}_i), \overline{\mathscr{L}_{c_j}}(\mathscr{M}_i))$ with $k = (0.33, 0.37, 0.3)^T$ and $l = (0.42, 0.58)^T$ is the weight vectors of experts $s_i$ and parameters $c_j$. Consider $t = (0.36, 0.34, 0.3)^T$ and $v = (0.55, 0.45)^T$ as the associated weight vectors of experts $s_i$ and parameters $c_j$. Then, the aggregated result for $\mathscr{L}^*_{\delta c_j}(\mathscr{M}_i)$ is given as

$$\begin{aligned}
&\text{q}-\text{ROFS}_t\text{RHG}\left(\mathscr{L}_{c_1}(\mathscr{M}_1), \ldots, \mathscr{L}_{c_n}(\mathscr{M}_m)\right) = \left[\oplus_{j=1}^2 \left\{\oplus_{i=1}^3 \left(\underline{\mathscr{L}_{\delta c_j}}(\mathscr{M}_i)\right)^{t_i}\right\}^{v_j}, \oplus_{j=1}^2 \left\{\oplus_{i=1}^3 \left(\overline{\mathscr{L}_{\delta c_j}}(\mathscr{M}_i)\right)^{t_i}\right\}^{v_j}\right], \\
&\text{q}-\text{ROFS}_t\text{RHG}\left(\mathscr{L}_{c_1}(\mathscr{M}_1), \ldots, \mathscr{L}_{c_n}(\mathscr{M}_m)\right) = ((0.602629, 0.250898), (0.479995, 0.210973)).
\end{aligned} \tag{55}$$

From the analysis of Theorem 11, the properties of q-ROFS$_t$RHG operator are given below.

**Theorem 11.** *Let* $\mathscr{L}_{c_j}(\mathscr{M}_i) = (\underline{\mathscr{L}_{c_j}}(\mathscr{M}_i), \overline{\mathscr{L}_{c_j}}(\mathscr{M}_i))$ $(i = 1, 2, \ldots, m,\ j = 1, 2, \ldots, n)$ *be the collection of q-ROFS$_t$RNs. Let the weight vectors* $k = (k_1, k_2, \ldots, k_m)^T$ *and* $l = (l_1, l_2, \ldots, l_n)^T$ *of experts* $s_i$ *and parameters* $c_j$ *with* $\sum_{i=1}^m t_i = 1$, $\sum_{j=1}^n v_j = 1$, *and* $0 \le k_i, l_j \le 1$. *Consider the associated weight vectors* $t = (t_1, t_2, \ldots, t_m)^T$ *and* $v = (v_1, v_2, \ldots, v_n)^T$ *of experts* $s_i$ *and parameters* $c_j$ *with* $\sum_{i=1}^m t_i = 1$, $\sum_{j=1}^n v_j = 1$, *and* $0 \le t_i, v_j \le 1$, *respectively. Then, q-ROFS$_t$RHG operator holds as follows:*

*(i) (Idempotency): if* $\mathscr{L}^*_{\delta c_j}(\mathscr{M}_i) = \mathscr{P}_c(\mathscr{N})$ *(for all* $i = 1, 2, \ldots, m$ *and* $j = 1, 2, \ldots, n$*) where* $\mathscr{P}_c(\mathscr{M}) = (\underline{\mathscr{P}_c}(\mathscr{M}), \overline{\mathscr{P}_c}(\mathscr{M})) = ((\underline{b}, \underline{d}), (\overline{b}, \overline{d}))$, *then*

$$\text{q}-\text{ROFS}_t\text{RHG}\left(\mathscr{L}_{c_1}(\mathscr{M}_1), \mathscr{L}_{c_2}(\mathscr{M}_2), \ldots, \mathscr{L}_{c_n}(\mathscr{M}_m)\right) = \mathscr{P}_c(\mathscr{M}). \tag{56}$$

*(ii) (Boundedness): let* $(\underline{\mathscr{L}^*_{\delta c_j}}(\mathscr{M}_i))^- = (\min_j \min_i \underline{\mathscr{L}^*_{\delta c_j}}(\mathscr{M}_i), \max_j \max_i \overline{\mathscr{L}^*_{\delta c_j}}(\mathscr{M}_i))$ *and* $(\mathscr{L}^*_{\delta c_j}(\mathscr{M}_i))^+ = (\max_j \max_i \underline{\mathscr{L}^*_{\delta c_j}}(\mathscr{M}_i), \min_j \min_i \overline{\mathscr{L}^*_{\delta c_j}}(\mathscr{M}_i))$. *Then,*

$$\left(\mathscr{L}^*_{\delta c_j}(\mathscr{M}_i)\right)^- \le \text{q}-\text{ROFS}_t\text{RHG}\left(\mathscr{L}_{c_1}(\mathscr{M}_1), \mathscr{L}_{c_2}(\mathscr{M}_2), \ldots, \mathscr{L}_{c_n}(\mathscr{M}_m)\right) \le \left(\mathscr{L}^*_{\delta c_j}(\mathscr{M}_i)\right)^+. \tag{57}$$

*(iii) (Monotonicity): let* $\mathscr{P}_{c_j}(\mathscr{N}_i) = (\underline{\mathscr{P}_{c_j}}(\mathscr{N}_i), \overline{\mathscr{P}_{c_j}}(\mathscr{N}_i))$ $(i = 1, 2, \ldots, m,\ j = 1, 2, \ldots, n)$ *be another collection of q-ROFS$_t$RNs such that* $\underline{\mathscr{P}_{c_j}}(\mathscr{N}_i) \le \underline{\mathscr{L}_{c_j}}(\mathscr{M}_i)$ *and* $\overline{\mathscr{P}_{c_j}}(\mathscr{N}_i) \le \overline{\mathscr{L}_{c_j}}(\mathscr{M}_i)$. *Then,*

$$\begin{aligned}
&\text{q}-\text{ROFS}_t\text{RHG}\left(\mathscr{P}_{c_1}(\mathscr{M}_1), \mathscr{P}_{c_2}(\mathscr{M}_2), \ldots, \mathscr{P}_{c_n}(\mathscr{M}_m)\right) \\
&\quad \le \text{q}-\text{ROFS}_t\text{RHG}\left(\mathscr{L}_{c_1}(\mathscr{M}_1), \mathscr{L}_{c_2}(\mathscr{M}_2), \ldots, \mathscr{L}_{c_n}(\mathscr{M}_m)\right).
\end{aligned} \tag{58}$$

*(iv) (Shift invariance): let* $\mathscr{P}_c(\mathscr{N}) = (\underline{\mathscr{P}_c}(\mathscr{N}), \overline{\mathscr{P}_c}(\mathscr{N})) = ((\underline{b}, \underline{d}), (\overline{b}, \overline{d}))$ *be any other* $q-ROFS_tRN$. *Then,*

$$\begin{aligned}&\text{q}-\text{ROFS}_t\text{RHG}\big(\mathscr{L}_{c_1}(\mathscr{M}_1)\oplus\mathscr{P}_c(\mathscr{N}),\mathscr{L}_{c_2}(\mathscr{M}_2)\\&\quad\oplus\mathscr{P}_c(\mathscr{N}),\ldots,\mathscr{L}_{c_n}(\mathscr{M}_m)\oplus\mathscr{P}_c(\mathscr{N})\big)=\text{q}-\text{ROFS}_t\text{RHG}\\&\quad\big(\mathscr{L}_{c_1}(\mathscr{M}_1),\mathscr{L}_{c_2}(\mathscr{M}_2),\ldots,\mathscr{L}_{c_n}(\mathscr{M}_m)\big)\oplus\mathscr{P}_c(\mathscr{N}).\end{aligned}\tag{59}$$

*(v) (Homogeneity): for any real number $\lambda > 0$,*

$$\begin{aligned}&\text{q}-\text{ROFS}_t\text{RHG}\big(\lambda\mathscr{L}_{c_1}(\mathscr{M}_1),\lambda\mathscr{L}_{c_2}(\mathscr{M}_2),\ldots,\\&\quad\lambda\mathscr{L}_{c_n}(\mathscr{M}_m)\big)=\lambda\text{q}-\text{ROFS}_t\text{RHG}\big(\mathscr{L}_{c_1}(\mathscr{M}_1),\mathscr{L}_{c_2}\\&\quad(\mathscr{M}_2),\ldots,\mathscr{L}_{c_n}(\mathscr{M}_m)\big).\end{aligned}\tag{60}$$

*Remark 7*

(a) If $q = 1$, then q-ROFS$_t$RHG operator reduces to IFS$_t$RHG operator

(b) If the value of rung $q = 2$, then q-ROFS$_t$RHG operator reduces to PFS$_t$RHG operator

(c) If the soft parameter $c_1$ is one (means $n = 1$), then q-ROFS$_t$RHG operator reduces to q-ROFRHG operator

(d) If $kl = ((1/n), (1/n), \ldots, (1/n))^T$, then the proposed q-ROFS$_t$RHG operator reduces to q-ROFS$_t$RWG operator

(e) If $tv = ((1/n), (1/n), \ldots, (1/n))^T$, then the proposed q-ROFS$_t$RHG operator reduces to q-ROFS$_t$ROWG operator

## 6. MCDM Based on Soft Rough Aggregation Operator by Using q-ROF Information

MCDM has a high potential and discipline process to improve and evaluate multiple conflicting criteria in all areas of decision-making. In this competitive environment, an enterprise needs a more accurate and more repaid response to change customer needs. So, MCDM has the ability to handle successfully the evaluation process of multiple contradictory criteria. For an intelligent decision, experts analyze each and every character of an alternative and, then, they take the decision. Further, we will present the model for MCDM and their basic steps of construction by utilizing the proposed aggregation operators under q-ROF soft rough information.

Suppose that $Y = \{s_1, s_2, s_3, \ldots, s_p\}$ be the initial set of various alternatives and $\mathscr{E} = \{c_1, c_2, c_3, \ldots, c_n\}$ be the set of $n$ parameters. Consider $\mathscr{Z} = \{z_1, z_2, z_3, \ldots, z_m\}$ as the set of $m$ professional experts of this area who present their assessment expertise for each alternative $s_k$ $(k = 1, 2, \ldots, p)$ corresponding to $n$ parameters. Let the weight vectors $t = (t_1, t_2, \ldots, t_m)^T$ of experts $z_i$ and $v = (v_1, v_2, \ldots, v_n)^T$ of parameters $c_j$ with $\sum_{i=1}^{m} t_i = 1$, $\sum_{j=1}^{n} v_j = 1$, and $0 \le t_i,\ v_j \le 1$, respectively. The professional experts express their preference evaluation for alternative $s_k$ with respect to parameter $c_j$ in the form of q-ROFS$_t$RNs. The collective preference information given by the professionals is managed in q-ROFS$_t$R decision matrix, which is $\mathbb{M} = [\mathscr{L}_{c_j}(\mathscr{M}_i)]_{n\times m}$ where $\mathscr{M} \subseteq \mathscr{E}$. Further, using the proposed models aggregate, the preferred choices of experts are to get the aggregated results $\mathscr{R}_k$ $(k = 1, \ldots, p)$ for each alternative $s_k$ against their parameter $c_j$. Finally, they utilize the score function on the aggregated results $\mathscr{R}_k = [(\underline{\beta}, \underline{\vartheta}), (\overline{\beta}, \overline{\vartheta})]$ and rank all the results in a specific order to get the most desirable option. The stepwise decision algorithm for the investigated operators:

Step (i): the professional experts express their preference evaluation for alternative $s_k$ with respect to parameter $c_j$ in the form of q-ROFS$_t$RNs. Then, they collect the preference information given by the professionals and manage them in q-ROFS$_t$R decision matrix, which is $\mathbb{M} = [\mathscr{L}_{c_j}(\mathscr{M}_i)]_{n\times m}$ where $\mathscr{M} \subseteq \mathscr{E}$.

Step (ii): apply the presented aggregation operators of each decision matrix $\mathbb{M} = [\mathscr{L}_{c_j}(\mathscr{M}_i)]_{n\times m}$ for each alternative $s_k$ $(k = 1, 2, \ldots, p)$ against parameter $c_j$ to get the aggregated results $\mathscr{R}_k = [(\underline{\beta}, \underline{\vartheta}), (\overline{\beta}, \overline{\vartheta})]$.

Step (iii): calculate the score value of aggregated results $\mathscr{R}_k = [(\underline{\beta}, \underline{\vartheta}), (\overline{\beta}, \overline{\vartheta})]$ for each object $s_k$.

Step (iv): rank the score value of $\mathscr{R}_k$ in a specific order to get the optimum option of professional experts.

## 7. Numerical Example

In this section, we will initiate an illustrative example to prove the quality and excellency of the developed operators. Let the Higher Education Commission (HEC) in Pakistan plans to introduce a selection board of four high potential and professional professors $\mathscr{Z} = \{z_1, z_2, z_3, z_4\}$ from home and abroad to select the most desirable applicant. Out of many applicants, three applicants $Y = \{s_1, s_2, s_3\}$ were called for interviews. The interview mainly judges the applicants against some parameters $\mathscr{M} = \{c_1 = \text{academic level}, c_2 = \text{development potential}, c_3 = \text{professional ethics}, c_4 = \text{research productivity}\} \subseteq \mathscr{E}$. Let the weight vector $t = (0.3, 0.28, 0.24, 0.18)^T$ for professional experts $z_i$ $(i = 1, \ldots, 4)$ and $v = (0.32, 0.17, 0.31, 0.2)^T$ be the weight vectors for parameters $c_j$ $(j = 1, 2, 3)$, respectively. The professional experts express their preference evaluation for candidate $s_k$ with respect to parameter $c_j$ in the form of q-ROFS$_t$RNs. Finally, follow the following steps by utilizing the proposed models to select the most desirable and suitable applicant $s_k$.

### 7.1. Aggregation Results Rendered by the q-ROFS$_t$RWA Method

Step (i): the professional experts express their preference evaluation for alternative $s_k$ with respect to parameter $c_j$ in the form of q-ROFS$_t$RNs. Then, they collect the preference information given by the professionals and manage them in q-ROFS$_t$R decision matrix, which is $\mathbb{M} = [\mathscr{L}_{c_j}(\mathscr{M}_i)]_{n\times m}$ where $\mathscr{M} \subseteq \mathscr{E}$ which is given in Tables 7–9.

Step (ii): apply the presented q-ROFS$_t$RWA aggregation operators on each decision matrix

Table 7: q-ROFS$_t$R matrix for candidate $s_1$.

| | $c_1$ | $c_2$ | $c_3$ | $c_4$ |
|---|---|---|---|---|
| $z_1$ | [(0.7, 0.2), (0.8, 0.1)] | [(0.65, 0.25), (0.3, 0.6)] | [(0.82, 0.18), (0.6, 0.4)] | [(0.5, 0.2), (0.4, 0.1)] |
| $z_2$ | [(0.6, 0.1), (0.5, 0.3)] | [(0.5, 0.1), (0.7, 0.15)] | [(0.3, 0.2), (0.2, 0.7)] | [(0.2, 0.3), (0.6, 0.2)] |
| $z_3$ | [(0.4, 0.5), (0.6, 0.2)] | [(0.75, 0.2), (0.4, 0.1)] | [(0.65, 0.3), (0.7, 0.25)] | [(0.5, 0.4), (0.1, 0.5)] |
| $z_4$ | [(0.5, 0.3), (0.3, 0.7)] | [(0.6, 0.4), (0.9, 0.1)] | [(0.78, 0.22), (0.45, 0.4)] | [(0.8, 0.1), (0.3, 0.1)] |

Table 8: q-ROFS$_t$R matrix for candidate $s_2$.

| | $c_1$ | $c_2$ | $c_3$ | $c_4$ |
|---|---|---|---|---|
| $z_1$ | [(0.6, 0.3), (0.9, 0.1)] | [(0.2, 0.4), (0.6, 0.1)] | [(0.5, 0.2), (0.9, 0.1)] | [(0.6, 0.2), (0.7, 0.2)] |
| $z_2$ | [(0.4, 0.25), (0.3, 0.5)] | [(0.5, 0.15), (0.7, 0.3)] | [(0.77, 0.1), (0.6, 0.35)] | [(0.4, 0.3), (0.5, 0.1)] |
| $z_3$ | [(0.3, 0.6), (0.65, 0.25)] | [(0.66, 0.2), (0.8, 0.17)] | [(0.8, 0.15), (0.55, 0.2)] | [(0.7, 0.1), (0.3, 0.6)] |
| $z_4$ | [(0.5, 0.15), (0.55, 0.2)] | [(0.8, 0.1), (0.4, 0.5)] | [(0.62, 0.3), (0.9, 0.1)] | [(0.2, 0.4), (0.5, 0.3)] |

Table 9: q-ROFS$_t$R matrix for candidate $s_3$.

| | $c_1$ | $c_2$ | $c_3$ | $c_4$ |
|---|---|---|---|---|
| $z_1$ | [(0.8, 0.13), (0.8, 0.1)] | [(0.5, 0.2), (0.6, 0.1)] | [(0.4, 0.1), (0.7, 0.2)] | [(0.7, 0.1), (0.6, 0.3)] |
| $z_2$ | [(0.5, 0.16), (0.4, 0.2)] | [(0.8, 0.12), (0.3, 0.4)] | [(0.6, 0.2), (0.4, 0.3)] | [(0.5, 0.2), (0.4, 0.1)] |
| $z_3$ | [(0.4, 0.5), (0.7, 0.3)] | [(0.5, 0.4), (0.5, 0.12)] | [(0.2, 0.4), (0.8, 0.14)] | [(0.3, 0.4), (0.2, 0.5)] |
| $z_4$ | [(0.3, 0.2), (0.5, 0.15)] | [(0.6, 0.25), (0.7, 0.3)] | [(0.8, 0.18), (0.75, 0.1)] | [(0.6, 0.2), (0.8, 0.1)] |

$\mathbb{M} = [\mathscr{L}_{c_j}(\mathscr{M}_i)]_{n\times m}$ for each alternative $s_k$ $(k = 1, 2, 3)$ against parameter $c_j$ to get the aggregated results $\mathscr{R}_k = [(\underline{\beta}, \underline{\vartheta}), (\overline{\beta}, \overline{\vartheta})]$; that is,

$$\begin{aligned}\mathscr{R}_1 &= [(0.640506, 0.218382), (0.600249, 0.254431)],\\ \mathscr{R}_2 &= [(0.607809, 0.217229), (0.730645, 0.202415)],\\ \mathscr{R}_3 &= [(0.606198, 0.198551), (0.649417, 0.183108)].\end{aligned} \tag{61}$$

Step (iii): calculate the score value of aggregated results $\mathscr{R}_k = [(\underline{\beta}, \underline{\vartheta}), (\overline{\beta}, \overline{\vartheta})]$ for each object $s_k$; that is,

$$\begin{aligned}\mathrm{SC}(\mathscr{R}_1) &= 0.226074,\\ \mathrm{SC}(\mathscr{R}_2) &= 0.298025,\\ \mathrm{SC}(\mathscr{R}_3) &= 0.241341.\end{aligned} \tag{62}$$

Step (iv): rank the score value of $\mathscr{R}_k$ in a specific order to get the optimum option of professional experts; that is,

$$\mathrm{SC}(\mathscr{R}_2) > \mathrm{SC}(\mathscr{R}_3) > \mathrm{SC}(\mathscr{R}_1). \tag{63}$$

Therefore, from the above analysis, it is observed that $s_2$ is a more suitable and desirable candidate against the given position.

### 7.2. Aggregation Results Rendered by the q-ROFS$_t$RWG Method

Step (i): similar to above.

Step (ii): apply the presented q-ROFS$_t$RWG aggregation operators on each decision matrix $\mathbb{M} = [\mathscr{L}_{c_j}(\mathscr{M}_i)]_{n\times m}$ for each alternative $s_k$ $(k = 1, 2, 3)$ against parameter $c_j$ to get the aggregated results $\mathscr{R}_k = [(\underline{\beta}, \underline{\vartheta}), (\overline{\beta}, \overline{\vartheta})]$; that is,

$$\begin{aligned}\mathscr{R}_1 &= [(0.540892, 0.29117), (0.446263, 0.448329)],\\ \mathscr{R}_2 &= [(0.50166, 0.321735), (0.598613, 0.331553)],\\ \mathscr{R}_3 &= [(0.491017, 0.291113), (0.545527, 0.264822)].\end{aligned} \tag{64}$$

Step (iii): calculate the score value of the proposed q-ROFS$_t$RWG aggregated results $\mathscr{R}_k = [(\underline{\beta}, \underline{\vartheta}), (\overline{\beta}, \overline{\vartheta})]$ for each object $s_k$; that is,

$$\begin{aligned}\mathrm{SC}(\mathscr{R}_1) &= 0.06616,\\ \mathrm{SC}(\mathscr{R}_2) &= 0.135502,\\ \mathrm{SC}(\mathscr{R}_3) &= 0.118744.\end{aligned} \tag{65}$$

Step (iv): rank the score value of $\mathscr{R}_k$ in a specific order to get the optimum option of professional experts; that is,

$$\mathrm{SC}(\mathscr{R}_2) > \mathrm{SC}(\mathscr{R}_3) > \mathrm{SC}(\mathscr{R}_1). \tag{66}$$

Therefore, from the above analysis, it is observed that $s_2$ is a more suitable and desirable candidate against the given position.

### 7.3. Aggregation Results Rendered by the q-ROFS$_t$ROWA Method

Step (i): similar to above.

Step (ii):

$$\mathscr{R}_1 = [(0.657933, 0.217542), (0.621125, 0.234448)],$$
$$\mathscr{R}_2 = [(0.612743, 0.216413), (0.744755, 0.193923)],$$
$$\mathscr{R}_3 = [(0.621268, 0.203392), (0.677814, 0.180796)]. \quad (67)$$

Step (iii): $\mathrm{SC}(\mathscr{R}_1) = 0.250625$, $\mathrm{SC}(\mathscr{R}_2) = 0.312857$, $\mathrm{SC}(\mathscr{R}_3) = 0.26844$.

Step (iv): $\mathrm{SC}(\mathscr{R}_2) > \mathrm{SC}(\mathscr{R}_3) > \mathrm{SC}(\mathscr{R}_1)$.

Therefore, from the above analysis, it is observed that $s_2$ is a more suitable and desirable candidate against the given position.

*7.4. Aggregation Results Rendered by the q-ROFS$_t$ROWG Method*

Step (i): similar to above.

Step (ii):

$$\mathscr{R}_1 = [(0.562833, 0.293072), (0.473363, 0.418793)],$$
$$\mathscr{R}_2 = [(0.517595, 0.319441), (0.614971, 0.321818)],$$
$$\mathscr{R}_3 = [(0.49826, 0.296096), (0.583543, 0.259781)]. \quad (68)$$

Step (iii): $\mathrm{SC}(\mathscr{R}_1) = 0.09287, \mathrm{SC}(\mathscr{R}_2) = 0.152658$, $\mathrm{SC}(\mathscr{R}_3) = 0.139459$.

Step (iv): $\mathrm{SC}(\mathscr{R}_2) > \mathrm{SC}(\mathscr{R}_3) > \mathrm{SC}(\mathscr{R}_1) \Rightarrow$

Therefore, from the above analysis, it is observed that $s_2$ is a more suitable and desirable candidate against the given position.

*7.5. Aggregation Results Rendered by the q-ROFS$_t$RHA Method*

Step (i): similar to above.

Step (ii): apply the presented q-ROFS$_t$RHA aggregation operators on each decision matrix $\mathbb{M} = [\mathscr{L}^*_{\delta c_j}(\mathscr{M}_i)]_{n\times m}$ for each alternative $s_k$ $(k = 1, 2, 3)$ against parameter $c_j$ to get the aggregated results $\mathscr{R}_k = [(\underline{\beta^*_\delta}, \underline{\vartheta^*_\delta}), (\overline{\beta^*_\delta}, \overline{\vartheta^*_\delta})]$, with $k = (0.25, 0.29, 0.3, 0.16)^T$ and $l = (0.27, 0.23, 0.32, 0.18)^T$ being the weight vectors of experts $s_i$ and parameters $c_j$. Then, the aggregated result for $\mathscr{L}^*_{\delta c_j}(\mathscr{M}_i)$ is given as

$$\mathscr{R}_1 = [(0.432025, 0.669215), (0.408108, 0.691652)],$$
$$\mathscr{R}_2 = [(0.417457, 0.662671), (0.503394, 0.6509)],$$
$$\mathscr{R}_3 = [(0.392067, 0.655437), (0.442525, 0.635251)]. \quad (69)$$

Step (iii): $\mathrm{SC}(\mathscr{R}_1) = -0.24099, \mathrm{SC}(\mathscr{R}_2) = -0.18323$, $\mathrm{SC}(\mathscr{R}_3) = -0.1955$.

Step (iv): $\mathrm{SC}(\mathscr{R}_2) > \mathrm{SC}(\mathscr{R}_3) > \mathrm{SC}(\mathscr{R}_1)$.

Therefore, from the above analysis, it is observed that $s_2$ is a more suitable and desirable candidate against the given position.

*7.6. Aggregation Results Rendered by the q-ROFS$_t$RHG Method*

Step (i): similar to above.

Step (ii): apply the presented q-ROFS$_t$RHG aggregation operators on each decision matrix $\mathbb{M} = [\mathscr{L}^*_{\delta c_j}(\mathscr{M}_i)]_{n\times m}$ for each alternative $s_k$ $(k = 1, 2, 3)$ against parameter $c_j$ to get the aggregated results $\mathscr{R}_k = [(\underline{\beta^*_\delta}, \underline{\vartheta^*_\delta}), (\overline{\beta^*_\delta}, \overline{\vartheta^*_\delta})]$, with $k = (0.25, 0.29, 0.3, 0.16)^T$ and $l = (0.27, 0.23, 0.32, 0.18)^T$ being the weight vectors of experts $s_i$ and parameters $c_j$. Then, the aggregated result for $\mathscr{L}^*_{\delta c_j}(\mathscr{M}_i)$ is given as

$$\mathscr{R}_1 = [(0.365745, 0.709626), (0.308693, 0.751049)],$$
$$\mathscr{R}_2 = [(0.332518, 0.713922), (0.407739, 0.706165)], \quad (70)$$
$$\mathscr{R}_3 = [(0.318639, 0.693983), (0.36813, 0.673448)].$$

Step (iii): $\mathrm{SC}(\mathscr{R}_1) = -0.35133, \mathrm{SC}(\mathscr{R}_2) = -0.30573, \mathrm{SC}(\mathscr{R}_3) = -0.27871$.

Step (iv): $\mathrm{SC}(\mathscr{R}_3) > \mathrm{SC}(\mathscr{R}_2) > \mathrm{SC}(\mathscr{R}_1)$.

Therefore, from the above analysis, it is observed that $s_3$ is a more suitable and desirable candidate against the given position. The higher the score value, the more optimist that value and the smaller the score value, the more pessimist that value. From the ranking results of the above proposed operators, it is clear that averaging operators' results are more optimist than geometric operators.

*7.7. Comparative Study.* To present the applicability and efficiency of the developed approach with some other existing methods based on IFS, PFS, and q-ROFS methods by using the same illustrative example. A comparative study has been made based on different aggregation operators (see [4, 5, 23, 27, 28, 54, 60, 61]). For this purpose, different parameters of the above numerical example are aggregated by utilizing the proposed aggregation operators having weight vectors $v = (0.32, 0.17, 0.31, 0.2)^T$, and their collective aggregated decision matrix for each candidate $s_i$ $(i = 1, 2, 3)$ is given in Table 10. Now, by using the information of the evaluated matrix, a comparative study of the investigated models with some existing aggregation operators is presented in Table 11. From Table 11, it is observed that the methods presented in [4, 5, 23, 27, 28, 54, 60, 61] are only capable of solving the q-ROFV of the form $(\beta, \vartheta)$ but are not capable of solving the q-ROFS$_t$RV of the form $[(\underline{\beta}, \underline{\vartheta}), (\overline{\beta}, \overline{\vartheta})]$. Thus, from the existing methods, it is clear that these existing methods have a lake of rough information and they are not capable of solving and ranking the given illustrative example. Therefore, from this

Table 10: Aggregated matrix for candidate.

| | $s_1$ | $s_2$ | $s_3$ |
|---|---|---|---|
| $z_1$ | [(0.719, 0.201), (0.651, 0.208)] | [(0.538, 0.256), (0.849, 0.115)] | [(0.670, 0.122), (0.713, 0.154)] |
| $z_2$ | [(0.471, 0.154), (0.534, 0.320)] | [(0.602, 0.179), (0.550, 0.298)] | [(0.616, 0.171), (0.387, 0.222)] |
| $z_3$ | [(0.598, 0.350), (0.582, 0.229)] | [(0.678, 0.226), (0.629, 0.260)] | [(0.370, 0.430), (0.682, 0.225)] |
| $z_4$ | [(0.703, 0.230), (0.616, 0.287)] | [(0.605, 0.211), (0.733, 0.205)] | [(0.650, 0.201), (0.703, 0.137)] |

Table 11: Comparative analysis of different methods.

| Methods | Score values $\mathscr{R}_1$ $\mathscr{R}_2$ $\mathscr{R}_3$ | Ranking |
|---|---|---|
| IFWA [4] | Unapproachable | × |
| IFS$_t$WA [54] | Unapproachable | × |
| PFS$_t$WA [60] | Unapproachable | × |
| q-ROFWA [23] | Unapproachable | × |
| q-ROFHOWAGA [61] | Unapproachable | × |
| q-ROFS$_t$WA [60] | Unapproachable | × |
| Cq-ROFWA [27] | Unapproachable | × |
| q-RONFWA [28] | Unapproachable | × |
| q-ROFS$_t$RWA (proposed) 0.226074, 0.298025, 0.241341 | | $\mathscr{R}_2 \succcurlyeq \mathscr{R}_3 \succcurlyeq \mathscr{R}_1$ |
| q-ROFS$_t$ROWA (proposed) 0.250625, 0.312857, 0.26844 | | $\mathscr{R}_2 \succcurlyeq \mathscr{R}_3 \succcurlyeq \mathscr{R}_1$ |
| q-ROFS$_t$RHA (proposed) | −0.24099, −0.18323, −0.1955 | $\mathscr{R}_2 \succcurlyeq \mathscr{R}_3 \succcurlyeq \mathscr{R}_1$ |
| IFWG [5] | Unapproachable | × |
| IFS$_t$WG [54] | Unapproachable | × |
| q-ROFWG [23] | Unapproachable | × |
| Cq-ROFWG [27] | Unapproachable | × |
| q-RONFWG [28] | Unapproachable | × |
| q-ROFS$_t$RWG (proposed) 0.06616, 0.135502, 0.118744 | | $\mathscr{R}_2 \succcurlyeq \mathscr{R}_3 \succcurlyeq \mathscr{R}_1$ |
| q-ROFS$_t$ROWG (proposed) 0.09287, 0.152658, 0.139459 | | $\mathscr{R}_2 \succcurlyeq \mathscr{R}_3 \succcurlyeq \mathscr{R}_1$ |
| q-ROFS$_t$RHG (proposed) | −0.35133, −0.30573, −0.27871 | $\mathscr{R}_3 \succcurlyeq \mathscr{R}_2 \succcurlyeq \mathscr{R}_1$ |

analysis, it is clear that the developed methods are more superior and capable than the existing methods.

## 8. Conclusion

MCDM has a high potential and discipline process to improve and evaluate multiple conflicting criteria in all areas of decision-making. For an intelligent decision, the experts analyze each and every character of an alternative and then they take the decision. For an intelligent and successful decision, the experts require a careful preparation and analysis of each and every character for an alternative and then they can take a good decision if they are armed with all the data and information that they need. The dominant notions of fuzzy sets, $S_t$Ss, and rough sets generalized the classical set theory to cope with uncertain information. Molodtsov investigated the pioneer notion of $S_t$S which is free from the inherent complexity which the contemporary theories faced. It is observed that $S_t$S has too close relation with fuzzy sets and rough sets. The $S_t$S theory is an effective mathematical tool for handling the uncertain, ambiguous, and imprecise data. The aim of our work is to investigate the hybrid concept of $S_t$S and rough set with the notion of q-ROFS to obtain the new notion of q-ROFS$_t$RS. In addition, some averaging aggregation operators such as q-ROFS$_t$RWA, q-ROFS$_t$ROWA, and q-ROFS$_t$RHA are presented. Then, important properties of these investigated averaging operators are given in detail. Moreover, we investigated the geometric aggregation operators such as q-ROFS$_t$RWG, q-ROFS$_t$ROWG, and q-ROFS$_t$RHG and proposed the basic desirable characteristics of investigated geometric operators. The technique for MCDM and stepwise algorithm for decision-making by utilizing the proposed approaches are demonstrated clearly. Finally, a numerical example for the developed approach is presented and a comparative study of the investigated models with some existing methods is brought to light in detail which shows that the proposed models are more effective and superior than existing approaches.

*8.1. Future Work.* In the future, we intend to further discuss the following topics:

(i) The investigation of q-ROF-entropy of soft rough sets

(ii) The investigation of the picture and spherical fuzzy information by using soft rough sets

(iii) Applying other decision-making methodology based q-ROFS$_t$RS to solve the MCDM problem

(iv) The discussions of other applied methods in information systems

## Abbreviations

| | |
|---|---|
| IFS: | Intuitionistic fuzzy set |
| PFS: | Pythagorean fuzzy set |
| q-ROFS: | q-Rung orthopair fuzzy set |
| $S_t$S: | Soft set |
| q-ROFS$_t$RS: | q-ROF soft rough set |
| q-ROFS$_t$RWA: | q-ROF soft rough weighted averaging |
| q-ROFS$_t$ROWA: | q-ROF soft rough ordered weighted averaging |
| q-ROFS$_t$RHA: | q-ROF soft rough hybrid averaging |
| q-ROFS$_t$RWG: | q-ROF soft rough weighted geometric |
| q-ROFS$_t$ROWG: | q-ROF soft rough ordered weighted geometric |
| q-ROFS$_t$RHG: | q-ROF soft rough hybrid geometric |
| MCDM: | Multicriteria decision-making |
| MemG: | Membership grade |
| NMemG: | Nonmembership grade |

IFWA: IF weighted averaging
IFOWA: IF ordered weighted averaging
IFHA: IF hybrid averaging
IFWG: IF weighted geometric
IFOWG: IF ordered weighted geometric
IFHG: IF hybrid geometric
q-ROFWA: q-ROF weighted averaging
q-ROFWG: q-ROF weighted geometric
IF$S_t$S: IF soft set
IF$S_t$WA: IF soft weighted averaging
IF$S_t$WG: IF soft weighted geometric
q-ROF$S_t$S: q-Rung orthopair fuzzy soft set
q-ROF$S_t$WA: q-ROF soft weighted averaging
q-ROF$S_t$OWA: q-ROF soft ordered weighted averaging
q-ROF$S_t$HA: q-ROF soft hybrid averaging
q-ROF$S_t$N: q-ROF soft number
q-ROF$S_t$RN: q-ROF soft rough number.

## Data Availability

The data used in this article are artificial and hypothetical, and anyone can use these data before prior permission by just citing this article.

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## Acknowledgments

This work was supported by Major Humanities and Social Sciences Research Projects in Zhejiang Universities (no. 2018QN058).

## References

[1] L. A. Zadeh, "Fuzzy sets," *Information and Control*, vol. 8, no. 3, pp. 338–353, 1965.
[2] K. T. Atanassov, "Intuitionistic fuzzy sets," *Fuzzy Sets Systems*, vol. 20, no. 1, pp. 87–96, 1986.
[3] M. I. Ali, F. Feng, T. Mahmood, I. Mahmood, and H. Faizan, "A graphical method for ranking Atanassov's intuitionistic fuzzy values using the uncertainty index and entropy," *International Journal of Intelligent Systems*, vol. 34, no. 10, pp. 2692–2712, 2019.
[4] Z. S. Xu, "Intuitionistic fuzzy aggregation operators," *IEEE Transection and Fuzzy Systems*, vol. 15, pp. 1179–1187, 2007.
[5] Z. Xu and R. R. Yager, "Some geometric aggregation operators based on intuitionistic fuzzy sets," *International Journal of General Systems*, vol. 35, no. 4, pp. 417–433, 2006.
[6] H. Zhao, Z. Xu, M. Ni, and S. Liu, "Generalized aggregation operators for intuitionistic fuzzy sets," *International Journal of Intelligent Systems*, vol. 25, no. 1, pp. 1–30, 2010.
[7] W. Wang and X. Liu, "Intuitionistic fuzzy geometric aggregation operators based on Einstein operations," *International Journal of Intelligent Systems*, vol. 26, no. 11, pp. 1049–1075, 2011.
[8] W. Wang and X. Liu, "Intuitionistic fuzzy information aggregation using Einstein operations," *IEEE Transactions on Fuzzy Systems*, vol. 20, no. 5, pp. 923–938, 2012.
[9] S. Zeng, S.-M. Chen, and L.-W. Kuo, "Multiattribute decision making based on novel score function of intuitionistic fuzzy values and modified VIKOR method," *Information Sciences*, vol. 488, pp. 76–92, 2019.
[10] R. R. Yager, "Pythagorean fuzzy subsets," in *Proceedings of the Annual Meeting of Joint IFSA World Congress and NAFIPS*, vol. 24, no. 28, pp. 57–61, Edmonton, Canada, 2013.
[11] R. R. Yager, "Pythagorean membership grades in multicriteria decision making," *IEEE Transactions on Fuzzy Systems*, vol. 22, no. 4, pp. 958–965, 2014.
[12] X. Peng and Y. Yang, "Some results for Pythagorean fuzzy sets," *International Journal of Intelligent Systems*, vol. 30, no. 11, pp. 1133–1160, 2015.
[13] X. Peng and Y. Yang, "Pythagorean fuzzy choquet integral based MABAC method for multiple attribute group decision making," *International Journal of Intelligent Systems*, vol. 31, no. 10, pp. 989–1020, 2016.
[14] H. Garg, "A new generalized Pythagorean fuzzy information aggregation using Einstein operations and its application to decision making," *International Journal of Intelligent Systems*, vol. 31, no. 9, pp. 886–920, 2016.
[15] H. Garg, "Generalized Pythagorean fuzzy geometric aggregation operators using Einstein *t*-norm and *t*-conorm for multicriteria decision-making process," *International Journal of Intelligent Systems*, vol. 32, no. 6, pp. 597–630, 2017.
[16] H. Garg, "Confidence levels based Pythagorean fuzzy aggregation operators and its application to decision-making process," *Computational and Mathematical Organization Theory*, vol. 23, no. 4, pp. 546–571, 2017.
[17] G. Wei and M. Lu, "Pythagorean fuzzy power aggregation operators in multiple attribute decision making," *International Journal of Intelligent Systems*, vol. 33, no. 1, pp. 169–186, 2018.
[18] G. Wei, "Pythagorean fuzzy interaction aggregation operators and their application to multiple attribute decision making," *Journal of Intelligent & Fuzzy Systems*, vol. 33, no. 4, pp. 2119–2132, 2017.
[19] S.-J. Wu and G.-W. Wei, "Pythagorean fuzzy Hamacher aggregation operators and their application to multiple attribute decision making," *International Journal of Knowledge-Based and Intelligent Engineering Systems*, vol. 21, no. 3, pp. 189–201, 2017.
[20] R. R. Yager, "Generalized orthopair fuzzy sets," *IEEE Transactions on Fuzzy Systems*, vol. 25, no. 5, pp. 1222–1230, 2017.
[21] R. R. Yager and N. Alajlan, "Approximate reasoning with generalized orthopair fuzzy sets," *Information Fusion*, vol. 38, pp. 65–73, 2017.
[22] M. I. Ali, "Another view on *q*-rung orthopair fuzzy sets," *International Journal of Intelligent Systems*, vol. 33, pp. 2139–2153, 2019.
[23] P. Liu and P. Wang, "Some q-rung orthopair fuzzy aggregation operators and their applications to multiple-attribute decision making," *International Journal of Intelligent Systems*, vol. 33, no. 2, pp. 259–280, 2018.
[24] P. Liu and J. Liu, "Some q-rung orthopai fuzzy Bonferroni mean operators and their application to multi-attribute group decision making," *International Journal of Intelligent Systems*, vol. 33, no. 2, pp. 315–347, 2018.
[25] C. Jana, G. Muhiuddin, and M. Pal, "Some Dombi aggregation of q-rung orthopair fuzzy numbers in multiple-attribute decision making," *International Journal of Intelligent Systems*, vol. 34, no. 12, pp. 3220–3240, 2019.

[26] J. Wang, R. Zhang, X. Zhu, Z. Zhou, X. Shang, and W. Li, "Some q-rung orthopair fuzzy Muirhead means with their application to multi-attribute group decision making," *Journal of Intelligent & Fuzzy Systems*, vol. 36, no. 2, pp. 1599–1614, 2019.

[27] B. P. Joshi and A. Gegov, "Confidence levels q-rung orthopair fuzzy aggregation operators and its applications to MCDM problems," *International Journal of Intelligent Systems*, vol. 35, no. 1, pp. 125–149, 2020.

[28] Z. Yang, X. Li, Z. Cao, and J. Li, "q-rung orthopair normal fuzzy aggregation operators and their application in multi-attribute decision-making," *Mathematics*, vol. 7, no. 12, p. 1142, 2019.

[29] A. Hussain, M. I. Ali, and T. Mahmood, "Hesitant q-rung orthopair fuzzy aggregation operators with their applications in multi-criteria decision making," *Iranian Journal of Fuzzy Systems*, vol. 17, no. 3, pp. 117–134, 2020.

[30] A. Hussain, M. I. Ali, T. Mahmood, and M. Munir, "Group-based generalized q-rung orthopair average aggregation operators and their applications in multi-criteria decision making," *Complex and Intelligent Systems*, 2020.

[31] Z. a. Pawlak, "Rough sets," *International Journal of Computer & Information Sciences*, vol. 11, no. 5, pp. 341–356, 1982.

[32] M. I. Ali, B. Davvaz, and M. Shabir, "Some properties of generalized rough sets," *Information Sciences*, vol. 224, pp. 170–179, 2013.

[33] D. Chen, C. Wang, and Q. Hu, "A new approach to attribute reduction of consistent and inconsistent covering decision systems with covering rough sets," *Information Sciences*, vol. 177, no. 17, pp. 3500–3518, 2007.

[34] T.-J. Li, Y. Leung, and W.-X. Zhang, "Generalized fuzzy rough approximation operators based on fuzzy coverings," *International Journal of Approximate Reasoning*, vol. 48, no. 3, pp. 836–856, 2008.

[35] M. I. Ali, T. Mahmood, and A. Hussain, "A study of generalized roughness in-fuzzy filters of ordered semigroups," *Journal of Taibah University for Science*, vol. 12, no. 2, pp. 163–172, 2018.

[36] T. Mahmood, M. I. Ali, and A. Hussain, "Generalized roughness in fuzzy filters and fuzzy ideals with thresholds in ordered semigroups," *Computational and Applied Mathematics*, vol. 37, no. 4, pp. 5013–5033, 2018.

[37] A. Hussain, M. I. Ali, and T. Mahmood, "Generalized roughness of q-fuzzy ideals in ordered semigroups," *Journal of New Theory*, vol. 26, pp. 32–53, 2019.

[38] W. Zhu, "Generalized rough sets based on relations," *Information Sciences*, vol. 177, no. 22, pp. 4997–5011, 2007.

[39] K. Chakrabarty, T. Gedeon, and L. Koczy, "Intuitionistic fuzzy rough set," in *Proceedings of 4th Joint Conference on Information Sciences (JCIS)*, pp. 211–214, Durham, NC, USA, 1998.

[40] B. Huang, D.-k. Wei, H.-x. Li, and Y.-l. Zhuang, "Using a rough set model to extract rules in dominance-based interval-valued intuitionistic fuzzy information systems," *Information Sciences*, vol. 221, pp. 215–229, 2013.

[41] S. P. Jena, S. K. Ghosh, and B. K. Tripathy, "Intuitionistic fuzzy rough sets," *Notes Intuitionistic Fuzzy Sets*, vol. 8, no. 1, pp. 1–18, 2002.

[42] S. K. Samanta and T. K. Mondal, "Intuitionistic fuzzy rough sets and rough intuitionistic fuzzy sets," *Journal of Fuzzy Mathematics*, vol. 9, no. 3, pp. 561–582, 2001.

[43] L. Zhou, W.-Z. Wu, and W.-X. Zhang, "On characterization of intuitionistic fuzzy rough sets based on intuitionistic fuzzy implicators," *Information Sciences*, vol. 179, no. 7, pp. 883–898, 2009.

[44] L. Zhou and W.-Z. Wu, "Characterization of rough set approximations in Atanassov intuitionistic fuzzy set theory," *Computers & Mathematics with Applications*, vol. 62, no. 1, pp. 282–296, 2011.

[45] L. Zhou and W. Z. Wu, "On generalized intuitionistic fuzzy rough approximation operators," *Information Sciences*, vol. 178, no. 11, pp. 2448–2465, 2008.

[46] A. Hussain, T. Mahmood, and M. I. Ali, "Rough Pythagorean fuzzy ideals in semigroups," *Computational and Applied Mathenatics*, vol. 38, no. 2, p. 67, 2019.

[47] S. Zeng, "Pythagorean fuzzy multiattribute group decision making with probabilistic information and OWA approach," *International Journal of Intelligent Systems*, vol. 32, no. 11, pp. 1136–1150, 2017.

[48] A. Hussain, M. Irfan Ali, and T. Mahmood, "Covering based q-rung orthopair fuzzy rough set model hybrid with TOPSIS for multi-attribute decision making," *Journal of Intelligent & Fuzzy Systems*, vol. 37, no. 1, pp. 981–993, 2019.

[49] D. Molodtsov, "Soft set theory-first results," *Computers & Mathematics with Applications*, vol. 37, no. 4-5, pp. 19–31, 1999.

[50] P. K. Maji, R. Biswas, and A. R. Roy, "Fuzzy soft sets," *Journal of Fuzzy Mathematics*, vol. 9, pp. 589–602, 2001.

[51] P. K. Maji, R. Biswas, and A. R. Roy, "Intuitionistic fuzzy soft sets," *Journal of Fuzzy Mathematics*, vol. 9, pp. 677–692, 2001.

[52] M. I. Ali, F. Feng, X. Liu, W. K. Min, and M. Shabir, "On some new operations in soft set theory," *Computers & Mathematics with Applications*, vol. 57, no. 9, pp. 1547–1553, 2009.

[53] M. Agarwal, K. K. Biswas, and M. Hanmandlu, "Generalized intuitionistic fuzzy soft sets with applications in decision-making," *Applied Soft Computing*, vol. 13, no. 8, pp. 3552–3566, 2013.

[54] R. Arora and H. Garg, "A robust aggregation operator for multi-criteria decision-making with intuitionistic fuzzy soft set environment," *Scientia Iranica*, vol. 25, no. 2, pp. 913–942, 2018.

[55] H. Garg and R. Arora, "Generalized intuitionistic fuzzy soft power aggregation operator based on *t*-norm and their application in multicriteria decision-making," *International Journal of Intelligent Systems*, vol. 34, no. 2, pp. 215–246, 2019.

[56] R. Arora, "Intuitionistic fuzzy soft aggregation operator based on Einstein norms and its applications in decision-making," in *Proceedings of the 2018 International Conference on Intelligent Systems Design and Applications*, Vellore, India, December 2018.

[57] F. Feng, H. Fujita, M. I. Ali, R. R. Yager, and X. Liu, "Another view on generalized intuitionistic fuzzy soft sets and related multi attribute decision making methods," *IEEE Transactions on Fuzzy Systems*, vol. 27, pp. 474–488, 2018.

[58] A. Hussain, M. I. Ali, and T. Mahmood, "Pythagorean fuzzy soft rough sets and their applications in decision-making," *Journal of Taibah University for Science*, vol. 14, no. 1, pp. 101–113, 2020.

[59] M. Riaz and M. R. Hashmi, "Soft rough Pythagorean m-polar fuzzy sets and Pythagorean m-polar fuzzy soft rough sets with application to decision-making," *Computational and Applied Mathematics*, vol. 39, no. 1, p. 16, 2020.

[60] A. Hussain, M. I. Ali, T. Mahmood, and M. Munir, "q-rung orthopair fuzzy soft average aggregation operators and their application in multicriteria decision-making," *International Journal of Intelligent Systems*, vol. 35, no. 4, pp. 571–599, 2020.

[61] M. Riaz, H. M. A. Farid, F. Karaaslan, and M. R. Hashmi, "Some q-rung orthopair fuzzy hybrid aggregation operators and TOPSIS method for multi-attribute decision-making," *Journal of Intelligent and Fuzzy Systems*, vol. 32, no. 2, pp. 259–280, 2020.

Hindawi
Mathematical Problems in Engineering
Volume 2020, Article ID 8859458, 8 pages
https://doi.org/10.1155/2020/8859458

*Research Article*

# Research on the Internal Financing Mechanism in the Innovation Chain

**Huihong Liu, Han Ding, and Xinmiao Zhou**

*Business School, Ningbo University, Ningbo 131521, Zhejiang, China*

Correspondence should be addressed to Xinmiao Zhou; zhouxinmiao@nbu.edu.cn

Received 15 September 2020; Revised 19 November 2020; Accepted 23 November 2020; Published 7 December 2020

Academic Editor: Chonghui Zhang



It is of great significance to build an efficient and cooperative innovation system and support system under the background of China persisting in leading the development with innovation and speeding up the construction of an innovation-oriented country. The collaborative relationship and profit distribution of each innovation subject will be analyzed through the study of the innovation activity in the innovation chain. With consideration of the characteristics of the innovation chain such as high risk, financing difficulty, cooperative partnership and so on and using the internal financing mode of supply chain for reference, the game model of internal financing between members of the innovation chain will be established. The game equilibriums will be solved and analyzed, and the optimal equilibriums will be achieved by designing the contracts to establish the internal financing mechanisms in the innovative chain. For the risk asymmetry at each stage of the innovation chain, the contracts will be optimized to construct the risk sharing mechanism to control the risk of the internal financing in the innovation chain. In order to make innovations, this paper creates the financial support mode of the innovation chain, improves the innovation support system and increases the efficiency of the innovation chain.

## 1. Introduction

China has entered an era of New Economic Normality, one characteristic of which is the turn from a factors-driven and investment-driven economy to an innovation-driven one. To achieve this goal, China continues to invest heavily in innovation. In 2018, R&D spending as a share of GDP was 2.18%, higher than the average level of OECD countries. However, such investment has not led to a rapid increase of TFP (total factor productivity), which, by contrast, is only half that of OECD countries. This shows that the innovation efficiency of China is still relatively low. The main reason is the lack of synergy of the innovation system, such as the disconnection between the research and development stage and the production and commercialization stage of the technological innovation, which renders many technological innovations really unable to meet the market demand and thus unable to be transformed. In addition, the support system for innovation is not perfect, such as the financial support system. Although the investment of R&D is increased, the investment is unevenly distributed among different industries, different regions and different subjects, and a large number of firms are still facing the constraints of innovation funds. To establish and improve the innovation system and support system is the most important task in the current scientific and technological innovation.

The innovation chain is an important means to realize the innovation-driven economy. The famous economist, Schumpeter, who put forward the concept of the "innovation," thought that in addition to the breakthrough in science and technology and the production of new invention, the innovation should also include their introduction into the firms, so as to improve the productivity of the firms and society [1]. That is, innovation refers not only to the R&D of the new technology but also to the whole process of converting it into the real productivity and the diffusion. It is a complex system process from the source of innovation to the successful transformation of technological achievements. It is difficult for an individual innovation subject to complete the whole process. The innovation chain, an efficient and

collaborative innovation system, needs to be constructed by innovation agents in different fields [2]. According to the functions in the innovation chain, the innovation chain can be divided into relatively independent three or five stages [3, 4]. There is no recognized definition of the innovation chain, but it is generally believed that the innovation chain will produce and market the technological innovation results in order to meet the needs of customers and make the technological innovation diffuse among the main members of innovation activities. Innovation subjects can improve innovation by improving the ability of cluster collaborative innovation [5]. This collaborative development relationship has been empirically proved [6–8].

The innovative activities in the innovation chain require financial support. Innovation is a high risk activity with a great risk of loss that can be prohibitive for most of the actors. In addition, the innovation activity needs a lot of financial support, and more capital is needed to transform the scientific research results into the real productivity and then to become a high-tech industry, so a good financial support is the necessary guarantee of the innovation chain. How to deploy the capital chain around the innovation chain, explore the new mechanism of the docking between science and technology and finance, and promote the capitalization and industrialization of scientific and technological achievements are of great value [9]. Scholars design mechanisms with the help of investment loans, project financing, securities issuance and other tools to finance innovative projects and support the operation of innovative science and technology business incubators [10] and build a relational loan model which is closely related to banks to solve the financing problems of innovation subjects in the innovation chain [11]. The innovation chain is often faced with a shortage of innovation fund, and different financial markets, different financial instruments, and different financial arrangements have heterogeneous promotion and support effect on each main body at different stages of the innovation industrialization [12] [13]; therefore, it is necessary to reasonably allocate the innovative funds. To optimize the allocation, the scholars allocate the financial resources in the innovation network according to the resource tree and the technology roadmap [14, 15], and put forward the innovation fund allocation mode in the innovation chain based on the financial service demand of the innovation-driven development and from the perspective of coordination and integration of the innovation chain, the industrial chain, the capital chain and the profit chain [16].

The internal financing mode between members of the innovation chain is a new financial support model to promote an innovation-driven economy by innovation chain. The innovation chain connects the subjects of innovation, forms a certain profit community through the partnership, and better realizes the process of knowledge economy and the optimization goal of innovation system. The internal financing mode between members of the innovation chain can perfect the resource allocation mode in accordance with the law of innovation and provide a new financial support mode for innovation activities. On the one hand, the joint fusion of the innovation chain and the capital chain is realized, which makes the capital chain be used for boosting the innovation chain and the innovation motivation be activated to realize the financing for innovation-driven economy by the innovation chain. On the other hand, the finance-assisting relationship in the innovation chain is established, which strengthens the leading role of the core innovation subject and the relationship between innovative subjects, improves the coordination and cooperation mechanism to strive to maximize the efficiency of scientific and technological innovation activities.

Although there is a lack of research on the internal financing of the innovation chain, the development of supply chain finance is mature, which provides a reference for the resource reconfiguration and cost optimization in the chain ecosystem [12, 17]. Considering the characteristics of innovation activities in the innovation chain, we mainly draw lessons from the downstream firms of the supply chain to provide internal financing for suppliers. When the supplier has a fund constraint, one of the main financing modes provided by the downstream firm to the supplier is the prepayment, the most important of which is to determine the payment time, the price discount, the price dependency, the economic order quantity, etc. [18–20], while the selling price, the cost of purchase, the uncertainty of the product, and the type of contract are also the influential factors of the financing [21–23]. The supplier's own funds affect internal financing decisions [24]. The core firms give full play to their coordination and control advantages in the supply chain and use the spare funds of upstream and downstream firms to provide financing for small and medium-sized firms in the chain [25, 26].

The present research is mainly focused on the financing model of a single innovation subject, and the research on the financing mode of the whole innovation chain is, however, inadequate. In fact, the profit of each innovative subject in the innovation chain is correlated, which makes the innovation chain to be a partnership and a benefit community, so the internal financing model of the innovation chain can be set up completely to relieve the financing constraint problem of the innovation chain. Although the internal financing model of the chain provides a reference for internal financing in the chain ecosystem, it is necessary to design the internal financing model of the innovation chain because of the special relationship between innovation subjects and the special operating mechanism of innovation subjects.

The objective of this paper is to contribute to the literature on innovation chains in two ways. First, it is among the first to focus on the quantitative research of the innovation chain. This paper is to construct models of innovation chain to study innovation activity and income distribution by game theory based on the analysis of composition and operation mechanism of innovation chain, solve game equilibrium and carry on optimization analysis to design internal financing mode of innovation chain. That provides quantitative research results for innovation chain research. Second, it provides the theoretical basis and operation guide for the financial support of the innovation chain. This paper is to design the internal financing model between members of the innovation chain and the risk control mechanism

based on the relationship and the profit distribution among the innovative subjects in the innovation chain, so as to reduce the financial constraints of innovation subjects and improve the innovation efficiency of the innovation chain.

The remainder of this paper is organized as follows. In Section 2, the operation mode and the profit distribution of the innovation chain are discussed. In Section 3, the internal financing in the innovation chain and the profit distribution mechanism are studied. In Section 4, the risk control mechanism of internal financing in the innovation chain is designed. Section 5 concludes the paper.

## 2. Operation Mode and Profit Distribution of Innovation Chain

In order to design the internal financing mechanism of the innovation chain, it is necessary to clarify the relationship between the innovative subjects in the innovation chain. In this section, we sort out the innovation process, analyze the relationships of innovation subjects at each stage, build the innovation chain to carry out the collaborative innovation, and make clear the profit distribution among the innovative subjects.

### *2.1. Innovative Activities and Profits in Noninnovation Chain*

*2.1.1. Assumptions of Innovative Activities.* The innovation process is divided into two stages: research and development; production and commercialization. There is only one the R&D institution at the R&D stage, and there are $n$ firms at the production stage, and $n$ is exogenous and fixed. At the first stage, the R&D institution has its own fund $C$ and invests $K$ in R&D, and innovation result is a function of R&D investment, $X = \sqrt{2K}$.The innovation result is bid in competition by $n$ firms, and firm $i$ bids $P_i$ for the innovation result, and it is only transferred to the highest bid firm $k$, $P_k = \max_i P_i$. At the second stage, the original marginal production cost of firm $i$ is $z_i$. If firm $k$ obtains the innovation result, its marginal cost becomes $z_k - X$; if there is gradual innovation in the innovation chain, that is to say $X(K) < \min_i z_i$. Given marginal cost condition, $n$ firms compete through quantities, which indicates Cournot competitiveness.

*2.1.2. Innovation Activities and Profit of Innovation Subjects.* For the dynamic problems of innovation process, the reverse solution analysis is adopted. At the second stage, after observing the marginal cost and the transfer result, firm $i$ chooses the quantity $q_i$ to maximize the profit $\pi_i$. Suppose the product inverse demand function is $p = 1 - \sum_{i=1}^{n} q_i$. The production decision-making problem of firm $k$, who obtains the innovative result, is as follows:

$$\max_{q_k} \pi_k = \left(1 - \sum_{j=1}^{n} q_j - (z_k - X)\right) q_k. \tag{1}$$

The production decision-making problems of other firms are as follows:

$$\max_{q_i} \pi_i = \left(1 - \sum_{j=1}^{n} q_j - z_i\right) q_i, \quad i \neq k. \tag{2}$$

By solving the optimization problems (1) and (2), the quantity $q_i^E(k)$ and the profit $\pi_i^E(k)$ of each firm can be obtained when the firm $k$ obtains the innovation result, in which the superscript $E$ represents the equilibrium result. The quantities are as follows:

$$\begin{aligned} q_k^E &= \frac{\left(1 + \sum_{j=1}^{n} z_j - (n+1)z_k + nX\right)}{(n+1)}, \\ q_i^E &= \frac{\left(1 + \sum_{j=1}^{n} z_j - (n+1)z_i - X\right)}{(n+1)}, \quad i \neq k. \end{aligned} \tag{3}$$

For all firms, the profit of firm $i$ is $\pi_i^E = (q_i^E)^2$.

Firm $i$ evaluates the innovation result by the difference between the profit when the innovation result is obtained and the one when the innovation result is not obtained. The difference is as follows:

$$v_i = \Delta\pi_i = X \frac{\left(2 + 2\sum_{j=1}^{n} z_j - 2(n+1)z_i + (n-1)X\right)}{(n+1)}. \tag{4}$$

Therefore, the difference in the evaluation by different firms is $v_i - v_j = 2X(z_j - z_i)$, which is entirely dependent on the difference in the original marginal production cost $z_i$. The less $z_i$ is, the higher the valuation $v_i$ is. Under the condition of complete information, it is clear that the bid of firm $i$ can be $P_i = \max_{j \neq i}\{v_j | v_j t \leq n v_i\}$.

Since the transfer price of innovation result mainly depends on the number of technology leaders but not the number of firms at the production stage, in order to focus on the internal financing of the innovation chain, the production stage is simplified to duopoly, that is, $n = 2$.

*(1) A Single Leader*. Assuming that $z_1 < z_2$, then $v_1 > v_2$, firm 1 obtains the innovation result at the price $P_1 = \max_{i \neq 1} v_i = v_2$. Then the profit of the R&D institution is as follows:

$$\pi_0 = P_1 - K = v_2 - K = \frac{1}{3}X(2 + 2z_1 - 4z_2 + X) - K. \tag{5}$$

The decision of the R&D institution is as follows:

$$\max_K \pi_0 = P_1 - K,$$
$$\text{s.t.} \begin{cases} K = \dfrac{X^2}{2}, \\ 0 \leq K \leq C. \end{cases} \tag{6}$$

The optimization problem (6) is solved to obtain the following:

(i) If $\sqrt{2C} < 2(1 + z_1 - 2z_2)$, then the optimal innovation result is $X_1^* = \sqrt{2C}$, the R&D investment is $K_1^E = C$, the profit of the R&D institution, firm 1 and the sum of both respectively are $\pi_{01}^E = 3/2$ $(1 + z_1 - 2z_2)\sqrt{2C} - 1/3C$, $\pi_{11}^E = 1/9(1 - 2z_1 + z_2 +$

$2\sqrt{2C})^2$ and $\Pi_1^E = \pi_{01}^E + \pi_{11}^E = 1/9(1 - 2z_1 + z_2)^2 + 2/9(5 - z_1 - 4z_2)\sqrt{2C} + 5/9C$.

(ii) If $\sqrt{2C} \geq 2(1 + z_1 - 2z_2)$, then the optimal innovation result is $X_1^* = 2(1 + z_1 - 2z_2)$, the R&D investment is $K_1{}^E = 2(1 + z_1 - 2z_2)^2$, the profit of the R&D institution, firm 1 and the sum of both respectively are $\pi_{01}^E = 2/3(1 + z_1 - 2z_2)^2$, $\pi_{11}^E = 1/9(5 + 2z_1 - 7z_2)^2$ and $\Pi_1^E = \pi_{01}^E + \pi_{11}^E = 10/9z_1^2 + 73/9z_2^2 - 52/9z_1z_2 + 22/9z_1 - 94/9z_2 + 31/9$.

*(2) Multiple Leaders.* Assuming that $z_1 = z_2$, then $v_1 = v_2$. The two firms have the same bid and the R&D institution randomly selects the trader, so firm 1 obtains the innovation result at the price $P_1 = v_1$ with the probability 1/2. Then the profit of the R&D institution is as follows:

$$\pi_0 = P_1 - K = v_1 - K = \frac{1}{3}X(2 - 4z_1 + 2z_2 + X) - K. \tag{7}$$

The decision of R&D institution still is optimization problem (6), and the solutions are as follows.

(i) If $\sqrt{2C} < 2(1 - z_1)$, then the optimal innovation result is $X_2^* = \sqrt{2C}$, the R&D investment is $K_2^E = C$, the profit of the R&D institution, firm 1 and the sum of both respectively are $\pi_{02}^E = 2/3(1 - z_1)\sqrt{2C} - 1/3C$, $\pi_{12}^E = 1/9(1 - 2z_1 + 2\sqrt{2C}))^2$ and $\Pi_2^E = \pi_{02}^E + \pi_{12}^E = 1/9(1 - 2z_1)^2 + 2/9(5 - 7z_1)\sqrt{2C} + 5/9C$.

(ii) If $\sqrt{2C} \geq 2(1 - z_1)$, then the optimal innovation result is $X_2^* = 2(1 - z_1)$, the R&D investment is $K_2^E = 2(1 - z_1)^2$, the profit of the R&D institution, firm 1 and the sum of both respectively are $\pi_{02}^E = 2/3(1 - z_1)^2$, $\pi_{12}^E = 5/9(1 - z_1)^2$ and $\Pi_2^E = \pi_{02}^E + \pi_{12}^E = 11/9(1 - z_1)^2$.

### 2.2. Innovative Activities and Profit Distribution in Innovation Chain

*2.2.1. Motivation to Set up Innovation Chain.* In the case of a single leader, firm 1 obtains the innovative result, resulting in a change in revenue $v_1 - P_1 = v_1 - v_2 > 0$, so firm 1 has the motivation to obtain the innovative result. In the case of multiple leaders, firm 1 obtains the innovative result, resulting in a change in revenue $v_1 - P_1 = v_1 - v_1 = 0$, but the innovation result can make it a single leader and dominance in the follow-up competition, so firm 1 still has the motivation to obtain the innovative result.

The R&D institution and firm $k$ establish an innovation chain to collaborate on innovation. The firm collects market information and provides it to the R&D institution, which makes the innovation activities of R&D institutions more targeted. At the same time, the R&D institution can arrange personnel to guide the firm to transform the innovation result, so as to enhance the contribution of it. The same innovation result $X$ can improve more of the marginal production cost of firm $k$. No matter how many leaders, firms have the motivation to obtain the innovation result, so they can establish an innovation chain with the R&D institution. Especially for multiple leaders, the leader has only a probability of $1/m$ to obtain the innovation result ($m$ is the number of neck-and-neck leaders). If the innovation chain is established, the probability of obtaining the innovation result can be improved, which is also one of the motivations to establish an innovation chain.

*2.2.2. Innovation Chain Activities and Profit Distribution.* The innovation activities in the innovation chain can also be expressed as a two-stage problem. At the first stage, the R&D institution cooperates to innovate with the firm and transfers the innovative result to the firm who pays the innovation result transfer price. At the second stage, the firm works with the R&D institution to transform the innovation result to improve its marginal productivity, compete with other firms, and finally allocate the excess profit of the innovation chain.

The R&D institutions and firm $k$ establish an innovation chain to collaborate on innovation, so as to enhance the contribution of the innovation result. The same innovation result $X$ can make the marginal production cost of firm $k$ become $z_k - \gamma X$, where $\gamma > 1$. Then the production decision of firm $k$ is as follows:

$$\max_{q_k} \pi_k = \left(1 - \sum_{i=1}^{n} q_i - (z_i - \gamma X))q_k\right). \tag{8}$$

Firms, which fail to obtain the innovative result, still make decisions according to problem (2). The quantities and profits can be obtained by solving optimization problems (8) and (2):

$$\begin{aligned} q_k^E &= \frac{\left(1 + \sum_{j=1}^{n} z_j - (n+1)z_k + n\gamma X)\right)}{(n+1)}, \\ \pi_k^E &= \left(q_k^E\right)^2. \end{aligned} \tag{9}$$

If $n = 2$ and the R&D institution forms the innovative chain with firm 1 (i.e., $k = 1$), then the solution can be simplified to the following:

$$\begin{aligned} q_1^E &= \frac{1}{3}(1 - 2z_1 + z_2 + 2\gamma X), \\ \pi_1^E &= \left(q_1^E\right)^2. \end{aligned} \tag{10}$$

In order to maximize the profit of the innovation chain, the decision-making problem of the R&D institutions is as follows:

$$\begin{aligned} &\max_K \Pi = \pi_1 - K, \\ &\text{s.t.} \begin{cases} K = \dfrac{X^2}{2}, \\ 0 \leq K \leq C, \end{cases} \end{aligned} \tag{11}$$

The optimization problem (11) is solved to obtain the following:

(1) If $8\gamma^2 - 9 < 0$ and $\sqrt{2C} \geq 4\gamma(1 - 2z_1 + z_2)/(9 - 8\gamma^2)$, then the optimal innovation result is $X_3^* = 4\gamma(1-$

$2z_1 + z_2)/(9 - 8\gamma^2)$, the R&D investment is $K_3^E = 8\gamma^2(1 - 2z_1 + z_2)^2/(9 - 8\gamma^2)^2$, and the profit of the innovation chain is $\Pi_3^E = (1 - 2z_1 + z_2)^2/(9 - 8\gamma^2)$. At this time, the R&D institution has sufficient funds and does not need financing.

(2) If $8\gamma^2 - 9 < 0$ and $\sqrt{2C} < 4\gamma(1 - 2z_1 + z_2)/(9 - 8\gamma^2)$, then the optimal innovation result is $X_3^* = \sqrt{2C}$, the R&D investment is $K_3^E = C$, the profit of the innovation chain is $\Pi_3^E = 1/9((8\gamma^2 - 9)C + 4\ (1 - 2z_1 + z_2)\ \gamma\sqrt{2C} + (1 - 2z_1 + z_2)^2)$. At this time, the R&D institution needs financing for lack of sufficient funds.

(3) If $8\gamma^2 - 9 \geq 0$, then the optimal innovation result is $X_3^* = \sqrt{2C}$, the R&D investment is $K_3^E = C$, the profit of the innovation chain is $\Pi_3^E = 1/9((8\gamma^2 - 9)\ C + 4(1 - 2z_1 + z_2)\gamma\sqrt{2C} + (1 - 2z_1 \quad + z_2)^2)$. At this time, the R&D institution needs financing for lack of sufficient funds.

The profit of the innovation chain need to distribute among the members of the innovation chain. The profit of the members of the innovation chain (R&D institution and firm (1) consists of two parts: one is the basic profit, that is, the profit in noninnovation chain; the other is the excess profit, that is, the excess profit of the innovation chain distributed according to the contribution of each member. There is only one member of each type in the simple innovation chain, so the excess profit of the innovation chain is distributed equally among the members. The total profit from the innovation chain of R&D institution and firm can be expressed as follows. If $s$ denotes the number of the leaders (i.e., $s = 1$, the case of a single leader; $s = 2$, the case of multiple leaders), then the profits of R&D institution and the firm respectively are $\pi_{03}^E = \pi_{0s}^E + (\Pi_3^E - \Pi_s^E)/2$ and $\pi_{13}^E = \pi_{1s}^E + (\Pi_3^E - \Pi_s^E)/2$.

## 3. Internal Financing and Profit Distribution Mechanism in Innovation Chain

The R&D institution in the innovation chain need more financing because of their own financial constraints, but since they lack the collateral, the external financing is more difficult, so the financing of the firms in the innovation chain can be considered. On the other hand, the innovation result can bring more profit to firms when the R&D investment of R&D institution increases, so firms also have the motivation to lend funds to the R&D institution. In this paper, we, with reference to the supply chain finance, design the innovation chain internal prepayment financing mode.

*3.1. Prepayment Financing Activities in Innovation Chain.* For the insufficient capital of the R&D institution, the constraint in the optimization problem (11) is a tight constraint, that is, $K^E = C$. At this time, increased R&D investment can boost the profits of the R&D institution and the firm, so the firm can pay the transfer price of the innovation result in advance, and increase the R&D investment for R&D institution financing. The purchase prices of production input elements(such as the innovative production of the upstream subject) influence the profit of the firm in the innovation chain, but the quality of the upstream subject's innovative product impacts more significantly the profit. Therefore when the firm in innovation chain pays the innovation result transfer price to the R&D institution in advance, it is more desirable for the R&D institution to deliver high-quality innovation, rather than hope to get a price discount. Considering cost control and risk control, the firm only pays part of the transfer price in advance.

The financing activities of prepayment in the innovation chain can be divided into three stages. At the first stage, the R&D institution finances $B$ from the firm. At the second stage, the R&D institution invests $K$ in R&D, obtains the innovation result $X$ and then transfers it to the firm, which in turn transfers the remaining cost, $P_1 - B$, to the R&D institution. The R&D institution obtains the first part of the profit, that is, noninnovation chain profit $P_1$. At the third stage, the firm converts the innovation result into the productive force, competes with other firms to get the profit, and finally distributes the excess profit of the innovation chain.

*3.2. Prepayment Financing Mechanism and Member Profit in Innovation Chain.* The third stage problem is the same as that in section 2.2.2, so the optimization problems (8) and (2) are used to describe the problem, and the quantity and profit of the firm are $q_1^E = 1/3(1 - 2z_1 + z_2 + 2\gamma X)$and $\pi_1^E = (q_1^E)^2$.

The second stage problem is similar to that in section 2.2.2, which can be expressed as follows:

$$\max_K \Pi = \pi_1 - K,$$
$$\text{s.t.} \begin{cases} K = X^2/2, \\ 0 \leq K \leq C + B, \end{cases} \tag{12}$$

By changing $C$ in the optimization problem (11) to $C + B$, we obtain the solution of the optimization problem (12) as follows:

(1) If $8\gamma^2 - 9 < 0$ and $\sqrt{2C} \geq 4\gamma(1 - 2z_1 + z_2)/(9 - 8\gamma^2)$, then the optimal innovation result is $X_4^* = 4\gamma(1 - 2z_1 + z_2)/(9 - 8\gamma^2)$, the R&D investment is $K_4^E = 8\gamma^2(1 - 2z_1 + z_2)^2/(9 - 8\gamma^2)^2$, the profit of the innovation chain is $\Pi_4^E = (1 - 2z_1 + z_2)^2/(9 - 8\gamma^2)$. At this time, the R&D institution has sufficient funds and does not need financing.

(2) If $8\gamma^2 - 9 < 0$ and $\sqrt{2C} < 4\gamma(1 - 2z_1 + z_2)/(9 - 8\gamma^2) \leq \sqrt{2(C + B)}$, then the optimal innovation result is $X_4^* = 4\gamma(1 - 2z_1 + z_2)/(9 - 8\gamma^2)$, the R&D investment is $K_4^E = 8\gamma^2(1 - 2z_1 + z_2)^2/(9 - 8\gamma^2)^2$, the profit of the innovation chain is $\Pi_4^E = (1 - 2z_1 + z_2)^2/(9 - 8\gamma^2)$. At this time, R&D

institution has insufficient funds of their own, but after financing, there will be a surplus of funds.

(3) If $8\gamma^2 - 9 < 0$ and $\sqrt{2(C+B)} < 4\gamma(1-2z_1+z_2)/(9-8\gamma^2)$, then the optimal innovation result is $X_4^* = \sqrt{2(C+B)}$, the R&D investment is $K_4^E = C + B$, the profit of the innovation chain is $\Pi_4^E = 1/9((8\gamma^2-9)(C+B) + 4(1-2z_1+z_2)\gamma\sqrt{2}(C+B) + (1-2z_1+z_2)^2)$. At this time, R&D institution has insufficient funds of their own, but after financing, there will still be a shortage of funds.

(4) If $8\gamma^2 - 9 \geq 0$, then the optimal innovation result is $X_4^* = \sqrt{2(C+B)}$, the R&D investment is $K_4^E = C + B$, the profit of the innovation chain is $\Pi_4^E = 1/9((8\gamma^2-9)(C+B) + 4(1-2z_1+z_2)\gamma\sqrt{2(C+B)} + (1-2z_1+z_2)^2)$. At this time, R&D institution has its insufficient own funds, and there is still a shortage of funds after financing.

At the first stage, the R&D institution in the innovation chain finances $B$ from the firm. With consideration of the capital cost of the firm, the innovation chain makes financing $B$ as small as possible while maximizing the overall profit in order to realize the internal financing of the innovation chain. The problem can be expressed as follows:

$$\begin{aligned} &\max_B \ (w\Pi - B), \\ &\text{s.t. } B \leq MB, \end{aligned} \tag{13}$$

where $w$ is a sufficiently large positive number, indicating that the target level of maximizing total profit is much higher than the goal of minimizing financing, and $MB$ indicates the maximum amount of financing that the firm can provide. The solution is as follows:

(1) If $8\gamma^2 - 9 < 0$ and $\sqrt{2C} \geq 4\gamma(1-2z_1+z_2)/(9-8\gamma^2)$, then the optimal financing is $B^* = 0$, the innovation result is $X_4^* = 4\gamma(1-2z_1+z_2)/(9-8\gamma^2)$, and the profit of the innovation chain is $\Pi_4^E = (1-2z_1+z_2)^2/(9-8\gamma^2)$.

(2) If $8\gamma^2 - 9 < 0$ and $\sqrt{2C} < 4\gamma(1-2z_1+z_2)/(9-8\gamma^2) \leq \sqrt{2(C+\mathrm{MB})}$, then the optimal financing is $B^* = 8\gamma^2(1-2z_1+z_2)^2/(9-8\gamma^2)^2 - C$, the innovation result is $X_4^* = 4\gamma(1-2z_1+z_2)/(9-8\gamma^2)$, and the profit of the innovation chain is $\Pi_4^E = (1-2z_1+z_2)^2/(9-8\gamma^2)$.

(3) If $8\gamma^2 - 9 < 0$ and $\sqrt{2(C+\mathrm{MB})} < 4\gamma(1-2z_1+z_2)/(9-8\gamma^2)$, then the optimal financing is $B^* = \mathrm{MB}$, the innovation result is $X_4^* = \sqrt{2(C+\mathrm{MB})}$, and the profit of the innovation chain is $\Pi_4^E = 1/9((8\gamma^2-9)(C+\mathrm{MB}) + 4(1-2z_1+z_2)\gamma\sqrt{2(C+\mathrm{MB})} + (1-2z_1+z_2)^2)$.

(4) If $8\gamma^2 - 9 \geq 0$, then the optimal financing is $B^* = MB$, the innovation result is $X_4{}^* = \sqrt{2(C+\mathrm{MB})}$, and the profit of the innovation chain is $\Pi_4^E = 1/9((8\gamma^2-9)(C+\mathrm{MB}) + 4(1-2z_1+z_2)\gamma\sqrt{2(C+\mathrm{MB})} + (1-2z_1+z_2)^2)$.

In the case (1), the R&D institution itself has sufficient funds and does not need financing. In the case (2), the R&D institution needs financing, the firm's capital is abundant, and the internal financing of the innovation chain can reach the optimal innovation and production level. In the case (3), the R&D institution needs financing, but the firm's capital is not sufficient, the innovation chain financing does not reach the optimal level, so the innovation chain needs external financing. In case (4), the more R&D investment in the innovation chain, the better, and both the R&D institution and the firm fully invest in R&D and seek external financing. However, in view of the assumption of gradual innovation $\gamma X < z_1$, case (4) is rare and needs to be discussed in the context of disruptive innovation. So it will not be investigated in this paper.

Finally, the total profit of the R&D institution and the firm obtained from the innovation chain can be expressed as follows. If $s$ denotes the number of the leaders, then the profits of the R&D institution and the firm respectively are $\pi_{04}^E = \pi_{0s}^E + (\Pi_4^E - \Pi_s^E)/2$ and $\pi_{14}^E = \pi_{1s}^E + (\Pi_4^E - \Pi_s^E)/2$.

Under the same circumstances, $\Pi_4^E \geq \Pi_3^E \geq \Pi_1^E \geq \Pi_2^E$, $\pi_{04}^E \geq \pi_{03}^E \geq \pi_{01}^E \geq \pi_{02}^E$ and $\pi_{14}^E \geq \pi_{13}^E \geq \pi_{11}^E \geq \pi_{12}^E$. It means that cooperation of innovation and production in the innovation chain can bring the excess profit to the members (R&D institution and firm) and the internal financing of the innovation chain can improve the excess profit.

## 4. Risk Control Mechanism of Internal Financing in Innovation Chain

*4.1. Risk Control Principles of Internal Financing in Innovation Chain.* The risk of the internal financing of innovation chain is derived from the risk of the R&D stage and the risk of the production stage. The main operation subject in the R&D stage is the R&D institution, and the main operation subject in the production stage is the firm. Due to the particularity of R&D activities, the risk of the R&D stage is much greater, and the risks of the two stages are seriously asymmetric. So if the risk is to be shared among the members of innovation chain, it may not be possible to attract the firm in the production stage to join the innovation chain. In order to construct the innovation chain and control the risk of the internal financing in innovation chain, it is necessary to let each member bear the risk of its own main activity stage, reduce the risk of each stage and control the chain contagion of the risk of the innovation chain so that the risk of the main subject in the internal financing of the innovation chain does not exceed the risk of the noninnovation chain. The risk control of the internal financing in innovation chain can be realized through the reasonable risk sharing mechanism.

In the internal financing mode of innovation chain, there are three capital transfers between the R&D institution and the firm. The first one is for the firm to finance R&D institution; the second one is for the firm to pay the remaining cost to the R&D institution when transferring innovation result, and the third one is for firm to pay the excess profit of

innovation chain to R&D institution at the end of production. The first capital transfer is the realization of the internal financing in innovation chain, and the risk of the internal financing in innovation chain is curbed by controlling the second and third capital transfers.

*4.2. Risk Sharing Mechanism at the R&D Stage.* The R&D institution shall bear the risk of R&D stage and reduce its transmission to production stage, namely, to ensure the profit of the firm. When the innovation result is transferred, the firm shall pay the remaining part of the price according to the value of the actual innovation result $\widehat{X}$, and determine the excess profit to be paid to the R&D institution.

*4.2.1. Superquality Innovation Result.* When the innovation result exceeds the plan, the transfer price remains unchanged, and the excess profit of the R&D institution is calculated according to the actual result. Therefore, if $\pi_{14}(\widehat{X}) - \pi_{14}(X_4^*) \geq 0$, the firm pays $P_1(C) - B$, and the excess profit of the R&D institution is $1/2(\pi_{14}(\widehat{X}) - \pi_{1s}(\widehat{X}))$, where $s$ denotes the number of the leaders. Then they will wait for the third capital transfer.

*4.2.2. Lower-Quality Innovation Result.* When the innovation result is lower than the plan, the total profit of the innovation chain is lost, and the loss is borne by the R&D institution. Firstly, the remaining cost is used to compensate for the loss. Secondly, the excess profit of the R&D institution is used to compensate. Finally, the amount of financing is used to compensate. The remaining loss is borne by the firm. We let $s = 1$ to denote the case of a single leader and $s = 2$ to denote the case of multiple leaders:

(1) If $\pi_{1s}(X_4^*) - \pi_{1s}(\widehat{X}) - 2(P_1 - B) < \pi_{14}(\widehat{X}) - \pi_{14}(X_4^*) < 0$, the R&D institution compensates the loss of the firm by repaying from the remaining cost, so the firm pays $(P_1 - B) - 1/2(\pi_{14}(X_4^*) - \pi_{14}(\widehat{X}) + \pi_{1s}(X_4^*) - \pi_{1s}(\widehat{X}))$, but the expected excess profit remains unchanged, which is still $1/2(\pi_{14}(\widehat{X}) - \pi_{1s}(\widehat{X}))$, and then they will wait for the third transfer of funds.

(2) If $1/2(\pi_{1s}(X_4^*) - \pi_{14}(X_4^*)) - (P_1 - B) < \pi_{14}(\widehat{X}) - \pi_{14}(X_4^*) < \pi_{1s}(X_4^*) - \pi_{1s}(\widehat{X}) - 2(P_1 - B)$, the R&D institution compensates the loss of the firm with the full remaining cost and part of the expected excess profit. The firm pays 0, the excess profit of the R&D institution is $1/2(2\pi_{14}(\widehat{X}) - \pi_{14}(X_4^*) - \pi_{1s}(X_4^*)) + (P_1 - B)$, and they will wait for the third capital transfer.

(3) If $1/2(\pi_{1s}(X_4^*) - \pi_{14}(X_4^*)) - P_1 < \pi_{14}(\widehat{X}) - \pi_{14}(X_4^*) < 1/2(\pi_{1s}(X_4^*) - \pi_{14}(X_4^*)) - (P_1 - B)$, the R&D institution repays the firm $B - P_1 - 1/2(2\pi_{14}(\widehat{X}) - \pi_{14}(X_4^*) - \pi_{1s}(X_4^*))$. The R&D institution cancels the remaining cost and the expected excess profit to compensate for the loss of the firm and at the same time repays part of the financing.

(4) If $\pi_{14}(\widehat{X}) - \pi_{14}(X_4^*) < 1/2(\pi_{1s}(X_4^*) - \pi_{14}(X_4^*)) - P_1$, the R&D institution repays the firm $B$. The R&D institution cancels the remaining cost and the expected excess profit, and repays the financing to compensate, but the firm still has to bear the remaining loss.

*4.3. Risk Sharing Mechanism at the Production Stage.* The firm shall bear the risk of production stage and reduce its transmission to R&D stage, that is, to ensure the profit of the R&D institution. At the end of the production, the firm should pay the actual excess profit to the R&D institution according to the actual profit $\widehat{\pi}$.

*4.3.1. Pay Total Excess Profit.* In the case of superquality innovation result or in the first case of lower-quality innovation result, the firm pays a total excess profit to the R&D institution. If $\widehat{\pi} \geq \pi_{14}(\widehat{X})$, the firm pays the excess profit $1/2(\widehat{\pi} - t\pi_{1s}n(\widehat{X}))$ to the R&D institution, if $1/2(\pi_{14}(\widehat{X}) - \pi_{1s}(\widehat{X})) \leq \widehat{\pi} \leq \pi_{14}(\widehat{X})$, the firm pays $1/2(\pi_{14}(\widehat{X}) - \pi_{1s}(\widehat{X}))$, if $0 \leq \widehat{\pi} \leq 1/2(\pi_{14}(\widehat{X}) - \pi_{1s}(\widehat{X}))$, the firm pays $\widehat{\pi}$, where $s$ denotes the number of the leaders.

*4.3.2. Pay Part of Excess Profit.* In the second case of lower-quality innovation result, the firm pays part of the excess profit. If $\widehat{\pi} \geq \pi_{14}(\widehat{X})$, the firm pays $1/2(2\widehat{\pi} - \pi_{14}(X_4^*) - \pi_{1s}(X_4^*)) + (P_1 - B)$, if $1/2(2\pi_{14}(\widehat{X}) - \pi_{14}(X_4^*) - \pi_{1s}(X_4^*)) + (P_1 - B) \leq \widehat{\pi} \leq \pi_{14}(\widehat{X})$, the firm pays $1/2(2\pi_{14}(\widehat{X}) - \pi_{14}(X_4^*) - \pi_{1s}(X_4^*)) + (P_1 - B)$, if $0 \leq \widehat{\pi} \leq 1/2(2\pi_{14}(\widehat{X}) - \pi_{14}(X_4^*) - \pi_{1s}(X_4^*)) + (P_1 - B)$, the firm pays $\widehat{\pi}$, where $s$ denotes the number of the leaders.

*4.3.3. Pay No Excess Profit.* In the third or fourth case of lower-quality innovation result, the firm pays no excess profit to the R&D institution.

## 5. Conclusion

Through the research-development stage and production-commercialization stage, we analyze the innovation activities and profits and find that the innovation chain which is established by the R&D institution and the firm can bring the excess profit. Based on the cooperative partnership between the R&D institution and the firm and the characteristics of innovation activities in the innovation chain, the internal financing mechanism between members of the innovation chain is constructed, in which the firm finances funds for the R&D institution. The optimal financing strategy is given to alleviate the constraints of innovation funds and improve the excess profit of the innovation chain. Finally, considering the risk property of the innovation activities, we set up the risk sharing mechanism of each stage to control the risk of the internal financing between members of the innovation chain, which ensures the operation of the internal financing mechanism.

In addition, the internal financing mechanism between members of the innovation chain establishes the mutual

finance-assisting relationship of the innovation subject in the innovation chain, which strengthens the leading role of the core innovation subject. If the R&D institution is the core, it can influence the firm through innovation results and its transfer. And if the firm is at the core, it can influence the R&D institution through financing. In a word, it could strengthen the relationship between innovation subjects and improve the coordination and cooperation mechanism, so that it would improve the efficiency of scientific and technological innovation activities.

## Data Availability

No data were used to support this study.

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## Acknowledgments

The authors gratefully acknowledge the financial support of the General Program of Natural Science Foundation of Zhejiang Province (LY20G030007), Humanities and Social Sciences Research Project of the Ministry of Education of China (20YJA790097), Major Supporting Project of Philosophy and Social Sciences in Zhejiang Province (20XXJC03ZD), Major Supporting Project of Academy of Longyuan Construction Financial Research Ningbo University (LYZDA2003), and Zhejiang Provincial Collaborative Innovation Center of Port Economy.

## References

[1] J. Schumpeter, *Capitalism, Socialism and Democracy*, Harper, New York, NY, USA, 3rd edition, 1950.

[2] M. T. Hansen and J. Birkinshaw, "The innovation value chain," *Harvard Business Review*, vol. 85, no. 6, pp. 121–135, 2007.

[3] N. Sen, "Innovation chain and CSIR," *Current Science*, vol. 85, no. 5, pp. 570–574, 2003.

[4] P. Bamfield, *The Innovation Chain, Research and Development Management in the Chemical and Pharmaceutical Industry*, WILrY VCH, Weinheim, Germany, 2nd edition, 2004.

[5] X. Lin, J. B. Liu, and X. Chen, "Industrial upgrading based on global innovation chains: a case study of Huawei technologies Co., Ltd. Shenzhen," *International Journal of Innovation Studies*, vol. 2, no. 3, pp. 81–90, 2018.

[6] R. Stephen, D. Jun, and H. L. James, "Modelling the innovation value chain," *Research Policy*, vol. 37, no. 6, pp. 961–977, 2008.

[7] G. Panagiotis and H. L. James, "The innovation value chain in new technology-based firms: evidence from the U.K." *Journal of Product Innovation Management*, vol. 29, no. 5, pp. 839–860, 2012.

[8] E. L. Olson, "Perspective: the green innovation value chain: a tool for evaluating the diffusion prospects of green products," *Journal of Product Innovation Management*, vol. 30, no. 4, pp. 782–793, 2013.

[9] L. Lai-Xin and Y. He, "The study on the cycle of technological innovation chain," in *Proceedings of the International Conference on Wireless Communications*, IEEE, Dalian, China, March 2008.

[10] K. C. Ekaterina, E. K. Ekaterina, D. K. Anna, and S. P. Julia, "Research of instruments for financing of innovation and investment construction projects," *Procedia Engineering*, vol. 153, pp. 112–117, 2016.

[11] B. Emanuele, "Innovation financing and the role of relationship lending for SMEs," *Small Business Economics*, vol. 44, no. 2, pp. 449–473, 2015.

[12] H. C. Pfohl and M. Gomm, "Supply chain finance: optimizing financial flows in supply chains," *Logistics Research*, vol. 1, no. 3-4, pp. 149–161, 2009.

[13] P.-H. Hsu, "development and innovation: cross-country evidence," *Journal of Financial Economics*, vol. 112, no. 1, pp. 116–135, 2014.

[14] J. F. Tu, "Construction of die and mould technology innovation chain based on resource tree," *Applied Mechanics and Materials*, vol. 201-202, pp. 1025–1028, 2012.

[15] L. Knut, M. S. Gordon, S. Jorg, and W. Arnold, "Financing innovations in uncertain networks-filling in roadmap gaps in the semiconductor industry," *Research Policy*, vol. 42, no. 3, pp. 647–661, 2013.

[16] H. Byun and J. Lee, "The study on the technology finance policy for technology development on the value chain based innovation system," in *Proceedings of the Portland International Center for Management of Engineering & Technology*, IEEE, Portland, Oregon, August 2007.

[17] L. M. Gelsomino, R. Mangiaracina, A. Perego et al., "Supply chain finance: a literature review," *International Journal of Physical Distribution & Logistics Management*, vol. 46, no. 4, pp. 348–366, 2016.

[18] A. K. Maiti, M. K. Maiti, and M. Maiti, "Inventory model with stochastic lead-time and price dependent demand incorporating advance payment," *Applied Mathematical Modelling*, vol. 33, no. 5, pp. 2433–2443, 2009.

[19] A. Thangam, "Optimal price discounting and lot-sizing policies for perishable items in a supply chain under advance payment scheme and two-echelon trade credits," *International Journal of Production Economics*, vol. 139, no. 2, pp. 459–472, 2012.

[20] A. A. Taleizadeh, "Lot-sizing model with advance payment pricing and disruption in supply under planned partial backordering," *International Transactions in Operational Research*, vol. 24, no. 4, pp. 783–800, 2017.

[21] G. P. Cachon, "The allocation of inventory risk in a supply chain: push, pull, and advance-purchase discount contracts," *Management Science*, vol. 50, no. 2, pp. 222–238, 2004.

[22] B. Zhang, D. Wu, L. Liang, and D. L. Olson, "Supply chain loss averse newsboy model with capital constraint," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 46, no. 5, pp. 646–658, 2016.

[23] A. Diabat, A. A. Taleizadeh, and M. Lashgari, "A lot sizing model with partial downstream delayed payment, partial upstream advance payment, and partial backordering for deteriorating items," *Journal of Manufacturing Systems*, vol. 45, pp. 322–342, 2017.

[24] M. Wandfluh, E. Hofmann, and P. Schoensleben, "Financing buyer-supplier dyads: an empirical analysis on financial collaboration in the supply chain," *International Journal of Logistics Research and Applications*, vol. 19, pp. 1–18, 2015.

[25] M. Oliveira, M. Kadapakkam, and M. Beyhaghi, "Effects of customer financial distress on supplier capital structure," *Journal of Corporate Finance*, vol. 42, pp. 131–149, 2017.

[26] H. Görg and E. Kersting, "Vertical integration and supplier finance," *Canadian Journal of Economics/Revue Canadienne D'économique*, vol. 50, no. 1, pp. 273–305, 2017.

Hindawi
Mathematical Problems in Engineering
Volume 2020, Article ID 6644465, 12 pages
https://doi.org/10.1155/2020/6644465

*Research Article*

# An Algorithm Combining Latent Dirichlet Allocation and Bimodal Network for Evaluating Goal Deviation of Intellectual Property Strategy Execution in China

**Bing Sun,[1] Mingxing Yu,[1] and Zaoli Yang[2]**

[1]*School of Economics and Management, Harbin Engineering University, Harbin 150001, China*
[2]*College of Economics and Management, Beijing University of Technology, Beijing 100124, China*

Correspondence should be addressed to Zaoli Yang; yangzaoli@hotmail.com

Received 3 October 2020; Revised 2 November 2020; Accepted 7 November 2020; Published 21 November 2020

Academic Editor: Shouzhen Zeng



China has implemented the intellectual property strategy since 2008 to support innovation-driven development. However, statistical data issued during the "12th Five-Year Plan" (2011–2015) showed that there are certain deviations between the actual and expected intellectual property strategy's goals. To effectively diagnose the goal deviation, an algorithm combining the latent Dirichlet allocation and bimodal network based on policy text was proposed. In this method, topics in intellectual property policy texts of China's provincial regions were extracted through the latent Dirichlet allocation model, and a bimodal network centered at provincial administration district-policy topics was constructed. Subsequently, the characteristics of the goal execution deviation of the IPS in the provincial government were explored based on the centrality of the bimodal network and singular value decomposition. Finally, some diagnosis results and conclusions were demonstrated to provide reasonable methods for evaluation of national strategic planning and promoting policy performance.

## 1. Introduction

China's economic development is characteristic of decreasing growth rate, the urgent needs of structural upgrading, and dynamic transformation since 2008. To get rid of resource restrictions, realize the transformation from resource element-driven development to innovation-driven development, and finally improve the economic development quality, the Chinese government issued the *Outline of National Intellectual Property Strategy* in 2008, which symbolized the application of intellectual property strategy (IPS) in innovation-driven development, to eliminate resource restrictions, transform resource element-driven development to innovation-driven development, and improve the economic development quality. IPS has become the overall strategy that has driven China to the knowledge-based economic development in the coming decades [1]. The central and provincial regions enacted more than 500 policy measures and documents for the smooth implementation of the IPS from 2009 to 2017. Hence, the IPS system under Chinese policies and laws was formed gradually. However, a large gap between the actual and expected goals of Chinese IPS was observed. In 2016, China released the *National Intellectual Property Protection and Application Planning* during the "13th Five-Year Plan," which indicated that Chinese intellectual property (IP) during the "12th Five-Year Plan" still faces numerous challenges, such as incoordination between the number and quality of intellectual properties, the small number of core patents, famous brands and high-quality copyrights, poor association with the market, low transformation rate of achievements, and imperfect management mechanism, as well as backward awareness of intellectual property protection, which restrict the improvement of China's intellectual property protection capability. The successful implementation of China's national overall strategy relies highly on the execution of provincial regions' supporting policy measures, which is the main reason for the deviation between the actual and

expected goals of Chinese IPS. Nevertheless, existing studies mainly focus on evaluating the IPS or other public policies' goal performance [2–7] rather than analyzing the deviation characteristics from the execution stages of provincial regions. Thus, they neither accurately evaluate the specific execution stage of the overall goal deviation of the national IPS nor disclose the causes of deviation. From this perspective, diagnosing the goal execution deviation of Chinese IPS in the provincial region's implementation stage, accurately positioning relevant weak stages, scientifically adjusting the implementation of the strategic plan, and assuring maximized implementation benefits are practically crucial.

The remainder of this paper is organized as follows. Section 3 introduces the methodologies, which mainly include data source and processing, LDA topic mining model, and bimodal network analysis. Section 4 describes the empirical analysis of the goal execution deviation characteristics of IPS in China's provincial regions based on policy texts. Section 5 presents the results and discussions. Section 6 draws the conclusions and proposes the future prospects. The framework of this study is shown in Figure 1.

## 2. State of the Art

Studies on the diagnosis and evaluation of IPS under the traditional paradigm can be divided into two types. First, a diagnosis or evaluation model was constructed on the basis of the goal-driven evaluation paradigm by using the open structural file information of public departments to evaluate the implementation effect of the IPS. Song et al. established a set of evaluation index systems and methods to measure the IP creation, IP application, IP protection, and IP management based on collected data in the statistical report of intellectual properties in China's scientific research institutions [2]. Additionally, statistical data related to patents were acquired to construct the patent application network model [3] and a "three stages-two dimensions" model [4]. These models were used to evaluate the patent value or quality and to diagnose the implementation effect of the IPS indirectly. Alternatively, the capability of patent industrialization was evaluated and the commercial effect of the IPS was diagnosed by using the collected transaction information in the technological market through the fuzzy multi-attribute comprehensive evaluation method [5] and the equation of patent standardized measurement [8]. Second, an econometric model was constructed on the basis of the theoretical hypothesis-driven diagnosis paradigm to test the present theoretical hypothesis by using data samples collected from field investigation or by summarizing statistical data of the document. Most of the associated studies focused on the influences of IPS or IP policies on product output [9–12] and innovation level [13, 14] or on enterprises' technological innovation [15, 16] and production efficiency [17–19]. However, the traditional evaluation models, which are based on the structural information of single-file data and driven by goals and theoretical hypothesis, can only diagnose the implementation results of IPS. National strategies or public policies cover multiple stages, such as the formation, execution, evaluation, feedback, and correction of public policies [20]. The traditional goal- and theoretical hypothesis-driven evaluation models based on structural information have a weak information basis. Therefore, the final evaluation results cannot involve the diagnosis of most policy execution stages or others.

Different from the traditional evaluation modes, the data-driven evaluation for policy effectiveness has become a new paradigm and been practiced in education policy [21, 22], energy policy [23–25], safety management [26, 27], building [28, 29], public medical policy [30–32], and so on. However, the aforementioned studies only achieve the local perfection of public policies on specific occasions with limited structured information; they have not discussed execution-stage evaluation problems that involve public policy processes. With the continuous development of the Internet in recent years, the public-derived big data based on the Internet unstructured text information may exert reconstructive influences on the national decision-making process [33]. The public policy evaluation driven by unstructured text information has started. However, existing associated studies only theoretically discuss the feasibility of such public policy evaluation and conduct a qualitative analysis of the influences of Internet unstructured text information on policy formulation and legislation changes [34] as well as the optimization of public policies [33]. Internet unstructured text information has not been applied in the field of public policy evaluation. Relevant studies remain in the theoretical exploration stage and associated empirical analysis must be extended urgently.

The goal of strategic planning is an important component of the national strategy and public policy [35]. The gap between the overall execution and expected goals has been expanding since China implemented the IPS in 2008. However, studies on the goal deviation of IPS, let alone goal execution deviation, are limited. Organizational behavioral science concerned "goal deviation" early. This field described goal deviation as follows: the organization deviates from the original preset goal and shifts to other goals in the actual process due to the influences of different factors [36]. Subsequently, connotations of goal deviation were further extracted. Representative scholars Resh and Marvel indicated that goal deviation occurs when the organization focuses on the replacement of output indexes, which are difficult to measure and have great relevance [37]. Liu et al. argued that goal deviation is an organizational phenomenon that pays further attention to the easy-to-measure explicit goal or uses the mean as the goal rather than the preset goal [38].

The study on the goal execution deviation of IPS can only be screened indirectly from previous investigations. For example, Wang and Hsieh [5] and Tamura [8] evaluated the application of IP and its commercial performance. The goal execution deviation is gained indirectly on the basis of a simple comparison between the evaluation results and the expected goal of IPS, but the characteristic of goal execution deviation remains difficult to obtain. As the goal of IPS is realized by specific policies and measures of the provincial region, the research scale on the goal execution deviation of

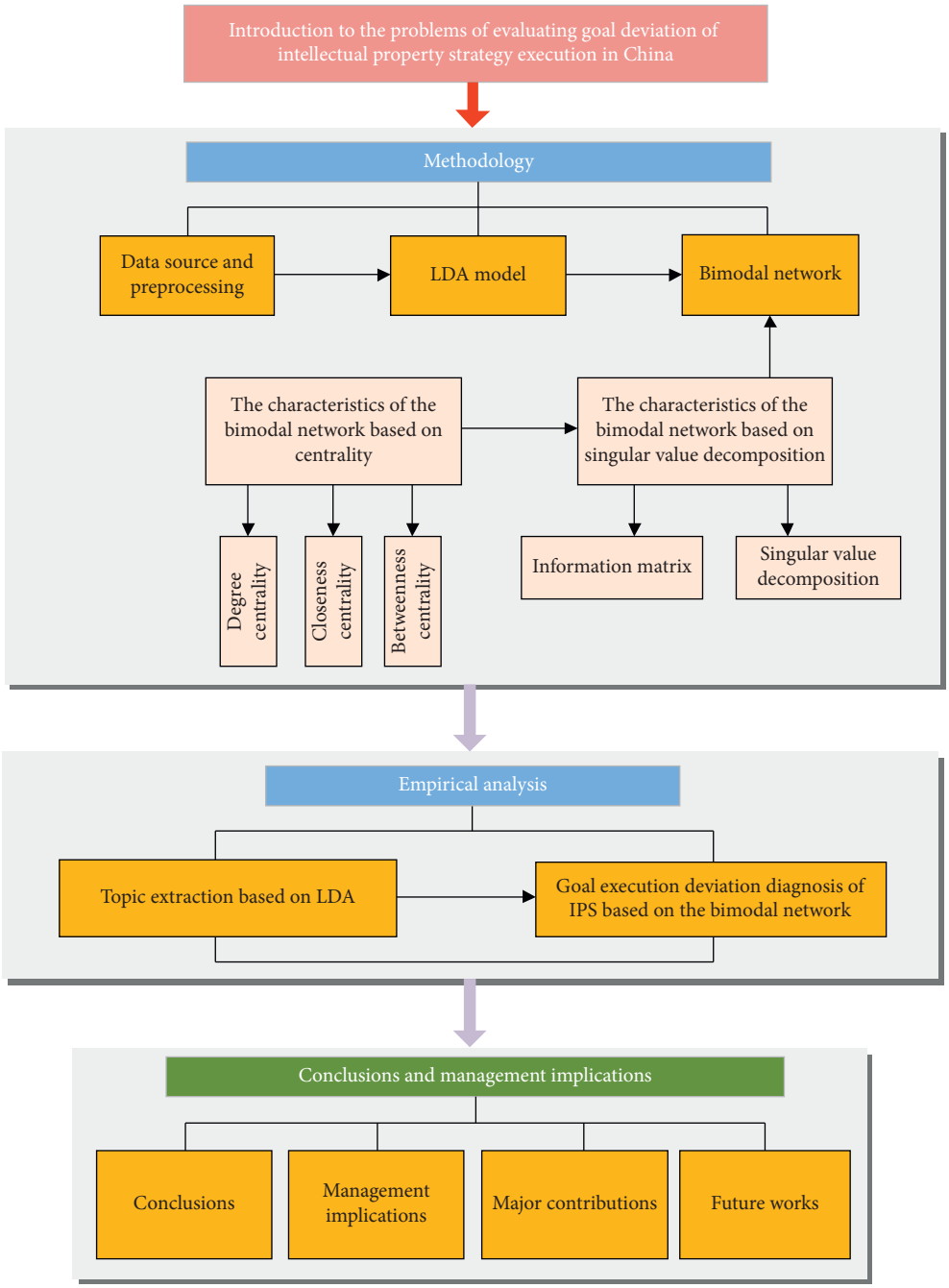



Figure 1: The framework of this study.

IP policies, which are important components of public policies, has been formed, thus laying the foundation for the goal execution deviation diagnosis. However, the goal execution deviation diagnosis of public policies sets different measurable variables, including the incompletion degree of the goal and the fuzzy degree of results 35 [39–42]. Subsequently, the degree of goal execution deviation is determined through empirical analysis based on a questionnaire survey or statistical data. Research perspectives remain limited in the implementation outcome of public policies due to the structured information, and the specific execution has not been discussed yet.

Overall, previous studies on goal execution deviation diagnoses of IPS and relevant policies are mainly based on the traditional goal- and theoretical hypothesis-driven paradigm. Diagnosis results only display the achievement of the overall goal but neglect the specific execution of policies. Moreover, the information source is limited in the second-handed data of government statistical report, and heterogeneous information lacks analysis based on policy texts, which hinder decision-makers from identifying multiple causes of goal execution deviation of IPS. For this reason, considering the advantages of the LDA model for accurately extracting text topics, the LDA model was introduced to

mine topics in the IP policy text of provincial regions. Then, since the bimodal network can analyze the relationship between two kinds of heterogeneous things, a bimodal network centered at topics and provinces in IP policy text was constructed. Through this network, the goal execution deviation of provincial regions was diagnosed in accordance with the structural characteristics of the bimodal network.

## 3. Methodology

*3.1. Data Source and Preprocessing.* In this study, the text files of policies, plans, and strategic plans related to IP from 2008 to 2017 were downloaded through artificial interpretation from IP websites of selected 30 provincial regions and of the Department of Science and Technology in China. A total of 449 effective policy texts were selected after the release of the *Outline of National Intellectual Property Strategy* in 2017 with considerations to the regionality and integrity of policy texts, which covered more than 220,000 Chinese characters.

The 449 original policy texts were transformed into the common CSV format, marked with provinces. IP policy texts also have specific writing norms. By combining the word-formation rules of the noun phrase and formal standards of policy texts, subject terms in policy texts were extracted in accordance with part-of-speech tagging and the extraction rule of noun phrases. This approach includes constructing a new stop words library (which is applicable to IP policy texts based on the stop words library of Harbin Institute of Technology) and adding empty words (which are inapplicable for extracting notional subject terms from the stop words library), such as "middle-level," "11th Five-Year Plan," and "copies." In addition, subject terms were extracted through the Jieba Chinese word segmentation. During the extraction, the dictionary named dict.txt in Jieba was used and a noun dictionary on IP policy analysis was constructed. For example, some professional terms in the IP policy text were added to the new dictionary, such as "IP creation," "IP use or application," "IP protection," "patent," and "trademark." Subject terms in policy texts were selected through Jieba word segmentation in Python, thus transforming an IP policy text into the eigenvectors comprising several subject terms about IP. These eigenvectors were used in the subsequent topic model analysis.

*3.2. LDA Model.* LDA is a statistical topic model constructed by Blei [43] for extracting potential topics. The basic principle of LDA is to calculate the conditional distribution of potential variables under the given observation variables in accordance with a joint probability distribution. The recognition method of potential document contents, which is based on the Bayes probability of words, topics, and documents, is typical unsupervised learning.

In the LDA model, documents were viewed as the mixed probability distribution of potential subjects, and topics were viewed as the probability of several words. Therefore, potential topics and documents were regarded as the probability distributions of vocabulary and potential topics, respectively, which is beneficial for projecting the document into the topic layer on a large scale. Figure 2 shows the process.

In Figure 1, $\theta$ and $Z$ are implicit variables, and $W$ denotes the observable vocabulary variable. A document set $D = \{d_1, d_2, \ldots, d_m\}$ covers $m$ documents and each document $d$ contains $n$ words. The document word set is $d = \{w_1, w_2, \ldots, w_n\}$. As assumed, $D$ is extracted into $T$ topics and $T_k$ represents the $k$-th potential topic. The specific process is introduced as follows [43]:

(1) The Dirichlet distribution with a $\beta$ parameter is calculated for $T$-dimensional vector $Z$, $Z \sim \text{Dirichlet}(\beta)$.

(2) The Dirichlet distribution with a $\alpha$ parameter is derived for $T$-dimensional vector $\theta$, $Z \sim \text{Dirichlet}(\alpha)$.

(3) With respect to vocabulary $w$ in the document: the multinominal distribution with a $\theta$ parameter is generated for topic $z$, and $P(w|z,\ \beta)$ is obtained for the vocabulary in accordance with the $\beta$ parameter:

$$P(w|z,\ \beta) = \int p(\theta|\alpha)\left(\prod_{i=1}^{n}\sum_{k=1}^{T} P(z_k|\theta)P(w_i|z_k,\ \beta)\right)\mathrm{d}\theta, \tag{1}$$

where $\theta$ and $Z$ are implicit variables and $W$ denotes the observation variable. The LDA model calculates the posterior and conditional distributions of implicit variables under the given observable document.

*3.3. Bimodal Network.* A bimodal network is formed by two types of heterogeneous nodes through certain connections. The dataset in the bimodal network is the entity that represents different types by rows and columns and is mainly utilized to describe the "binary" structural relationship between a group of actors and the active events. In this study, a bimodal network that uses provincial and policy topics as the heterogeneous nodes was constructed by extracting relevant topics from all IP policy texts of provincial government departments through LDA. Subsequently, the goal execution deviation of the national IPS of provincial regions was reflected by the structural characteristics of the bimodal network.

In the bimodal network, the two types of heterogeneous nodes are actors and events [44]. In this study, actors and events represent the provinces and policy topics of IP, respectively. The characteristic indexes of the bimodal network are as follows.

*3.3.1. The Characteristics of the Bimodal Network Based on Centrality.* The node centrality of the bimodal network is a quantitative analysis of individual power, which explores the positional characteristics of all nodes in the network from the perspective of microscopic nodes [45]. This model

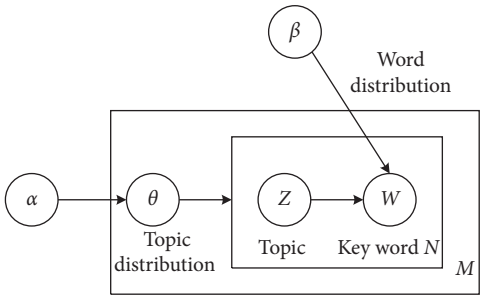



Figure 2: Principle of the LDA topic model.

mainly includes degree, closeness, and betweenness centralities.

*(1) Degree Centrality*. Degree centrality refers to the degree of connections between the actor and other actors in the network. The actor with more connections has a stronger influence than that with fewer connections [45]. The degree centrality of provincial nodes is the number of topics concerned by the provincial nodes, whereas the degree centrality of topic nodes refers to the number of involved provinces. If $g$ provinces form the actor set $n = (n_1, n_2, \ldots, n_g)$ and $h$ topics form the event set $z = (z_1, z_2, \ldots, z_h)$, then the degree centralities of province $i$ and topic $k$ can be expressed as

$$C_D^{\mathrm{NZ}}(n_i) = \sum_{k=1}^{g+h} x_{ik}^{\mathrm{NZ}}, \quad (i = 1, 2, \ldots, g;\ k = 1, 2, \ldots, h), \tag{2}$$

$$C_D^{\mathrm{NZ}}(z_k) = \sum_{i=1}^{g+h} x_{ik}^{\mathrm{NZ}}, \quad (i = 1, 2, \ldots, g; k = 1, 2, \ldots, h), \tag{3}$$

where NZ is the bimodal network comprising $g$ provinces and $h$ topics. $x_{ik}^{\mathrm{NZ}}$ denotes the affiliation relationship data of the bimodal network of $g \times h$ bipartite matrix $x^{\mathrm{NZ}}$.

*(2) Closeness Centrality*. Closeness centrality is applied to represent the time for information diffusing from one node to another [44]. Different from closeness centrality in the single-mode network, the closeness centrality of actors in the bimodal network is proportional to the sum of the distance between one node and other nodes in the network and between the node and all event nodes. The closeness centrality of events is similar. The calculation formulas for the closeness centralities of actors and events are as follows:

$$C_C^{\mathrm{NZ}}(n_i) = \left[1 + \frac{\sum_{j=1}^{g+h} \min k\, d(k, j)}{g + h - 1}\right]^{-1}, (k = 1, 2, \ldots, h), \tag{4}$$

$$C_C^{\mathrm{NZ}}(m_k) = \left[1 + \frac{\sum_{i=1}^{g+h} \min j\, d(i, j)}{g + h - 1}\right]^{-1}, (i = 1, 2, \ldots, g). \tag{5}$$

In equation (3), connections exist between event point $k$ and actor $i$. In equation (4), actor $j$ and event node $k$ are connected. $d(k, j)$ indicates the distance between event node $k$ and node $j$ in the bimodal network. Here, $j$ can be used to represent the event or actor node.

*(3) Betweenness Centrality*. Betweenness centrality represents the positional importance of nodes in the network. In the bimodal network, connections between every two actor nodes must run through all event nodes concerning them [45]. Therefore, the event nodes are in the shortest route between actors. Similarly, actor nodes are in the shortest route between event nodes. For this reason, calculating the betweenness centrality of event nodes in the bimodal network must consider all actor nodes concerning the event node. The calculation formulas of the betweenness centrality of actor and event nodes are as follows:

$$C_B^{\mathrm{NZ}}(n_i) = \frac{1}{2} \sum_{m_k, m_l \in n_i}^{g+h} \frac{1}{x_{kl}^M}, \quad (i = 1, 2, \ldots, g; k = 1, 2, \ldots, h), \tag{6}$$

$$C_B^{\mathrm{NZ}}(m_k) = \frac{1}{2} \sum_{n_i, n_j \in m_k} \frac{1}{x_{ij}^N}, \quad (i = 1, 2, \ldots, g; k = 1, 2, \ldots, h). \tag{7}$$

In equation (7), actors $n_i$ and $n_j$ concern a total of $x_{ij}^N$ events. For any pair of actors $(n_i, n_j)$ in $m_k$, the betweenness centrality contribution of $m_k$ is $1/x_{ij}^N$ units. $x^N$ and $x^M$ represent the data matrixes of single nodes of actors and events, respectively.

*3.3.2. The Characteristics of the Bimodal Network Based on Singular Value Decomposition (SVD).* Singular value analysis is used to reduce the dimension of the bimodal network. The hierarchical classification of the network structure is conducted, and the classification factors hidden in the bimodal network are recognized. The network characteristics are close to the overall structure of the bimodal network by analyzing the hidden classification factors. The calculation principle of the SVD method is as follows [44].

An $n \times m$ information matrix of the bimodal network $X$ $(n \geq m)$ is available. Matrixes $U$, $D$, and $V$ are simultaneously obtained through SVD. $X = U \times D \times V'$, where $V'$ is the transposed matrix of $V$ and $D$ denotes the $r \times r$ diagonal matrix and has $r$ singular values. Matrixes $U$ and $V$ are $n \times r$ and $m \times r$ matrixes, which cover the $r$ eigenvector of matrix $XX'$. The corresponding eigenvalues are arranged in descending order. A real matrix $A$ with an order of $r$ can be decomposed as follows:

$$A = U \begin{bmatrix} \begin{bmatrix} a_1 & & & \\ & a_2 & & \\ & & \cdots & \\ & & & a_r \end{bmatrix} & 0 \\ 0 & 0 \end{bmatrix} V'. \tag{8}$$

Equation (8) is defined as the SVD of matrix $A$. In equation (8), matrixes $U$ and $V\prime$ are orthogonal matrixes and

$a_i > 0\,(i = 1, 2, \ldots, r)$, where $a_1, a_2, \ldots, a_r$ is the $A$ singular value of the matrix. Dimensions of topics in IP policy texts are reduced by the bimodal network constructed through SVD. Topics are classified into factors with common characteristics, which is conducive to analysis loads of provincial topics on different factors. Hence, the preference of different provinces in knowledge property strategy is evaluated from the perspective of the overall network.

## 4. Empirical Results and Discussions

*4.1. Topic Extraction Based on LDA.* LDA modeling is performed through Gibbs sampling. Topics are analyzed by using Gensim source packages based on the *Python* software. Moreover, 4–20 topics and 1000 iterations are set, thereby obtaining the document-topic and word-topic probability matrixes according to formula (1). Topic clustering data under different numbers of topics are compared repeatedly. Finally, nine topics are selected in accordance with the nine task goals in *Outline of National Intellectual Property Strategy*. Table 1 lists the 10 major characteristic words of each topic.

Table 1 shows that nine subjects are selected on the basis of the LDA topic model, namely, IP culture, IP intermediary services, IP transformation and applications, IP legal system, IP administration management, IP overseas cooperation, IP talents, IP law enforcement, and IP creation. These topics are the key strategic goals proposed in the *Outline of National Intellectual Property Strategy*. The provincial topics are classified in accordance with previous marks of provinces, thereby further exploring the frequency distributions of each provincial policy text in the nine topics. Then, the number of topics involved in all policy texts of each province is calculated, which provides a data basis for the subsequent construction of the bimodal network matrix and explores contents hidden in the province-topic information matrix.

*4.2. Goal Execution Deviation Diagnosis of IPS Based on the Bimodal Network.* The bimodal data are visualized automatically by using Ucinet 6.0 software on the basis of the data matrix of the constructed bimodal network. Figure 2 presents the results.

Figure 3 intuitively displays the province-topic binary affiliation relationship. On the basis of the size of the topic center point, IP transformation and applications, IP administration management, IP overseas cooperation, and IP legal system are topics with the highest concern frequency of the provincial region, whereas IP culture and IP intermediary services are topics of less concern. This finding indicates the general emphasis of the provincial region to implement IPS.

Furthermore, the bimodal network forms the single-property concurrence network from the perspectives of topics and provinces to explore the single-property concurrence connotations (Figure 3).

The single-property concurrence network of topics in Figure 4(a) reflects two or more topics simultaneously concerned by provincial IP policy. On the basis of the thick connections of different topics, IP administration management, IP legal system, IP overseas cooperation, and IP creation are highly concerned by provincial regions in IPS implementation. The single-property concurrence network of topics in Figure 4(b) reflects the number of topics concerned by more than two provinces. Similarly, numerous topics are simultaneously concerned by Shanghai, Beijing, Tianjin, and Shanxi, but only a few topics are concerned by Hebei, Guizhou, and Ningxia.

The characteristics of nodes are further generated by using Ucinet 6.0. In Table 2, great hierarchical differences are observed among 30 provinces in terms of degree and closeness centralities. Beijing, Tianjin, Shanghai, Chongqing, Sichuan, Gansu, Qinghai, Hunan, and Guangdong rank the top positions in view of the degree centrality of provincial nodes. Hence, these provinces are concerned with extensive and comprehensive scopes of IP policy topics. The content scope and execution speed of supporting policies in these provinces are inferior to those of other provinces. With respect to closeness centrality, Beijing, Tianjin, Shanghai, Chongqing, and Sichuan occupy the first layer of the network, indicating their quick responses to the topics in national IP policies. This result implies that these provinces can formulate supporting policies in accordance with local situations immediately after the release of the national IPS. Nevertheless, all selected provinces show relatively similar betweenness centrality, indicating the unsubstantial performance as the bridge for information diffusion and transportation. This finding conforms to the actual situation of the bimodal network of policies. Data sources of the policy network of different provinces are mainly contributed by the number of concerns of topic information. The characteristics do not highlight the social affiliation. Hence, using provincial nodes as the bridge in the single-mode network is difficult.

In addition, the microstructural characteristics of topic nodes in Table 3 reflect that the three types of microscopic characteristics show a consistent variation trend. IP transformation and applications, IP legal system, IP administration management, and IP overseas cooperation show the highest degree centrality. This result shows that provincial region's supporting policies provide the highest concerns on the five topics. Topics ranking the top of closeness centrality also belong to the five topics, which confirms that these topics are easy to mention and emphasize through local supporting policies. This finding is closely related to the subsequent successive release of details and directed optimal adjustment of the central government. Similarly, the five topics have the highest betweenness centrality, indicating the prominent universal effect among IP transformation and applications, IP legal system, IP administration management, and IP overseas cooperation. The five topics also serve as the "bridge" in the topic node network. When one or several provincial supporting policies concern these topics, it can be easily "followed" by other provinces. The "followers" may pay great attention to the related supporting policy contents of other provinces to a large extent when they formulate supporting policies. On the contrary, microstructural data of talents, IP law enforcement, IP creation, IP

Table 1: Distribution of topics-characteristic words.

| Topics | IP culture | IP intermediary services | IP transformation and applications | IP legal system |
|---|---|---|---|---|
| Characteristic words | Culture<br>Patents<br>Trademark<br>Request<br>Review<br>Service<br>Talents<br>Intellectual property<br>Rewards | Intellectual property<br>Comments<br>Rights protection<br>Aids<br>Service<br>Patents<br>Pledge<br>Copyright<br>Enterprises | Intellectual property<br>Patents<br>Cultivation<br>Scientific and technological achievements<br>Scientific technology<br>Pilot project<br>Industry<br>Construction<br>Demonstrations | Intellectual property<br>Cases<br>Administrative enforcement<br>Rights protection<br>Complaints<br>Open<br>Law enforcement<br>Infringement<br>Intellectual property protection |
| IP administration management | IP overseas cooperation | IP talents | IP law enforcement | IP creation |
| Methods | Projects | Intellectual property | Intellectual property | Patents |
| Standards | Patent technology | Popularization | Mediation | Patents |
| Law enforcement | Industry | Talents | Law enforcement | Innovation |
| Reward | Financial aid | Cultivation | Copyright | Patent technology |
| Confirmation | Overseas | Scientific technology | Administration and law enforcement | Products |
| Talent | Progress | Construction | Intellectual property protection | Inventor |
| Patents | Autonomous intellectual property | Innovation | Impact | Design |
| Intellectual property | Intellectual property office | Applications | Administration department | Knowledge property |
| Financial aid | Strategy | Engineering | Intellectual property office | Scientific and technologic type |

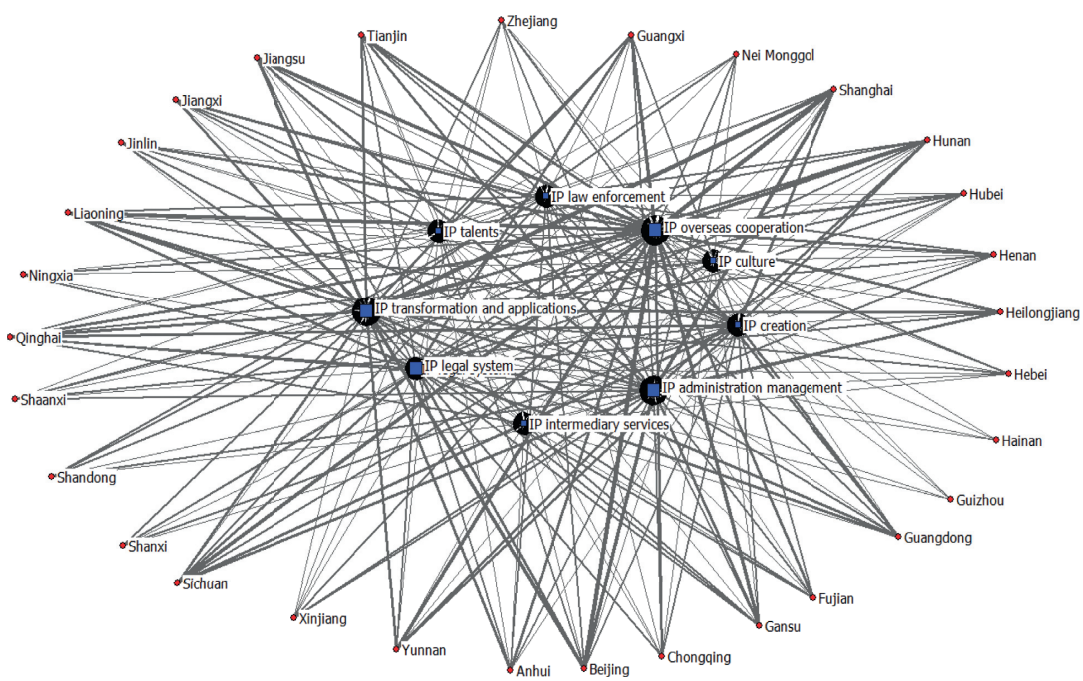



Figure 3: Visualized diagram of the province-topic bimodal network.

culture, and IP intermediary services are relatively small, especially the microstructural data of culture and intermediary services. The degree, closeness, and betweenness centralities of IP culture and IP intermediary services are lower than those of the seven remaining topics, which reflects less attention by provincial regions on these two topics. Moreover, the optimization efficiency and performance quantification of the two topics are considerably poorer than those of superior IP transformation and applications. This finding explains the weak positions of IP

Table 2: Characteristics of provincial nodes.

| Province | Degree centrality | Closeness centrality | Betweenness centrality |
|---|---|---|---|
| Beijing | 0.231 | 1.716 | 0.001 |
| Tianjin | 0.231 | 1.716 | 0.001 |
| Hebei | 0.179 | 1.62 | 0.001 |
| Shanxi | 0.154 | 1.575 | 0 |
| Nei Mongol | 0.154 | 1.575 | 0.001 |
| Heilongjiang | 0.179 | 1.62 | 0.001 |
| Jinlin | 0.205 | 1.667 | 0.001 |
| Liaoning | 0.205 | 1.667 | 0.001 |
| Shanghai | 0.231 | 1.716 | 0.001 |
| Jiangsu | 0.205 | 1.667 | 0.001 |
| Chongqing | 0.231 | 1.716 | 0.001 |
| Sichuan | 0.231 | 1.716 | 0.001 |
| Yunnan | 0.205 | 1.667 | 0.001 |
| Shaanxi | 0.179 | 1.62 | 0.001 |
| Gansu | 0.231 | 1.716 | 0.001 |
| Qinghai | 0.231 | 1.716 | 0.001 |
| Ningxia | 0.128 | 1.533 | 0 |
| Xinjiang | 0.179 | 1.62 | 0.001 |
| Hainan | 0.103 | 1.494 | 0 |
| Guizhou | 0.128 | 1.533 | 0 |
| Zhejiang | 0.205 | 1.667 | 0.001 |
| Anhui | 0.205 | 1.667 | 0.001 |
| Fujian | 0.205 | 1.667 | 0.001 |
| Jiangxi | 0.205 | 1.667 | 0.001 |
| Shandong | 0.154 | 1.575 | 0 |
| Henan | 0.205 | 1.667 | 0.001 |
| Hubei | 0.205 | 1.667 | 0.001 |
| Hunan | 0.231 | 1.716 | 0.001 |
| Guangdong | 0.231 | 1.716 | 0.001 |
| Guangxi | 0.205 | 1.667 | 0.001 |

Table 3: Characteristics of topic nodes.

| Topic | Degree centrality | Closeness centrality | Betweenness centrality |
|---|---|---|---|
| IP culture | 0.462 | 1.643 | 0.009 |
| IP intermediary services | 0.513 | 1.742 | 0.009 |
| IP transformation and applications | 0.744 | 2.396 | 0.024 |
| IP legal system | 0.718 | 2.3 | 0.022 |
| IP administration management | 0.744 | 2.396 | 0.024 |
| IP overseas cooperation | 0.769 | 2.5 | 0.027 |
| IP talents | 0.615 | 1.983 | 0.015 |
| IP law enforcement | 0.615 | 1.983 | 0.015 |
| IP creation | 0.692 | 2.212 | 0.02 |

culture and IP intermediary services in the supporting policies of provincial regions for IPS.

The overall features of the province-topic bimodal network are further explored through the SVD, which is used for the hierarchical classification of topics and provinces. The interpretation capability of the first three factors accounted for 76.5%, which is calculated by using Ucinet 6.0, indicating that these factors can interpret the macroscopic relation characteristics between provinces and topics. In addition, the SVD data of topic nodes are obtained, through which the topic factors are calculated as shown in Table 4.

Table 3 shows that characteristic factor 1 is a public factor, which indicates that the topic of characteristic factor 1 is used as the public and universal major content in provincial knowledge property policies. Data in Table 4 reflect that IP overseas cooperation and IP transformation and applications are public and universal topics highly concerned by local provinces in formulating the supporting IP policies. For factors 2 and 3, topics with high values in different factors are clustered into one type. In factor 2, IP culture, IP intermediary services, IP law enforcement, and IP creation are important topics and can be classified into the IP development environment. IP transformation and

Table 4: SVD data of topic nodes.

| | Factor 1 | Factor 2 | Factor 3 |
|---|---|---|---|
| IP culture | 0.052 | 0.153 | −0.14 |
| IP intermediary services | 0.077 | 0.189 | −0.181 |
| IP transformation and applications | 0.473 | −0.305 | 0.27 |
| IP legal system | 0.302 | 0.354 | 0.443 |
| IP administration management | 0.393 | −0.575 | −0.52 |
| IP overseas cooperation | 0.644 | 0.074 | 0.152 |
| IP talents | 0.102 | 0.085 | 0.206 |
| IP law enforcement | 0.129 | 0.092 | 0.072 |
| IP creation | 0.283 | 0.609 | −0.58 |

applications, IP legal system, IP administration management, IP overseas cooperation, and IP talents are five topics belonging to factor 3 and are classified as the IP operation system.

On the basis of the classification of factors 2 and 3, SVD data of nodes in provinces were investigated to classify the policy topics issued by each province accurately (Table 5).

Table 5 exhibits that the supporting IP policies of Shanxi, Chongqing, Sichuan, Yunnan, Shaanxi, Gansu, Qinghai, Xinjiang, Henan, Fujian, Hunan, and Guangdong pay further attention to factor 2, that is, the IP development environment. On the contrary, other provinces pay further attention to IP operation system. With regard to loads of public factors, Beijing, Tianjin, Heilongjiang, Jilin, Shanghai, and Jiangsu rely on public factors. The supporting IP policies in these provinces attach further attention to IP overseas cooperation and IP transformation and applications.

For a further intuitive display of distributions of the three factors in different provinces, the relation visualization diagram between two eigenvectors comprising factor numerical values is illustrated to describe the eigenvalue relationship among different factors (Figure 5).

In Figure 4, the $x$-axis is factor 1 and the $y$-axis represents factors 2 and 3. In factor 1, IP overseas cooperation, IP transformation and applications, and IP administration management are prominent topics and have the largest contribution to public factors. IP culture and IP intermediary services are easy to cluster (Figure 5(a)). Shanxi, Chongqing, and Guangdong are regions in the same dimensions. IP transformation and applications and IP legal system are easy to cluster (Figure 5(b)), which involves Beijing, Tianjin, and Hebei. The factor clusters of provinces reflected by relationships among eigenvalues of different factors agree with the province-topic clustering results of the SVD.

## 5. Conclusions and Management Implications

*5.1. Conclusions.* In this study, topics in IP policy texts of the provincial region were extracted by using the LDA model, and the goal execution deviation of the IPS of the provincial region was diagnosed by the bimodal network. The following conclusions can be drawn.

First, in the process of implementing national intellectual property strategic goals, provincial regions have paid much attention to topics such as IP transformation and applications, IP administration, IP overseas cooperation, IP creation, and IP legal system, whereas they have paid very little attention to topics such as IP culture, IP intermediary services, IP talents, and IP law enforcement. These overlooked topics present the provincial regions' execution deviation.

Second, when it comes to establishing the supporting policies, Beijing, Tianjin, Shanghai, Chongqing, Sichuan, Gansu, Qinghai, Hunan, and Guangdong provinces played the role as forerunners, while other provinces played the role as followers. It is also indicated that the topics that these forerunner governments have concerned about are more timely and comprehensive.

Third, the topics of IP transformation and applications, IP administration, IP overseas cooperation, IP creation, and IP legal system played the roles as "bridge" in the network. When the supporting policies of one or several provinces have paid attention to the above five topics, it is more likely to cause other provinces to focus on these five topics as well when establishing supporting policies. In contrast, the topics of IP talents, IP law enforcement, IP creation, IP culture, and IP intermediary services are at the edge of the network.

*5.2. Management Implications.* Based on the above findings, we offered the following management implications to correct provincial regions' IP strategy execution deviation.

First, when provincial regions implement the IP strategic goals, it is better to refer to the nine major goals in the *Outline of National Intellectual Property Strategy*. Provincial regions should make effort to balance the nine goals of IP culture, IP intermediary services, IP transformation and applications, IP legal system, IP administration, IP overseas cooperation, IP talents, and IP law enforcement in order to avoid execution deviation.

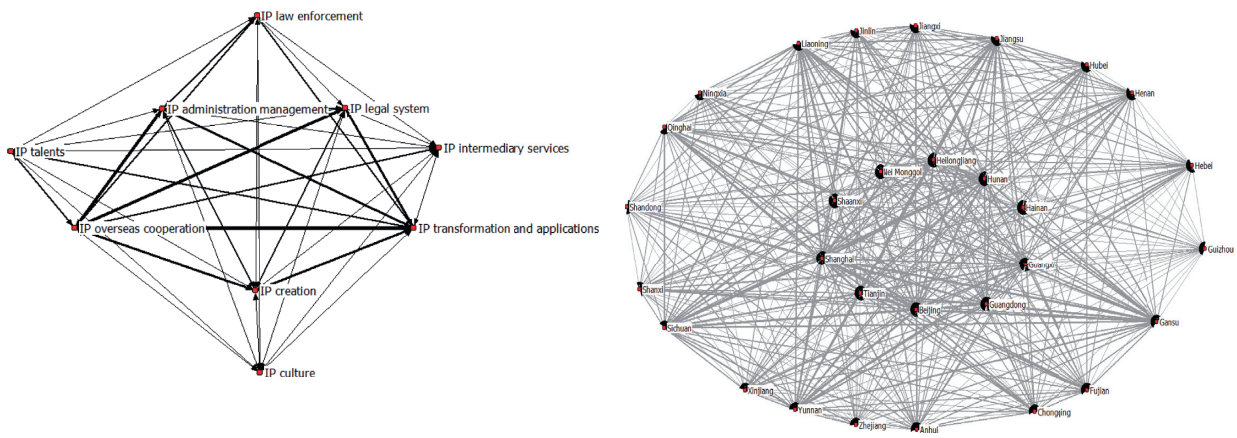
Second, provincial regions should learn rationally rather than blindly accepting and following other provinces. It is suggested that provincial regions can propose the corresponding policies by taking into consideration their regional and developmental phase characteristics.

Third, provincial regions should cooperate more often when implementing policies. They should learn from each other's experience and consider their own logical path of implementing the nine strategic goals. Therefore, the IP strategy execution deviation or omission can be prevented as much as possible. It is not only about improving the patent quality but also about commercial efficiency. Provincial regions should not only strengthen the legal system and law enforcement but also improve the environment and culture for IP creation and operation.

The major contributions of this paper are discussed as follows.

First, the combination of LDA topic mining and bimodal network can offset the shortcomings of traditional models that cannot diagnose nonstructured policy texts. This condition can reflect the goal deviation among different execution stages of the national strategic planning.

Second, the relation network of heterogeneous nodes, which cannot be constructed in the single-mode network,




(a) (b)

Figure 4: Single-property concurrence network generated by the bimodal network. (a) Single-property concurrence network of topics. (b) Single-property concurrence network of provinces.

Table 5: SVD data of nodes in provinces.

| Province | Factor 1 | Factor 2 | Factor 3 |
|---|---|---|---|
| Beijing | 0.338 | −0.302 | 0.078 |
| Tianjin | 0.204 | −0.068 | −0.053 |
| Hebei | 0.085 | −0.045 | −0.001 |
| Shanxi | 0.086 | 0.081 | 0.029 |
| Nei Mongol | 0.073 | −0.107 | 0.115 |
| Heilongjiang | 0.207 | −0.352 | −0.071 |
| Jinlin | 0.126 | −0.244 | −0.047 |
| Liaoning | 0.261 | −0.151 | 0.167 |
| Shanghai | 0.353 | −0.286 | −0.129 |
| Jiangsu | 0.245 | −0.138 | 0.116 |
| Chongqing | 0.15 | 0.036 | −0.154 |
| Sichuan | 0.246 | 0.334 | −0.227 |
| Yunnan | 0.236 | 0.014 | −0.223 |
| Shaanxi | 0.062 | 0.078 | −0.153 |
| Gansu | 0.234 | 0.121 | −0.184 |
| Qinghai | 0.191 | 0.467 | −0.208 |
| Ningxia | 0.065 | 0.117 | 0.217 |
| Xinjiang | 0.091 | 0.039 | −0.345 |
| Hainan | 0.026 | −0.005 | −0.118 |
| Guizhou | 0.022 | 0.007 | 0.089 |
| Zhejiang | 0.07 | −0.061 | −0.049 |
| Anhui | 0.167 | −0.042 | 0.25 |
| Fujian | 0.156 | 0.086 | −0.068 |
| Jiangxi | 0.167 | 0.202 | 0.337 |
| Shandong | 0.092 | −0.057 | −0.038 |
| Henan | 0.119 | −0.017 | 0.142 |
| Hubei | 0.116 | 0.071 | 0.269 |
| Hunan | 0.242 | 0.131 | −0.068 |
| Guangdong | 0.212 | 0.22 | 0.013 |
| Guangxi | 0.204 | 0.287 | 0.466 |

can be realized on the basis of the LDA topic mining and bimodal network model. This relation network can reflect the roles of provincial administration regions and policy topics in the network, diagnose the accuracy of goal execution deviations, and accurately recognize key topics.

Third, "concurrence" and "marginality" of single-property nodes of policy topics can be measured accurately on the basis of the LDA topic mining and bimodal network model. The goal execution deviation is diagnosed objectively and comprehensively during the implementation of the

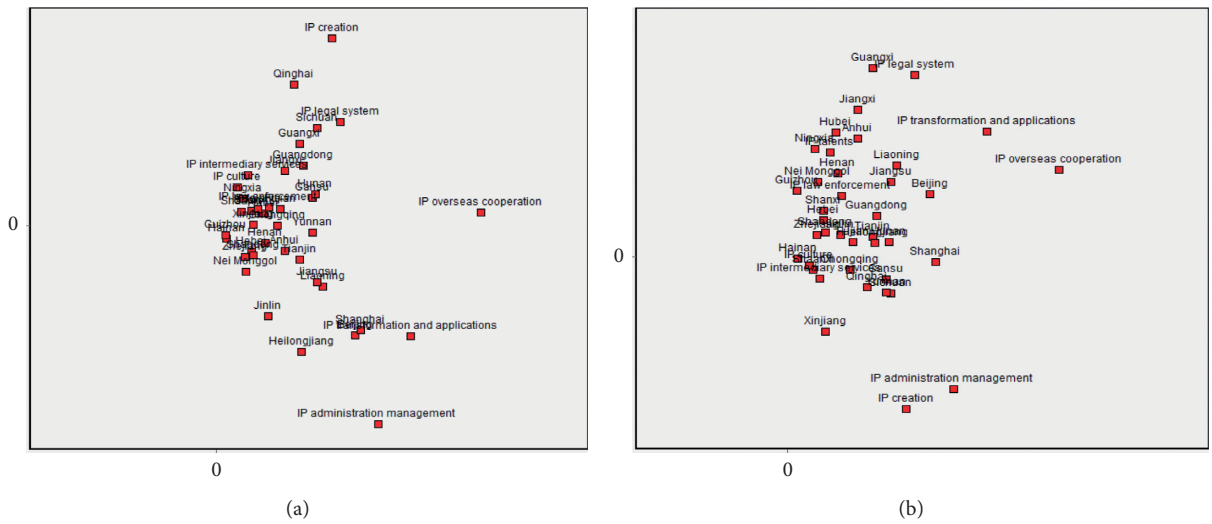



Figure 5: Analysis of singular values. (a) Relationship between factors 1 and 2. (b) Relationship between factors 1 and 3.

strategic plan by the provincial region. A simple and reliable method is offered.

The goal execution deviation for the Chinese national IPS is diagnosed by policy text information, and meaningful results are obtained. This study can further extend the research contents in the following aspects: (1) Popular topics of IP policies are analyzed by combining heterogeneous information on network media, forums, social software, and instant music video, which provides a further extensive basis for the goal execution deviation diagnosis. (2) The occurrence mechanism of deviation is comprehensively analyzed, which is beneficial for constructing the deviation correction mechanism.

## Data Availability

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

## Conflicts of Interest

There are no conflicts of interest.

## Acknowledgments

This research was supported by the Natural Science Foundation of China (Grant no. 71704007), the Beijing Social Science Foundation of China (Grant no. 18GLC082), and the Beijing Financial Foundation of China (Grant no. PXM2020-178216-000001).

## References

[1] T. Dong, "On the relationship between the NIPS and Chinese economic development," *Studies in Science of Science*, vol. 5, pp. 641–652, 2009.

[2] H. F. Song, Y. Li, and W. Qu, "The intellectual property capacity measurement indicator system and method based on equivalent relationship of kinds of intellectual property and kinds of research institute," *Studies in Science of Science*, vol. 12, pp. 1826–1834+1825, 2013.

[3] G.-C. Yang, G. Li, C.-Y. Li et al., "Using the comprehensive patent citation network (CPC) to evaluate patent value," *Scientometrics*, vol. 105, no. 3, pp. 1319–1346, 2015.

[4] Y. T. Sun and Q. Luan, "A three stage-two dimensions' model of patent quality measuring and its empirical study: a case study of C9 league," *Science of Science and Management of S.&T.*vol. 6, pp. 23–32, 2016.

[5] B. Wang and C.-H. Hsieh, "Measuring the value of patents with fuzzy multiple criteria decision making: insight into the practices of the Industrial Technology Research Institute," *Technological Forecasting and Social Change*, vol. 92, pp. 263–275, 2015.

[6] A. Flynn, "Investigating the implementation of SME-friendly policy in public procurement," *Policy Studies*, vol. 39, no. 4, pp. 422–443, 2018.

[7] L. Iain and B. B. Bella, "Distinctive policy diffusion patterns, processes and actors: drawing implications from the case of sport in international development," *Policy Studies*, vol. 39, no. 4, pp. 444–464, 2018.

[8] S. Tamura, "Effects of integrating patents and standards on intellectual property management and corporate innovativeness in Japanese electric machine corporations," *International Journal of Technology Management*, vol. 59, no. 3-4, pp. 180–202, 2016.

[9] P. Alfons, H. J. Paul, and W. Elizabeth, "The effect of patents on trade," *Journal of International Economics*, vol. 105, pp. 1–9, 2017.

[10] C. Mercedes and D. Marco, "Intellectual property rights, trade agreements, and international trade," *Research Policy*, vol. 48, no. 3, pp. 531–545, 2019.

[11] M. Lee, J. D. Alba, and D. Park, "Intellectual property rights, informal economy, and FDI into developing countries," *Journal of Policy Modeling*, vol. 40, no. 5, pp. 1067–1081, 2018.

[12] E. Billette de Villemeur, R. Ruble, and B. Versaevel, "Dynamic competition and intellectual property rights in a model of product development," *Journal of Economic Dynamics and Control*, vol. 100, pp. 270–296, 2019.

[13] B. Julia, C. Paolo, M. Lukas, and B. Kilian, "Intellectual property rights hinder sequential innovation. Experimental

evidence," *Research Policy*, vol. 45, no. 10, pp. 2054–2068, 2016.

[14] A. C. Chu, H. Fan, G. Shen, and X. Zhang, "Effects of international trade and intellectual property rights on innovation in China," *Journal of Macroeconomics*, vol. 57, pp. 110–121, 2018.

[15] M. Grimaldi, M. Greco, and L. Cricelli, "A framework of intellectual property protection strategies and open innovation," *Journal of Business Research*, vol. 123, pp. 156–164, 2021.

[16] Z. Liu, R. Mu, S. Hu, L. Wang, and S. Wang, "Intellectual property protection, technological innovation and enterprise value-An empirical study on panel data of 80 advanced manufacturing SMEs," *Cognitive Systems Research*, vol. 52, pp. 741–746, 2018.

[17] R. Smeets and A. de Vaal, "Intellectual property rights and the productivity effects of MNE affiliates on host-country firms," *International Business Review*, vol. 25, no. 1, pp. 419–434, 2016.

[18] B. Kim, E. Kim, D. J. Miller, and J. T. Mahoney, "The impact of the timing of patents on innovation performance," *Research Policy*, vol. 45, no. 4, pp. 914–928, 2016.

[19] I. Nemlioglu, "A comparative analysis of intellectual property rights: a case of developed versus developing countries," *Procedia Computer Science*, vol. 158, pp. 988–998, 2019.

[20] L. Xue and Z. L. Lin, "Three perspectives to understand public policy process and their implications for China study," *Chinese Public Administration*, vol. 5, pp. 41–46, 2013.

[21] G. Alison and L. Bob, "NAPLAN data: a new policy assemblage and mode of governance in Australian schooling," *Policy Studies*, vol. 37, no. 6, pp. 568–582, 2016.

[22] E. Kurilovas, "On data-driven decision-making for quality education," *Computers in Human Behavior*, vol. 107, Article ID 105774, 2020.

[23] D. Lee, M. Kim, and J. Lee, "Adoption of green electricity policies: investigating the role of environmental attitudes via big data-driven search-queries," *Energy Policy*, vol. 90, pp. 187–201, 2016.

[24] X. Fu, G. Li, X. Zhang, and Z. Qiao, "Failure probability estimation of the gas supply using a data-driven model in an integrated energy system," *Applied Energy*, vol. 232, pp. 704–714, 2018.

[25] S. Zeng, X. Peng, T. Baležentis, and D. Streimikiene, "Prioritization of low-carbon suppliers based on Pythagorean fuzzy group decision making with self-confidence level," *Economic Research-Ekonomska Istraživanja*, vol. 32, no. 1, pp. 1073–1087, 2019.

[26] Z. L. Yang, H. Garg, J. Li, G. Srivastavad, and Z. Cao, "Investigation of multiple heterogeneous relationships using a q-rung orthopair fuzzy multi-criteria decision algorithm," *Neural Computing and Applications*, pp. 1–22, 2020.

[27] B. Wang, C. Wu, L. Huang, and L. Kang, "Using data-driven safety decision-making to realize smart safety management in the era of big data: a theoretical perspective on basic questions and their answers," *Journal of Cleaner Production*, vol. 210, pp. 1595–1604, 2019.

[28] Z. Yang, J. Roth, and R. K. Jain, "DUE-B: data-driven urban energy benchmarking of buildings using recursive partitioning and stochastic frontier analysis," *Energy and Buildings*, vol. 163, pp. 58–69, 2018.

[29] T. Ahmad, H. Chen, Y. Guo, and J. Wang, "A comprehensive overview on the data driven and large scale based approaches for forecasting of building energy demand: a review," *Energy and Buildings*, vol. 165, pp. 301–320, 2018.

[30] S. Burris, L. Hitchcock, J. Ibrahim, M. Penn, and T. Ramanathan, "Policy surveillance: a vital public health practice comes of age: table 1," *Journal of Health Politics, Policy and Law*, vol. 41, no. 6, pp. 1151–1173, 2016.

[31] S. Zeng, D. Luo, C. Zhang, and X. Li, "A correlation-based TOPSIS method for multiple attribute decision making with single-valued neutrosophic information," *International Journal of Information Technology & Decision Making*, vol. 19, no. 01, pp. 343–358, 2020.

[32] Z. Yang, X. Li, H. Garg, and M. Qi, "Decision support algorithm for selecting an antivirus mask over COVID-19 pandemic under spherical normal fuzzy environment," *International Journal of Environmental Research and Public Health*, vol. 17, no. 10, p. 3407, 2020.

[33] N. Zhang, "Analyzing public generated big data and restructuring government decision Making process: review and prospect," *Chinese Public Administration*, vol. 10, pp. 19–24, 2015.

[34] A. Clarke and H. Margetts, "Governments and citizens getting to know each other open, closed, and big data in public management reform," *Policy & Internet*, vol. 6, no. 4, pp. 393–417, 2015.

[35] F. H. Zheng, L. Deng, and J. T. Lin, "Why the performance objectives of subsidy fiscal policy will be aliasing? — based on the third-party performance evaluation of the special fiscal funds from Guangdong province," *Journal of Public Management*, vol. 3, pp. 122–134+159, 2016.

[36] R. Michels, *Political Party*, Free Press, New York, NY, USA, 1968.

[37] W. G. Resh and J. D. Marvel, "Loopholes to load-shed: contract management capacity, representative bureaucracy, and goal displacement in federal procurement decisions," *International Public Management Journal*, vol. 15, no. 4, pp. 525–547, 2012.

[38] H. Liu, J. N. Wu, and M. M. Xu, "A review of goal deviation and its influence factors from different theoretical perspectives," *Journal of Public Administration*, vol. 1, pp. 151–171+186, 2016.

[39] J. Liang and L. Langbein, "Performance management, high-powered incentives, and environmental policies in China," *International Public Management Journal*, vol. 18, no. 3, pp. 346–385, 2015.

[40] S. Zeng, S.-M. Chen, and K.-Y. Fan, "Interval-valued intuitionistic fuzzy multiple attribute decision making based on nonlinear programming methodology and TOPSIS method," *Information Sciences*, vol. 506, pp. 424–442, 2020.

[41] Z. Yang, T. Ouyang, X. Fu, and X. Peng, "A decision–making algorithm for online shopping using deep–learning-based opinion pairs mining and q –rung orthopair fuzzy interaction Heronian mean operators," *International Journal of Intelligent Systems*, vol. 35, no. 5, pp. 783–825, 2020.

[42] D. Luo, S. Zeng, and J. Chen, "A probabilistic linguistic multiple attribute decision making based on a new correlation coefficient method and its application in hospital assessment," *Mathematics*, vol. 8, no. 3, p. 340, 2020.

[43] M. D. Blei, Y. A. Ng, and I. M. Jordan, "Latent dirichlet allocation," *Journal of Machine Learning Research*, vol. 3, pp. 993–1022, 2003.

[44] L. C. Freeman, "Centrality in social networks conceptual clarification," *Social Networks*, vol. 1, no. 3, pp. 215–239, 1978.

[45] K. Faust, "Centrality in affiliation networks," *Social Networks*, vol. 19, no. 2, pp. 157–191, 1997.

Hindawi
Mathematical Problems in Engineering
Volume 2020, Article ID 5920432, 25 pages
https://doi.org/10.1155/2020/5920432

*Research Article*

# Jaccard and Dice Similarity Measures Based on Novel Complex Dual Hesitant Fuzzy Sets and Their Applications

**Tahir Mahmood,[1] Ubaid Ur Rehman,[1] Zeeshan Ali,[1] and Ronnason Chinram[2,3]**

[1]*Department of Mathematics and Statistics, International Islamic University, Islamabad, Pakistan*
[2]*Algebra and Applications Research Unit, Division of Computational Science, Faculty of Science, Prince of Songkla University, Hat Yai, Songkhla 90110, Thailand*
[3]*Centre of Excellence in Mathematics, Si Ayuthaya Road, Bangkok 10400, Thailand*

Correspondence should be addressed to Ronnason Chinram; ronnason.c@psu.ac.th

Received 16 August 2020; Revised 12 September 2020; Accepted 9 October 2020; Published 11 November 2020

Academic Editor: Chonghui Zhang



Complex dual hesitant fuzzy set (CDHFS) is a combination of two modifications, called complex fuzzy set (CFS) and dual hesitant fuzzy set (DHFS). CDHFS makes two degrees, called membership valued and nonmembership valued in the form of a finite subset of a unit disc in the complex plane, and is a capable method to solve uncertain and unpredictable information in real-life problems. The goal of this study is to describe the notion of CDHFS and its operational laws. The novel approach of the complex interval-valued dual hesitant fuzzy set (CIvDHFS) and its fundamental laws are also described and defended with the help of an example. Further, the vector similarity measures (VSMs), weighted vector similarity measures (WVSMs), hybrid vector similarity measure, and weighted hybrid vector similarity measure are additionally explored. These similarity measures (SM) are applied to the environment of pattern recognition and medical diagnosis to assess the capability and feasibility of the interpreted measures. We additionally solved some numerical examples utilizing the established measures. We examine the dependability and validity of the proposed measures by comparing them with some existing measures. The advantages, comparative analysis, and graphical portrayal of the investigated interpreted measures and existing measures are additionally described in detail.

## 1. Introduction

Since fuzzy set [1] was presented, various extensions of this concept have been proposed, for example, interval-valued fuzzy set [2–4], intuitionistic fuzzy set [5], interval-valued intuitionistic fuzzy set [6], linguistic fuzzy set [7], type-2 fuzzy set [8], type-$n$ fuzzy set [8], and fuzzy multiset [9]. Baležentis and Zeng [10] interpreted group multicriteria decision making based upon an interval-valued fuzzy number. Su et al. [11] explored the intuitionistic fuzzy decision making with SMs and OWA operator. Peng et al. [12] described some intuitionistic fuzzy weighted geometric distance measures and their application to group decision making. An intuitionistic fuzzy MULTIMOORA approach for multicriteria assessment of the energy storage technologies was described by Zhang et al. [13]. Zhou et al. [14] interpreted the normalized weighted Bonferroni harmonic mean-based intuitionistic fuzzy operators and their application to the sustainable selection of search and rescue robots. An interval-valued intuitionistic fuzzy multiple attribute decision making based on nonlinear programming methodology and TOPSIS method was described by Zeng et al. [15]. The theory of linear Diophantine fuzzy set (LDFS) was explored by Riaz and Hashmi [16]. The proposed model has a resemblance with the well-known linear Diophantine equation $ax + by = c$ in the number theory. Since IFSs, PFSs, and q-ROFSs have some limitations on membership/nonmembership grades. In order to get rid of such limitations, the theory of LDFS was explored with the addition of reference parameters. This idea removes the restrictions of membership/nonmembership grades and the decision maker can freely choose the grades without any limitation.

This structure also categorizes the problem by choosing different types of reference parameters. The structure of LDFS and its graphical representation explaining these concepts with the help of illustrations are given in reference [16]. In the real world, we often encounter fuzzy situations in which it is hard to determine the membership of an element to a set because of the doubts between a few different values. To resolve this sort of issue, Torra [17] presented the idea of the hesitant fuzzy set (HFS), which allows the membership to have a set of possible finite values. Since its appearance, more and more multiple decision-making theories and methods have been proposed under a hesitant fuzzy environment [18–21]. Xu and Xia [22] described the distance measures for HFS and compared similarity measures. Distance and SMs among HFSs and their application in pattern recognition were expressed by Zeng et al. [23]. Mu et al. [24] described a novel aggregation principle for hesitant fuzzy elements. Decomposition theorems and extension principles for HFSs are explored by Alcantud and Torra [25]. Liao and Xu [26] described subtraction and division operation over HFSs. Bisht and Kumar [27] characterized the fuzzy time series forecasting method based on HFSs. Alcantud and Giarlotta [28] proposed the expansion of Torra's idea of HFSs. Novel distance and SMs on HFSs with application to clustering analysis were introduced by Zhang and Xu [29]. Farhadinia and Herrera-Viedma [30] described multiple criteria group decision-making methods based on extended HFSs with unknown weight information. Chen et al. [31] interpreted the idea of interval-valued hesitant fuzzy sets (IvHFSs) which is the generalization of HFS.

Dual hesitant fuzzy set (DHFS), as another augmentation of HFS, was proposed by Zhu et al. [32], in which the participation degree and nonenrollment level of a component to a given set are meant by two sets of a few fresh qualities. It includes a fuzzy set, fuzzy multiset, intuitionistic fuzzy set, and hesitant fuzzy set as uncommon cases and has gotten increasingly more consideration from scientists as of late. Wang et al. [33] characterized distance and comparability proportions of DHFSs with their applications to different quality dynamics. Zhu and Xu [34] characterized a few outcomes for DHFSs. Tyagi [35] introduced the relationship coefficient of DHFS and its application. Considering such capacities gives us more praiseworthy and adaptable access to allocate esteems for every component in the space. Obviously, DHFSs can mirror the human's reluctance more dispassionately than the other existing augmentations of the fuzzy set (AIFSs, IVAIFSs, HFSs, and so on.). In any case, in the soul of what has been accomplished for IVFSs, Farhadinia [36] presented a dual interval-valued hesitant fuzzy set (DIVHFS) where its principal trademark is that the value of the membership and nonmembership function are set of intervals as opposed to a set of definite numbers. Many scholars raised the question, what happened when a decision-maker changes the range of the fuzzy set into a unit disc. Ramot et al. [37] presented the possibility of complex FS (CFS), which contains membership grade as a complex number of a unit disc in the complex plane. CFS manages two dimensions in a solitary set. CFS is an amazing strategy to outline the conviction of a person in the development of evaluations. Many scholars have work on multiattribute decision making [38–44].

In real-life issues, we run over various conditions where we need to measure the weakness existing in the data to choose perfect decisions. Vector similarity measures (VSMs) and hybrid vector similarity measure are significant apparatuses for dealing with dubious data present in our day-to-day life problems. Different measures, for instance, similarity, exponential, separation, entropy, and incorporation process the flawed information and engage us to show up at some goal. Starting late, these measures have expanded a great deal of thought from various makers due to their wide applications in various fields, for instance, pattern recognition, medical diagnosis, and decision making. All the current techniques of chiefs, considering VSM and hybrid vector similarity measures, in FS, CFS, and HFS speculations, manage participation capacities having a place with a unit span as a subset in the idea of HFS. In the CDHFS speculation, membership and nonmembership degrees are astoundingly regarded and are addressed in polar ways. These all thoughts worked feasibly, in any case, when a pioneer stood up to such kinds of information which contains two-dimensional information in a singular set. For instance, $\{\{0.8e^{i2\pi(0.4)},\ 0.5e^{i2\pi(0.7)}\}, \{0.2e^{i2\pi(0.2)},\ 0.1e^{i2\pi(0.5)}\}, \{0.15e^{i2\pi(0.35)}, 0.7e^{i2\pi(0.56)}\}\}$ at that point all the current ideas are fizzled. For adapting to such sorts of issues, the CDHFS is a capable method to resolve realistic decision problems in the environment of the FS theory. CDHFS is more impressive and broader than existing ideas such as HFS, CFS, and FS to adapt to awkward and confusing data, in real-world decisions. Since these all thoughts are the particular cases of the investigated CDHFS, the upsides of the introduced CDHFS are talked about underneath:

(1) At the point when we pick the imaginary parts of the CDHFS as zero, the CDHFS is diminished into DHFS which is as $\{\{0.8, 0.5\}, \{0.2, 0.1\}, \{0.15, 0.7\}\}$

(2) At the point when we pick the CDHFS as a singleton set, the CDHFS is diminished into CIFS which is as $\{0.8e^{i2\pi(0.4)}, 0.5e^{i2\pi(0.7)}\}$

(3) At the point when we pick the CDHFS as a singleton set and the imaginary parts as zero, the CDHFS is decreased into IFS which is as $\{0.8, 0.5\}$

The theory of CDHFS is a proficient technique to cope with uncertain and awkward information in the real-life problems. CDHFS contains the grade of membership and the grade of nonmembership in the form of complex number belonging to the complex plane in a unit disc, whose real part and imaginary part are the subset of unit interval. CDHFS is more generalized than many existing notions such as fuzzy sets, intuitionistic fuzzy sets, hesitant fuzzy sets, dual hesitant fuzzy sets, complex fuzzy sets, complex intuitionistic fuzzy sets, and complex hesitant fuzzy sets . For example, if we choose the value of the imaginary part to be zero, then the CDHFS is converted for dual hesitant fuzzy sets, and if we choose the value of nonmembership to be zero, then the CDHFS is converted for HFS. Similarly, if we choose the value of nonmembership to be zero, then the

explored approach is converted for complex hesitant fuzzy sets. Inspired by the above difficulties and maintaining the benefits of the CDHFS, in this paper, some key contributions are made as follows:

(1) Complex dual hesitant fuzzy set (CDHFS) is a mix of two alterations, called complex fuzzy set (CFS) and dual hesitant fuzzy set (DHFS). CDHFS makes two degrees, called membership valued and nonmembership valued in the form of subset of a unit disc in a complex plane, and is a capable method to adapt to questionable and capricious real-world issues. The aim of the paper is to investigate the idea of a CDHFS and its operational laws.

(2) The novel approach of CIvDHFS and its essential laws are investigated and advocated with the assistance of an example.

(3) Further, the VSMs, WVSMs, hybrid VSM, and weighted hybrid VSM and their significant attributes are likewise investigated.

(4) These SMs are applied to the environment of pattern recognition and medical diagnosis to assess the capability and plausibility of the interpreted measures. We likewise illuminated some numerical examples utilizing the interpreted measures.

(5) We examine the dependability and legitimacy of the proposed measures by contrasting them with existing measures.

(6) The advantages, relative investigation, and graphical portrayal of the investigated measures and existing measures are additionally talked about in detail.

The rest of the paper is organized as follows: in Section 2, we review some basic definitions such as FS, CFS, HFS, DHFS, and IvDHFS. In Section 3, we investigate the thought of a CDHFS and its operational laws. The epic methodology of CIvDHFS and its basic laws are additionally investigated and defended with the assistance of models. In Section 4, the vector similarity measures (VSMs), weighted vector similarity measures (WVSMs), hybrid vector similarity measure, and weighted hybrid vector similarity measure are investigated. In Section 5, these measures are applied to the environment of pattern recognition and medical diagnosis to assess the capability and practicality of the explored measures. We likewise tackled some numerical models utilizing the explored measures. We examine the dependability and legitimacy of the proposed measures by contrasting them with existing measures given in Section 6. The points of interest, relative investigation, and graphical portrayal of the investigated gauges and existing measures are likewise talked about in detail. Section 7 concludes the paper.

## 2. Preliminaries

In this section, we survey fundamental definitions such as FS, CFS, HFS, DHFS, and IvDHFS. Through this article, $X$ speaks to a fix set.

*Definition 1* (see [1])A FS $\mathcal{S}$ is of the structure:

$$\mathcal{S} = \{(x, \mu_{\mathcal{S}}(x)) \mid x \in X\}, \tag{1}$$

with a condition $0 \le \mu_{\mathcal{S}}(x) \le 1$, where $\mu_{\mathcal{S}}(x)$ represents the grade of truth. Through this article, the collection of all FSs on $X$ is represented by $F\mathcal{S}(X)$. The pair $\mathcal{S} = (x, \mu_{\mathcal{S}}(x))$ is known as fuzzy number (FN).

*Definition 2* (see [37])A CFS $\mathcal{S}$ is of the structure:

$$\mathcal{S} = \{(x, \mu_{\mathcal{S}}(x)) \mid x \in X\}, \tag{2}$$

where $\mu_{\mathcal{S}}(x) = \gamma_{\mathcal{S}}(x) \cdot e^{i2\pi(\omega_{\gamma_{\mathcal{S}}}(x))}$ represents the complex-valued truth grade in the form of polar coordinate, where $\gamma_{\mathcal{S}}(x), \omega_{\gamma_{\mathcal{S}}}(x) \in [0,1]$. Additionally, the pair $(x, \gamma_{\mathcal{S}}(x) \cdot e^{i2\pi(\omega_{\gamma_{\mathcal{S}}}(x))})$ is known as complex fuzzy number (CFN).

*Definition 3* (see [17])An HFS $\mathcal{S}$ is of the structure:

$$\mathcal{S} = \{(x, \mu_{\mathcal{S}}(x)) | x \in X\}, \tag{3}$$

where $\mu_{\mathcal{S}}(x)$ is the set of different finite values in $[0,1]$ representing the grade of truth for each element $x \in X$. Further, the pair $\mathcal{S} = (x, \mu_{\mathcal{S}}(x))$ is known as hesitant fuzzy number (HFN).

*Definition 4* (see [32])A DHFS $\mathcal{S}$ is of the structure:

$$\mathcal{S} = \{(x, (\mu_{\mathcal{S}}(x), \nu_{\mathcal{S}}(x))) | x \in X\}, \tag{4}$$

where $\mu_{\mathcal{S}}(x)$ and $\nu_{\mathcal{S}}(x)$ are two finite subsets in $[0,1]$, representing the membership grade and nonmembership grade of the component $x \in X$, respectively, with the conditions $0 \le \gamma_{\mathcal{S}}(x), \delta_{\mathcal{S}}(x) \le 1$, $0 \le \gamma_{\mathcal{S}}^{+}(x) + \delta_{\mathcal{S}}^{+}(x) \le 1$, where $\gamma_{\mathcal{S}}(x) \in \mu_{\mathcal{S}}(x)$, $\delta_{\mathcal{S}}(x) \in \nu_{\mathcal{S}}(x)$, $\gamma_{\mathcal{S}}^{+}(x) \in \mu_{\mathcal{S}}^{+}(x) = \cup_{\gamma_{\mathcal{S}}(x) \in \mu_{\mathcal{S}}(x)} \max\{\gamma_{\mathcal{S}}(x)\}$, and $\delta_{\mathcal{S}}^{+}(x) \in \nu_{\mathcal{S}}^{+}(x) = \cup_{\delta_{\mathcal{S}}(x) \in \nu_{\mathcal{S}}(x)} \max\{\delta_{\mathcal{S}}(x)\}$.

*Definition 5* (see [33])For any two DHFSs $\mathcal{S}$ and $\mathcal{T}$, the similarity measure $\mathbb{S}(\mathcal{S}, \mathcal{T})$ satisfies the following axioms:

(1) $0 \le \mathbb{S}(\mathcal{S}, \mathcal{T}) \le 1$

(2) $\mathbb{S}(\mathcal{S}, \mathcal{T}) = 1 \Longleftrightarrow \mathcal{S} = \mathcal{T}$

(3) $\mathbb{S}(\mathcal{S}, \mathcal{T}) = \mathbb{S}(\mathcal{T}, \mathcal{S})$

*Definition 6* (see [33])For any two DHFSs $\mathcal{S}$ and $\mathcal{T}$, the distance measure $\mathbb{D}(\mathcal{S}, \mathcal{T})$ satisfies the following axioms:

(1) $0 \le \mathbb{D}(\mathcal{S}, \mathcal{T}) \le 1$

(2) $\mathbb{D}(\mathcal{S}, \mathcal{T}) = 1 \Longleftrightarrow \mathcal{S} = \mathcal{T}$

(3) $\mathbb{D}(\mathcal{S}, \mathcal{T}) = \mathbb{D}(\mathcal{T}, \mathcal{S})$

From the above discussion, we obtain that $\mathbb{S}(\mathcal{S}, \mathcal{T}) = 1 - \mathbb{D}(\mathcal{S}, \mathcal{T})$.

*Definition 7* (see [36])An IvDHFS $\mathcal{S}$ is of the structure:

$$\mathcal{S} = \{(x, (\mu_{\mathcal{S}}(x), \nu_{\mathcal{S}}(x))) | x \in X\}, \tag{5}$$

where $\mu_{\mathcal{S}}(x) = \cup_{[\gamma_{\mathcal{S}}^L(x)\gamma_{\mathcal{S}}^U]\in\mu_{\mathcal{S}}(x)}\max\{[\gamma_{\mathcal{S}}^L(x), \gamma_{\mathcal{S}}^U(x)]\}$ and $\nu_{\mathcal{S}}(x) = \cup_{[\delta_{\mathcal{S}}^L(x)\delta_{\mathcal{S}}^U]\in\nu_{\mathcal{S}}(x)}\max\{[\delta_{\mathcal{S}}^L(x), \delta_{\mathcal{S}}^U(x)]\}$ are two finite subsets of some interval values in $[0,1]$, representing the membership grade and nonmembership grade of the component $x \in X$, respectively, with the conditions $[\gamma_{\mathcal{S}}^L(x)\gamma_{\mathcal{S}}^U(x)], [\delta_{\mathcal{S}}^L(x), \delta_{\mathcal{S}}^U(x)] \subseteq [0,1]$, and $0 \le (\gamma_{\mathcal{S}}^U(x))^+ + (\delta_{\mathcal{S}}^U(x))^+ \le 1$, where $[\gamma_{\mathcal{S}}^L(x)\gamma_{\mathcal{S}}^U(x)] \in \mu_{\mathcal{S}}(x)$, $[\delta_{\mathcal{S}}^L(x), \delta_{\mathcal{S}}^U(x)] \in \nu_{\mathcal{S}}(x)$, $(\gamma_{\mathcal{S}}^U(x))^+ \in \mu_{\mathcal{S}}^+(x) = \cup_{[\gamma_{\mathcal{S}}^L(x)\gamma_{\mathcal{S}}^U]\in\mu_{\mathcal{S}}(x)}\max\{\gamma_{\mathcal{S}}^L(x)\}$, and $(\delta_{\mathcal{S}}^U(x))^+ \in \nu_{\mathcal{S}}^+(x) = \cup_{[\delta_{\mathcal{S}}^L(x), \delta_{\mathcal{S}}^U]\in\nu_{\mathcal{S}}(x)}\max\{\delta_{\mathcal{S}}(x)\}$ for all $x \in X$.

## 3. Complex Dual Hesitant Fuzzy Sets and Complex Interval-Valued Dual Hesitant Fuzzy Sets

In this section, we have two subsections in which we defined the idea of CDHFS, CIvDHFS, and their operational laws.

*3.1. Complex Dual Hesitant Fuzzy Sets.* In this section, we investigated the notion of CDHFS which is the alteration of CFS and DHFS. We additionally investigated its operational laws.

*Definition 8.* A CDHFS $\mathcal{S}$ is of the structure:

$$\mathcal{S} = \{(x, (\mu_{\mathcal{S}}(x), \nu_{\mathcal{S}}(x)))|x \in X\}, \tag{6}$$

where $\mu_{\mathcal{S}}(x) = \left\{(x, \gamma_{\mathcal{S}_p}(x) \cdot e^{i2\pi(\omega_{\gamma_{\mathcal{S}_p}})}), p = 1, 2, 3, \ldots, g\right\}$ and $\nu_{\mathcal{S}}(x) = \left\{(x, \delta_{\mathcal{S}_q}(x) \cdot e^{i2\pi(\omega_{\gamma_{\mathcal{S}_q}})}), q = 1, 2, 3, \ldots, h\right\}$ represented the complex-valued membership grade and nonmembership grade, which are subsets of a unit disc in the complex plane with conditions $\gamma_{\mathcal{S}_p}(x), \omega_{\gamma_{\mathcal{S}_p}}(x), \delta_{\mathcal{S}_q}(x), \overline{\omega}_{\delta_{\mathcal{S}_q}}(x) \in [0,1]$, $0 \le \gamma_{\mathcal{S}}^+(x) + \delta_{\mathcal{S}}^+(x) \le 1$, and $0 \le \omega_{\gamma_{\mathcal{S}}}^+(x) + \overline{\omega}_{\delta_{\mathcal{S}}}^+(x) \le 1$, where $\gamma_{\mathcal{S}}^+(x) = \cup_{\gamma_{\mathcal{S}_p}(x)\in\mu_{\mathcal{S}}(x)}\max\left\{\gamma_{\mathcal{S}_p}(x)\right\}$, $\omega_{\gamma_{\mathcal{S}}}^+(x) = \cup_{\omega_{\mathcal{S}_p}(x)\in\mu_{\mathcal{S}}(x)}\max\left\{\omega_{\gamma_{\mathcal{S}_p}}(x)\right\}$, $\delta_{\mathcal{S}}^+(x) = \cup_{\delta_{\mathcal{S}_q}(x)\in\nu_{\mathcal{S}}(x)}\max\left\{\delta_{\mathcal{S}_q}(x)\right\}$, and $\overline{\omega}_{\delta_{\mathcal{S}}}^+(x) = \cup_{\overline{\omega}_{\delta_{\mathcal{S}_q}}(x)\in\nu_{\mathcal{S}}(x)}\max\left\{\overline{\omega}_{\delta_{\mathcal{S}_q}}(x)\right\}$ for $p = 1, 2, \ldots, g$ and $q = 1, 2, \ldots, h$. Further, $\mathcal{S} = (x, (\gamma_{\mathcal{S}_p}(x) \cdot e^{i2\pi(\omega_{\gamma_{\mathcal{S}_p}}(x))}, \delta_{\mathcal{S}_q}(x) \cdot e^{i2\pi(\overline{\omega}_{\delta_{\mathcal{S}_q}}(x))}))$ is called complex dual hesitant fuzzy number (CDHFN).

*Definition 9.* Let $\mathcal{S} = (x, \gamma_{\mathcal{S}_p}(x) \cdot e^{i2\pi(\omega_{\gamma_{\mathcal{S}_p}}(x))}, \delta_{\mathcal{S}_q}(x) \cdot e^{i2\pi(\overline{\omega}_{\delta_{\mathcal{S}_q}}(x))})$ and $\mathcal{T} = (x, \gamma_{\mathcal{T}_p}(x) \cdot e^{i2\pi(\omega_{\gamma_{\mathcal{T}_p}}(x))}, \delta_{\mathcal{T}_q}(x) \cdot e^{i2\pi(\overline{\omega}_{\delta_{\mathcal{T}_q}}(x))})$ be two CDHFNs. Then their complement, union, and intersection are defined as follows:

(1) $\mathcal{S}^c = \{(x, (\nu_{\mathcal{S}}(x), \mu_{\mathcal{S}}(x)))\}$

(2) $\mathcal{S} \cup \mathcal{T} = \left\{(x, (\max(\gamma_{\mathcal{S}_p}(x), \gamma_{\mathcal{T}_p}(x)) \cdot e^{i2\pi(\max(\omega_{\gamma_{\mathcal{S}_p}}(x), \omega_{\gamma_{\mathcal{T}_p}}(x)))}, \min(\delta_{\mathcal{S}_q}(x), \delta_{\mathcal{T}_q}(x)) \cdot e^{i2\pi(\max(\overline{\omega}_{\delta_{\mathcal{S}_q}}(x), \overline{\omega}_{\delta_{\mathcal{T}_q}}(x)))}))\right\}$

(3) $\mathcal{S} \cap \mathcal{T} = \left\{(x, (\min(\gamma_{\mathcal{S}_p}(x), \gamma_{\mathcal{T}_p}(x)) \cdot e^{i2\pi(\max(\omega_{\gamma_{\mathcal{S}_p}}(x), \omega_{\gamma_{\mathcal{T}_p}}(x)))}, \max(\delta_{\mathcal{S}_q}(x), \delta_{\mathcal{T}_q}(x)) \cdot e^{i2\pi(\max(\overline{\omega}_{\delta_{\mathcal{S}_q}}(x), \overline{\omega}_{\delta_{\mathcal{T}_q}}(x)))}))\right\}$

*Example 1.* Let

$$\mathcal{S} = \left\{\begin{array}{l} (x_1, \{\{\{0.7e^{i2\pi(0.4)}, 0.3e^{i2\pi(0.6)}\}, \\ \quad \{0.2e^{i2\pi(0.2)}, 0.1e^{i2\pi(0.3)}\}\}\}), \\ (x_2, \{\{0.7e^{i2\pi(0.6)}\}, \{0.2e^{i2\pi(0.3)}\}\}), \\ (x_3, \{\{0.6e^{i2\pi(0.5)}, 0.5e^{i2\pi(0.4)}, 0.4e^{i2\pi(0.7)}\}, \\ \quad \{0.3e^{i2\pi(0.1)}0.1e^{i2\pi(0.2)}\}\}), \\ (x_4, \{\{0.4e^{i2\pi(0.5)}, 0.6e^{i2\pi(0.8)}\}, \{0.515e^{i2\pi(0.1)}\}\}), \\ (x_5, \{\{0.5e^{i2\pi(0.6)}, 0.4e^{i2\pi(0.5)}, 0.2e^{i2\pi(0.4)}\}, \\ \quad \{0.3e^{i2\pi(0.2)}, 0.1e^{i2\pi(0.2)}, 0.2e^{i2\pi(0.3)}\}\}) \end{array}\right\}$$

and

$$\mathcal{T} = \left\{\begin{array}{l} (x_1, \{\{0.3e^{i2\pi(0.2)}\}\{0.65e^{i2\pi(0.7)}, 0.5e^{i2\pi(0.6)}\}\}), \\ (x_2, \{\{0.8e^{i2\pi(0.6)}, 0.55e^{i2\pi(0.55)}, 0.45e^{i2\pi(0.8)}\}, \\ \quad \{0.1e^{i2\pi(0.15)}, 0.15e^{i2\pi(0.1)}, 0.1e^{i2\pi(0.5)}\}\}), \\ (x_3, \{\{0.3e^{i2\pi(0.55)}\}, \{0.25e^{i2\pi(0.35)}\}\}), \\ (x_4\{\{0.35e^{i2\pi(0.65)}, 0.65e^{i2\pi(0.5)}\}, \\ \quad \{0.15e^{i2\pi(0.25)}, 0.3e^{i2\pi(0.3)}\}\}), \\ (x_5, \{\{0.5e^{i2\pi(0.3)}, 0.3e^{i2\pi(0.6)}\}, \{0.2e^{i2\pi(0.25)}\}\}) \end{array}\right\}$$

be two CDHFSs. Then their operational laws are given as follows:

(1)

$$\mathcal{S}^c = \left\{\begin{array}{l} (x_1, \{\{\{0.2e^{i2\pi(0.2)}, 0.1e^{i2\pi(0.3)}\}, \\ \quad \{0.7e^{i2\pi(0.4)}, 0.3e^{i2\pi(0.6)}\}\}\}), \\ (x_2, \{\{0.2e^{i2\pi(0.3)}\}, \{0.7e^{i2\pi(0.6)}\}\}), \\ (x_3, \{\{0.3e^{i2\pi(0.1)}, 0.1e^{i2\pi(0.2)}\}, \\ \quad \{0.6e^{i2\pi(0.5)}, 0.5e^{i2\pi(0.4)}, 0.4e^{i2\pi(0.7)}\}\}), \\ (x, \{\{0.15e^{i2\pi(0.1)}\}, \{0.4e^{i2\pi(0.5)}, 0.6e^{i2\pi(0.8)}\}\}) \\ (x_5, \{\{0.3e^{i2\pi(0.2)}, 0.2e^{i2\pi(0.3)}\}, \\ \quad \{0.5e^{i2\pi(0.6)}, 0.4e^{i2\pi(0.5)}, 0.2e^{i2\pi(0.4)}\}\}) \end{array}\right\}$$

(2) $\mathcal{S} \cup \mathcal{T} =$

$$\left\{\begin{array}{l} (x_1, \{\{\{0.7e^{i2\pi(0.4)}, 0.3e^{i2\pi(0.1)}\}, \{0.2e^{i2\pi(0.2)}, 0.1e^{i2\pi(0.3)}\}\}\}) \\ (x_2, \{\{0.8e^{i2\pi(0.6)}, 0.55e^{i2\pi(0.55)}, 0.45e^{i2\pi(0.8)}\}, \{0.1e^{i2\pi(0.15)}\}\}), \\ (x_3, \{\{0.6e^{i2\pi(0.55)}, 0.5e^{i2\pi(0.4)}, 0.4e^{i2\pi(0.7)}\}, \{0.25e^{i2\pi(0.1)}\}\}), \\ (x_4\{\{0.4e^{i2\pi(0.65)}, 0.65e^{i2\pi(0.8)}\}, \{0.15e^{i2\pi(0.1)}\}\}) \\ (x_5, \{\{0.5e^{i2\pi(0.6)}, 0.4e^{i2\pi(0.6)}0.2e^{i2\pi(0.4)}\}, \{0.2e^{i2\pi(0.2)}\}\}) \end{array}\right\}$$

(3) $\mathcal{S} \cap \mathcal{T} =$

$$\left\{\begin{array}{l} (x_1, \{\{0.3e^{i2\pi(0.2)}\}\{0.65e^{i2\pi(0.7)}, 0.5e^{i2\pi(0.6)}\}\}), \\ (x_2, \{\{0.7e^{i2\pi(0.6)}\}, \{0.2e^{i2\pi(0.3)}, 0.15e^{i2\pi(0.1)}, 0.1e^{i2\pi(0.5)}\}\}), \\ (x_3, \{\{0.45e^{i2\pi(0.5)}\}, \{0.3e^{i2\pi(0.35)}, 0.1e^{i2\pi(0.2)}\}\}), \\ (x_4, \{\{0.35e^{i2\pi(0.5)}, 0.6e^{i2\pi(0.5)}\}, \{0.15e^{i2\pi(0.25)}, 0.3e^{i2\pi(0.3)}\}\}) \\ (x_5, \{\{0.5e^{i2\pi(0.3)}, 0.3e^{i2\pi(0.5)}\}, \\ \qquad \{0.3e^{i2\pi(0.25)}, 0.1e^{i2\pi(0.2)}, 0.2e^{i2\pi(0.3)}\}\}) \end{array}\right\}$$

*3.2. Complex Interval-Valued Dual Hesitant Fuzzy Sets.* In this section, we investigated the notion of CIvDHFS and additionally investigated its operational laws.

*Definition 10.* A CIvDHFS $\mathcal{S}$ is of the structure:

$$\mathcal{S} = \{(x, (\mu_{\mathcal{S}}(x), \nu_{\mathcal{S}}(x)))|x \in X\}, \tag{7}$$

where

$$\begin{aligned} \mu_{\mathcal{S}}(x) &= \left[\mu_{\mathcal{S}}^L(x), \mu_{\mathcal{S}}^U(x)\right] \\ &= \left\{\left(x, \left[\gamma_{\mathcal{S}_p}^L(x), \gamma_{\mathcal{S}_p}^U(x)\right] \cdot e^{ie\pi\left(\left[\omega_{\gamma_{\mathcal{S}_p}}^L(x), \omega_{\gamma_{\mathcal{S}_p}}^U(x)\right]\right)}\right), \right. \\ &\qquad \left. p = 1, 2, 3, \ldots, g\right\}, \\ \nu_{\mathcal{S}}(x) &= \left[\nu_{\mathcal{S}}^L(x), \nu_{\mathcal{S}}^U(x)\right] \\ &= \left\{\left(x, \left[\delta_{\mathcal{S}_q}^L(x), \delta_{\mathcal{S}_q}^U(x)\right] \cdot e^{ie\pi\left(\left[\bar{\omega}_{\delta_{\mathcal{S}_q}}^L(x), \bar{\omega}_{\delta_{\mathcal{S}_q}}^U(x)\right]\right)}\right), \right. \\ &\qquad \left. p = 1, 2, 3, \ldots, h\right\}, \end{aligned} \tag{8}$$

represented the complex-valued membership grade and nonmembership grade, which are finite subsets of different interval values of a unit disc in complex plane with conditions $\gamma_{\mathcal{S}_p}^L(x), \gamma_{\mathcal{S}_p}^U(x), \omega_{\gamma_{\mathcal{S}_p}}^L(x), \omega_{\gamma_{\mathcal{S}_p}}^U(x), \delta_{\mathcal{S}_q}^L(x), \delta_{\mathcal{S}_q}^U(x), \bar{\omega}_{\gamma_{\mathcal{S}_p}}^L(x), \bar{\omega}_{\gamma_{\mathcal{S}_p}}^U(x) \in [0, 1]$, $0 \le (\gamma_{\mathcal{S}}^U(x))^+ + (\delta_{\mathcal{S}}^U(x))^+ \le 1$, and $0 \le (\omega_{\gamma_{\mathcal{S}}}^U(x))^+ + (\bar{\omega}_{\delta_{\mathcal{S}}}^U(x))^+ \le 1$, where $(\gamma_{\mathcal{S}}^U(x))^+ = \cup_{[\gamma_{\mathcal{S}_p}^L(x), \gamma_{\mathcal{S}_p}^U(x)] \in \mu_{\mathcal{S}}(x)} \max\left\{\gamma_{\mathcal{S}_p}^L(x)\right\}$, $(\omega_{\gamma_{\mathcal{S}}}^U(x))^+ = \cup_{[\omega_{\gamma_{\mathcal{S}_p}}^L(x), \omega_{\gamma_{\mathcal{S}_p}}^U(x)] \in \mu_{\mathcal{S}}(x)} \max\left\{\omega_{\gamma_{\mathcal{S}_p}}^U(x)\right\}$, $(\delta_{\mathcal{S}}^U(x))^U = \cup_{[\delta_{\mathcal{S}_q}^L(x), \gamma_{\mathcal{S}_q}^U(x)] \in \nu_{\mathcal{S}}(x)} \max\left\{\delta_{\mathcal{S}_p}^L(x)\right\}$, and $(\bar{\omega}_{\delta_{\mathcal{S}}}^U(x))^+ = \cup_{[\bar{\omega}_{\delta_{\mathcal{S}_q}}^L(x), \bar{\omega}_{\delta_{\mathcal{S}_q}}^U(x)] \in \nu_{\mathcal{S}}(x)} \max\left\{\bar{\omega}_{\delta_{\mathcal{S}_q}}^U(x)\right\}$ for $p = 1, 2, \ldots, g$ and $q = 1, 2, \ldots, h$. Further, $([\gamma_{\mathcal{S}_p}^L, \gamma_{\mathcal{S}_p}^U] \cdot e^{i2\pi([\omega_{\gamma_{\mathcal{S}_p}}^L(x), \omega_{\gamma_{\mathcal{S}_p}}^U])}, [\delta_{\mathcal{S}_q}^L(x), \delta_{\mathcal{S}_q}^U(x)] \cdot e^{i2\pi([\bar{\omega}_{\delta_{\mathcal{S}_q}}^L(x), \bar{\omega}_{\delta_{\mathcal{S}_q}}^U])})$ is called complex interval-valued dual hesitant fuzzy number (CIvDHFN).

*Definition 11.* Let $\mathcal{S} = (x, ([\gamma_{\mathcal{S}_p}^L, \gamma_{\mathcal{S}_p}^U] \cdot e^{i2\pi([\omega_{\gamma_{\mathcal{S}_p}}^L(x), \omega_{\gamma_{\mathcal{S}_p}}^U])}, [\delta_{\mathcal{S}_q}^L(x), \delta_{\mathcal{S}_q}^U(x)] \cdot e^{i2\pi([\bar{\omega}_{\delta_{\mathcal{S}_q}}^L(x), \bar{\omega}_{\delta_{\mathcal{S}_q}}^U])}))$ and $\mathcal{T} = (x, ([\gamma_{\mathcal{T}_p}^L, \gamma_{\mathcal{T}_p}^U] \cdot e^{i2\pi([\omega_{\gamma_{\mathcal{T}_p}}^L(x), \omega_{\gamma_{\mathcal{T}_p}}^U])}, [\delta_{\mathcal{T}_q}^L(x), \delta_{\mathcal{T}_q}^U(x)] \cdot e^{i2\pi([\bar{\omega}_{\delta_{\mathcal{T}_q}}^L(x), \bar{\omega}_{\delta_{\mathcal{T}_q}}^U])}))$ be two CIvDHFNs. Then their complement, union, and intersection are defined as follows:

(1) $\mathcal{S}^c = \{(x, ((\nu_{\mathcal{S}}), \mu_{\mathcal{S}}(x)))\}$

(2) $\mathcal{S} \cup \mathcal{T} =$

$$\left\{\left(x, \left(\begin{array}{c} [\max(\gamma_{\mathcal{S}_p}^L(x), \gamma_{\mathcal{T}_p}^L(x)), \max(\gamma_{\mathcal{S}_p}^U(x), \gamma_{\mathcal{T}_p}^U(x))] \\ \cdot e^{i2\pi([\max(\omega_{\gamma_{\mathcal{S}_p}}^L(x), \omega_{\gamma_{\mathcal{T}_p}}^L(x)), \max(\omega_{\gamma_{\mathcal{S}_p}}^U(x), \omega_{\gamma_{\mathcal{T}_p}}^U(x))])}, \\ [\min(\delta_{\mathcal{S}_q}^L(x), \delta_{\mathcal{T}_q}^L(x)), \min(\delta_{\mathcal{S}_q}^U(x), \delta_{\mathcal{T}_q}^U(x))] \\ \cdot e^{i2\pi([\min(\bar{\omega}_{\delta_{\mathcal{S}_p}}^L(x), \bar{\omega}_{\delta_{\mathcal{T}_p}}^L(x)), \min(\bar{\omega}_{\delta_{\mathcal{S}_p}}^U(x), \bar{\omega}_{\delta_{\mathcal{T}_p}}^U(x))])} \end{array}\right)\right)\right\}$$

(3) $\mathcal{S} \cap \mathcal{T} =$

$$\left\{\left(x, \left(\begin{array}{c} [\min(\gamma_{\mathcal{S}_p}^L(x), \gamma_{\mathcal{T}_p}^L(x)), \min(\gamma_{\mathcal{S}_p}^U(x), \gamma_{\mathcal{T}_p}^U(x))] \\ \cdot e^{i2\pi([\min(\omega_{\gamma_{\mathcal{S}_p}}^L(x), \omega_{\gamma_{\mathcal{T}_p}}^L(x)), \min(\omega_{\gamma_{\mathcal{S}_p}}^U(x), \omega_{\gamma_{\mathcal{T}_p}}^U(x))])}, \\ [\max(\delta_{\mathcal{S}_q}^L(x), \delta_{\mathcal{T}_q}^L(x)), \max(\delta_{\mathcal{S}_q}^U(x), \delta_{\mathcal{T}_q}^U(x))] \\ \cdot e^{i2\pi([\max(\bar{\omega}_{\delta_{\mathcal{S}_p}}^L(x), \bar{\omega}_{\delta_{\mathcal{T}_p}}^L(x)), \max(\bar{\omega}_{\delta_{\mathcal{S}_p}}^U(x), \bar{\omega}_{\delta_{\mathcal{T}_p}}^U(x))])} \end{array}\right)\right)\right\}$$

*Example 2.* Let

$$\mathcal{S} = \left\{ \begin{array}{c} \left(x_1, \left\{ \left\{ \begin{array}{c} [0.2, 0.6]e^{i2\pi([0.4,0.8])}, \\ [0.1, 0.4]e^{i2\pi([0.2,0.5])} \end{array} \right\}, \left\{ [0.1, 0.2]e^{i2\pi([0.05,0.1])} \right\} \right\} \right), \left(x_2, \left\{ \begin{array}{c} \left\{ [0.4, 0.7]e^{i2\pi([0.6,0.7])} \right\}, \\ \left\{ [0.1, 0.25]e^{i2\pi([0.15,0.2])} \right\} \end{array} \right\} \right), \\ \left(x_3, \left\{ \left\{ \begin{array}{c} [0.5, 0.6]e^{i2\pi([0.3,0.5])}, \\ [0.5, 0.55]e^{i2\pi([0.4,0.6])}, \\ [0.1, 0.4]e^{i2\pi([0.2,0.7])} \end{array} \right\}, \left\{ \begin{array}{c} [0.15, 0.3]e^{i2\pi([0.1,0.2])}, \\ [0.1, 0.2]e^{i2\pi([0.2,0.25])} \end{array} \right\} \right\} \right), \left(x_4, \left\{ \begin{array}{c} \left\{ \begin{array}{c} [0.4, 0.5]e^{i2\pi([0.4,0.6])}, \\ [0.2, 0.4]e^{i2\pi([0.3,0.5])} \end{array} \right\} \\ \left\{ \begin{array}{c} [0.2, 0.3]e^{i2\pi([0.1,0.3])}, \\ [0.1, 0.4]e^{i2\pi([0.2,0.25])} \end{array} \right\} \end{array} \right\} \right) \\ \left(x_5, \left\{ \left\{ [0.3, 0.4]e^{i2\pi([0.1,0.5])} \right\}, \left\{ [0.05, 0.2]e^{i2\pi([0.1,0.2])} \right\} \right\} \right) \end{array} \right\}$$

and

$$\mathcal{T} = \left\{ \begin{array}{c} \left(x_1, \left\{ \left\{ \begin{array}{c} [0.3, 0.6]e^{i2\pi([0.2,0.6])}, \\ [0.3, 0.55]e^{i2\pi([0.15,0.7])}, \\ [0.45, 0.6]e^{i2\pi([0.3,0.8])} \end{array} \right\}, \left\{ \begin{array}{c} [0.1, 0.2]e^{i2\pi([0.05,0.1])}, \\ [0.15, 0.3]e^{i2\pi([0.1,0.15])}, \\ [0.2, 0.3]e^{i2\pi([0.15,0.2])} \end{array} \right\} \right\} \right), \left(x_2, \left\{ \begin{array}{c} \left\{ [0.2, 0.3]e^{i2\pi([0.1,0.2])} \right\}, \\ \left\{ \begin{array}{c} [0.3, 0.6]e^{i2\pi([0.5,0.7])}, \\ [0.2, 0.5]e^{i2\pi([0.3,0.6])} \end{array} \right\} \end{array} \right\} \right), \\ \left(x_3, \left\{ \left\{ \begin{array}{c} [0.35, 0.5]e^{i2\pi([0.4,0.65])}, \\ [0.5, 0.65]e^{i2\pi([0.5,0.6])} \end{array} \right\}, \left\{ \begin{array}{c} [0.1, 0.15]e^{i2\pi([0.2,0.25])}, \\ [0.2, 0.3]e^{i2\pi([0.15,0.3])} \end{array} \right\} \right\} \right), \left(x_4, \left\{ \begin{array}{c} \left\{ [0.4, 0.45]e^{i2\pi([0.3,0.55])} \right\}, \\ \left\{ [0.1, 0.25]e^{i2\pi([0.2,0.45])} \right\} \end{array} \right\} \right) \\ \left(x_5, \left\{ \left\{ [0.05, 0.2]e^{i2\pi([0.1,0.2])} \right\}, \left\{ [0.3, 0.4]e^{i2\pi([0.1,0.5])} \right\} \right\} \right) \end{array} \right\}$$

be two CIvDHFSs. Then its operational laws are given as follows:

(1)

$$\mathcal{S}^c = \left\{ \begin{array}{c} \left(x_1, \left\{ \left\{ [0.1, 0.2]e^{i2\pi([0.05,0.1])} \right\}, \left\{ \begin{array}{c} [0.2, 0.6]e^{i2\pi([0.4,0.8])}, \\ [0.1, 0.4]e^{i2\pi([0.2,0.5])} \end{array} \right\} \right\} \right), \left(x_2, \left\{ \begin{array}{c} \left\{ [0.1, 0.25]e^{i2\pi([0.1,0.2])} \right\}, \\ \left\{ [0.4, 0.7]e^{i2\pi([0.6,0.7])} \right\} \end{array} \right\} \right) \\ \left(x_3, \left\{ \left\{ \begin{array}{c} [0.15, 0.3]e^{i2\pi([0.1,0.2])} \\ [0.1, 0.2]e^{i2\pi([0.2,0.25])} \end{array} \right\}, \left\{ \begin{array}{c} [0.5, 0.6]e^{i2\pi([0.3,0.5])}, \\ [0.5, 0.55]e^{i2\pi([0.4,0.6])}, \\ [0.1, 0.4]e^{i2\pi([0.2,0.7])} \end{array} \right\} \right\} \right), \left(x_4, \left\{ \begin{array}{c} \left\{ \begin{array}{c} [0.4, 0.45]e^{i2\pi([0.3,0.55])}, \\ [0.1, 0.4]e^{i2\pi([0.2,0.25])} \end{array} \right\}, \\ \left\{ [0.1, 0.2]e^{i2\pi([0.2,0.45])} \right\} \end{array} \right\} \right) \\ \left(x_5, \left\{ \left\{ [0.05, 0.2]e^{i2\pi([0.1,0.2])} \right\}, \left\{ [0.3, 0.4]e^{i2\pi([0.1,0.5])} \right\} \right\} \right) \end{array} \right\}$$

(2)

$$\mathcal{S}\cup\mathcal{T}=\left\{\begin{array}{c}\left(x_1,\left\{\left\{\begin{array}{c}[0.3,0.6]e^{i2\pi([0.4,0.8])},\\ [0.3,0.55]e^{i2\pi([0.5,0.7])},\\ [0.45,0.6]e^{i2\pi([0.3,0.8])}\end{array}\right\},\left\{[0.1,0.2]e^{i2\pi([0.05,0.1])}\right\}\right\}\right),\left(x_2,\left\{\begin{array}{c}\left\{[0.4,0.7]e^{i2\pi([0.6,0.7])}\right\},\\ \left\{[0.1,0.25]e^{i2\pi([0.15,0.2])}\right\}\end{array}\right\}\right),\\ \left(x_3,\left\{\left\{\begin{array}{c}[0.5,0.6]e^{i2\pi([0.4,0.65])},\\ [0.5,0.65]e^{i2\pi([0.5,0.6])},\\ [0.1,0.4]e^{i2\pi([0.2,0.7])}\end{array}\right\},\left\{\begin{array}{c}[0.1,0.15]e^{i2\pi([0.1,0.2])},\\ [0.1,0.2]e^{i2\pi([0.15,0.3])}\end{array}\right\}\right\}\right),\left(x_4\left\{\begin{array}{c}\left\{\begin{array}{c}[0.4,0.5]e^{i2\pi([0.4,0.6])},\\ [0.2,04]e^{i2\pi([0.3,0.5])}\end{array}\right\},\\ \left\{\begin{array}{c}[0.2,0.3]e^{i2\pi([0.1,0.3])},\\ [0.1,0.4]e^{i2\pi([0.2,0.25])}\end{array}\right\}\end{array}\right\}\right)\\ \left(x_5,\left\{\left\{[0.3,0.4]e^{i2\pi([0.1,0.5])}\right\},\left\{[0.05,0.2]e^{i2\pi([0.1,0.2])}\right\}\right\}\right)\end{array}\right\}$$

(3)

$$\mathcal{S}\cap\mathcal{T}=\left\{\begin{array}{c}\left(x_1,\left\{\left\{\begin{array}{c}[0.2,0.6]e^{i2\pi([0.2,0.6])},\\ [0.1,0.4]e^{i2\pi([0.2,0.5])}\end{array}\right\},\left\{\begin{array}{c}[0.1,0.2]e^{i2\pi([0.05,0.1])},\\ [0.15,0.3]e^{i2\pi([0.1,0.15])},\\ [0.2,0.3]e^{i2\pi([0.15,0.2])}\end{array}\right\}\right\}\right)\left(x_2,\left\{\begin{array}{c}\left\{[0.2,0.3]e^{i2\pi([0.1,0.2])}\right\},\\ \left\{\begin{array}{c}[0.3,0.6]e^{i2\pi([0.5,0.7])},\\ [0.2,0.5]e^{i2\pi([0.3,0.6])}\end{array}\right\}\end{array}\right\}\right)\\ \left(x_3,\left\{\left\{\begin{array}{c}[0.35,0.5]e^{i2\pi([0.3,0.5])},\\ [0.5,0.55]e^{i2\pi([0.4,0.6])}\end{array}\right\},\left\{\begin{array}{c}[0.15,0.3]e^{i2\pi([0.2,0.25])}\\ [0.2,0.3]e^{i2\pi([0.2,0.3])}\end{array}\right\}\right\}\right),\left(x_4,\left\{\begin{array}{c}\left\{[0.4,0.45]e^{i2\pi([0.3,0.55])}\right\},\\ \left\{\begin{array}{c}[0.2,0.3]e^{i2\pi([0.2,0.45])},\\ [0.1,0.4]e^{i2\pi([0.2,0.25])}\end{array}\right\}\end{array}\right\}\right)\\ \left(x_5,\left\{\left\{[0.05,0.2]e^{i2\pi([0.1,0.2])}\right\},\left\{[0.3,0.4]e^{i2\pi([0.1,0.5])}\right\}\right\}\right)\end{array}\right\}$$

## 4. The Similarity Measures For CDHFSs

The theory of similarity measures (SM) is an essential idea of the human point of view. SM assumes a significant job in numerous fields, for example, machine learning and decision making. In this section, we have two subsections, firstly, vector similarity measures (VSMs) and secondly, hybrid vector similarity measure.

*4.1. Vector Similarity Measures of CDHFSs.* The VSM is one of the important tools for the similarity degree between objects. We straightforwardly utilized Jaccard, Dice, and Cosine SM. Presently, in this segment, we characterize VSMs and weighted VSMs (WVSMs) for CDHFSs.

*Definition 12.* Suppose that $\mathcal{S}$ and $\mathcal{T}$ are two CDHFSs on $X$, then the Jaccard similarity measure (JSM) between $\mathcal{S}$ and $\mathcal{T}$ is denoted and defined as follows:

$$\mathrm{Jac}(\mathcal{S},\mathcal{T})=\frac{1}{n}\sum_{k=1}^{n}\left(\frac{(1/g)\sum_{p=1}^{g}\gamma_{\mathcal{S}_p}(x_k)\cdot\gamma_{\mathcal{T}_p}(x_k)+(1/g)\sum_{p=1}^{g}\omega_{\gamma_{\mathcal{S}_p}}(x_k)\cdot\omega_{\gamma_{\mathcal{T}_p}}(x_k)+(1/h)\sum_{q=1}^{h}\delta_{\mathcal{S}_q}(x_k)\cdot\delta_{\mathcal{T}_q}(x_k)+(1/h)\sum_{q=1}^{h}\varpi_{\delta_{\mathcal{S}_q}}(x_k)\cdot\varpi_{\delta_{\mathcal{T}_q}}(x_k)}{\begin{array}{c}(1/g)\sum_{p=1}^{g}\gamma^2_{\mathcal{S}_p}(x_k)+(1/g)\sum_{p=1}^{g}\gamma^2_{\mathcal{T}_p}(x_k)+(1/g)\sum_{p=1}^{g}\omega^2_{\gamma_{\mathcal{S}_p}}(x_k)+(1/g)\sum_{p=1}^{g}\omega^2_{\gamma_{\mathcal{T}_p}}(x_k)+(1/h)\sum_{q=1}^{h}\delta^2_{\mathcal{S}_q}(x_k)+(1/h)\sum_{q=1}^{h}\delta^2_{\mathcal{T}_q}(x_k)+(1/h)\sum_{q=1}^{h}\varpi^2_{\delta_{\mathcal{S}_q}}(x_k)+\\ (1/h)\sum_{q=1}^{h}\varpi^2_{\delta_{\mathcal{T}_q}}(x_k)-\left((1/g)\sum_{p=1}^{g}\gamma_{\mathcal{S}_p}(x_k)\cdot\gamma_{\mathcal{T}_p}(x_k)+(1/g)\sum_{p=1}^{g}\omega_{\gamma_{\mathcal{S}_p}}(x_k)\cdot\gamma_{\gamma_{\mathcal{T}_p}}(x_k)+(1/h)\sum_{q=1}^{h}\delta_{\mathcal{S}_q}(x_k)\cdot\delta_{\mathcal{T}_q}(x_k)+(1/h)\sum_{q=1}^{h}\varpi_{\delta_{\mathcal{S}_q}}(x_k)\cdot\varpi_{\delta_{\mathcal{T}_q}}(x_k)\right)\end{array}}\right). \tag{9}$$

JSMs fulfill the following axioms:

(1) $0\le\mathrm{Jac}(\mathcal{S},\mathcal{T})\le 1$

(2) $\mathrm{Jac}(\mathcal{S},\mathcal{T})=\mathrm{Jac}(\mathcal{T},\mathcal{S})$

(3) $\mathrm{Jac}(\mathcal{S},\mathcal{T})=1$, if $\mathcal{S}=\mathcal{T}$

**Theorem 1.** *Prove that equation (9) holds the above three conditions.*

*Proof*

(1) Since $(1/g)\sum_{p=1}^{g}\gamma_{\mathcal{S}_p}(x_k)\cdot\gamma_{\mathcal{T}_p}(x_k)\in[0,1]$, $(1/g)\sum_{p=1}^{g}\omega_{\gamma_{\mathcal{S}_p}}(x_k)\cdot\omega_{\gamma_{\mathcal{T}_p}}(x_k)\in[0,1]$, $(1/h)\sum_{q=1}^{h}\delta_{\mathcal{S}_q}(x_k)\cdot\delta_{\mathcal{T}_q}(x_k)\in[0,1]$, $(1/h)\sum_{q=1}^{h}\varpi_{\delta_{\mathcal{S}_q}}(x_k)\cdot\varpi_{\delta_{\mathcal{T}_q}}(x_k)\in[0,1]$, $(1/g)\sum_{p=1}^{g}\gamma^2_{\mathcal{S}_p}(x_k)\in[0,1]$, $(1/g)\sum_{q=1}^{g}\delta^2_{\mathcal{T}_p}(x_k)\in[0,1]$, $(1/g)\sum_{p=1}^{g}\omega^2_{\gamma_{\mathcal{S}_p}}(x_k)\in[0,1]$, $(1/g)\sum_{p=1}^{g}\omega^2_{\gamma_{\mathcal{T}_p}}(x_k)\in[0,1]$, $(1/h)\sum_{q=1}^{h}\delta^2_{\mathcal{S}_q}(x_k)\in[0,1]$, $(1/h)\sum_{q=1}^{h}\delta^2_{\mathcal{T}_q}(x_k)\in[0,1]$, $(1/h)\sum_{q=1}^{h}\varpi^2_{\delta_{\mathcal{S}_q}}(x_k)\in[0,1]$, $(1/h)\sum_{q=1}^{h}\varpi^2_{\delta_{\mathcal{T}_q}}(x_k)\in[0,1]$, and denominator will always remain greater than nominator, then for $k=1$,

$$\left(\frac{(1/g)\sum_{p=1}^{g}\gamma_{\mathcal{S}_p}(x_1)\cdot\gamma_{\mathcal{T}_p}(x_1)+(1/g)\sum_{p=1}^{g}\omega_{\gamma_{\mathcal{S}_p}}(x_1)\cdot\omega_{\gamma_{\mathcal{T}_p}}(x_1)+(1/h)\sum_{q=1}^{h}\delta_{\mathcal{S}_q}(x_1)\cdot\delta_{\mathcal{T}_q}(x_1)+(1/h)\sum_{q=1}^{h}\varpi_{\delta_{\mathcal{S}_q}}(x_1)\cdot\varpi_{\delta_{\mathcal{T}_q}}(x_1)}{(1/g)\sum_{p=1}^{g}\gamma^2_{\mathcal{S}_p}(x_1)+(1/g)\sum_{p=1}^{g}\gamma^2_{\mathcal{T}_p}(x_1)+(1/g)\sum_{p=1}^{g}\omega^2_{\gamma_{\mathcal{S}_p}}(x_1)+(1/g)\sum_{p=1}^{g}\omega^2_{\gamma_{\mathcal{T}_p}}(x_1)+(1/h)\sum_{q=1}^{h}\delta^2_{\mathcal{S}_q}(x_1)+(1/h)\sum_{q=1}^{h}\delta^2_{\mathcal{T}_q}(x_1)+(1/h)\sum_{q=1}^{h}\varpi^2_{\delta_{\mathcal{S}_q}}(x_1)+(1/h)\sum_{q=1}^{h}\varpi^2_{\delta_{\mathcal{T}_q}}(x_1)-\left((1/g)\sum_{p=1}^{g}\gamma_{\mathcal{S}_p}(x_1)\cdot\gamma_{\mathcal{T}_p}(x_1)+(1/g)\sum_{p=1}^{g}\omega_{\gamma_{\mathcal{S}_p}}(x_1)\cdot\gamma_{\gamma_{\mathcal{T}_p}}(x_1)+(1/h)\sum_{q=1}^{h}\delta_{\mathcal{S}_q}(x_1)\cdot\delta_{\mathcal{T}_q}(x_1)+(1/h)\sum_{q=1}^{h}\varpi_{\delta_{\mathcal{S}_q}}(x_1)\cdot\varpi_{\delta_{\mathcal{T}_q}}(x_1)\right)}\right)\in[0,1]. \tag{10}$$

For $k=2$,

$$\left(\frac{(1/g)\sum_{p=1}^{g}\gamma_{\mathcal{S}_p}(x_2)\cdot\gamma_{\mathcal{T}_p}(x_2)+(1/g)\sum_{p=1}^{g}\omega_{\gamma_{\mathcal{S}_p}}(x_2)\cdot\omega_{\gamma_{\mathcal{T}_p}}(x_2)+(1/h)\sum_{q=1}^{h}\delta_{\mathcal{S}_q}(x_2)\cdot\delta_{\mathcal{T}_q}(x_2)+(1/h)\sum_{q=1}^{h}\varpi_{\delta_{\mathcal{S}_q}}(x_2)\cdot\varpi_{\delta_{\mathcal{T}_q}}(x_2)}{(1/g)\sum_{p=1}^{g}\gamma^2_{\mathcal{S}_p}(x_2)+(1/g)\sum_{p=1}^{g}\gamma^2_{\mathcal{T}_p}(x_2)+(1/g)\sum_{p=1}^{g}\omega^2_{\gamma_{\mathcal{S}_p}}(x_2)+(1/g)\sum_{p=1}^{g}\omega^2_{\gamma_{\mathcal{T}_p}}(x_2)+(1/h)\sum_{q=1}^{h}\delta^2_{\mathcal{S}_q}(x_2)+(1/h)\sum_{q=1}^{h}\delta^2_{\mathcal{T}_q}(x_2)+(1/h)\sum_{q=1}^{h}\varpi^2_{\delta_{\mathcal{S}_q}}(x_2)+(1/h)\sum_{q=1}^{h}\varpi^2_{\delta_{\mathcal{T}_q}}(x_2)-\left((1/g)\sum_{p=1}^{g}\gamma_{\mathcal{S}_p}(x_2)\cdot\gamma_{\mathcal{T}_p}(x_2)+(1/g)\sum_{p=1}^{g}\omega_{\gamma_{\mathcal{S}_p}}(x_2)\cdot\gamma_{\gamma_{\mathcal{T}_p}}(x_2)+(1/h)\sum_{q=1}^{h}\delta_{\mathcal{S}_q}(x_2)\cdot\delta_{\mathcal{T}_q}(x_2)+(1/h)\sum_{q=1}^{h}\varpi_{\delta_{\mathcal{S}_q}}(x_2)\cdot\varpi_{\delta_{\mathcal{T}_q}}(x_2)\right)}\right)\in[0,1]. \tag{11}$$

By doing this process, we obtain

$$\sum_{k=1}^{n}\left(\frac{(1/g)\sum_{p=1}^{g}\gamma_{\mathcal{S}_p}(x_k)\cdot\gamma_{\mathcal{T}_p}(x_k)+(1/g)\sum_{p=1}^{g}\omega_{\gamma_{\mathcal{S}_p}}(x_k)\cdot\omega_{\gamma_{\mathcal{T}_p}}(x_k)+(1/h)\sum_{q=1}^{h}\delta_{\mathcal{S}_q}(x_k)\cdot\delta_{\mathcal{T}_q}(x_k)+(1/h)\sum_{q=1}^{h}\varpi_{\delta_{\mathcal{S}_q}}(x_k)\cdot\varpi_{\delta_{\mathcal{T}_q}}(x_k)}{(1/g)\sum_{p=1}^{g}\gamma^2_{\mathcal{S}_p}(x_k)+(1/g)\sum_{p=1}^{g}\gamma^2_{\mathcal{T}_p}(x_k)+(1/g)\sum_{p=1}^{g}\omega^2_{\gamma_{\mathcal{S}_p}}(x_k)+(1/g)\sum_{p=1}^{g}\omega^2_{\gamma_{\mathcal{T}_p}}(x_k)+(1/h)\sum_{q=1}^{h}\delta^2_{\mathcal{S}_q}(x_k)+(1/h)\sum_{q=1}^{h}\delta^2_{\mathcal{T}_q}(x_k)+(1/h)\sum_{q=1}^{h}\varpi^2_{\delta_{\mathcal{S}_q}}(x_k)}\right.$$
$$\left.+(1/h)\sum_{q=1}^{h}\varpi^2_{\delta_{\mathcal{T}_q}}(x_k)-\left((1/g)\sum_{p=1}^{g}\gamma_{\mathcal{S}_p}(x_k)\cdot\gamma_{\mathcal{T}_p}(x_k)+(1/g)\sum_{p=1}^{g}\omega_{\gamma_{\mathcal{S}_p}}(x_k)\cdot\gamma_{\gamma_{\mathcal{T}_p}}(x_k)+(1/h)\sum_{q=1}^{h}\delta_{\mathcal{S}_q}(x_k)\cdot\delta_{\mathcal{T}_q}(x_k)+(1/h)\sum_{q=1}^{h}\varpi_{\delta_{\mathcal{S}_q}}(x_k)\cdot\varpi_{\delta_{\mathcal{T}_q}}(x_k)\right)\right)$$
$$\in n[0,1],$$

$$0\le\sum_{k=1}^{n}\left(\frac{(1/g)\sum_{p=1}^{g}\gamma_{\mathcal{S}_p}(x_k)\cdot\gamma_{\mathcal{T}_p}(x_k)+(1/g)\sum_{p=1}^{g}\omega_{\gamma_{\mathcal{S}_p}}(x_k)\cdot\omega_{\gamma_{\mathcal{T}_p}}(x_k)+(1/h)\sum_{q=1}^{h}\delta_{\mathcal{S}_q}(x_k)\cdot\delta_{\mathcal{T}_q}(x_k)+(1/h)\sum_{q=1}^{h}\varpi_{\delta_{\mathcal{S}_q}}(x_k)\cdot\varpi_{\delta_{\mathcal{T}_q}}(x_k)}{(1/g)\sum_{p=1}^{g}\gamma^2_{\mathcal{S}_p}(x_k)+(1/g)\sum_{p=1}^{g}\gamma^2_{\mathcal{T}_p}(x_k)+(1/g)\sum_{p=1}^{g}\omega^2_{\gamma_{\mathcal{S}_p}}(x_k)+(1/g)\sum_{p=1}^{g}\omega^2_{\gamma_{\mathcal{T}_p}}(x_k)+(1/h)\sum_{q=1}^{h}\delta^2_{\mathcal{S}_q}(x_k)+(1/h)\sum_{q=1}^{h}\delta^2_{\mathcal{T}_q}(x_k)+(1/h)\sum_{q=1}^{h}\varpi^2_{\delta_{\mathcal{S}_q}}(x_k)}\right.$$
$$\left.+(1/h)\sum_{q=1}^{h}\varpi^2_{\delta_{\mathcal{T}_q}}(x_k)-\left((1/g)\sum_{p=1}^{g}\gamma_{\mathcal{S}_p}(x_k)\cdot\gamma_{\mathcal{T}_p}(x_k)+(1/g)\sum_{p=1}^{g}\omega_{\gamma_{\mathcal{S}_p}}(x_k)\cdot\gamma_{\gamma_{\mathcal{T}_p}}(x_k)+(1/h)\sum_{q=1}^{h}\delta_{\mathcal{S}_q}(x_k)\cdot\delta_{\mathcal{T}_q}(x_k)+(1/h)\sum_{q=1}^{h}\varpi_{\delta_{\mathcal{S}_q}}(x_k)\cdot\varpi_{\delta_{\mathcal{T}_q}}(x_k)\right)\right)$$
$$\le n,$$

$$0\le\frac{1}{n}\sum_{k=1}^{n}\left(\frac{(1/g)\sum_{p=1}^{g}\gamma_{\mathcal{S}_p}(x_k)\cdot\gamma_{\mathcal{T}_p}(x_k)+(1/g)\sum_{p=1}^{g}\omega_{\gamma_{\mathcal{S}_p}}(x_k)\cdot\omega_{\gamma_{\mathcal{T}_p}}(x_k)+(1/h)\sum_{q=1}^{h}\delta_{\mathcal{S}_q}(x_k)\cdot\delta_{\mathcal{T}_q}(x_k)+(1/h)\sum_{q=1}^{h}\varpi_{\delta_{\mathcal{S}_q}}(x_k)\cdot\varpi_{\delta_{\mathcal{T}_q}}(x_k)}{(1/g)\sum_{p=1}^{g}\gamma^2_{\mathcal{S}_p}(x_k)+(1/g)\sum_{p=1}^{g}\gamma^2_{\mathcal{T}_p}(x_k)+(1/g)\sum_{p=1}^{g}\omega^2_{\gamma_{\mathcal{S}_p}}(x_k)+(1/g)\sum_{p=1}^{g}\omega^2_{\gamma_{\mathcal{T}_p}}(x_k)+(1/h)\sum_{q=1}^{h}\delta^2_{\mathcal{S}_q}(x_k)+(1/h)\sum_{q=1}^{h}\delta^2_{\mathcal{T}_q}(x_k)+(1/h)\sum_{q=1}^{h}\varpi^2_{\delta_{\mathcal{S}_q}}(x_k)}\right.$$
$$\left.+(1/h)\sum_{q=1}^{h}\varpi^2_{\delta_{\mathcal{T}_q}}(x_k)-\left((1/g)\sum_{p=1}^{g}\gamma_{\mathcal{S}_p}(x_k)\cdot\gamma_{\mathcal{T}_p}(x_k)+(1/g)\sum_{p=1}^{g}\omega_{\gamma_{\mathcal{S}_p}}(x_k)\cdot\gamma_{\gamma_{\mathcal{T}_p}}(x_k)+(1/h)\sum_{q=1}^{h}\delta_{\mathcal{S}_q}(x_k)\cdot\delta_{\mathcal{T}_q}(x_k)+(1/h)\sum_{q=1}^{h}\varpi_{\delta_{\mathcal{S}_q}}(x_k)\cdot\varpi_{\delta_{\mathcal{T}_q}}(x_k)\right)\right)$$
$$\le 1,$$
$$0\le\mathrm{Jac}(\mathcal{S},\mathcal{T})\le 1. \tag{12}$$

(2) By definition of JSM, we have

$$\mathrm{Jac}(\mathcal{S},\mathcal{T})=\frac{1}{n}\sum_{k=1}^{n}\left(\frac{(1/g)\sum_{p=1}^{g}\gamma_{\mathcal{S}_p}(x_k)\cdot\gamma_{\mathcal{T}_p}(x_k)+(1/g)\sum_{p=1}^{g}\omega_{\gamma_{\mathcal{S}_p}}(x_k)\cdot\omega_{\gamma_{\mathcal{T}_p}}(x_k)+(1/h)\sum_{q=1}^{h}\delta_{\mathcal{S}_q}(x_k)\cdot\delta_{\mathcal{T}_q}(x_k)+(1/h)\sum_{q=1}^{h}\varpi_{\delta_{\mathcal{S}_q}}(x_k)\cdot\varpi_{\delta_{\mathcal{T}_q}}(x_k)}{(1/g)\sum_{px3d;1}^{g}\gamma^2_{\mathcal{S}_p}(x_k)+(1/g)\sum_{p=1}^{g}\gamma^2_{\mathcal{T}_p}(x_k)+(1/g)\sum_{p=1}^{g}\omega^2_{\gamma_{\mathcal{S}_p}}(x_k)+(1/g)\sum_{p=1}^{g}\omega^2_{\gamma_{\mathcal{T}_p}}(x_k)+(1/h)\sum_{q=1}^{h}\delta^2_{\mathcal{S}_q}(x_k)+(1/h)\sum_{q=1}^{h}\delta^2_{\mathcal{T}_q}(x_k)+(1/h)\sum_{q=1}^{h}\varpi^2_{\delta_{\mathcal{S}_q}}(x_k)}\right.$$
$$\left.+(1/h)\sum_{q=1}^{h}\varpi^2_{\delta_{\mathcal{T}_q}}(x_k)-\left((1/g)\sum_{p=1}^{g}\gamma_{\mathcal{S}_p}(x_k)\cdot\gamma_{\mathcal{T}_p}(x_k)+(1/g)\sum_{p=1}^{g}\omega_{\gamma_{\mathcal{S}_p}}(x_k)\cdot\gamma_{\gamma_{\mathcal{T}_p}}(x_k)+(1/h)\sum_{q=1}^{h}\delta_{\mathcal{S}_q}(x_k)\cdot\delta_{\mathcal{T}_q}(x_k)+(1/h)\sum_{q=1}^{h}\varpi_{\delta_{\mathcal{S}_q}}(x_k)\cdot\varpi_{\delta_{\mathcal{T}_q}}(x_k)\right)\right)$$
$$=\frac{1}{n}\sum_{k=1}^{n}\left(\frac{(1/g)\sum_{p=1}^{g}\gamma_{\mathcal{T}_p}(x_k)\cdot\gamma_{\mathcal{S}_p}(x_k)+(1/g)\sum_{p=1}^{g}\omega_{\gamma_{\mathcal{T}_p}}(x_k)\cdot\omega_{\gamma_{\mathcal{S}_p}}(x_k)+(1/h)\sum_{q=1}^{h}\delta_{\mathcal{T}_q}(x_k)\cdot\delta_{\mathcal{S}_q}(x_k)+(1/h)\sum_{q=1}^{h}\varpi_{\delta_{\mathcal{T}_q}}(x_k)\cdot\varpi_{\delta_{\mathcal{S}_q}}(x_k)}{(1/g)\sum_{p=1}^{g}\gamma^2_{\mathcal{T}_p}(x_k)+(1/g)\sum_{p=1}^{g}\gamma^2_{\mathcal{S}_p}(x_k)+(1/g)\sum_{p=1}^{g}\omega^2_{\gamma_{\mathcal{T}_p}}(x_k)+(1/g)\sum_{p=1}^{g}\omega^2_{\gamma_{\mathcal{S}_p}}(x_k)+(1/h)\sum_{q=1}^{h}\delta^2_{\mathcal{T}_q}(x_k)+(1/h)\sum_{q=1}^{h}\delta^2_{\mathcal{S}_q}(x_k)+(1/h)\sum_{q=1}^{h}\varpi^2_{\delta_{\mathcal{T}_q}}(x_k)}\right.$$
$$\left.+(1/h)\sum_{q=1}^{h}\varpi^2_{\delta_{\mathcal{S}_q}}(x_k)-\left((1/g)\sum_{p=1}^{g}\gamma_{\mathcal{T}_p}(x_k)\cdot\gamma_{\mathcal{S}_p}(x_k)+(1/g)\sum_{p=1}^{g}\omega_{\gamma_{\mathcal{T}_p}}(x_k)\cdot\gamma_{\gamma_{\mathcal{S}_p}}(x_k)+(1/h)\sum_{q=1}^{h}\delta_{\mathcal{T}_q}(x_k)\cdot\delta_{\mathcal{S}_q}(x_k)+(1/h)\sum_{q=1}^{h}\varpi_{\delta_{\mathcal{T}_q}}(x_k)\cdot\varpi_{\delta_{\mathcal{S}_q}}(x_k)\right)\right)$$
$$=\mathrm{Jac}(\mathcal{T},\mathcal{S}). \tag{13}$$

(3) By definition, we have

$$\mathrm{Jac}(\mathcal{S},\mathcal{T}) = \frac{1}{n}\sum_{k=1}^{n}\left(\frac{(1/g)\sum_{p=1}^{g}\gamma_{\mathcal{S}_p}(x_k)\cdot\gamma_{\mathcal{T}_p}(x_k)+(1/g)\sum_{p=1}^{g}\omega_{\gamma_{\mathcal{S}_p}}(x_k)\cdot\omega_{\gamma_{\mathcal{T}_p}}(x_k)+(1/h)\sum_{q=1}^{h}\delta_{\mathcal{S}_q}(x_k)\cdot\delta_{\mathcal{T}_q}(x_k)+(1/h)\sum_{q=1}^{h}\varpi_{\delta_{\mathcal{S}_q}}(x_k)\cdot\varpi_{\delta_{\mathcal{T}_q}}(x_k)}{(1/g)\sum_{p=1}^{g}\gamma^2_{\mathcal{S}_p}(x_k)+(1/g)\sum_{p=1}^{g}\gamma^2_{\mathcal{T}_p}(x_k)+(1/g)\sum_{p=1}^{g}\omega^2_{\gamma_{\mathcal{S}_p}}(x_k)+(1/g)\sum_{p=1}^{g}\omega^2_{\gamma_{\mathcal{T}_p}}(x_k)+(1/h)\sum_{q=1}^{h}\delta^2_{\mathcal{S}_q}(x_k)+(1/h)\sum_{q=1}^{h}\delta^2_{\mathcal{T}_q}(x_k)+(1/h)\sum_{q=1}^{h}\varpi^2_{\delta_{\mathcal{S}_q}}(x_k)} + (1/h)\sum_{q=1}^{h}\varpi^2_{\delta_{\mathcal{T}_q}}(x_k) - \left((1/g)\sum_{p=1}^{g}\gamma_{\mathcal{S}_p}(x_k)\cdot\gamma_{\mathcal{T}_p}(x_k)+(1/g)\sum_{p=1}^{g}\omega_{\gamma_{\mathcal{S}_p}}(x_k)\cdot\gamma_{\gamma_{\mathcal{T}_p}}(x_k)+(1/h)\sum_{q=1}^{h}\delta_{\mathcal{S}_q}(x_k)\cdot\delta_{\mathcal{T}_q}(x_k)+(1/h)\sum_{q=1}^{h}\varpi_{\delta_{\mathcal{S}_q}}(x_k)\cdot\varpi_{\delta_{\mathcal{T}_q}}(x_k)\right)\right). \tag{14}$$

Now as $\mathcal{S}=\mathcal{T}\Longleftrightarrow\mu_{\mathcal{S}}(x_k)=\mu_{\mathcal{T}}(x_k)$ and $\nu_{\mathcal{S}}(x_k)=\nu_{\mathcal{T}}(x_k)$ for $k=1,2,\ldots,n\Longleftrightarrow\gamma_{\mathcal{S}_p}(x_k)e^{i2\pi}(\omega_{\gamma_{\mathcal{S}_p}}(x_k))=\gamma_{\mathcal{T}_p}(x_k)e^{i2\pi}(\omega_{\gamma_{\mathcal{T}_p}}(x_k))$ and $\delta_{\mathcal{S}_q}(x_k)e^{i2\pi}(\varpi_{\gamma_{\mathcal{S}_q}})=\delta_{\mathcal{T}_q}(x_k)e^{i2\pi}(\varpi_{\gamma_{\mathcal{T}_q}})$, $k=1,2,\ldots,n\Longleftrightarrow\gamma_{\mathcal{S}_p}(x_k)=\gamma_{\mathcal{T}_p}(x_k),\delta_{\mathcal{S}_q}(x_k)=\delta_{\mathcal{T}_q}(x_k),e^{i2\pi}(\omega_{\gamma_{\mathcal{S}_p}}(x_k))=e^{i2\pi}(\varpi_{\gamma_{\mathcal{T}_p}}(x_k))$ and $e^{i2\pi}(\varpi_{\delta_{\mathcal{S}_q}}(x_k))=e^{i2\pi}(\varpi_{\delta_{\mathcal{T}_q}}(x_k))$, for $k=1,2,\ldots,n$. Then

$$\mathrm{Jac}(\mathcal{S},\mathcal{T}) = \frac{1}{n}\sum_{k=1}^{n}\left(\frac{(1/g)\sum_{p=1}^{g}\gamma^2_{\mathcal{S}_p}(x_k)+(1/g)\sum_{p=1}^{g}\omega^2_{\mathcal{S}_p}(x_k)+(1/h)\sum_{q=1}^{h}\delta^2_{\mathcal{S}_q}(x_k)+(1/h)\sum_{q=1}^{h}\varpi^2_{\delta_{\mathcal{S}_q}}(x_k)}{(2/g)\sum_{p=1}^{g}\gamma^2_{\mathcal{S}_p}(x_k)+(2/g)\sum_{p=1}^{g}\omega^2_{\gamma_{\mathcal{S}_p}}(x_k)+(2/h)\sum_{q=1}^{h}\delta^2_{\mathcal{S}_q}(x_k)+(2/h)\sum_{q=1}^{h}\varpi^2_{\delta_{\mathcal{S}_q}}(x_k)+(1/h)\sum_{q=1}^{h}\varpi^2_{\delta_{\mathcal{T}_q}}(x_k)} - \left((1/g)\sum_{p=1}^{g}\gamma^2_{\mathcal{S}_p}(x_k)+(1/g)\sum_{p=1}^{g}\omega^2_{\gamma_{\mathcal{S}_p}}(x_k)+(1/h)\sum_{q=1}^{h}\delta^2_{\mathcal{S}_q}(x_k)+(1/h)\sum_{q=1}^{h}\varpi^2_{\delta_{\mathcal{S}_q}}(x_k)\right)\right). \tag{15}$$

□

*Definition 13.* Suppose that $\mathcal{S}$ and $\mathcal{T}$ are two CDHFSs on $X$, then the weighted JSM (WJSM) between $\mathcal{S}$ and $\mathcal{T}$ is denoted and defined as follows:

$$\mathrm{Jac}_w(\mathcal{S},\mathcal{T}) = \frac{1}{n}\sum_{k=1}^{n}w_k\left(\frac{(1/g)\sum_{p=1}^{g}\gamma_{\mathcal{S}_p}(x_k)\cdot\gamma_{\mathcal{T}_p}(x_k)+(1/g)\sum_{p=1}^{g}\omega_{\gamma_{\mathcal{S}_p}}(x_k)\cdot\omega_{\gamma_{\mathcal{T}_p}}(x_k)+(1/h)\sum_{q=1}^{h}\delta_{\mathcal{S}_q}(x_k)\cdot\delta_{\mathcal{T}_q}(x_k)+(1/h)\sum_{q=1}^{h}\varpi_{\delta_{\mathcal{S}_q}}(x_k)\cdot\varpi_{\delta_{\mathcal{T}_q}}(x_k)}{(1/g)\sum_{p=1}^{g}\gamma^2_{\mathcal{S}_p}(x_k)+(1/g)\sum_{p=1}^{g}\gamma^2_{\mathcal{T}_p}(x_k)+(1/g)\sum_{p=1}^{g}\omega^2_{\gamma_{\mathcal{S}_p}}(x_k)+(1/g)\sum_{p=1}^{g}\omega^2_{\gamma_{\mathcal{T}_p}}(x_k)+(1/h)\sum_{q=1}^{h}\delta^2_{\mathcal{S}_q}(x_k)+(1/h)\sum_{q=1}^{h}\delta^2_{\mathcal{T}_q}(x_k)+(1/h)\sum_{q=1}^{h}\varpi^2_{\delta_{\mathcal{S}_q}}(x_k)} + (1/h)\sum_{q=1}^{h}\varpi^2_{\delta_{\mathcal{T}_q}}(x_k) - \left((1/g)\sum_{p=1}^{g}\gamma_{\mathcal{S}_p}(x_k)\cdot\gamma_{\mathcal{T}_p}(x_k)+(1/g)\sum_{p=1}^{g}\omega_{\gamma_{\mathcal{S}_p}}(x_k)\cdot\gamma_{\gamma_{\mathcal{T}_p}}(x_k)+(1/h)\sum_{q=1}^{h}\delta_{\mathcal{S}_q}(x_k)\cdot\delta_{\mathcal{T}_q}(x_k)+(1/h)\sum_{q=1}^{h}\varpi_{\delta_{\mathcal{S}_q}}(x_k)\cdot\varpi_{\delta_{\mathcal{T}_q}}(x_k)\right)\right), \tag{16}$$

where $w=(w_1,w_2,\ldots,w_n)^T$ speaks to the weight vector of every component $x_k\,(k=1,2,3,\ldots,n)$ contained in CDHFS and the weight vector fulfills $w_k\in[0,1]$ for each $k=1,2,\ldots,n$, $\sum_{k=1}^{n}w_k=1$. When we assume the weight vector be $w=((1/n),(1/n),\ldots,(1/n))$, at that point the WJSM will change into JSM. Otherwise, speaking when $w_k=(1/n), k=1,2,3,\ldots,n$, then $\mathrm{Jac}_w(\mathcal{S},\mathcal{T})=\mathrm{Jac}(\mathcal{S},\mathcal{T})$.

*Definition 14.* Suppose that $\mathcal{S}$ and $\mathcal{T}$ are two CDHFSs on $X$, then the Dice similarity measure (DSM) between $\mathcal{S}$ and $\mathcal{T}$ is denoted and defined as follows:

$$\mathrm{Dic}(\mathcal{S},\mathcal{T}) = \frac{1}{n}\sum_{k=1}^{n}\left(\frac{(2/g)\sum_{p=1}^{g}\gamma_{\mathcal{S}_p}^2(x_k)\cdot\gamma_{\mathcal{T}_p}^2(x_k)+(2/g)\sum_{p=1}^{g}\omega_{\gamma_{\mathcal{S}_p}}(x_k)\cdot\omega_{\gamma_{\mathcal{T}_p}}(x_k)+(2/h)\sum_{q=1}^{h}\delta_{\mathcal{S}_q}(x_k)\cdot\delta_{\mathcal{T}_q}(x_k)+(2/h)\sum_{q=1}^{h}\varpi_{\delta_{\mathcal{S}_q}}(x_k)\cdot\varpi_{\delta_{\mathcal{T}_q}}(x_k)}{\begin{array}{c}(1/g)\sum_{p=1}^{g}\gamma_{\mathcal{S}_p}^2(x_k)+(1/g)\sum_{p=1}^{g}\gamma_{\mathcal{T}_p}^2(x_k)+(1/g)\sum_{q=1}^{g}\omega_{\gamma_{\mathcal{S}_q}}^2(x_k)+(1/g)\sum_{q=1}^{g}\omega_{\gamma_{\mathcal{T}_q}}^2(x_k)\\ +(1/h)\sum_{q=1}^{h}\delta_{\mathcal{S}_q}^2(x_k)+(1/h)\sum_{q=1}^{h}\delta_{\mathcal{T}_q}^2(x_k)+(1/h)\sum_{q=1}^{h}\varpi_{\delta_{\mathcal{S}_q}}^2(x_k)+(1/h)\sum_{q=1}^{h}\varpi_{\delta_{\mathcal{T}_q}}^2(x_k)\end{array}}\right). \tag{17}$$

DSMs fulfills the following axioms:

(1) $0 \le \mathrm{Dic}(\mathcal{S},\mathcal{T}) \le 1$

(2) $\mathrm{Dic}(\mathcal{S},\mathcal{T}) = \mathrm{Dic}(\mathcal{T},\mathcal{S})$

(3) $\mathrm{Dic}(\mathcal{S},\mathcal{T}) = 1$, if $\mathcal{S} = \mathcal{T}$

**Theorem 2.** *Prove that equation (17) holds the above three conditions.*

*Proof.*

(1) Since $(1/g)\sum_{p=1}^{g}\gamma_{\mathcal{S}_p}(x_k)\cdot\gamma_{\mathcal{T}_p}(x_k)\in[0,1]$, $(1/g)\sum_{p=1}^{g}\gamma_{\mathcal{S}_p}(x_k)\cdot\gamma_{\gamma_{\mathcal{T}_p}}(x_k)\in[0,1]$, $(1/h)\sum_{q=1}^{h}\delta_{\mathcal{S}_q}(x_k)\cdot\delta_{\mathcal{T}_q}(x_k)\in[0,1]$, $(1/h)\sum_{q=1}^{h}\varpi_{\delta_{\mathcal{S}_q}}(x_k)\cdot\varpi_{\delta_{\mathcal{T}_q}}(x_k)\in[0,1]$, $(1/g)\sum_{p=1}^{g}\gamma_{\mathcal{S}_p}^2(x_k)\in[0,1]$, $(1/g)\sum_{q=1}^{g}\delta_{\mathcal{T}_p}^2(x_k)\in[0,1]$, $(1/g)\sum_{p=1}^{g}\omega_{\gamma_{\mathcal{S}_p}}^2(x_k)\in[0,1]$, $(1/g)\sum_{p=1}^{g}\omega_{\gamma_{\mathcal{T}_p}}^2(x_k)\in[0,1]$, $(1/h)\sum_{q=1}^{h}\delta_{\mathcal{S}_q}^2(x_k)\in[0,1]$, $(1/h)\sum_{q=1}^{h}\delta_{\mathcal{T}_q}^2(x_k)\in[0,1]$, $(1/h)\sum_{q=1}^{h}\varpi_{\delta_{\mathcal{S}_q}}^2(x_k)\in[0,1]$, $(1/h)\sum_{q=1}^{h}\varpi_{\delta_{\mathcal{T}_q}}^2(x_k)\in[0,1]$, and denominator will always remain greater than nominator, then for $k=1$,

$$\left(\frac{(2/g)\sum_{p=1}^{g}\gamma_{\mathcal{S}_p}^2(x_1)\cdot\gamma_{\mathcal{T}_p}^2(x_1)+(2/g)\sum_{p=1}^{g}\omega_{\gamma_{\mathcal{S}_p}}(x_1)\cdot\omega_{\gamma_{\mathcal{T}_p}}(x_1)+(2/h)\sum_{q=1}^{h}\delta_{\mathcal{S}_q}(x_1)\cdot\delta_{\mathcal{T}_q}(x_1)+(2/h)\sum_{q=1}^{h}\varpi_{\delta_{\mathcal{S}_q}}(x_1)\cdot\varpi_{\delta_{\mathcal{T}_q}}(x_1)}{(1/g)\sum_{p=1}^{g}\gamma_{\mathcal{S}_p}^2(x_1)+(1/g)\sum_{p=1}^{g}\gamma_{\mathcal{T}_p}^2(x_1)+(1/g)\sum_{q=1}^{g}\omega_{\gamma_{\mathcal{S}_q}}^2(x_1)+(1/g)\sum_{q=1}^{g}\omega_{\gamma_{\mathcal{T}_q}}^2(x_1)+(1/h)\sum_{q=1}^{h}\delta_{\mathcal{S}_q}^2(x_1)+(1/h)\sum_{q=1}^{h}\delta_{\mathcal{T}_q}^2(x_1)+(1/h)\sum_{q=1}^{h}\varpi_{\delta_{\mathcal{S}_q}}^2(x_1)+(1/h)\sum_{q=1}^{h}\varpi_{\delta_{\mathcal{T}_q}}^2(x_1)}\right) \in n[0,1]. \tag{18}$$

For $k=2$,

$$\left(\frac{(2/g)\sum_{p=1}^{g}\gamma_{\mathcal{S}_p}(x_1)\cdot\gamma_{\mathcal{T}_p}(x_2)+(2/g)\sum_{p=1}^{g}\omega_{\gamma_{\mathcal{S}_p}}(x_2)\cdot\omega_{\gamma_{\mathcal{T}_p}}(x_2)+(2/h)\sum_{q=1}^{h}\delta_{\mathcal{S}_q}(x_2)\cdot\delta_{\mathcal{T}_q}(x_2)+(2/h)\sum_{q=1}^{h}\varpi_{\delta_{\mathcal{S}_q}}(x_2)\cdot\varpi_{\delta_{\mathcal{T}_q}}(x_2)}{(1/g)\sum_{p=1}^{g}\gamma_{\mathcal{S}_p}^2(x_2)+(1/g)\sum_{p=1}^{g}\gamma_{\mathcal{T}_p}^2(x_2)+(1/g)\sum_{q=1}^{g}\omega_{\gamma_{\mathcal{S}_q}}^2(x_2)+(1/g)\sum_{q=1}^{g}\omega_{\gamma_{\mathcal{T}_q}}^2(x_2)+(1/h)\sum_{q=1}^{h}\delta_{\mathcal{S}_q}^2(x_2)+(1/h)\sum_{q=1}^{h}\delta_{\mathcal{T}_q}^2(x_2)+(1/h)\sum_{q=1}^{h}\varpi_{\delta_{\mathcal{S}_q}}^2(x_2)+(1/h)\sum_{q=1}^{h}\varpi_{\delta_{\mathcal{T}_q}}^2(x_2)}\right) \in [0,1]. \tag{19}$$

By doing this process, we obtain

$$0 \le \sum_{k=1}^{n}\left(\frac{(2/g)\sum_{p=1}^{g}\gamma_{\mathcal{S}_p}(x_k)\cdot\gamma_{\mathcal{T}_p}(x_k)+(2/g)\sum_{p=1}^{g}\omega_{\gamma_{\mathcal{S}_p}}(x_k)\cdot\omega_{\gamma_{\mathcal{T}_p}}(x_k)+(2/h)\sum_{q=1}^{h}\delta_{\mathcal{S}_q}(x_k)\cdot\delta_{\mathcal{T}_q}(x_k)+(2/h)\sum_{q=1}^{h}\varpi_{\delta_{\mathcal{S}_q}}(x_k)\cdot\varpi_{\delta_{\mathcal{T}_q}}(x_k)}{(1/g)\sum_{p=1}^{g}\gamma^2_{\mathcal{S}_p}(x_k)+(1/g)\sum_{p=1}^{g}\gamma^2_{\mathcal{T}_p}(x_k)+(1/g)\sum_{q=1}^{g}\omega^2_{\gamma_{\mathcal{S}_q}}(x_k)+(1/g)\sum_{q=1}^{g}\omega^2_{\gamma_{\mathcal{T}_q}}(x_k)+(1/h)\sum_{q=1}^{h}\delta^2_{\mathcal{S}_q}(x_k)+(1/h)\sum_{q=1}^{h}\delta^2_{\mathcal{T}_q}(x_k)+(1/h)\sum_{q=1}^{h}\varpi^2_{\delta_{\mathcal{S}_q}}(x_k)+(1/h)\sum_{q=1}^{h}\varpi^2_{\delta_{\mathcal{T}_q}}(x_k)}\right)\le n,$$

$$0 \le \sum_{k=1}^{n}\left(\frac{(2/g)\sum_{p=1}^{g}\gamma_{\mathcal{S}_p}(x_k)\cdot\gamma_{\mathcal{T}_p}(x_k)+(2/g)\sum_{p=1}^{g}\omega_{\gamma_{\mathcal{S}_p}}(x_k)\cdot\omega_{\gamma_{\mathcal{T}_p}}(x_k)+(2/h)\sum_{q=1}^{h}\delta_{\mathcal{S}_q}(x_k)\cdot\delta_{\mathcal{T}_q}(x_k)+(2/h)\sum_{q=1}^{h}\varpi_{\delta_{\mathcal{S}_q}}(x_k)\cdot\varpi_{\delta_{\mathcal{T}_q}}(x_k)}{(1/g)\sum_{p=1}^{g}\gamma^2_{\mathcal{S}_p}(x_k)+(1/g)\sum_{p=1}^{g}\gamma^2_{\mathcal{T}_p}(x_k)+(1/g)\sum_{q=1}^{g}\omega^2_{\gamma_{\mathcal{S}_q}}(x_k)+(1/g)\sum_{q=1}^{g}\omega^2_{\gamma_{\mathcal{T}_q}}(x_k)+(1/h)\sum_{q=1}^{h}\delta^2_{\mathcal{S}_q}(x_k)+(1/h)\sum_{q=1}^{h}\delta^2_{\mathcal{T}_q}(x_k)+(1/h)\sum_{q=1}^{h}\varpi^2_{\delta_{\mathcal{S}_q}}(x_k)+(1/h)\sum_{q=1}^{h}\varpi^2_{\delta_{\mathcal{T}_q}}(x_k)}\right)\le n,$$

$$0 \le \sum_{k=1}^{n}\left(\frac{(2/g)\sum_{p=1}^{g}\gamma_{\mathcal{S}_p}(x_k)\cdot\gamma_{\mathcal{T}_p}(x_k)+(2/g)\sum_{p=1}^{g}\omega_{\gamma_{\mathcal{S}_p}}(x_k)\cdot\omega_{\gamma_{\mathcal{T}_p}}(x_k)+(2/h)\sum_{q=1}^{h}\delta_{\mathcal{S}_q}(x_k)\cdot\delta_{\mathcal{T}_q}(x_k)+(2/h)\sum_{q=1}^{h}\varpi_{\delta_{\mathcal{S}_q}}(x_k)\cdot\varpi_{\delta_{\mathcal{T}_q}}(x_k)}{(1/g)\sum_{p=1}^{g}\gamma^2_{\mathcal{S}_p}(x_k)+(1/g)\sum_{p=1}^{g}\gamma^2_{\mathcal{T}_p}(x_k)+(1/g)\sum_{q=1}^{g}\omega^2_{\gamma_{\mathcal{S}_q}}(x_k)+(1/g)\sum_{q=1}^{g}\omega^2_{\gamma_{\mathcal{T}_q}}(x_k)+(1/h)\sum_{q=1}^{h}\delta^2_{\mathcal{S}_q}(x_k)+(1/h)\sum_{q=1}^{h}\delta^2_{\mathcal{T}_q}(x_k)+(1/h)\sum_{q=1}^{h}\varpi^2_{\delta_{\mathcal{S}_q}}(x_k)+(1/h)\sum_{q=1}^{h}\varpi^2_{\delta_{\mathcal{T}_q}}(x_k)}\right)\le 1,$$

$$0 \le \mathrm{Dic}(\mathcal{S},\mathcal{T}) \le 1. \tag{20}$$

(2) By definition of DSM, we have

$$\begin{aligned}
\mathrm{pac}(\mathcal{S},\mathcal{T}) &= \frac{1}{n}\sum_{k=1}^{n}\left(\frac{(2/g)\sum_{p=1}^{g}\gamma_{\mathcal{S}_p}(x_k)\cdot\gamma_{\mathcal{T}_p}(x_k)+(2/g)\sum_{p=1}^{g}\omega_{\gamma_{\mathcal{S}_p}}(x_k)\cdot\omega_{\gamma_{\mathcal{T}_p}}(x_k)+(2/h)\sum_{q=1}^{h}\delta_{\mathcal{S}_q}(x_k)\cdot\delta_{\mathcal{T}_q}(x_k)+(2/h)\sum_{q=1}^{h}\varpi_{\delta_{\mathcal{S}_q}}(x_k)\cdot\varpi_{\delta_{\mathcal{T}_q}}(x_k)}{(1/g)\sum_{p=1}^{g}\gamma^2_{\mathcal{S}_p}(x_k)+(1/g)\sum_{p=1}^{g}\gamma^2_{\mathcal{T}_p}(x_k)+(1/g)\sum_{q=1}^{g}\omega^2_{\gamma_{\mathcal{S}_q}}(x_k)+(1/g)\sum_{q=1}^{g}\omega^2_{\gamma_{\mathcal{T}_q}}(x_k)+(1/h)\sum_{q=1}^{h}\delta^2_{\mathcal{S}_q}(x_k)+(1/h)\sum_{q=1}^{h}\delta^2_{\mathcal{T}_q}(x_k)+(1/h)\sum_{q=1}^{h}\varpi^2_{\delta_{\mathcal{S}_q}}(x_k)+(1/h)\sum_{q=1}^{h}\varpi^2_{\delta_{\mathcal{T}_q}}(x_k)}\right)\\
&= \frac{1}{n}\sum_{k=1}^{n}\mathrm{pac}(\mathcal{S},\mathcal{T})\\
&= \frac{1}{n}\sum_{k=1}^{n}\left(\frac{(2/g)\sum_{p=1}^{g}\gamma_{\mathcal{T}_p}(x_1)\cdot\gamma_{\mathcal{S}_p}(x_k)+(2/g)\sum_{p=1}^{g}\omega_{\gamma_{\mathcal{T}_p}}(x_k)\cdot\omega_{\gamma_{\mathcal{S}_p}}(x_k)+(2/h)\sum_{q=1}^{h}\delta_{\mathcal{T}_q}(x_k)\cdot\delta_{\mathcal{S}_q}(x_k)+(2/h)\sum_{q=1}^{h}\varpi_{\delta_{\mathcal{T}_q}}(x_k)\cdot\varpi_{\delta_{\mathcal{S}_q}}(x_k)}{(1/g)\sum_{p=1}^{g}\gamma^2_{\mathcal{T}_p}(x_k)+(1/g)\sum_{p=1}^{g}\gamma^2_{\mathcal{S}_p}(x_k)+(1/g)\sum_{q=1}^{g}\omega^2_{\gamma_{\mathcal{T}_q}}(x_k)+(1/g)\sum_{q=1}^{g}\omega^2_{\gamma_{\mathcal{S}_q}}(x_k)+(1/h)\sum_{q=1}^{h}\delta^2_{\mathcal{T}_q}(x_k)+(1/h)\sum_{q=1}^{h}\delta^2_{\mathcal{S}_q}(x_k)+(1/h)\sum_{q=1}^{h}\varpi^2_{\delta_{\mathcal{T}_q}}(x_k)+(1/h)\sum_{q=1}^{h}\varpi^2_{\delta_{\mathcal{S}_q}}(x_k)}\right)\\
&= \mathrm{Dic}(\mathcal{T},\mathcal{S}).
\end{aligned} \tag{21}$$

(3) By definition, we have

$$\begin{aligned}
&\mathrm{Dic}(\mathcal{S},\mathcal{T})\\
&= \frac{1}{n}\sum_{k=1}^{n}\left(\frac{(2/g)\sum_{p=1}^{g}\gamma_{\mathcal{S}_p}(x_k)\cdot\gamma_{\mathcal{T}_p}(x_k)+(2/g)\sum_{p=1}^{g}\omega_{\gamma_{\mathcal{S}_p}}(x_k)\cdot\omega_{\gamma_{\mathcal{T}_p}}(x_k)+(2/h)\sum_{q=1}^{h}\delta_{\mathcal{S}_q}(x_k)\cdot\delta_{\mathcal{T}_q}(x_k)+(2/h)\sum_{q=1}^{h}\varpi_{\delta_{\mathcal{S}_q}}(x_k)\cdot\varpi_{\delta_{\mathcal{T}_q}}(x_k)}{(1/g)\sum_{p=1}^{g}\gamma^2_{\mathcal{S}_p}(x_k)+(1/g)\sum_{p=1}^{g}\gamma^2_{\mathcal{T}_p}(x_k)+(1/g)\sum_{q=1}^{g}\omega^2_{\gamma_{\mathcal{S}_q}}(x_k)+(1/g)\sum_{q=1}^{g}\omega^2_{\gamma_{\mathcal{T}_q}}(x_k)+(1/h)\sum_{q=1}^{h}\delta^2_{\mathcal{S}_q}(x_k)+(1/h)\sum_{q=1}^{h}\delta^2_{\mathcal{T}_q}(x_k)+(1/h)\sum_{q=1}^{h}\varpi^2_{\delta_{\mathcal{S}_q}}(x_k)+(1/h)\sum_{q=1}^{h}\varpi^2_{\delta_{\mathcal{T}_q}}(x_k)}\right).
\end{aligned} \tag{22}$$

Now as $\mathcal{S}=\mathcal{T} \Longleftrightarrow \mu_{\mathcal{S}}(x_k)=\mu_{\mathcal{T}}(x_k)$ and $\nu_{\mathcal{S}}(x_k)=\nu_{\mathcal{T}}(x_k)$ for $k=1,2,\ldots,n \Longleftrightarrow \gamma_{\mathcal{S}_p}(x_k)e^{i2\pi}(\omega_{\gamma_{\mathcal{S}_p}}(x_k))=\gamma_{\mathcal{T}_p}(x_k)e^{i2\pi}(\omega_{\gamma_{\mathcal{T}_p}}(x_k))$ and $\delta_{\mathcal{S}_q}(x_k)e^{i2\pi}(\varpi_{\gamma_{\mathcal{S}_q}})=\delta_{\mathcal{T}_q}(x_k)e^{i2\pi}(\varpi_{\gamma_{\mathcal{T}_q}})$, $k=1,2,\ldots,n \Longleftrightarrow \gamma_{\mathcal{S}_p}(x_k)=\gamma_{\mathcal{T}_p}(x_k), \delta_{\mathcal{S}_q}(x_k)=\delta_{\mathcal{T}_q}(x_k), e^{i2\pi}(\omega_{\gamma_{\mathcal{S}_p}}(x_k))=e^{i2\pi}(\omega_{\gamma_{\mathcal{T}_p}}(x_k))$ and $e^{i2\pi}(\varpi_{\delta_{\mathcal{S}_q}}(x_k))=e^{i2\pi}(\varpi_{\delta_{\mathcal{T}_q}}(x_k))$ for $k=1,2,\ldots,n$, then

$$\mathrm{Dic}(\mathcal{S},\mathcal{T})=\frac{1}{n}\sum_{k=1}^{n}\left(\frac{(2/g)\sum_{p=1}^{g}\gamma^2_{\mathcal{S}_p}(x_k)+(2/g)\sum_{p=1}^{g}\omega^2_{\mathcal{S}_p}(x_k)+(2/h)\sum_{q=1}^{h}\delta^2_{\mathcal{S}_q}(x_k)+(2/h)\sum_{q=1}^{h}\varpi^2_{\delta_{\mathcal{S}_q}}(x_k)}{(2/g)\sum_{p=1}^{g}\gamma^2_{\mathcal{S}_p}(x_k)+(2/g)\sum_{p=1}^{g}\omega^2_{\mathcal{S}_p}(x_k)+(2/h)\sum_{q=1}^{h}\delta^2_{\mathcal{S}_q}(x_k)+(2/h)\sum_{q=1}^{h}\varpi^2_{\delta_{\mathcal{S}_q}}(x_k)}\right)=1. \tag{23}$$

□

*Definition 15.* Suppose that $\mathcal{S}$ and $\mathcal{T}$ are two CDHFSs on $X$, then the weighted DSM (WJSM) between $\mathcal{S}$ and $\mathcal{T}$ is denoted and defined as follows:

$$\mathrm{Dic}(\mathcal{S},\mathcal{T}) = \sum_{k=1}^{n} w_k \left( \frac{(2/g)\sum_{p=1}^{g}\gamma_{\mathcal{S}_p}(x_k)\cdot\gamma_{\mathcal{T}_p}(x_k) + (2/g)\sum_{p=1}^{g}\omega_{\gamma_{\mathcal{S}_p}}(x_k)\cdot\omega_{\gamma_{\mathcal{T}_p}}(x_k) + (2/h)\sum_{q=1}^{h}\delta_{\mathcal{S}_q}(x_k)\cdot\delta_{\mathcal{T}_q}(x_k) + (2/h)\sum_{q=1}^{h}\varpi_{\delta_{\mathcal{S}_q}}(x_k)\cdot\varpi_{\delta_{\mathcal{T}_q}}(x_k)}{(1/g)\sum_{p=1}^{g}\gamma^2_{\mathcal{S}_p}(x_k) + (1/g)\sum_{p=1}^{g}\gamma^2_{\mathcal{T}_p}(x_k) + (1/g)\sum_{q=1}^{g}\omega^2_{\gamma_{\mathcal{S}_q}}(x_k) + (1/g)\sum_{q=1}^{g}\omega^2_{\gamma_{\mathcal{T}_q}}(x_k) + (1/h)\sum_{q=1}^{h}\delta^2_{\mathcal{S}_q}(x_k) + (1/h)\sum_{q=1}^{h}\delta^2_{\mathcal{T}_q}(x_k) + (1/h)\sum_{q=1}^{h}\varpi^2_{\delta_{\mathcal{S}_q}}(x_k) + (1/h)\sum_{q=1}^{h}\varpi^2_{\delta_{\mathcal{T}_q}}(x_k)} \right), \tag{24}$$

where $w = (w_1, w_2, \ldots, w_n)^T$ speaks to the weight vector of every component $x_k (k = 1, 2, 3, \ldots, n)$ contained in CDHFS and the weight vector fulfills $w_k \in [0, 1]$ for each $k = 1, 2, 3, \ldots, n$, $\sum_{k=1}^{n} w_k = 1$. When we assume the weight vector be $w = ((1/n), (1/n), \ldots, (1/n))^T$, at that point the WDSM will change into DSM. Otherwise, speaking when $w_k = (1/n), k = 1, 2, 3, \ldots, n$, then $\mathrm{Dic}_w(\mathcal{S},\mathcal{T}) = \mathrm{Dic}(\mathcal{S},\mathcal{T})$.

*Definition 16.* Suppose that $\mathcal{S}$ and $\mathcal{T}$ are two CDHFSs on $X$, then the Cosine similarity measure (CSM) between $\mathcal{S}$ and $\mathcal{T}$ is denoted and defined as follows:

$$\mathrm{Cos}(\mathcal{S},\mathcal{T}) = \frac{1}{n}\sum_{k=1}^{n} w_k \left( \frac{(1/g)\sum_{p=1}^{g}\gamma_{\mathcal{S}_p}(x_k)\cdot\gamma_{\mathcal{T}_p}(x_k) + (1/g)\sum_{p=1}^{g}\omega_{\gamma_{\mathcal{S}_p}}(x_k)\cdot\omega_{\gamma_{\mathcal{T}_p}}(x_k) + (1/h)\sum_{q=1}^{h}\delta_{\mathcal{S}_q}(x_k)\cdot\delta_{\mathcal{T}_q}(x_k) + (1/h)\sum_{q=1}^{h}\varpi_{\delta_{\mathcal{S}_q}}(x_k)\cdot\varpi_{\delta_{\mathcal{T}_q}}(x_k)}{\sqrt{(1/g)\sum_{p=1}^{g}\gamma^2_{\mathcal{S}_p}(x_k) + (1/g)\sum_{p=1}^{g}\omega^2_{\gamma_{\mathcal{S}_p}}(x_k) + (1/h)\sum_{q=1}^{h}\delta^2_{\mathcal{S}_q}(x_k) + (1/h)\sum_{q=1}^{h}\varpi^2_{\delta_{\mathcal{S}\mathcal{S}}}(x_k)}\sqrt{(1/h)\sum_{q=1}^{h}\delta^2_{\mathcal{T}_q}(x_k) + (1/h)\sum_{q=1}^{h}\varpi^2_{\delta_{\mathcal{T}_q}}(x_k) + (1/g)\sum_{q=1}^{g}\gamma^2_{\mathcal{T}_q}(x_k) + (1/g)\sum_{q=1}^{g}\varpi^2_{\gamma_{\mathcal{T}_q}}(x_k)}} \right). \tag{25}$$

CSMs fulfill the following axioms:

(1) $0 \le \mathrm{Cos}(\mathcal{S},\mathcal{T}) \le 1$

(2) $\mathrm{Cos}(\mathcal{S},\mathcal{T}) = \mathrm{Cos}(\mathcal{T},\mathcal{S})$

(3) $\mathrm{Cos}(\mathcal{S},\mathcal{T}) = 1$, if $\mathcal{S} = \mathcal{T}$

**Theorem 3.** *Prove that equation (25), holds the above three conditions.*

*Proof.*

(1) Since $(1/g)\sum_{p=1}^{g}\gamma_{\mathcal{S}_p}(x_k)\cdot\gamma_{\mathcal{T}_p}(x_k) \in [0,1]$, $(1/g)\sum_{p=1}^{g}\omega_{\gamma_{\mathcal{S}_p}}(x_k)\cdot\omega_{\gamma_{\mathcal{T}_p}}(x_k) \in [0,1]$, $(1/h)\sum_{q=1}^{h}\delta_{\mathcal{S}_q}(x_k)\cdot\delta_{\mathcal{T}_q}(x_k) \in [0,1]$, $(1/h)\sum_{q=1}^{h}\varpi_{\delta_{\mathcal{S}_q}}(x_k)\cdot\varpi_{\delta_{\mathcal{T}_q}}(x_k) \in [0,1]$, $(1/g)\sum_{p=1}^{g}\gamma^2_{\mathcal{S}_p}(x_k) \in [0,1]$, $(1/g)\sum_{q=1}^{g}\delta^2_{\mathcal{T}_p}(x_k) \in [0,1]$, $(1/g)\sum_{p=1}^{g}\omega^2_{\gamma_{\mathcal{S}_p}}(x_k) \in [0,1]$, $(1/g)\sum_{p=1}^{g}\omega^2_{\gamma_{\mathcal{T}_p}}(x_k) \in [0,1]$, $(1/h)\sum_{q=1}^{h}\delta^2_{\mathcal{S}_q}(x_k) \in [0,1]$, $(1/h)\sum_{q=1}^{h}\delta^2_{\mathcal{T}_q}(x_k) \in [0,1]$, $(1/h)\sum_{q=1}^{h}\varpi^2_{\delta_{\mathcal{S}_q}}(x_k) \in [0,1]$, $(1/h)\sum_{q=1}^{h}\varpi^2_{\delta_{\mathcal{T}_q}}(x_k) \in [0,1]$, and denominator will always remain greater than nominator, then for $k=1$,

$$\left( \frac{(1/g)\sum_{p=1}^{g}\gamma_{\mathcal{S}_p}(x_1)\cdot\gamma_{\mathcal{T}_p}(x_1) + (1/g)\sum_{p=1}^{g}\omega_{\gamma_{\mathcal{S}_p}}(x_1)\cdot\omega_{\gamma_{\mathcal{T}_p}}(x_1) + (1/h)\sum_{q=1}^{h}\delta_{\mathcal{S}_q}(x_1)\cdot\delta_{\mathcal{T}_q}(x_1) + (1/h)\sum_{q=1}^{h}\varpi_{\delta_{\mathcal{S}_q}}(x_1)\cdot\varpi_{\delta_{\mathcal{T}_q}}(x_1)}{\sqrt{(1/g)\sum_{p=1}^{g}\gamma^2_{\mathcal{S}_p}(x_1) + (1/g)\sum_{p=1}^{g}\omega^2_{\gamma_{\mathcal{S}_p}}(x_1) + 1/h\sum_{q=1}^{h}\delta^2_{\mathcal{S}_q}(x_1) + 1/h\sum_{q=1}^{h}\varpi^2_{\delta_{\mathcal{S}_q}}(x_1)}\sqrt{(1/h)\sum_{q=1}^{h}\delta^2_{\mathcal{T}_q}(x_1) + (1/h)\sum_{q=1}^{h}\varpi^2_{\delta_{\mathcal{T}_q}}(x_1) + (1/g)\sum_{q=1}^{g}\gamma^2_{\mathcal{T}_q}(x_1) + (1/g)\sum_{q=1}^{g}\varpi^2_{\gamma_{\mathcal{T}_q}}(x_1)}} \right) \in [0,1]. \tag{26}$$

For $k = 2$,

$$\left(\frac{(1/g)\sum_{p=1}^{g}\gamma_{\mathcal{S}_p}(x_2)\cdot\gamma_{\mathcal{T}_p}(x_2)+(1/g)\sum_{p=1}^{g}\omega_{\gamma_{\mathcal{S}_p}}(x_2)\cdot\omega_{\gamma_{\mathcal{T}_p}}(x_2)+(1/h)\sum_{q=1}^{h}\delta_{\mathcal{S}_q}(x_2)\cdot\delta_{\mathcal{T}_q}(x_2)+(1/h)\sum_{q=1}^{h}\varpi_{\delta_{\mathcal{S}_q}}(x_2)\cdot\varpi_{\delta_{\mathcal{T}_q}}(x_2)}{\sqrt{(1/g)\sum_{p=1}^{g}\gamma^2_{\mathcal{S}_p}(x_2)+(1/g)\sum_{p=1}^{g}\omega^2_{\gamma_{\mathcal{S}_p}}(x_2)+(1/h)\sum_{q=1}^{h}\delta^2_{\mathcal{S}_q}(x_2)+(1/h)\sum_{q=1}^{h}\varpi^2_{\delta_{\mathcal{S}_q}}(x_2)}\sqrt{(1/h)\sum_{q=1}^{h}\delta^2_{\mathcal{T}_q}(x_2)+(1/h)\sum_{q=1}^{h}\varpi^2_{\delta_{\mathcal{T}_q}}(x_2)+(1/g)\sum_{q=1}^{g}\gamma^2_{\mathcal{T}_q}(x_2)+(1/g)\sum_{q=1}^{g}\varpi^2_{\gamma_{\mathcal{T}_q}}(x_2)}}\right)$$

$$\in [0,1]. \tag{27}$$

By doing this process, we obtain

$$\sum_{k=1}^{n}\left(\frac{(1/g)\sum_{p=1}^{g}\gamma_{\mathcal{S}_p}(x_k)\cdot\gamma_{\mathcal{T}_p}(x_k)+(1/g)\sum_{p=1}^{g}\omega_{\gamma_{\mathcal{S}_p}}(x_k)\cdot\omega_{\gamma_{\mathcal{T}_p}}(x_k)+(1/h)\sum_{q=1}^{h}\delta_{\mathcal{S}_q}(x_{2k})\cdot\delta_{\mathcal{T}_q}(x_k)+(1/h)\sum_{q=1}^{h}\varpi_{\delta_{\mathcal{S}_q}}(x_k)\cdot\varpi_{\delta_{\mathcal{T}_q}}(x_k)}{\sqrt{(1/g)\sum_{p=1}^{g}\gamma^2_{\mathcal{S}_p}(x_k)+(1/g)\sum_{p=1}^{g}\omega^2_{\gamma_{\mathcal{S}_p}}(x_k)+(1/h)\sum_{q=1}^{h}\delta^2_{\mathcal{S}_q}(x_k)+(1/h)\sum_{q=1}^{h}\varpi^2_{\delta_{\mathcal{S}_q}}(x_k)}\sqrt{(1/h)\sum_{q=1}^{h}\delta^2_{\mathcal{T}_q}(x_k)+(1/h)\sum_{q=1}^{h}\varpi^2_{\delta_{\mathcal{T}_q}}(x_k)+(1/g)\sum_{q=1}^{g}\gamma^2_{\mathcal{T}_q}(x_k)+(1/g)\sum_{q=1}^{g}\omega^2_{\gamma_{\mathcal{T}_q}}(x_k)}}\right)\in n[0,1],$$

$$0\le\sum_{k=1}^{n}\left(\frac{(1/g)\sum_{p=1}^{g}\gamma_{\mathcal{S}_p}(x_k)\cdot\gamma_{\mathcal{T}_p}(x_k)+(1/g)\sum_{p=1}^{g}\omega_{\gamma_{\mathcal{S}_p}}(x_k)\cdot\omega_{\gamma_{\mathcal{T}_p}}(x_k)+(1/h)\sum_{q=1}^{h}\delta_{\mathcal{S}_q}(x_{2k})\cdot\delta_{\mathcal{T}_q}(x_k)+(1/h)\sum_{q=1}^{h}\varpi_{\delta_{\mathcal{S}_q}}(x_k)\cdot\varpi_{\delta_{\mathcal{T}_q}}(x_k)}{\sqrt{(1/g)\sum_{p=1}^{g}\gamma^2_{\mathcal{S}_p}(x_k)+(1/g)\sum_{p=1}^{g}\omega^2_{\gamma_{\mathcal{S}_p}}(x_k)+(1/h)\sum_{q=1}^{h}\delta^2_{\mathcal{S}_q}(x_k)+(1/h)\sum_{q=1}^{h}\varpi^2_{\delta_{\mathcal{S}_q}}(x_k)}\sqrt{(1/h)\sum_{q=1}^{h}\delta^2_{\mathcal{T}_q}(x_k)+(1/h)\sum_{q=1}^{h}\varpi^2_{\delta_{\mathcal{T}_q}}(x_k)+(1/g)\sum_{q=1}^{g}\gamma^2_{\mathcal{T}_q}(x_k)+(1/g)\sum_{q=1}^{g}\omega^2_{\gamma_{\mathcal{T}_q}}(x_k)}}\right)\le n,$$

$$0\le\frac{1}{n}\sum_{k=1}^{n}\left(\frac{(1/g)\sum_{p=1}^{g}\gamma_{\mathcal{S}_p}(x_k)\cdot\gamma_{\mathcal{T}_p}(x_k)+(1/g)\sum_{p=1}^{g}\omega_{\gamma_{\mathcal{S}_p}}(x_k)\cdot\omega_{\gamma_{\mathcal{T}_p}}(x_k)+(1/h)\sum_{q=1}^{h}\delta_{\mathcal{S}_q}(x_{2k})\cdot\delta_{\mathcal{T}_q}(x_k)+(1/h)\sum_{q=1}^{h}\varpi_{\delta_{\mathcal{S}_q}}(x_k)\cdot\varpi_{\delta_{\mathcal{T}_q}}(x_k)}{\sqrt{(1/g)\sum_{p=1}^{g}\gamma^2_{\mathcal{S}_p}(x_k)+(1/g)\sum_{p=1}^{g}\omega^2_{\gamma_{\mathcal{S}_p}}(x_k)+(1/h)\sum_{q=1}^{h}\delta^2_{\mathcal{S}_q}(x_k)+(1/h)\sum_{q=1}^{h}\varpi^2_{\delta_{\mathcal{S}_q}}(x_k)}\sqrt{(1/h)\sum_{q=1}^{h}\delta^2_{\mathcal{T}_q}(x_k)+(1/h)\sum_{q=1}^{h}\varpi^2_{\delta_{\mathcal{T}_q}}(x_k)+(1/g)\sum_{q=1}^{g}\gamma^2_{\mathcal{T}_q}(x_k)+(1/g)\sum_{q=1}^{g}\omega^2_{\gamma_{\mathcal{T}_q}}(x_k)}}\right)\le 1,$$

$$0\le \mathrm{Cos}(\mathcal{S},\mathcal{T})\le 1. \tag{28}$$

(2) By definition of CSM, we have

$$\begin{aligned}&\mathrm{Cos}(\mathcal{S},\mathcal{T})\\&=\frac{1}{n}\sum_{k=1}^{n}\left(\frac{(1/g)\sum_{p=1}^{g}\gamma_{\mathcal{S}_p}(x_k)\cdot\gamma_{\mathcal{T}_p}(x_k)+(1/g)\sum_{p=1}^{g}\omega_{\gamma_{\mathcal{S}_p}}(x_k)\cdot\omega_{\gamma_{\mathcal{T}_p}}(x_k)+(1/h)\sum_{q=1}^{h}\delta_{\mathcal{S}_q}(x_{2k})\cdot\delta_{\mathcal{T}_q}(x_k)+(1/h)\sum_{q=1}^{h}\varpi_{\delta_{\mathcal{S}_q}}(x_k)\cdot\varpi_{\delta_{\mathcal{T}_q}}(x_k)}{\sqrt{(1/g)\sum_{p=1}^{g}\gamma^2_{\mathcal{S}_p}(x_k)+(1/g)\sum_{p=1}^{g}\omega^2_{\gamma_{\mathcal{S}_p}}(x_k)+(1/h)\sum_{q=1}^{h}\delta^2_{\mathcal{S}_q}(x_k)+(1/h)\sum_{q=1}^{h}\varpi^2_{\delta_{\mathcal{S}_q}}(x_k)}\sqrt{(1/h)\sum_{q=1}^{h}\delta^2_{\mathcal{T}_q}(x_k)+(1/h)\sum_{q=1}^{h}\varpi^2_{\delta_{\mathcal{T}_q}}(x_k)+(1/g)\sum_{q=1}^{g}\gamma^2_{\mathcal{T}_q}(x_k)+(1/g)\sum_{q=1}^{g}\omega^2_{\gamma_{\mathcal{T}_q}}(x_k)}}\right)\\&=\mathrm{Cos}(\mathcal{T},\mathcal{S}).\end{aligned} \tag{29}$$

(3) By definition, we have

$$\begin{aligned}&\mathrm{Cos}(\mathcal{S},\mathcal{T})\\&=\frac{1}{n}\sum_{k=1}^{n}\left(\frac{(1/g)\sum_{p=1}^{g}\gamma_{\mathcal{S}_p}(x_k)\cdot\gamma_{\mathcal{T}_p}(x_k)+(1/g)\sum_{p=1}^{g}\omega_{\gamma_{\mathcal{S}_p}}(x_k)\cdot\omega_{\gamma_{\mathcal{T}_p}}(x_k)+(1/h)\sum_{q=1}^{h}\delta_{\mathcal{S}_q}(x_{2k})\cdot\delta_{\mathcal{T}_q}(x_k)+(1/h)\sum_{q=1}^{h}\varpi_{\delta_{\mathcal{S}_q}}(x_k),\ldots,(x_k)}{\sqrt{(1/g)\sum_{p=1}^{g}\gamma^2_{\mathcal{S}_p}(x_k)+(1/g)\sum_{p=1}^{g}\omega^2_{\gamma_{\mathcal{S}_p}}(x_k)+(1/h)\sum_{q=1}^{h}\delta^2_{\mathcal{S}_q}(x_k)+(1/h)\sum_{q=1}^{h}\varpi^2_{\delta_{\mathcal{S}_q}}(x_k)}\sqrt{(1/h)\sum_{q=1}^{h}\delta^2_{\mathcal{T}_q}(x_k)+(1/h)\sum_{q=1}^{h}\varpi^2_{\delta_{\mathcal{T}_q}}(x_k)+(1/g)\sum_{q=1}^{g}\gamma^2_{\mathcal{T}_q}(x_k)+(1/g)\sum_{q=1}^{g}\omega^2_{\gamma_{\mathcal{T}_q}}(x_k)}}\right).\end{aligned} \tag{30}$$

Now as $\mathcal{S}=\mathcal{T}\Longleftrightarrow\mu_{\mathcal{S}}(x_k)=\mu_{\mathcal{T}}(x_k)$ and $\nu_{\mathcal{S}}(x_k)=\nu_{\mathcal{T}}(x_k)$ for $k=1,2,\ldots,n\Longleftrightarrow\gamma_{\mathcal{S}_p}(x_k)e^{i2\pi}(\omega_{\gamma_{\mathcal{S}_p}}(x_k))=\gamma_{\mathcal{T}_p}(x_k)e^{i2\pi}(\omega_{\gamma_{\mathcal{T}_p}}(x_k))$ and $\delta_{\mathcal{S}_q}(x_k)e^{i2\pi}(\varpi_{\gamma_{\mathcal{S}_q}})=\delta_{\mathcal{T}_q}(x_k)e^{i2\pi}(\varpi_{\gamma_{\mathcal{T}_q}})$, $k=1,2,\ldots,n\Longleftrightarrow\gamma_{\mathcal{S}_p}(x_k)=\gamma_{\mathcal{T}_p}(x_k),\delta_{\mathcal{S}_q}(x_k)=\delta_{\mathcal{T}_q}(x_k),e^{i2\pi}(\omega_{\gamma_{\mathcal{S}_p}}(x_k))=e^{i2\pi}(\varpi_{\gamma_{\mathcal{T}_p}}(x_k))$ and $e^{i2\pi}(\varpi_{\delta_{\mathcal{S}_q}}(x_k))=e^{i2\pi}(\varpi_{\delta_{\mathcal{T}_q}}(x_k))$ for $k=1,2,\ldots,n$, then

$$\mathrm{Cos}(\mathcal{S},\mathcal{T}) = \frac{1}{n}\sum_{k=1}^{n}\left(\frac{(1/g)\sum_{p=1}^{g}\gamma_{\mathcal{S}_p}^2(x_k)+(1/g)\sum_{p=1}^{g}\omega_{\mathcal{S}_p}^2(x_k)+(1/h)\sum_{q=1}^{h}\delta_{\mathcal{S}_q}^2(x_{2k})+(1/h)\sum_{q=1}^{h}\varpi_{\delta_{\mathcal{S}_q}}^2(x_k)}{\left(\sqrt{(1/g)\sum_{p=1}^{g}\gamma_{\mathcal{S}_p}^2(x_k)+(1/g)\sum_{p=1}^{g}\omega_{\gamma_{\mathcal{S}_p}}^2(x_k)+(1/h)\sum_{q=1}^{h}\delta_{\mathcal{S}_q}^2(x_k)+(1/h)\sum_{q=1}^{h}\varpi_{\delta_{\mathcal{S}_q}}^2(x_k)}\right)^2}\right) = 1. \tag{31}$$

□

*Definition 17.* Suppose that $\mathcal{S}$ and $\mathcal{T}$ are two CDHFSs on $X$, then the weighted CSM (WCSM) between $\mathcal{S}$ and $\mathcal{T}$ is denoted and defined as follows:

$$\mathrm{Cos}(\mathcal{S},\mathcal{T}) = \sum_{k=1}^{n} w_k\left(\frac{(1/g)\sum_{p=1}^{g}\gamma_{\mathcal{S}_p}(x_k)\cdot\gamma_{\mathcal{T}_p}(x_k)+(1/g)\sum_{p=1}^{g}\omega_{\gamma_{\mathcal{S}_p}}(x_k)\cdot\omega_{\gamma_{\mathcal{T}_p}}(x_k)+(1/h)\sum_{q=1}^{h}\delta_{\mathcal{S}_q}(x_{2k})\cdot\delta_{\mathcal{T}_q}(x_k)+(1/h)\sum_{q=1}^{h}\varpi_{\delta_{\mathcal{S}_q}}(x_k)\cdot\varpi_{\delta_{\mathcal{T}_q}}(x_k)}{\sqrt{(1/g)\sum_{p=1}^{g}\gamma_{\mathcal{S}_p}^2(x_k)+(1/g)\sum_{p=1}^{g}\omega_{\gamma_{\mathcal{S}_p}}^2(x_k)+(1/h)\sum_{q=1}^{h}\delta_{\mathcal{S}_q}^2(x_k)+(1/h)\sum_{q=1}^{h}\varpi_{\delta_{\mathcal{S}_q}}^2(x_k)}\sqrt{(1/h)\sum_{q=1}^{h}\delta_{\mathcal{T}_q}^2(x_k)+(1/h)\sum_{q=1}^{h}\varpi_{\delta_{\mathcal{T}_q}}^2(x_k)+(1/g)\sum_{q=1}^{g}\gamma_{\mathcal{T}_q}^2(x_k)+(1/g)\sum_{q=1}^{g}\omega_{\gamma_{\mathcal{T}_q}}^2(x_k)}}\right), \tag{32}$$

where $w = (w_1, w_2, \ldots, w_n)^T$ speaks to the weight vector of every component $x_k\,(k = 1, 2, 3, \ldots, n)$ contained in CDHFS and the weight vector fulfills $w_k \in [0, 1]$ for each $k = 1, 2, 3, \ldots, n$, $\sum_{k=1}^{n} w_k = 1$. When we assume the weight vector be $w = ((1/n), (1/n), \ldots, (1/n))^T$, at that point the WCSM will change into CSM. Otherwise, speaking when $w_k = (1/n), k = 1, 2, 3, \ldots, n$, then $\mathrm{Dic}_w(\mathcal{S},\mathcal{T}) = \mathrm{Dic}(\mathcal{S},\mathcal{T})$.

*4.2. Hybrid Vector Similarity Measure.* In this section, we characterize hybrid vector similarity measure and weighted hybrid vector similarity measure for complex dual hesitant fuzzy sets.

*Definition 18.* Suppose that $\mathcal{S}$ and $\mathcal{T}$ are two CDHFSs on $X$, then the hybrid VSM between $\mathcal{S}$ and $\mathcal{T}$ is denoted and defined as follows:

$$\mathrm{Hyb}(\mathcal{S},\mathcal{T}) = \frac{1}{n}\left[\begin{array}{l}\lambda\sum_{k=1}^{n}\left(\frac{(2/g)\sum_{p=1}^{g}\gamma_{\mathcal{S}_p}(x_k)\cdot\gamma_{\mathcal{T}_p}(x_k)+(2/g)\sum_{p=1}^{g}\omega_{\gamma_{\mathcal{S}_p}}(x_k)\cdot\omega_{\gamma_{\mathcal{T}_p}}(x_k)+(2/h)\sum_{q=1}^{h}\delta_{\mathcal{S}_q}(x_{2k})\cdot\delta_{\mathcal{T}_q}(x_k)+(2/h)\sum_{q=1}^{h}\varpi_{\delta_{\mathcal{S}_q}}(x_k)\cdot\varpi_{\delta_{\mathcal{T}_q}}(x_k)}{\sqrt{(1/g)\sum_{p=1}^{g}\gamma_{\mathcal{S}_p}^2(x_k)+(1/g)\sum_{p=1}^{g}\omega_{\gamma_{\mathcal{S}_p}}^2(x_k)+(1/h)\sum_{q=1}^{h}\delta_{\mathcal{S}_q}^2(x_k)+(1/h)\sum_{q=1}^{h}\omega_{\delta_{\mathcal{S}_q}}^2(x_k)}\sqrt{(1/h)\sum_{q=1}^{h}\delta_{\mathcal{T}_q}^2(x_k)+(1/h)\sum_{q=1}^{h}\varpi_{\delta_{\mathcal{T}_q}}^2(x_k)+(1/g)\sum_{q=1}^{g}\gamma_{\mathcal{T}_q}^2(x_k)+(1/g)\sum_{q=1}^{g}\omega_{\gamma_{\mathcal{T}_q}}^2(x_k)}}\right)\\ +(1-\lambda)\sum_{k=1}^{n}\left(\frac{(1/g)\sum_{p=1}^{g}\gamma_{\mathcal{S}_p}(x_k)\cdot\gamma_{\mathcal{T}_p}(x_k)+(1/g)\sum_{p=1}^{g}\omega_{\gamma_{\mathcal{S}_p}}(x_k)\cdot\omega_{\gamma_{\mathcal{T}_p}}(x_k)+(1/h)\sum_{q=1}^{h}\delta_{\mathcal{S}_q}(x_{2k})\cdot\delta_{\mathcal{T}_q}(x_k)+(1/h)\sum_{q=1}^{h}\varpi_{\delta_{\mathcal{S}_q}}(x_k)\cdot\varpi_{\delta_{\mathcal{T}_q}}(x_k)}{\sqrt{(1/g)\sum_{p=1}^{g}\gamma_{\mathcal{S}_p}^2(x_k)+(1/g)\sum_{p=1}^{g}\omega_{\gamma_{\mathcal{S}_p}}^2(x_k)+(1/h)\sum_{q=1}^{h}\delta_{\mathcal{S}_q}^2(x_k)+(1/h)\sum_{q=1}^{h}\varpi_{\delta_{\mathcal{S}_q}}^2(x_k)}\sqrt{(1/h)\sum_{q=1}^{h}\delta_{\mathcal{T}_q}^2(x_k)+(1/h)\sum_{q=1}^{h}\varpi_{\delta_{\mathcal{T}_q}}^2(x_k)+(1/g)\sum_{q=1}^{g}\gamma_{\mathcal{T}_q}^2(x_k)+(1/g)\sum_{q=1}^{g}\omega_{\gamma_{\mathcal{T}_q}}^2(x_k)}}\right)\end{array}\right]. \tag{33}$$

*Definition 19.* Suppose that $\mathcal{S}$ and $\mathcal{T}$ are two CDHFSs on $X$, then the weighted hybrid VSM between $\mathcal{S}$ and $\mathcal{T}$ is denoted and defined as follows:

$$\mathrm{Hyb}_w(\mathcal{S},\mathcal{T}) = \left[\begin{array}{l}\lambda\sum_{k=1}^{n}\left(\frac{(2/g)\sum_{p=1}^{g}\gamma_{\mathcal{S}_p}(x_k)\cdot\gamma_{\mathcal{T}_p}(x_k)+(2/g)\sum_{p=1}^{g}\omega_{\gamma_{\mathcal{S}_p}}(x_k)\cdot\omega_{\gamma_{\mathcal{T}_p}}(x_k)+(2/h)\sum_{q=1}^{h}\delta_{\mathcal{S}_q}(x_{2k})\cdot\delta_{\mathcal{T}_q}(x_k)+(2/h)\sum_{q=1}^{h}\varpi_{\delta_{\mathcal{S}_q}}(x_k)\cdot\varpi_{\delta_{\mathcal{T}_q}}(x_k)}{\sqrt{(1/g)\sum_{p=1}^{g}\gamma_{\mathcal{S}_p}^2(x_k)+(1/g)\sum_{p=1}^{g}\omega_{\gamma_{\mathcal{S}_p}}^2(x_k)+(1/h)\sum_{q=1}^{h}\delta_{\mathcal{S}_q}^2(x_k)+(1/h)\sum_{q=1}^{h}\omega_{\delta_{\mathcal{S}_q}}^2(x_k)}\sqrt{(1/h)\sum_{q=1}^{h}\delta_{\mathcal{T}_q}^2(x_k)+(1/h)\sum_{q=1}^{h}\varpi_{\delta_{\mathcal{T}_q}}^2(x_k)+(1/g)\sum_{q=1}^{g}\gamma_{\mathcal{T}_q}^2(x_k)+(1/g)\sum_{q=1}^{g}\omega_{\gamma_{\mathcal{T}_q}}^2(x_k)}}\right)\\ +(1-\lambda)\sum_{k=1}^{n}\left(\frac{(1/g)\sum_{p=1}^{g}\gamma_{\mathcal{S}_p}(x_k)\cdot\gamma_{\mathcal{T}_p}(x_k)+(1/g)\sum_{p=1}^{g}\omega_{\gamma_{\mathcal{S}_p}}(x_k)\cdot\omega_{\gamma_{\mathcal{T}_p}}(x_k)+(1/h)\sum_{q=1}^{h}\delta_{\mathcal{S}_q}(x_{2k})\cdot\delta_{\mathcal{T}_q}(x_k)+(1/h)\sum_{q=1}^{h}\varpi_{\delta_{\mathcal{S}_q}}(x_k)\cdot\varpi_{\delta_{\mathcal{T}_q}}(x_k)}{\sqrt{(1/g)\sum_{p=1}^{g}\gamma_{\mathcal{S}_p}^2(x_k)+(1/g)\sum_{p=1}^{g}\omega_{\gamma_{\mathcal{S}_p}}^2(x_k)+(1/h)\sum_{q=1}^{h}\delta_{\mathcal{S}_q}^2(x_k)+(1/h)\sum_{q=1}^{h}\omega_{\delta_{\mathcal{S}_q}}^2(x_k)}\sqrt{(1/h)\sum_{q=1}^{h}\delta_{\mathcal{T}_q}^2(x_k)+(1/h)\sum_{q=1}^{h}\varpi_{\delta_{\mathcal{T}_q}}^2(x_k)+(1/g)\sum_{q=1}^{g}\gamma_{\mathcal{T}_q}^2(x_k)+(1/g)\sum_{q=1}^{g}\omega_{\gamma_{\mathcal{T}_q}}^2(x_k)}}\right)\end{array}\right]. \tag{34}$$

The weighted hybrid VSM of CDHFS fulfill the following axioms:

(1) $0 \leq \mathrm{Hyb}_w(\mathcal{S}, \mathcal{T}) \leq 1$

(2) $\mathrm{Hy}(\mathcal{S}, \mathcal{T}) = \mathrm{Hyb}_w(\mathcal{T}, \mathcal{S})$

(3) $\mathrm{Hyb}_w(\mathcal{S}, \mathcal{T}) = 1$, if $\mathcal{S} = \mathcal{T}$

**Theorem 4.** *Prove that equation (34) holds the above three conditions.*

*Proof* (1) From equations (24) and (32), we have $0 \leq \mathrm{Dic}_w \leq 1$ and $0 \leq \mathrm{Cos}_w \leq 1$ for all $k = 1, 2, \ldots, n$. Thus, equation (34) becomes

$$\begin{aligned} \mathrm{Hyb}_w(\mathcal{S}, \mathcal{T}) &= \lambda \mathrm{Dic}_w(\mathcal{S}, \mathcal{T}) + (1-\lambda)\mathrm{Cos}_w(\mathcal{S}, \mathcal{T}) \\ &\leq \lambda + (1-\lambda) = 1, \end{aligned} \tag{35}$$

because $\mathrm{Dic}_w(\mathcal{S}, \mathcal{T}) \geq 0$ and $\mathrm{Cos}_w(\mathcal{S}, \mathcal{T}) \geq 0$; this implies that $\mathrm{Hyb}_w(\mathcal{S}, \mathcal{T}) \geq 0$ for any value of $\lambda \in [0, 1]$. This completes the prove of 1.

(2) Since $\mathrm{Dic}_w(\mathcal{S}, \mathcal{T}) = \mathrm{Dic}_w(\mathcal{T}, \mathcal{S})$ and $\mathrm{Cos}_w(\mathcal{S}, \mathcal{T}) = \mathrm{Cos}_w(\mathcal{T}, \mathcal{S})$, this implies that.

(3) If $\mathcal{S} = \mathcal{T}$, then from equation (24) and (32) we have $\mathrm{Dic}_w(\mathcal{S}, \mathcal{T}) = 1$ and $\mathrm{Cos}_w(\mathcal{S}, \mathcal{T}) = 1$; this implies that $\mathrm{Hyb}_w(\mathcal{S}, \mathcal{T}) = \lambda + (1-\lambda) = 1$. □

## 5. Applications

In this section of the article, we portray two applications, for example, pattern recognition and medical diagnosis to show the viability and value of the deciphered SMS. We applied the explored SMs to the environment of CDHFSs in pattern recognition and medical diagnosis.

*5.1. Pattern Recognition.* The tools of similarity measures have applications in pattern classification. In such a marvel, the class of an obscure pattern or object discovered abuse some closeness estimating instruments and scarcely any inclinations of decision makers. In this section, the similarity measures developed up to this point in Section 4 are applied to a pattern recognition (building pattern recognition) drawback any place the class of an obscure building substance has been evaluated. The results procured using the SMs of CDHFSs are then explored for delineation of the upsides of proposed SMs and the containments of existing work. To explain the wonder, the underneath explained is examined.

*Example 3.* Its feed that amounts from developments by an organization is legitimately corresponding to the standard of building substances they use. An appropriate review of the building substance before development is the confirmation of good building measures. The building substances to be utilized ought to be carefully checked before applying. The best possible check and equalization arrangement of investigation approve the manufacturers to utilize the correct substances for developments to improve the standard of their task. Five known building substances $\mathcal{S}_r$ $(r = 1, 2, 3, 4, 5)$ which are given in the CDHFSs structure are as follows:

$$\begin{aligned}
\mathcal{S}_1 = \{&\left(x_1, \left\{\left\{0.2e^{i2\pi(0.15)}, 0.3e^{i2\pi(0.2)}, 0.5e^{i2\pi(0.4)}\right\}, \left\{0.3e^{i2\pi(0.5)}, 0.2e^{i2\pi(0.25)}\right\}\right\}\right), \left(x_2, \left\{\left\{0.3e^{i2\pi(0.2)}\right\}, \left\{0.6e^{i2\pi(0.5)}, 0.4e^{i2\pi(0.35)}\right\}\right\}\right), \\
&\left(x_3, \left\{\left\{0.6e^{i2\pi(0.1)}\right\}, \left\{0.2e^{i2\pi(0.4)}\right\}\right\}\right), \left(x_4, \left\{\left\{0.15e^{i2\pi(0.25)}, 0.6e^{i2\pi(0.5)}\right\}, \left\{0.35e^{i2\pi(0.15)}\right\}\right\}\right), \left(x_5, \left\{\left\{0.4e^{i2\pi(0.65)}\right\}, \left\{0.3e^{i2\pi(0.2)}\right\}\right\}\right)\}, \\
\mathcal{S}_2 = \{&\left(x_1, \left\{\left\{0.25e^{i2\pi(0.1)}\right\}, \left\{0.6e^{i2\pi(0.7)}, 0.3e^{i2\pi(0.5)}\right\}\right\}\right), \left(x_2, \left\{\left\{0.35e^{i2\pi(0.3)}\right\}, \left\{0.5e^{i2\pi(0.4)}\right\}\right\}\right), \\
&\left(x_3, \left\{\left\{0.35e^{i2\pi(0.4)}, 0.45e^{i2\pi(0.5)}, 0.6e^{i2\pi(0.35)}\right\}, \left\{0.3e^{i2\pi(0.45)}, 0.1e^{i2\pi(0.25)}\right\}\right\}\right), \\
&\left(x_4, \left\{\left\{0.6e^{i2\pi(0.33)}, 0.5e^{i2\pi(0.45)}\right\}, \left\{0.2e^{i2\pi(0.5)}, 0.25e^{i2\pi(0.3)}\right\}\right\}\right)\left(x_5, \left\{\left\{0.5e^{i2\pi(0.4)}\right\}, \left\{0.25e^{i2\pi(0.45)}, 0.4e^{i2\pi(0.35)}\right\}\right\}\right)\}, \\
\mathcal{S}_3 = \{&\left(x_1, \left\{\left\{0.3e^{i2\pi(0.4)}, 0.2e^{i2\pi(0.45)}\right\}, \left\{0.4e^{i2\pi(0.3)}, 0.5e^{i2\pi(0.1)}\right\}\right\}\right), \left(x_2, \left\{\left\{0.45e^{i2\pi(0.7)}, 0.5e^{i2\pi(0.65)}\right\}, \left\{0.15e^{i2\pi(0.2)}\right\}\right\}\right), \\
&\left(x_3, \left\{\left\{0.6e^{i2\pi(0.5)}\right\}, \left\{0.35e^{i2\pi(0.4)}\right\}\right\}\right), \left(x_4, \left\{\left\{0.7e^{i2\pi(0.5)}\right\}, \left\{0.25e^{i2\pi(0.3)}, 0.15e^{i2\pi(0.1)}\right\}\right\}\right), \\
&\left(x_5, \left\{\left\{0.1e^{i2\pi(0.05)}, 0.25e^{i2\pi(0.15)}, 0.4e^{i2\pi(0.45)}\right\}, \left\{0.5e^{i2\pi(0.5)}, 0.2e^{i2\pi(0.3)}\right\}\right\}\right)\},
\end{aligned}$$

Table 1: Estimations of the interpreted VSMs.

| Similarity measures | $(\mathcal{S}, \mathcal{S}_1)$ | $(\mathcal{S}, \mathcal{S}_2)$ | $(\mathcal{S}, \mathcal{S}_3)$ | $(\mathcal{S}, \mathcal{S}_4)$ | $(\mathcal{S}, \mathcal{S}_5)$ | Ranking |
|---|---|---|---|---|---|---|
| $\mathrm{Jac}(\mathcal{S}, \mathcal{S}_r)$ | 0.3936 | 0.4455 | 0.2748 | 0.6742 | 0.4087 | $\mathcal{S}_4 \geq \mathcal{S}_2 \geq \mathcal{S}_5 \geq \mathcal{S}_1 \geq \mathcal{S}_3$ |
| $\mathrm{Dic}(\mathcal{S}, \mathcal{S}_r)$ | 0.5353 | 0.581 | 0.4253 | 0.7755 | 0.5504 | $\mathcal{S}_4 \geq \mathcal{S}_2 \geq \mathcal{S}_5 \geq \mathcal{S}_1 \geq \mathcal{S}_3$ |
| $\mathrm{Cos}(\mathcal{S}, \mathcal{S}_r)$ | 0.5465 | 0.6048 | 0.447 | 0.7444 | 0.557 | $\mathcal{S}_4 \geq \mathcal{S}_2 \geq \mathcal{S}_5 \geq \mathcal{S}_1 \geq \mathcal{S}_3$ |

Table 2: Estimations of the interpreted WVSMs.

| Similarity measures | $(\mathcal{S}, \mathcal{S}_1)$ | $(\mathcal{S}, \mathcal{S}_2)$ | $(\mathcal{S}, \mathcal{S}_3)$ | $(\mathcal{S}, \mathcal{S}_4)$ | $(\mathcal{S}, \mathcal{S}_5)$ | Ranking |
|---|---|---|---|---|---|---|
| $\mathrm{Jac}_w(\mathcal{S}, \mathcal{S}_r)$ | 0.3697 | 0.4635 | 0.285 | 0.6618 | 0.3711 | $\mathcal{S}_4 \geq \mathcal{S}_2 \geq \mathcal{S}_5 \geq \mathcal{S}_1 \geq \mathcal{S}_3$ |
| $\mathrm{Dic}_w(\mathcal{S}, \mathcal{S}_r)$ | 0.5131 | 0.6042 | 0.4382 | 0.7468 | 0.5116 | $\mathcal{S}_4 \geq \mathcal{S}_2 \geq \mathcal{S}_5 \geq \mathcal{S}_1 \geq \mathcal{S}_3$ |
| $\mathrm{Cos}_w(\mathcal{S}, \mathcal{S}_\mathbf{r})$ | 0.5239 | 0.6342 | 0.4561 | 0.7199 | 0.5173 | $\mathcal{S}_4 \geq \mathcal{S}_2 \geq \mathcal{S}_1 \geq \mathcal{S}_2 \geq \mathcal{S}_3$ |

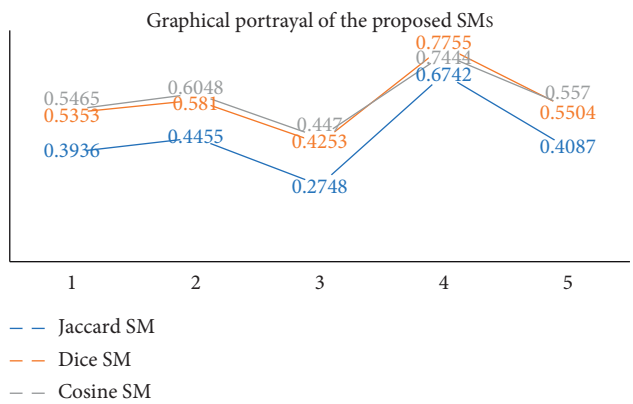



Figure 1: The graphical depiction of the deciphered VSMs between $\mathcal{S}$ and $\mathcal{S}_r$ $(r = 1, 2, 3, 4, 5)$.

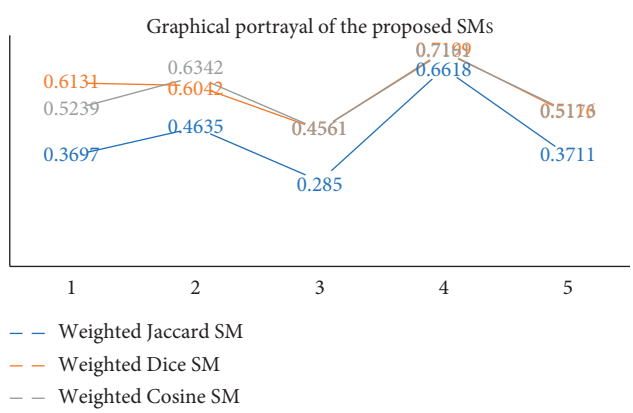



Figure 2: The graphical depiction of the deciphered WVSMs between $\mathcal{S}$ and $\mathcal{S}_r$ $(r = 1, 2, 3, 4, 5)$.

$$\begin{aligned}\mathcal{S}_4 = \{&\left(x_1, \{\{0.4e^{i2\pi(0.6)}, 0.6e^{i2\pi(0.7)}\}, \{0.15e^{i2\pi(0.1)}, 0.3e^{i2\pi(0.2)}\}\}\right), \left(x_2, \{\{0.3e^{i2\pi(0.4)}\}, \{0.6e^{i2\pi(0.5)}\}\}\right),\\ &\left(x_3, \{\{0.4e^{i2\pi(0.3)}, 0.2e^{i2\pi(0.4)}, 0.6e^{i2\pi(0.5)}\}, \{0.2e^{i2\pi(0.4)}, 0.15e^{i2\pi(0.3)}\}\}\right), \left(x_4, \{\{0.5e^{i2\pi(0.7)}, 0.3e^{i2\pi(0.6)}\}, \{0.4e^{i2\pi(0.2)}\}\}\right),\\ &\left(x_5, \{\{0.6e^{i2\pi(0.4)}, 0.55e^{i2\pi(0.55)}, 0.3e^{i2\pi(0.35)}\}, \{0.35e^{i2\pi(0.25)}, 0.15e^{i2\pi(0.3)}, 0.1e^{i2\pi(0.2)}\}\}\right)\},\\ \mathcal{S}_5 = \{&\left(x_1, \{\{0.3e^{i2\pi(0.2)}, 0.2e^{i2\pi(0.45)}\}, \{0.5e^{i2\pi(0.1)}, 0.4e^{i2\pi(0.3)}\}\}\right), \left(x_2, \{\{0.45e^{i2\pi(0.6)}, 0.7e^{i2\pi(0.55)}\}, \{0.25e^{i2\pi(0.35)}\}\}\right),\\ &\left(x_3, \{\{0.2e^{i2\pi(0.35)}\}, \{0.6e^{i2\pi(0.6)}\}\}\right), \left(x_4, \{\{0.7e^{i2\pi(0.5)}\}, \{0.15e^{i2\pi(0.2)}, 0.25e^{i2\pi(0.3)}\}\}\right),\\ &\left(x_5, \{\{0.4e^{i2\pi(0.3)}, 0.25e^{i2\pi(0.45)}, 0.1e^{i2\pi(0.5)}\}, \{0.5e^{i2\pi(0.4)}, 0.5e^{i2\pi(0.3)}, 0.3e^{i2\pi(0.35)}\}\}\right)\}.\end{aligned} \tag{36}$$

Now an unknown building substance which needs to be identified is as follows:

$$\begin{aligned}\mathcal{S} = \{&\left(x_1, \{\{\{0.7e^{i2\pi(0.4)}, 0.3e^{i2\pi(0.6)}\}, \{0.2e^{i2\pi(0.2)}, 0.1e^{i2\pi(0.3)}\}\}\}\right), \left(x_2, \{\{0.7e^{i2\pi(0.6)}\}, \{0.2e^{i2\pi(0.3)}\}\}\right),\\ &\left(x_3, \{\{0.6e^{i2\pi(0.5)}, 0.5e^{i2\pi(0.4)}, 0.4e^{i2\pi(0.7)}\}, \{0.3e^{i2\pi(0.1)}, 0.1e^{i2\pi(0.2)}\}\}\right), \left(x_4, \{\{0.4e^{i2\pi(0.5)}, 0.6e^{i2\pi(0.8)}\}, \{0.15e^{i2\pi(0.1)}\}\}\right),\\ &\left(x_5, \{\{0.5e^{i2\pi(0.6)}, 0.4e^{i2\pi(0.5)}, 0.2e^{i2\pi(0.4)}\}, \{0.3e^{i2\pi(0.2)}, 0.1e^{i2\pi(0.2)}, 0.2e^{i2\pi(0.3)}\}\}\right)\}.\end{aligned} \tag{37}$$

The goal of this reliance is to sort the unknown building substance $\mathcal{S}$ in one of the arrangements $\mathcal{S}_r$ $(r = 1, 2, 3, 4, 5)$. For this, the deciphered VSMs (Jaccard, Dice, and Cosine) are used to check the similarity from $\mathcal{S}$ to $\mathcal{S}_r$ $(r = 1, 2, 3, 4, 5)$ and calculations are given in Tables 1 and 2.

As indicated by the above-figured estimations provided in Table 1, we basically notice that the similarity level between $\mathcal{S}$ and $\mathcal{S}_4$ is the best one as implantation by all VSMs. This characterizes that VSMs (Jaccard, Dice, and Cosine) give out the unknown building substance $\mathcal{S}$ to the known building substance $\mathcal{S}_4$ reliant on the standard of the best degree of similarity. Situating of the researched VSMs between and $\mathcal{S}_r$ $(r = 1, 2, 3, 4, 5)$ is likewise given in Table 1. The graphical depiction of the deciphered VSMs between $\mathcal{S}$ and $\mathcal{S}_r$ $(r = 1, 2, 3, 4, 5)$ is exhibited in Figure 1.

The weight of segments has exceptional importance to consider in certifiable powerful issues. If we surmise the greatness of segments $x_k$ $(k = 1, 2, 3, 4, 5)$ be $w_k$ $(0.2, 0.25, 0.3, 0.15, 0.1)$, respectively. Now the interpreted WVSMs are used to choose the similarity from $\mathcal{S}$ and $\mathcal{S}_r$ $(r = 1, 2, 3, 4, 5)$ and estimations are given in Table 2.

As indicated by the above-figured estimations provided in Table 2, we basically notice that the similarity level between $\mathcal{S}$ and $\mathcal{S}_4$ is the best one as implantation by all WVSMs. This characterizes that WVSMs give out the unknown building substance $\mathcal{S}$ to the known building substance $\mathcal{S}_4$ reliant on the standard of the best degree of similarity. Situating of the researched VSMs between $\mathcal{S}$ and $\mathcal{S}_r$ $(r = 1, 2, 3, 4, 5)$ is likewise given in Table 2. The graphical depiction of the deciphered VSMs between $\mathcal{S}$ and $\mathcal{S}_r$ $(r = 1, 2, 3, 4, 5)$ is exhibited in Figure 2.

*5.2. Medical Diagnosis.* In this section, an altered algorithm for medical diagnosis is planned, which is improved depending on Xiao's algorithm [38]. The new algorithm uses the complex dual hesitant fuzzy similarity and distance measures that acquire glorious leads to application.

*Example 4.* Particular diseases have different signs. The medical diagnosis trust in the victim's signs to assess what kind of disease the victim has. The victim's signs are a set of signs and unknown diseases are a set of diagnostic diseases. A set of diagnoses is given by $\mathbb{D} = \{\mathbb{D}_1$ (Diabetes), $\mathbb{D}_2$ (Malaria), $\mathbb{D}_3$ (Heart problem), $\mathbb{D}_4$ (Flu), $\mathbb{D}_5$ (Hepatitis A)$\}$ and a set of signs is given by $X = \{x_1$ (Fatigue), $x_2$ (Fever), $x_3$ (Heart pain), $x_4$ (Cough), $x_5$ (Abdominal pain)$\}$. The victim's signs can be given in the structure of CDHFSs as follows:

Table 3: Estimations of the interpreted VSMs.

| Similarity measures | $(\mathbb{D}, \mathbb{D}_1)$ | $(\mathbb{D}, \mathbb{D}_2)$ | $(\mathbb{D}, \mathbb{D}_3)$ | $(\mathbb{D}, \mathbb{D}_4)$ | $(\mathbb{D}, \mathbb{D}_5)$ | Ranking |
|---|---|---|---|---|---|---|
| $\mathrm{Jac}(\mathbb{P}, \mathbb{D}_r)$ | 0.1975 | 0.749 | 0.3781 | 0.3133 | 0.7702 | $\mathbb{D}_2 \geq \mathbb{D}_5 \geq \mathbb{D}_3 \geq \mathbb{D}_4 \geq \mathbb{D}_1$ |
| $\mathrm{Dic}(\mathbb{P}, \mathbb{D}_r)$ | 0.3154 | 0.7977 | 0.4825 | 0.5178 | 0.6759 | $\mathbb{D}_2 \geq \mathbb{D}_5 \geq \mathbb{D}_4 \geq \mathbb{D}_3 \geq \mathbb{D}_1$ |
| $\mathrm{Cos}(\mathbb{P}, \mathbb{D}_r)$ | 0.3328 | 0.8189 | 0.4865 | 0.4854 | 0.7159 | $\mathbb{D}_2 \geq \mathbb{D}_5 \geq \mathbb{D}_3 \geq \mathbb{D}_4 \geq \mathbb{D}_1$ |

Table 4: Estimations of the interpreted WVSMs.

| Similarity measures | $(\mathbb{D}, \mathbb{D}_1)$ | $(\mathbb{D}, \mathbb{D}_2)$ | $(\mathbb{D}, \mathbb{D}_3)$ | $(\mathbb{D}, \mathbb{D}_4)$ | $(\mathbb{D}, \mathbb{D}_5)$ | Ranking |
|---|---|---|---|---|---|---|
| $\mathrm{Jac}_w(\mathbb{P}, \mathbb{D}_r)$ | 0.2154 | 0.9221 | 0.3307 | 0.3136 | 0.7831 | $\mathbb{D}_2 \geq \mathbb{D}_5 \geq \mathbb{D}_3 \geq \mathbb{D}_4 \geq \mathbb{D}_1$ |
| $\mathrm{Dic}_w(\mathbb{P}, \mathbb{D}_r)$ | 0.3369 | 0.8173 | 0.4358 | 0.5009 | 0.7423 | $\mathbb{D}_2 \geq \mathbb{D}_5 \geq \mathbb{D}_4 \geq \mathbb{D}_3 \geq \mathbb{D}_1$ |
| $\mathrm{Cos}_w(\mathbb{P}, \mathbb{D}_r)$ | 0.3488 | 0.8407 | 0.4405 | 0.4829 | 0.7794 | $\mathbb{D}_2 \geq \mathbb{D}_5 \geq \mathbb{D}_4 \geq \mathbb{D}_3 \geq \mathbb{D}_1$ |

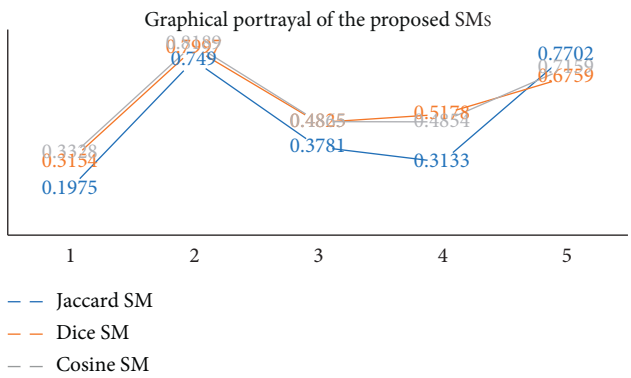



Figure 3: The graphical depiction of the deciphered VSMs between $\mathbb{P}$ and $\mathbb{D}_r$ $(r = 1, 2, 3, 4, 5)$.

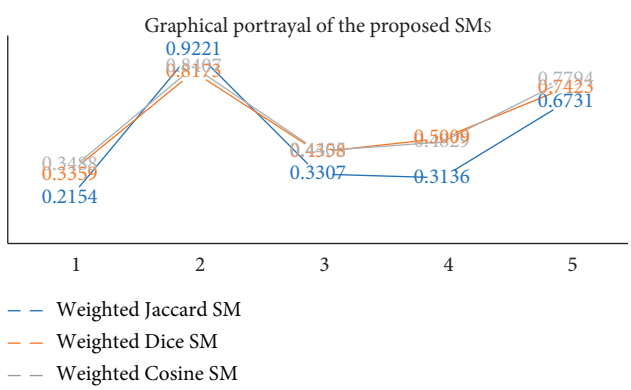



Figure 4: The graphical depiction of the deciphered WVSMs between $\mathbb{P}$ and $\mathbb{D}_r$ $(r = 1, 2, 3, 4, 5)$.

$$\begin{aligned}\mathbb{P}(\text{Victim}) = \Big\{&\Big(x_1, \Big\{\Big\{0.3e^{i2\pi(0.2)}\Big\}\Big\{0.65e^{i2\pi(0.7)}, 0.5e^{i2\pi(0.6)}\Big\}\Big\}\Big), \Big(x_2, \Big\{\Big\{0.8e^{i2\pi(0.6)}, 0.55e^{i2\pi(0.55)}, 0.45e^{i2\pi(0.8)}\Big\},\\ &\Big\{0.1e^{i2\pi(0.15)}, 0.15e^{i2\pi(0.1)}, 0.1e^{i2\pi(0.05)}\Big\}\Big\}\Big), \Big(x_3, \Big\{\Big\{0.45e^{i2\pi(0.55)}\Big\}, \Big\{0.25e^{i2\pi(0.35)}\Big\}\Big\}\Big),\\ &\Big(x_4, \Big\{\Big\{0.35e^{i2\pi(0.65)}, 0.65e^{i2\pi(0.5)}\Big\}, \Big\{0.15e^{i2\pi(0.25)}, 0.3e^{i2\pi(0.3)}\Big\}\Big\}\Big), \Big(x_5, \Big\{\Big\{0.5e^{i2\pi(0.3)}, 0.3e^{i2\pi(0.6)}\Big\}, \Big\{0.2e^{i2\pi(0.25)}\Big\}\Big\}\Big)\Big\}.\end{aligned} \tag{38}$$

The signs of each disease $\mathbb{D}_r\,(r = 1, 2, 3, 4, 5)$ can be represented in CDHFSs as follows:

$$\begin{aligned}
\mathbb{D}_1\,(\text{Diabetes}) = \{&\big(x_1, \{\{0.65e^{i2\pi(0.7)}, 0.5e^{i2\pi(0.6)}, 0.4e^{i2\pi(0.5)}\}, \{0.15e^{i2\pi(0.2)}, 0.25e^{i2\pi(0.15)}, 0.3e^{i2\pi(0.25)}\}\}\big),\\
&\big(x_2, \{\{0.05e^{i2\pi(0.15)}\}, \{0.55e^{i2\pi(0.8)}\}\}\big), \big(x_3, \{\{0.1e^{i2\pi(0.01)}\}, \{0.15e^{i2\pi(0.65)}\}\}\big),\\
&\big(x_4, \{\{0.5e^{i2\pi(0.1)}\}, \{0.5e^{i2\pi(0.3)}, 0.1e^{i2\pi(0.2)}\}\}\big), \big(x_5, \{\{0.01e^{i2\pi(0.05)}\}, \{0.05e^{i2\pi(0.34)}, 0.15e^{i2\pi(0.25)}\}\}\big)\},\\
\mathbb{D}_2\,(\text{Malaria}) = \{&\big(x_1, \{\{0.3e^{i2\pi(0.4)}\}\{0.45e^{i2\pi(0.5)}, 0.6e^{i2\pi(0.45)}\}\}\big), \big(x_2, \{\{0.7e^{i2\pi(0.55)}, 0.65e^{i2\pi(0.6)}, 0.7e^{i2\pi(0.5)}\},\\
&\{0.25e^{i2\pi(0.1)}, 0.2e^{i2\pi(0.15)}, 0.25e^{i2\pi(0.35)}\}\}\big), \big(x_3, \{\{0.5e^{i2\pi(0.4)}\}, \{0.3e^{i2\pi(0.4)}\}\}\big),\\
&\big(x_4, \{\{0.75e^{i2\pi(0.45)}, 0.55e^{i2\pi(0.6)}\}, \{0.2e^{i2\pi(0.25)}, 0.15e^{i2\pi(0.3)}\}\}\big),\\
&\big(x_5, \{\{0.7e^{i2\pi(0.4)}, 0.2e^{i2\pi(0.3)}\}, \{0.25e^{i2\pi(0.5)}\}\}\big)\},\\
\mathbb{D}_3\,(\text{Heart problem}) = \{&\big(x_1, \{\{0.45e^{i2\pi(0.3)}\}, \{0.35e^{i2\pi(0.5)}\}\}\big), \big(x_2, \{\{0.05e^{i2\pi(0.15)}\}, \{0.55e^{i2\pi(0.2)}\}\}\big),\\
&\big(x_3, \{\{0.6e^{i2\pi(0.5)}, 0.8e^{i2\pi(0.6)}, 0.7e^{i2\pi(0.4)}\}, \{0.1e^{i2\pi(0.35)}, 0.15e^{i2\pi(0.25)}\}\}\big),\\
&\big(x_4, \{\{0.4e^{i2\pi(0.5)}, 0.7e^{i2\pi(0.6)}\}, \{0.2e^{i2\pi(0.1)}, 0.25e^{i2\pi(0.35)}\}\}\big),\\
&\big(x_5, \{\{0.35e^{i2\pi(0.1)}\}, \{0.5e^{i2\pi(0.35)}, 0.3e^{i2\pi(0.65)}\}\}\big)\},\\
\mathbb{D}_4\,(\text{Flu}) = \{&\big(x_1, \{\{0.4e^{i2\pi(0.3)}\}, \{0.2e^{i2\pi(0.35)}\}\}\big), \big(x_2, \{\{0.5e^{i2\pi(0.7)}, 0.4e^{i2\pi(0.6)}\}, \{0.4e^{i2\pi(0.2)}, 0.3e^{i2\pi(0.15)}\}\}\big),\\
&\big(x_3, \{\{0.1e^{i2\pi(0.05)}\}, \{0.35e^{i2\pi(0.5)}, 0.5e^{i2\pi(0.3)}\}\}\big), \big(x_4, \{\{0.6e^{i2\pi(0.3)}, 0.55e^{i2\pi(0.45)}, 0.5e^{i2\pi(0.55)}\},\\
&\{0.2e^{i2\pi(0.35)}, 0.3e^{i2\pi(0.25)}, 0.15e^{i2\pi(0.2)}\}\}\big), \big(x_5, \{\{0.25e^{i2\pi(0.4)}, 0.5e^{i2\pi(0.36)}\}, \{0.2e^{i2\pi(0.3)}\}\}\big)\},\\
\mathbb{D}_5\,(\text{Hepatitis A}) = \{&\big(x_1, \{\{0.2e^{i2\pi(0.45)}\}, \{0.7e^{i2\pi(0.35)}, 0.6e^{i2\pi(0.3)}\}\}\big), \big(x_2, \{\{0.5e^{i2\pi(0.05)}\}, \{0.4e^{i2\pi(0.45)}\}\}\big),\\
&\big(x_3, \{\{0.25e^{i2\pi(0.5)}\}, \{0.6e^{i2\pi(0.25)}\}\}\big), \big(x_4, \{\{0.2e^{i2\pi(0.35)}, 0.3e^{i2\pi(0.6)}\}, \{0.2e^{i2\pi(0.2)}\}\}\big),\\
&\big(x_5, \{\{0.8e^{i2\pi(0.5)}, 0.7e^{i2\pi(0.7)}, 0.5e^{i2\pi(0.55)}\}, \{0.15e^{i2\pi(0.25)}, 0.1e^{i2\pi(0.2)}, 0.05e^{i2\pi(0.25)}\}\}\big)\}.
\end{aligned} \tag{39}$$

The goal of this reliance is to sort the disease of the victim $\mathbb{P}$ in one of the diseases $\mathbb{D}_r\,(r = 1, 2, 3, 4, 5)$. For this, the deciphered VSMs (Jaccard, Dice, and Cosine) are used to check the similarity from $\mathbb{P}$ to $\mathbb{D}_r\,(r = 1, 2, 3, 4, 5)$ and calculations are given in Tables 3 and 4.

As indicated by the above-figured estimations provided in Table 3, we basically notice that the similarity level between $\mathbb{P}$ and $\mathbb{D}_2$ is the best one as implantation by all VSMs. This characterizes that VSMs (Jaccard, Dice, and Cosine) give out that the victim has malaria reliant on the standard of the best degree of similarity. Situating of the researched VSMs between $\mathbb{P}$ and $\mathbb{D}_r\,(r = 1, 2, 3, 4, 5)$ is likewise given in Table 3. The graphical depiction of the deciphered VSMs between $\mathbb{P}$ and $\mathbb{D}_r\,(r = 1, 2, 3, 4, 5)$ is exhibited in Figure 3.

The weight of segments has exceptional importance to consider in certifiable powerful issues. If we surmise the greatness of segments $x_k\,(k = 1, 2, 3, 4, 5)$ be $w_k\,(0.2, 0.25, 0.3, 0.15, 0.1)$, respectively. Now the interpreted WVSMs are used to choose the similarity from $\mathbb{P}$ and $\mathbb{D}_r\,(r = 1, 2, 3, 4, 5)$ and estimations are given in Table 4.

As indicated by the above-figured estimations provided in Table 4, we basically notice that the similarity level between $\mathbb{P}$ and $\mathbb{D}_2$ is the best one as implantation by all WVSMs. This characterizes that WVSMs give out that the victim has malaria reliant on the standard of the best degree of similarity. Situating of the researched WVSMs between $\mathbb{P}$ and $\mathbb{D}_r\,(r = 1, 2, 3, 4, 5)$ is likewise given in Table 4. The graphical depiction of the deciphered VSMs between $\mathbb{P}$ and $\mathbb{D}_r\,(r = 1, 2, 3, 4, 5)$ is exhibited in Figure 4.

Table 5: Comparison among interpreted and some existing SMs for Example 5.

| Method | Score value | Ranking |
|---|---|---|
| Wang et al. [33] | $s^1(\mathcal{S},\mathcal{S}_1)=0.8273$, $s^1(\mathcal{S},\mathcal{S}_2)=0.914$, $s^1(\mathcal{S},\mathcal{S}_3)=0.8277$, $s^1(\mathcal{S},\mathcal{S}_4)=0.8945$, $s^1(\mathcal{S},\mathcal{S}_5)=0.797$ | $\mathcal{S}_2\geq\mathcal{S}_4\geq\mathcal{S}_3\geq\mathcal{S}_1\geq\mathcal{S}_5$ |
| Wang et al. [33] | $s^2(\mathcal{S},\mathcal{S}_1)=0.69$, $s^2(\mathcal{S},\mathcal{S}_2)=0.691$, $s^2(\mathcal{S},\mathcal{S}_3)=0.68$, $s^2(\mathcal{S},\mathcal{S}_4)=0.70$, $s^2(\mathcal{S},\mathcal{S}_5)=0.63$ | $\mathcal{S}_4\geq\mathcal{S}_2\geq\mathcal{S}_1\geq\mathcal{S}_3\geq\mathcal{S}_5$ |
| Wang et al. [33] | $s^3(\mathcal{S},\mathcal{S}_1)=0.7525$, $s^3(\mathcal{S},\mathcal{S}_2)=0.7216$, $s^3(\mathcal{S},\mathcal{S}_3)=0.6958$, $s^3(\mathcal{S},\mathcal{S}_4)=0.7425$, $s^3(\mathcal{S},\mathcal{S}_5)=0.6725$ | $\mathcal{S}_1\geq\mathcal{S}_4\geq\mathcal{S}_2\geq\mathcal{S}_3\geq\mathcal{S}_5$ |
| Explored SM | $\mathrm{Jac}(\mathcal{S},\mathcal{S}_1)=0.4148$, $\mathrm{Jac}(\mathcal{S},\mathcal{S}_2)=0.4765$, $\mathrm{Jac}(\mathcal{S},\mathcal{S}_3)=0.2716$, $\mathrm{Jac}(\mathcal{S},\mathcal{S}_4)=0.587$, $\mathrm{Jac}(\mathcal{S},\mathcal{S}_5)=0.3582$ | $\mathcal{S}_4\geq\mathcal{S}_2\geq\mathcal{S}_1\geq\mathcal{S}_5\geq\mathcal{S}_3$ |
| Explored SM | $\mathrm{Dic}(\mathcal{S},\mathcal{S}_1)=0.5624$, $\mathrm{Dic}(\mathcal{S},\mathcal{S}_2)=0.6103$, $\mathrm{Dic}(\mathcal{S},\mathcal{S}_3)=0.426$, $\mathrm{Dic}(\mathcal{S},\mathcal{S}_4)=0.7291$, $\mathrm{Dic}(\mathcal{S},\mathcal{S}_5)=0.5123$ | $\mathcal{S}_4\geq\mathcal{S}_2\geq\mathcal{S}_1\geq\mathcal{S}_5\geq\mathcal{S}_3$ |
| Explored SM | $\mathrm{Cos}(\mathcal{S},\mathcal{S}_1)=0.5696$, $\mathrm{Cos}(\mathcal{S},\mathcal{S}_2)=0.6322$, $\mathrm{Cos}(\mathcal{S},\mathcal{S}_3)=0.4539$, $\mathrm{Cos}(\mathcal{S},\mathcal{S}_4)=0.7062$, $\mathrm{Cos}(\mathcal{S},\mathcal{S}_5)=0.5248$ | $\mathcal{S}_4\geq\mathcal{S}_2\geq\mathcal{S}_1\geq\mathcal{S}_5\geq\mathcal{S}_3$ |

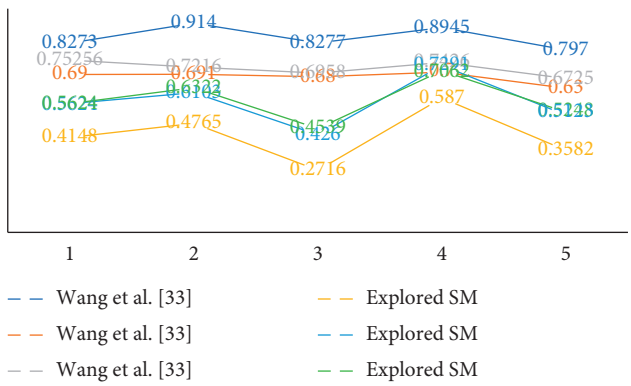



Figure 5: The graphical depiction between the proposed and existing SMs for Example 5.

Table 6: Comparison among interpreted and some existing SMs for Example 3.

| Method | Score value | Ranking |
|---|---|---|
| Wang et al. [33] | Failed | Failed |
| Wang et al. [33] | Failed | Failed |
| Wang et al. [33] | Failed | Failed |
| Explored SM | $\mathrm{Jac}(\mathcal{S},\mathcal{S}_1)=0.3936$, $\mathrm{Jac}(\mathcal{S},\mathcal{S}_2)=0.4455$, $\mathrm{Jac}(\mathcal{S},\mathcal{S}_3)=0.2748$, $\mathrm{Jac}(\mathcal{S},\mathcal{S}_4)=0.6742$, $\mathrm{Jac}(\mathcal{S},\mathcal{S}_5)=0.4087$ | $\mathcal{S}_4\geq\mathcal{S}_2\geq\mathcal{S}_5\geq\mathcal{S}_1\geq\mathcal{S}_3$ |
| Explored SM | $\mathrm{Dic}(\mathcal{S},\mathcal{S}_1)=0.5353$, $\mathrm{Dic}(\mathcal{S},\mathcal{S}_2)=0.581$, $\mathrm{Dic}(\mathcal{S},\mathcal{S}_3)=0.4253$, $\mathrm{Dic}(\mathcal{S},\mathcal{S}_4)=0.7755$, $\mathrm{Dic}(\mathcal{S},\mathcal{S}_5)=0.5504$, | $\mathcal{S}_4\geq\mathcal{S}_2\geq\mathcal{S}_5\geq\mathcal{S}_1\geq\mathcal{S}_3$ |
| Explored SM | $\mathrm{Cos}(\mathcal{S},\mathcal{S}_1)=0.5465$, $\mathrm{Cos}(\mathcal{S},\mathcal{S}_1)=0.6048$, $\mathrm{Cos}(\mathcal{S},\mathcal{S}_3)=0.447$, $\mathrm{Cos}(\mathcal{S},\mathcal{S}_4)=0.7444$, $\mathrm{Cos}(\mathcal{S},\mathcal{S}_5)=0.557$ | $\mathcal{S}_4\geq\mathcal{S}_2\geq\mathcal{S}_5\geq\mathcal{S}_1\geq\mathcal{S}_3$ |

## 6. Comparison

In this section, we showed the feasibility and focal points of the deciphered SMs by differentiating some existing SMs.

*Example 5.* Its feed that amounts from developments by an organization is legitimately corresponding to the standard of building substances they use. An appropriate review of the building substance before development is the confirmation of good building measures. The building substances to be utilized ought to be carefully checked before applying. The best possible check and equalization arrangement of investigation approve the manufacturers to utilize the correct substances for developments to improve the standard of their task. Five known building substances $\mathcal{S}_r\,(r=1,2,3,4,5)$ which are given in the DHFSs structure are as follows:

$$\begin{aligned}
\mathcal{S}_1 &= \{(x_1, \{\{0.2, 0.3, 0.5\}, \{0.3, 0.2\}\}), (x_2, \{\{0.3\}, \{0.6, 0.4\}\}), (x_3, \{\{0.6\}, \{0.2\}\}), (x_4, \{\{0.15, 0.6\}, \{0.35\}\}), (x_5, \{\{0.4\}, \{0.3\}\})\},\\
\mathcal{S}_2 &= \{(x_1, \{\{0.25\}, \{0.6, 0.3\}\}), (x_2, \{\{0.35\}, \{0.5\}\}), (x_3, \{\{0.35, 0.45, 0.6\}, \{0.3, 0.1\}\}), (x_4, \{\{0.6, 0.5\}, \{0.2, 0.25\}\}),\\
&\quad (x_5, \{\{0.5\}, \{0.25, 0.4\}\})\},\\
\mathcal{S}_3 &= \{(x_1, \{\{0.3, 0.2\}, \{0.4, 0.5\}\}), (x_2, \{\{0.45, 0.5\}, \{0.15\}\}), (x_3, \{\{0.6\}, \{0.35\}\}), (x_4, \{\{0.7\}, \{0.25, 0.15\}\}),\\
&\quad (x_5, \{\{0.1, 0.25, 0.4\}, \{0.5, 0.2\}\})\},\\
\mathcal{S}_4 &= \{(x_1, \{\{0.4, 0.6\}, \{0.15, 0.3\}\}), (x_2, \{\{0.3\}, \{0.6\}\}), (x_3, \{\{0.4, 0.2, 0.6\}, \{0.2, 0.15\}\}), (x_4, \{\{0.5, 0.3\}, \{0.4\}\}),\\
&\quad (x_5, \{\{0.6, 0.55, 0.3\}, \{0.35, 0.15, 0.1\}\})\},\\
\mathcal{S}_5 &= \{(x_1, \{\{0.3, 0.2\}, \{0.5, 0.4\}\}), (x_2, \{\{0.45, 0.7\}, \{0.25\}\}), (x_3, \{\{0.2\}, \{0.6\}\}), (x_4, \{\{0.7\}, \{0.15, 0.25\}\}),\\
&\quad (x_5, \{\{0.4, 0.25, 0.1\}, \{0.5, 0.5, 0.3\}\})\},\\
\mathcal{S} &= \{(x_1, \{\{\{0.7, 0.3\}, \{0.2, 0.1\}\}\}), (x_2, \{\{0.7\}, \{0.2\}\}), (x_3, \{\{0.6, 0.5, 0.4\}, \{0.3, 0.1\}\}), (x_4, \{0.4, 0.6\}, \{0.15\}),\\
&\quad (x_5, \{\{0.5, 0.4, 0.2\}, \{0.3, 0.1, 0.2\}\})\}.
\end{aligned} \tag{40}$$

We convert the above information which is given in the form of DHFSs into CDHFSs by putting $1 = e^0$ as presented below:

$$\begin{aligned}
\mathcal{S}_1 &= \{(x_1, \{\{0.2e^{i2\pi(0.0)}, 0.3e^{i2\pi(0.0)}, 0.5e^{i2\pi(0.0)}\}, \{0.3e^{i2\pi(0.0)}, 0.2e^{i2\pi(0.0)}\}\}), (x_2, \{\{0.3e^{i2\pi(0.0)}\}, \{0.6e^{i2\pi(0.0)}, 0.4e^{i2\pi(0.0)}\}\}),\\
&\quad (x_3, \{\{0.6e^{i2\pi(0.0)}\}, \{0.2e^{i2\pi(0.0)}\}\}), (x_4, \{\{0.15e^{i2\pi(0.0)}, 0.6e^{i2\pi(0.0)}\}, \{0.35e^{i2\pi(0.0)}\}\}), (x_5, \{\{0.4e^{i2\pi(0.0)}\}, \{0.3e^{i2\pi(0.0)}\}\})\},\\
\mathcal{S}_2 &= \{(x_1, \{\{0.25e^{i2\pi(0.0)}\}, \{0.6e^{i2\pi(0.0)}, 0.3e^{i2\pi(0.0)}\}\}), (x_2, \{\{0.35e^{i2\pi(0.0)}\}, \{0.5e^{i2\pi(0.0)}\}\}),\\
&\quad (x_3, \{\{0.35e^{i2\pi(0.0)}, 0.45e^{i2\pi(0.0)}, 0.6e^{i2\pi(0.0)}\}, \{0.3e^{i2\pi(0.0)}, 0.1e^{i2\pi(0.0)}\}\}), (x_4, \{\{0.6e^{i2\pi(0.0)}, 0.5e^{i2\pi(0.0)}\},\\
&\quad \{0.2e^{i2\pi(0.0)}, 0.25e^{i2\pi(0.0)}\}\}), (x_5, \{\{0.5e^{i2\pi(0.0)}\}, \{0.25e^{i2\pi(0.0)}, 0.4e^{i2\pi(0.0)}\}\})\},\\
\mathcal{S}_3 &= \{(x_1, \{\{0.3e^{i2\pi(0.0)}, 0.2e^{i2\pi(0.0)}\}, \{0.4e^{i2\pi(0.0)}, 0.5e^{i2\pi(0.0)}\}\}), (x_2, \{\{0.45e^{i2\pi(0.0)}, 0.5e^{i2\pi(0.0)}\}, \{0.15e^{i2\pi(0.0)}\}\}),\\
&\quad (x_3, \{\{0.6e^{i2\pi(0.0)}\}, \{0.35e^{i2\pi(0.0)}\}\}), (x_4, \{\{0.7e^{i2\pi(0.0)}\}, \{0.25e^{i2\pi(0.0)}, 0.15e^{i2\pi(0.0)}\}\}),\\
&\quad (x_5, \{\{0.1e^{i2\pi(0.0)}, 0.25e^{i2\pi(0.0)}, 0.4e^{i2\pi(0.0)}\}, \{0.5e^{i2\pi(0.0)}, 0.2e^{i2\pi(0.0)}\}\})\},\\
\mathcal{S}_4 &= \{(x_1, \{\{0.4e^{i2\pi(0.0)}, 0.6e^{i2\pi(0.0)}\}, \{0.15, 0.3\}\}), (x_2, \{\{0.3e^{i2\pi(0.0)}\}, \{0.6e^{i2\pi(0.0)}\}\}),\\
&\quad (x_3, \{\{0.4e^{i2\pi(0.0)}, 0.2e^{i2\pi(0.0)}, 0.6e^{i2\pi(0.0)}\}, \{0.2e^{i2\pi(0.0)}, 0.15e^{i2\pi(0.0)}\}\}), (x_4, \{\{0.5e^{i2\pi(0.0)}, 0.3e^{i2\pi(0.0)}\}, \{0.4e^{i2\pi(0.0)}\}\}),\\
&\quad (x_5, \{\{0.6e^{i2\pi(0.0)}, 0.55e^{i2\pi(0.0)}, 0.3e^{i2\pi(0.0)}\}, \{0.35e^{i2\pi(0.0)}, 0.15e^{i2\pi(0.0)}, 0.1e^{i2\pi(0.0)}\}\})\},\\
\mathcal{S}_5 &= \{(x_1, \{\{0.3e^{i2\pi(0.0)}, 0.2e^{i2\pi(0.0)}\}, \{0.5e^{i2\pi(0.0)}, 0.4e^{i2\pi(0.0)}\}\}), (x_2, \{\{0.45e^{i2\pi(0.0)}, 0.7e^{i2\pi(0.0)}\}, \{0.25e^{i2\pi(0.0)}\}\}),\\
&\quad (x_3, \{\{0.2e^{i2\pi(0.0)}\}, \{0.6e^{i2\pi(0.0)}\}\}), (x_4, \{\{0.7e^{i2\pi(0.0)}\}, \{0.15e^{i2\pi(0.0)}, 0.25e^{i2\pi(0.0)}\}\}),\\
&\quad (x_5, \{\{0.4e^{i2\pi(0.0)}, 0.25e^{i2\pi(0.0)}, 0.1e^{i2\pi(0.0)}\}, \{0.5e^{i2\pi(0.0)}, 0.5e^{i2\pi(0.0)}, 0.3e^{i2\pi(0.0)}\}\})\},\\
\mathcal{S} &= \{(x_1, \{\{\{0.7e^{i2\pi(0.0)}, 0.3e^{i2\pi(0.0)}\}, \{0.2e^{i2\pi(0.0)}, 0.1e^{i2\pi(0.0)}\}\}\}), (x_2, \{\{0.7e^{i2\pi(0.0)}\}, \{0.2e^{i2\pi(0.0)}\}\}),\\
&\quad (x_3, \{0.6, 0.5, 0.4\}, \{0.3e^{i2\pi(0.0)}, 0.1e^{i2\pi(0.0)}\}\}), (x_4, \{0.4e^{i2\pi(0.0)}, 0.6e^{i2\pi(0.0)}\}, \{0.15e^{i2\pi(0.0)}\}),\\
&\quad (x_5, \{\{0.5e^{i2\pi(0.0)}, 0.4e^{i2\pi(0.0)}, 0.2e^{i2\pi(0.0)}\}, \{0.3e^{i2\pi(0.0)}, 0.1e^{i2\pi(0.0)}, 0.2e^{i2\pi(0.0)}\}\})\}.
\end{aligned} \tag{41}$$

For instance (25), we have to find that the obscure building substance $\mathcal{S}$ has a spot with the foreordained building material $\mathcal{S}_r$ $(r = 1, 2, 3, 4, 5)$. The data of model (25) introduced looking like DHFSs. By some current SMs for DHFSs, we identified the closeness among $\mathcal{S}$ and $\mathcal{S}_r$ $(r = 1, 2, 3, 4, 5)$ as shown in Table 5. As we have $1 = e^0$, then the

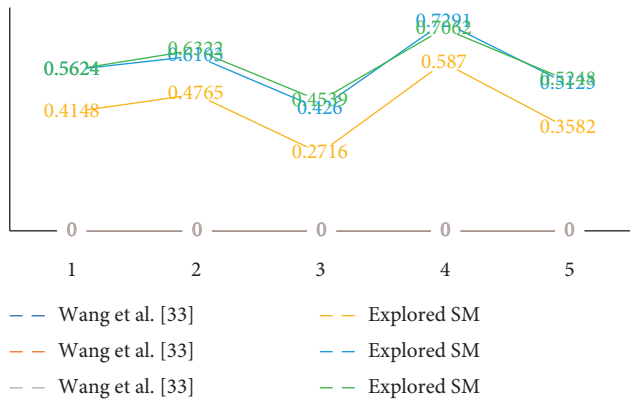



Figure 6: The graphical depiction between the proposed and existing SMs for Example 3.

data, given in model (25), were transformed into CDHFSs. Now through deciphered SMs, we identified the closeness among $\mathcal{S}$ and $r\,(r = 1, 2, 3, 4, 5)$ in like manner as shown in Table 5. Our deciphered SMs (Jaccard, Dice, and Cosine) exhibited that obscure building substance $\mathcal{S}$ has a spot with predefined building substance $\mathcal{S}_4$ in the light of the fact that the similitude among $\mathcal{S}$ and $\mathcal{S}_4$ is most conspicuous one. The positioning of the explored and existing SMs is similarly presented in Table 5. The graphical depiction of the correlation of the proposed and existing SMs is shown in Figure 5.

By and by we analyze the correlation among deciphered and existing SMs for instance (17). For instance (17), the data are given in the structure of CDHFSs. We understand that no SM exists in the composition to comprehend such data. The current SMs are insufficient to find the closeness among $\mathcal{S}$ and $\mathcal{S}_r\,(r = 1, 2, 3, 4, 5)$ as shown in Table 6. From Table 6, we see that the data introduced in model (17) are just plausible by the investigated SMs. The deciphered SMs find the similitude among $\mathcal{S}$ and $\mathcal{S}_r\,(r = 1, 2, 3, 4, 5)$ as shown in Table 6. Our deciphered SMs (Jaccard, Dice, and Cosine) exhibited that dubious building material $\mathcal{S}$ has a spot with the foreordained building material $\mathcal{S}_4$ in light of the face that the closeness among $\mathcal{S}$ and $\mathcal{S}_4$ is most imperative one. The situating of the interpreted SMs is furthermore presented in Table 6. The graphical depiction of the assessment of proposed and existing SMs is shown in Figure 6.

In this correlation, we stood out our researched SMs from some current SMs described by Wang et al. [33]. From the above discussion, indisputably, our proposed SMs can talk extra fluffy data and put it extensively in conditions, all things considered, issues. In the view of CDHFS, we investigated the SMs, our SMs are dynamically pleasing for authentic issues than the current SMs, and our SMs are more expansive than the current SMs.

## 7. Conclusion

The notion of CDHFS is a mixture of two alterations called CFS and DHFS. CDHFS creates two degrees called truth valued and falsely valued as a subset of a unit disc in a complex plane and is a capable method to adapt to questionable and unusual data, in real-life decisions. In this paper, we investigated the notion of CDHFS and its operational laws. The novel approach of CIvDHFS and its fundamental laws are also investigated and furthermore advocated with the assistance of models. Further, the VSMs, WVSMs, hybrid vector similarity measures, and weighted hybrid vector similarity measures are also explored. These SMs are applied in the environment of pattern recognition and medical diagnosis to assess the proficiency and feasibility of the interpreted measures. We additionally unraveled some numerical examples utilizing the setup measures. Moreover, we examined the dependability and legitimacy of the proposed SMs by comparing them with some existing SMs. In this paper, the point of interest, relative examination, and graphical portrayal of the investigated SMs and existing SMs are also discussed in detail.

In future, we will work on linear Diophantine fuzzy sets (LDFS) [16, 45, 46], complex fuzzy sets [47], complex q-rung orthopair fuzzy sets [48–52], etc. [53–59].

## Data Availability

The data used in this article are artificial and hypothetical, and anyone can use these data before prior permission by just citing this article.

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## Acknowledgments

This paper was supported by "Algebra and Applications Research Unit, Faculty of Science, Prince of Songkla University."

## References

[1] L. A. Zadeh, "Fuzzy sets," *Information and Control*, vol. 8, no. 3, pp. 338–353, 1965.

[2] L. A. Zadeh, "The concept of a linguistic variable and its application to approximate reasoning—I," *Information Sciences*, vol. 8, no. 3, pp. 199–249, 1975.

[3] L. A. Zadeh, "The concept of a linguistic variable and its application to approximate reasoning—II," *Information Sciences*, vol. 8, no. 4, pp. 301–357, 1975.

[4] L. A. Zadeh, "The concept of a linguistic variable and its application to approximate reasoning—III," *Information Sciences*, vol. 9, no. 1, pp. 43–80, 1975.

[5] K. T. Atanassov, "Intuitionistic fuzzy sets," in *Intuitionistic Fuzzy Sets*, pp. 1–137, Springer, Berlin, Germany, 1999.

[6] K. T. Atanassov, "Interval valued intuitionistic fuzzy sets," in *Intuitionistic Fuzzy Sets*, pp. 139–177, Springer, Berlin, Germany, 1999.

[7] Z. Xu, "A method based on linguistic aggregation operators for group decision making with linguistic preference relations," *Information Sciences*, vol. 166, no. 1–4, pp. 19–30, 2004.

[8] D. J. Dubois, *Fuzzy Sets and Systems: Theory and Applications*, Vol. 144, Academic Press, Cambridge, MA, USA, 1980.

[9] R. R. Yager, "On the theory of bags," *International Journal of General Systems*, vol. 13, no. 1, pp. 23–37, 1986.

[10] T. Baležentis and S. Zeng, "Group multi-criteria decision making based upon interval-valued fuzzy numbers: an extension of the MULTIMOORA method," *Expert Systems with Applications*, vol. 40, no. 2, pp. 543–550, 2013.

[11] W. Su, Y. Yang, C. Zhang, and S. Zeng, "Intuitionistic fuzzy decision-making with similarity measures and OWA operator," *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, vol. 21, no. 2, pp. 245–262, 2013.

[12] B. Peng, C. Ye, and S. Zeng, "Some intuitionist fuzzy weighted geometric distance measures and their application to group decision making," *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, vol. 22, no. 5, pp. 699–715, 2014.

[13] C. Zhang, C. Chen, D. Streimikiene, and T. Balezentis, "Intuitionistic fuzzy MULTIMOORA approach for multi-criteria assessment of the energy storage technologies," *Applied Soft Computing*, vol. 79, pp. 410–423, 2019.

[14] J. Zhou, T. Baležentis, and D. Streimikiene, "Normalized weighted Bonferroni harmonic mean-based intuitionistic fuzzy operators and their application to the sustainable selection of search and rescue robots," *Symmetry*, vol. 11, no. 2, p. 218, 2019.

[15] S. Zeng, S.-M. Chen, and K.-Y. Fan, "Interval-valued intuitionistic fuzzy multiple attribute decision making based on nonlinear programming methodology and TOPSIS method," *Information Sciences*, vol. 506, pp. 424–442, 2020.

[16] M. Riaz and M. R. Hashmi, "Linear diophantine fuzzy set and its applications towards multi-attribute decision-making problems," *Journal of Intelligent & Fuzzy Systems*, vol. 37, no. 4, pp. 5417–5439, 2019.

[17] V. Torra, "Hesitant fuzzy sets," *International Journal of Intelligent Systems*, vol. 25, pp. 529–539, 2010.

[18] Z. Xu and M. Xia, "On distance and correlation measures of hesitant fuzzy information," *International Journal of Intelligent Systems*, vol. 26, no. 5, pp. 410–425, 2011.

[19] D. Li, W. Zeng, and J. Li, "New distance and similarity measures on hesitant fuzzy sets and their applications in multiple criteria decision making," *Engineering Applications of Artificial Intelligence*, vol. 40, pp. 11–16, 2015.

[20] S. Zeng and Y. Xiao, "A method based on TOPSIS and distance measures for hesitant fuzzy multiple attribute decision making," *Technological and Economic Development of Economy*, vol. 24, no. 3, pp. 969–983, 2018.

[21] J. Hu, Y. Yang, X. Zhang, and X. Chen, "Similarity and entropy measures for hesitant fuzzy sets," *International Transactions in Operational Research*, vol. 25, no. 3, pp. 857–886, 2018.

[22] Z. Xu and M. Xia, "Distance and similarity measures for hesitant fuzzy sets," *Information Sciences*, vol. 181, no. 11, pp. 2128–2138, 2011.

[23] W. Zeng, D. Li, and Q. Yin, "Distance and similarity measures between hesitant fuzzy sets and their application in pattern recognition," *Pattern Recognition Letters*, vol. 84, pp. 267–271, 2016.

[24] Z. Mu, S. Zeng, and T. Baležentis, "A novel aggregation principle for hesitant fuzzy elements," *Knowledge-Based Systems*, vol. 84, pp. 134–143, 2015.

[25] J. C. R. Alcantud and V. Torra, "Decomposition theorems and extension principles for hesitant fuzzy sets," *Information Fusion*, vol. 41, pp. 48–56, 2018.

[26] H. Liao and Z. Xu, "Subtraction and division operations over hesitant fuzzy sets," *Journal of Intelligent & Fuzzy Systems*, vol. 27, no. 1, pp. 65–72, 2014.

[27] K. Bisht and S. Kumar, "Fuzzy time series forecasting method based on hesitant fuzzy sets," *Expert Systems with Applications*, vol. 64, pp. 557–568, 2016.

[28] J. C. R. Alcantud and A. Giarlotta, "Necessary and possible hesitant fuzzy sets: a novel model for group decision making," *Information Fusion*, vol. 46, pp. 63–76, 2019.

[29] X. Zhang and Z. Xu, "Novel distance and similarity measures on hesitant fuzzy sets with applications to clustering analysis," *Journal of Intelligent & Fuzzy Systems*, vol. 28, no. 5, pp. 2279–2296, 2015.

[30] B. Farhadinia and E. Herrera-Viedma, "Multiple criteria group decision making method based on extended hesitant fuzzy sets with unknown weight information," *Applied Soft Computing*, vol. 78, pp. 310–323, 2019.

[31] N. Chen, Z. Xu, and M. Xia, "Interval-valued hesitant preference relations and their applications to group decision making," *Knowledge-Based Systems*, vol. 37, pp. 528–540, 2013.

[32] B. Zhu, Z. Xu, and M. Xia, "Dual hesitant fuzzy sets," *Journal of Applied Mathematics*, vol. 2012, Article ID 879629, 13 pages, 2012.

[33] L. Wang, Q. Wang, S. Xu, and M. Ni, "Distance and similarity measures of dual hesitant fuzzy sets with their applications to multiple attribute decision making," in *Proceedings of the 2014 IEEE International Conference on Progress in Informatics and Computing*, Shanghai, China, May 2014.

[34] B. Zhu and Z. Xu, "Some results for dual hesitant fuzzy sets," *Journal of Intelligent & Fuzzy Systems*, vol. 26, no. 4, pp. 1657–1668, 2014.

[35] S. K. Tyagi, "Correlation coefficient of dual hesitant fuzzy sets and its applications," *Applied Mathematical Modelling*, vol. 39, no. 22, pp. 7082–7092, 2015.

[36] B. Farhadinia, "Correlation for dual hesitant fuzzy sets and dual interval-valued hesitant fuzzy sets," *International Journal of Intelligent Systems*, vol. 29, no. 2, pp. 184–205, 2014.

[37] D. Ramot, R. Milo, M. Friedman, and A. Kandel, "Complex fuzzy sets," *IEEE Transactions on Fuzzy Systems*, vol. 10, no. 2, pp. 171–186, 2002.

[38] F. Xiao and W. Ding, "Divergence measure of Pythagorean fuzzy sets and its application in medical diagnosis," *Applied Soft Computing*, vol. 79, pp. 254–267, 2019.

[39] M. Munir, H. Kalsoom, K. Ullah, T. Mahmood, and Y.-M. Chu, "T-spherical fuzzy Einstein hybrid aggregation operators and their applications in multi-attribute decision making problems," *Symmetry*, vol. 12, no. 3, p. 365, 2020.

[40] K. Ullah, T. Mahmood, and H. Garg, "Evaluation of the performance of search and rescue robots using T-spherical

fuzzy hamacher aggregation operators," *International Journal of Fuzzy Systems*, vol. 22, no. 2, pp. 570–582, 2020.

[41] T. Mahmood, K. Ullah, Q. Khan, and N. Jan, "An approach toward decision-making and medical diagnosis problems using the concept of spherical fuzzy sets," *Neural Computing and Applications*, vol. 31, no. 11, pp. 7041–7053, 2019.

[42] Z. Ali and T. Mahmood, "Complex neutrosophic generalised dice similarity measures and their application to decision making," *CAAI Transactions on Intelligence Technology*, vol. 5, no. 2, pp. 78–87, 2020.

[43] Z. Ali and T. Mahmood, "Maclaurin symmetric mean operators and their applications in the environment of complex q-rung orthopair fuzzy sets," *Computational and Applied Mathematics*, vol. 39, p. 161, 2020.

[44] P. Liu, T. Mahmood, and Z. Ali, "Complex Q-rung orthopair fuzzy aggregation operators and their applications in multi-attribute group decision making," *Information*, vol. 11, no. 1, p. 5, 2020.

[45] M. Riaz, K. Naeem, and D. Afzal, "Pythagorean m-polar fuzzy soft sets with TOPSIS method for MCGDM," *Punjab University Journal of Mathematics*, vol. 52, no. 3, pp. 21–46, 2020.

[46] M. Riaz and M. R. Hashmi, "Soft rough Pythagorean m-polar fuzzy sets and Pythagorean m-polar fuzzy soft rough sets with application to decision-making," *Computational and Applied Mathematics*, vol. 39, no. 1, p. 16, 2020.

[47] P. Liu, Z. Ali, and T. Mahmood, "The distance measures and cross-entropy based on complex fuzzy sets and their application in decision making," *Journal of Intelligent & Fuzzy Systems*, vol. 39, no. 3, pp. 3351–3374, 2020.

[48] P. Liu, Z. Ali, and T. Mahmood, "A method to multi-attribute group decision-making problem with complex q-rung orthopair linguistic information based on Heronian mean operators," *International Journal of Computational Intelligence Systems*, vol. 12, no. 2, pp. 1465–1496, 2019.

[49] H. Garg, J. Gwak, T. Mahmood, and Z. Ali, "Power aggregation operators and VIKOR methods for complex q-rung orthopair fuzzy sets and their applications," *Mathematics*, vol. 8, no. 4, p. 538, 2020.

[50] P. Liu, Z. Ali, T. Mahmood, and N. Hassan, "Group decision-making using complex q-rung orthopair fuzzy Bonferroni mean," *International Journal of Computational Intelligence Systems*, vol. 13, no. 1, pp. 822–851, 2020.

[51] T. Mahmood and Z. Ali, "Entropy measure and TOPSIS method based on correlation coefficient using complex q-rung orthopair fuzzy information and its application to multi-attribute decision making," *Soft Computing*, pp. 1–27, 2020.

[52] H. Garg, Z. Ali, and T. Mahmood, "Algorithms for complex interval-valued q-rung orthopair fuzzy sets in decision making based on aggregation operators, AHP, and TOPSIS," *Expert Systems*, Article ID e12609, 2020.

[53] Z. Ali, T. Mahmood, and M.-S. Yang, "Complex T-spherical fuzzy aggregation operators with application to multi-attribute decision making," *Symmetry*, vol. 12, no. 8, p. 1311, 2020.

[54] T. Mahmood, U. Ur Rehman, Z. Ali, and T. Mahmood, "Hybrid vector similarity measures based on complex hesitant fuzzy sets and their applications to pattern recognition and medical diagnosis," *Journal of Intelligent & Fuzzy Systems*, pp. 1–22, 2020.

[55] M. A. Ramadan, H. S. Osheba, and A. R. Hadhoud, "A highly efficient and accurate finite iterative method for solving linear two-dimensional Fredholm fuzzy integral equations of the second kind using triangular functions," *Mathematical Problems in Engineering*, vol. 2020, Article ID 2028763, 16 pages, 2020.

[56] J. Hamaydi and N. Qatanani, "Computational methods for solving linear fuzzy Volterra integral equation," *Journal of Applied Mathematics*, vol. 2017, Article ID 2417195, 12 pages, 2017.

[57] R. M. Shabestari, R. Ezzati, and T. Allahviranloo, "Solving fuzzy Volterra integrodifferential equations of fractional order by Bernoulli wavelet method," *Advances in Fuzzy Systems*, vol. 2018, Article ID 5603560, 11 pages, 2018.

[58] S. Zeng, X. Peng, T. Baležentis, and D. Streimikiene, "Prioritization of low-carbon suppliers based on Pythagorean fuzzy group decision making with self-confidence level," *Economic Research-Ekonomska Istraživanja*, vol. 32, no. 1, pp. 1073–1087, 2019.

[59] S. Zeng, D. Luo, C. Zhang, and X. Li, "A correlation-based TOPSIS method for multiple attribute decision making with single-valued neutrosophic information," *International Journal of Information Technology & Decision Making*, vol. 19, no. 1, pp. 343–358, 2020.

Hindawi
Mathematical Problems in Engineering
Volume 2020, Article ID 8829948, 10 pages
https://doi.org/10.1155/2020/8829948

*Research Article*

# A Multivariate Grey Prediction Model Using Neural Networks with Application to Carbon Dioxide Emissions Forecasting

**Yu-Jing Chiu**,[1] **Yi-Chung Hu**,[1] **Peng Jiang**,[2] **Jingci Xie**,[3] **and Yen-Wei Ken**[1]

[1]*Department of Business Administration, Chung Yuan Christian University, Taoyuan City, Taiwan*
[2]*School of Business, Shandong University, Weihai, China*
[3]*School of Management, Shandong University, Jinan, China*

Correspondence should be addressed to Jingci Xie; xjc@sdu.edu.cn

Received 8 September 2020; Revised 10 October 2020; Accepted 12 October 2020; Published 29 October 2020

Academic Editor: Tomas Balezentis



The forecast of carbon dioxide ($CO_2$) emissions has played a significant role in drawing up energy development policies for individual countries. Since data about $CO_2$ emissions are often limited and do not conform to the usual statistical assumptions, this study attempts to develop a novel multivariate grey prediction model (MGPM) for $CO_2$ emissions. Compared with other MGPMs, the proposed model has several distinctive features. First, both feature selection and residual modification are considered to improve prediction accuracy. For the former, grey relational analysis is used to filter out the irrelevant features that have weaker relevance with $CO_2$ emissions. For the latter, predicted values obtained from the proposed MGPM are further adjusted by establishing a neural-network-based residual model. Prediction accuracies of the proposed MGPM were verified using real $CO_2$ emission cases. Experimental results demonstrated that the proposed MGPM performed well compared with other MGPMs considered.

## 1. Introduction

Carbon dioxide ($CO_2$) is mainly produced from fossil fuel combustion [1], and reducing the impact that energy consumption and economic growth have on $CO_2$ emissions has become a global challenge [2]. According to the International Energy Agency (IEA) [3], total emissions of greenhouse gas in 2018 were a record 33.1 billion tons, along with a global economic growth rate that increased by 3.2%. Despite a $CO_2$ emission plateau from 2014 to 2016, the IEA reported that China and USA were the highest energy-using and carbon-emitting countries, and $CO_2$ emissions went up in each country by 2.5% and 3.1%, respectively, mainly arising from an increased use of fossil fuel to meet the energy demand. In fact, $CO_2$ emissions can significantly give rise to climate change and have a negative impact on economic growth. Therefore, to keep a green economic growth, the national authorities make an effort to devise energy development policies that reduce the impact of $CO_2$ emissions. An accurate forecast of $CO_2$ emissions becomes a remarkable issue when public sectors set up policies.

From the viewpoint of the grey system theory [4–6], the prediction of $CO_2$ emissions can be viewed as a grey system problem because although the relevant features, such as energy consumption, population, and gross domestic product (GDP) [7–9], influence $CO_2$ emissions, the precise relationship between these features and emissions is not clear. Furthermore, it is possible that emissions data do not conform to any statistical assumptions [7]. Compared with the prediction models implemented by artificial intelligence techniques [10–15], statistical models including logistic models [16], multivariate regression [17, 18], and time series analysis [19, 20], MGPMs have the advantage of characterizing an unknown system using limited samples [6], without requiring conformance with statistical assumptions. Despite huge amount of data we can collect, only a few sample data points are required to achieve reliable and acceptable prediction accuracy [21, 22]. Therefore, it is

interesting to apply the multivariate grey prediction models (MGPMs) to $CO_2$ emissions.

In contrast to the frequently used GM (1, 1), the first-order grey differential equation with one variable, which does not consider the influence of relevant factors on the system [23], the GM (1, $N$), which consists of a system's characteristic sequence and several relevant factor sequences, is fundamental to MGPMs and has been widely applied to time series forecasting, such as the forecasting of traffic flow [24], number of motor vehicles [25], production in high-tech industries [26, 27], energy consumption [28, 29], integrated circuit output [30, 31], OFDI analysis [32], and pattern classification [33, 34]. Many GM (1, $N$) variants have been introduced to improve the prediction accuracy of the traditional GM (1, $N$), such as a rolling multivariable model [8], background value optimization [27], the optimization of GM (1, $N$) (OGM (1, $N$)) using grey differential equations with linear correction and grey quantity terms [25, 35], nonlinear grey models (NGM (1, $N$)) using grey differential equations with power exponents, and transformed NGM (1, $N$) (TNGM (1, $N$)) [28]. These MGPMs have shown their superior prediction performance when used with time series problems.

Grey prediction has demonstrated its effectiveness on $CO_2$ emissions forecasting, such as univariate grey prediction models by Pao et al. [36], Wu et al. [2], and Xu et al. [37], prediction models based on trends of driving coefficients (TDVGM (1, $N$)) by Ding et al. [7], TNGM (1, $N$) by Wang and Ye [28], multikernel nonlinear multivariable grey model by Duan et al. [23], a forecasting method for the traffic-related emissions by Xie et al. [38], a nonlinear grey power model (DGPM (1, $N$)) by Ding et al. [39], and a nonequigap grey Verhulst model by Wang and Li [40]. This study contributed to developing a distinctive MGPM with feature selection and residual modification performed to improve prediction accuracy. For the former, since system performance can be improved by feature selection [41], the proposed MGPM performs grey relational analysis (GRA) [4–6, 42] to estimate relevance between independent variables and $CO_2$ emissions. For the latter, it has been suggested that the prediction accuracy of the GM (1, 1) can be improved by residual modification [4]. However, it is very interesting to extend residual modification to the GM (1, $N$). A neural-network-based residual model is thus created for the proposed MGPM to adjust predicted values from the GM (1, $N$). To sum up, the proposed MGPM with feature selection and residual modification can be treated as a residual modification model. Genetic algorithms (GAs) are employed to determine required parameters of the GRA and GM (1, $N$) to construct the proposed MGPM with high prediction accuracy. Experimental results have indicated that the proposed MGPM performs well compared with the other MGPMs considered. It is noted that, to estimate the correlation between the system behavior variable and the influential factors, the proposed prediction model used GA to determine a threshold value that is not easily prespecified. In contrast, Xie et al. [38] applied GRA to identify relevant factors from passenger cars per 1000 inhabitants, stock of vehicles, volume of freight transport relative to GDP (VFTRG), and volume of passenger transport relative to GDP, but only VFTRG (with time lags four) with maximum correlation was considered. Meanwhile Ding et al. [39] used a prespecified threshold value to evaluate the relationship among factors.

The remainder of the paper is organized as follows. Section 2 introduces the traditional GM (1, $N$), grey residual modification model, and Section 3 introduces the proposed MGPM with residual modification on the basis of a neural network. Section 4 examines prediction performances of the proposed MGPM using two real cases of $CO_2$ emissions. Section 5 discusses the outcomes and presents conclusions.

## 2. The GM (1, $N$) Model

The GM (1, $N$) is a grey prediction model with $N$ variables, including a dependent variable (system characteristic), $x_1$, and $N-1$ explanatory variables (relevant factors), $x_2, x_3, \ldots, x_N$ [5]. Then, an original sequence or a time series $x_i^{(0)} = (x_i^{(0)}(1), x_i^{(0)}(2), \ldots, x_i^{(0)}(m))$ is associated with $x_i$, where $i = 1, 2, \ldots, N$, and $m$ is the number of samples. $x_i^{(0)}$ is usually a sequence taken at successive equally spaced points in time.

*Step 1.* Present an original and nonnegative sequence $x_i^{(0)}$.

*Step 2.* Perform the accumulated generating operation (AGO).

A new sequence $x_i^{(1)} = (x_i^{(1)}(1), x_i^{(1)}(2), \ldots, x_i^{(1)}(m))$ can be generated from $x_i^{(0)}$ by the AGO as follows:

$$x_i^{(1)}(k) = \sum_{j=1}^{k} x_0^{(i)}, \quad k = 1, 2, \ldots, m. \tag{1}$$

The AGO can help identify potential hidden regularities in data sequences [4, 43]. $(x_i^{(1)}(1), x_i^{(1)}(2), \ldots, x_i^{(1)}(m))$ can be approximated by an exponential function, which is a first-order whitening equation,

$$\frac{\mathrm{d}x_1^{(1)}}{\mathrm{d}t} + a x_1^{(1)} = \sum_{j=2}^{N} b_j x_j^{(1)}. \tag{2}$$

where $a$ and $b_j$ are the development and the driving coefficients, respectively. The time response expression of $\hat{x}_k^{(1)}(k)$ ($k = 2, 3, \ldots, m$) can be obtained by solving the differential equation with the initial condition $x_1^{(1)}(1) = x_1^{(0)}(1)$:

$$\hat{x}_1^{(1)}(k) = \left( x_1^{(0)}(1) - \frac{1}{a} \sum_{j=2}^{N} b_j x_j^{(1)}(k) \right) e^{-ak} + \frac{1}{a} \sum_{j=2}^{N} b_j x_j^{(1)}(k). \tag{3}$$

*Step 3.* Determine the development and the driving coefficients.

A grey differential equation of the GM (1, $N$) is as follows:

$$x_1^{(0)}(k) + az_1^{(1)}(k) = \sum_{j=2}^{N} b_j x_j^{(1)}(k) \quad k = 2, 3, \ldots, m. \tag{4}$$

where the background value $z_1^{(1)}(k)$ with the generating coefficient $\alpha$ $(0 \le \alpha \le 1)$ being usually set to 0.5 is formulated as

$$z_1^{(1)}(k) = ax_1^{(1)}(k) + (1-\alpha)x_1^{(1)}(k-1), \quad k = 2, 3, \ldots, m. \tag{5}$$

Thus, a linear regression model consisting of grey differential equations is used to estimate $a$, $b_2$,. . ., and $b_N$ through the ordinary least squares (OLS) method:

$$[a, b_2, \ldots, b_N] = \left(B^T B\right)^{-1} B^T y. \tag{6}$$

where

$$B = \begin{bmatrix} -z_1^{(1)}(2) & x_2^{(1)}(2) & \cdots & x_N^{(1)}(2) \\ -z_1^{(1)}(3) & x_2^{(1)}(3) & \cdots & x_N^{(1)}(3) \\ \vdots & \vdots & \ddots & \vdots \\ -z_1^{(1)}(m) & x_2^{(1)}(m) & \cdots & x_N^{(1)}(m) \end{bmatrix}, \quad y = \begin{bmatrix} x_1^{(0)}(2) \\ x_1^{(0)}(3) \\ \vdots \\ x_1^{(0)}(m) \end{bmatrix}. \tag{7}$$

*Step 4.* Perform the inverse accumulated generating operation (IAGO).

When $x_i^{(1)}$ varies slightly, $\hat{x}_k^{(0)}$ can be generated by means of the IAGO:

$$\hat{x}_1^{(0)}(k) = \hat{x}_1^{(1)}(k) - \hat{x}_1^{(1)}(k), \quad k = 2, 3, \ldots, m. \tag{8}$$

where $\hat{x}_1^{(1)}(1) = x_1^{(0)}(1)$.

To improve the prediction performance of the traditional GM (1, $N$), several improved versions of MGPMs have been proposed by deriving new whitening and grey differential equations:

(1) The transformed model of nonlinear GM (1, $N$) (TNGM (1, $N$)) [28]: the TNGM (1, $N$) has been applied to forecast Chinese carbon emissions. The whitening equation of the TNGM (1, $N$) is defined as

$$\frac{dx_1^{(1)}}{dt} + ax_1^{(1)} = \sum_{j=2}^{N} b_j \left(x_j^{(1)}\right)^{\gamma_j}. \tag{9}$$

The solution of the whitening equation $\hat{x}_1^{(0)}(k)$ is thus given by

$$\hat{x}_1^{(0)}(k) = \frac{1}{1+0.5a} \sum_{j=2}^{N} b_j \left(x_j^1(k)\right)^{\gamma_j} - \frac{a}{1+0.5a} \sum_{j=2}^{N} x_1^{(1)}(k-1). \tag{10}$$

$a$, $b_2$,. . ., and $b_N$ can be further derived by OLS as

$$[a, b_2, \ldots, b_N] = \left(B^T B\right)^{-1} B^T y. \tag{11}$$

where

$$B = \begin{bmatrix} -z_1^{(1)}(2) & \left(x_2^{(1)}(2)\right)^{\gamma_2} & \cdots & \left(x_N^{(1)}(2)\right)^{\gamma_N} \\ -z_1^{(1)}(3) & \left(x_2^{(1)}(3)\right)^{\gamma_2} & \cdots & \left(x_N^{(1)}(3)\right)^{\gamma_N} \\ \vdots & \vdots & \ddots & \vdots \\ -z_1^{(1)}(m) & \left(x_2^{(1)}(m)\right)^{\gamma_2} & \cdots & \left(x_N^{(1)}(m)\right)^{\gamma_N} \end{bmatrix}. \tag{12}$$

Any optimization technique such as a GA can be used to derive the optimal $\gamma_2$, $\gamma_3$,. . ., and $\gamma_N$.

(2) The optimization of the GM (1, $N$) (OGM (1, $N$)) [35]: the grey difference equation of the OGM (1, $N$) is defined as

$$x_1^{(0)}(k) + az_1^{(1)}(k) = \sum_{j=2}^{N} b_j x_j^{(1)}(k) + h_1(k-1) + h_2, \quad k = 2, 3, \ldots, m. \tag{13}$$

The time response expression of $\hat{x}_1^{(1)}(k)$ is thus given by

$$\hat{x}_1^{(1)}(k) = \sum_{j=2}^{k} \left[\mu_1 \sum_{m=2}^{N} \mu_2^{j-1} b_m x_m^{(1)}(k-j+1)\right] + \mu_2^{k-1} x_1^{(0)}(1) + \sum_{j=0}^{k-2} \mu_2^j \left[(k-j)\mu_3 + \mu_4\right]. \tag{14}$$

where

$$\mu_1 = \frac{1}{1+0.5a}, \quad \mu_2 = \frac{1-0.5a}{1+0.5a}, \quad \mu_3 = \frac{h_1}{1+0.5a}, \quad \mu_4 = \frac{h_2 - h_1}{1+0.5a}. \tag{15}$$

$a$, $b_2$,. . ., $b_N$, $h_1$, and $h_2$ can be obtained by OLS as

$$[a, b_2, \ldots, b_N, h_1, h_2] = \left(B^T B\right)^{-1} B^T y, \tag{16}$$

where

$$B = \begin{bmatrix} -z_1^{(1)}(2) & x_2^{(1)}(2) & \cdots & x_N^{(1)}(2) & 1 & 1 \\ -z_1^{(1)}(3) & x_2^{(1)}(3) & \cdots & x_N^{(1)}(3) & 2 & 1 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ -z_1^{(1)}(m) & x_2^{(1)}(m) & \cdots & x_N^{(1)}(m) & m-1 & 1 \end{bmatrix}. \tag{17}$$

Zeng et al. [25] applied a variant of the OGM (1, $N$) to forecast the number of motor vehicles in Beijing.

(3) The MGPM based on trends of driving coefficients (TDVGM (1, $N$)) [7]: the development of the TDVGM (1, $N$) addressed an issue of forecasting Chinese $CO_2$ emissions. This prediction model initially divided data from the first year by data from each year. Then, it predicted trends of driving coefficients by defining the grey differential equation for $b_j$ ($j = 2, 3, \ldots, N$) as

$$x_j^{(0)}(k) + a_j z_j^{(1)}(k) = b_j. \tag{18}$$

The time response expression of $\hat{x}_j^{(1)}(k)$ is formulated as

$$\hat{x}_j^{(1)}(k) = \left(1 - \frac{b_j}{a_j}\right) e^{-aj(k-1)} + \frac{b_j}{a_j}, \quad k = 2, 3, \ldots, m. \tag{19}$$

$a_j$ and $b_j$ were obtained by OLS as

$$\left[a_j, b_j\right] = \left(B^T B\right)^{-1} B^T y, \tag{20}$$

where

$$B = \begin{bmatrix} -z_1^{(1)}(2), & 1, \\ -z_1^{(1)}(3), & 1, \\ \vdots & \vdots \\ -z_1^{(1)}(m), & 1. \end{bmatrix} \tag{21}$$

Then, a time response expression of $\hat{x}_1^{(1)}(k)$ can be derived as

$$\hat{x}_1^{(1)}(k) = ce^{-ak} + \left(\sum_{j=2}^{N} \frac{d_j\left(\left(1 - b_j\right)/a_j\right)}{a - a_j} e^{-a_j(k-1)}\right) e^{-ak} + \sum_{j=2}^{N} \frac{d_j b_j}{a a_j}. \tag{22}$$

where

$$c = ce^{-ak} + \frac{\sum_{k=1}^{m} e^{-ak}\left(x_1^{(0)}(k) - \sum_{j=2}^{N} d_j\left(1 - b_j/a_j\right)/a - a_j\left(1 - e^{a_j}\right)e^{-a_j(k-1)}\right)}{\sum_{k=1}^{n}\left(1 - e^{a}\right)^2 e^{-2ak}}. \tag{23}$$

## 3. The Proposed Multivariate Grey Prediction Model

For the proposed MGPM, GRA was first used to keep explanatory variables that are more relevant to $CO_2$ emissions. Then, the proposed MGPM can be constructed by GAs. Subsequently, a residual GM (1, 1) model can be embedded into the proposed MGPM by establishing a functional-link net to adjust $\hat{x}_1^{(0)}(k)$.

*3.1. Feature Selection by Grey Relational Analysis.* In contrast to statistical correlation analysis that measures the relationship between any two random variables, GRA can effectively measure the relationships between one reference sequence and the other comparative sequences by viewing the reference sequence as the desired goal [44]. The grey relational coefficient, $\xi_{jk}$, for the time period $k$ ($1 \le j \le N-1$, $1 \le k \le m$) is addressed by the discriminative coefficient $\rho$ ($0 \le \rho \le 1$) to indicate the relationship between $x_j^{(0)}(k)$ and $x_1^{(0)}(k)$:

$$\xi jk = \frac{\Delta_{\min} + \rho\Delta_{\max}}{\Delta_{jk} + \rho\Delta_{\max}}, \tag{24}$$

where $\rho$ is usually specified as 0.5 and

$$\begin{aligned} \Delta\min &= \min_{j=1,\ldots,N-1} \min_{k=1,\ldots,m} \left|x_1^{(0)}(k) - x_j^{(0)}(k)\right|, \\ \Delta\min &= \max_{j=1,\ldots,N-1} \max_{k=1,\ldots,m} \left|x_1^{(0)}(k) - x_j^{(0)}(k)\right|, \\ \Delta jk &= \left|x_1^{(0)}(k) - x_j^{(0)}(k)\right|. \end{aligned} \tag{25}$$

It is seen that $\xi_{jk}$ lies in [0, 1] and approaches 1 if $\Delta_{jk}$ approaches $\Delta_{\min}$.

To measure the degree of proximity between $x_j$ and $x_1$, the grey relational grade (GRG) $\gamma_j$ can be calculated as follows:

$$\gamma_j = \frac{1}{m} \sum_{k=1}^{m} \xi_{ik}, \tag{26}$$

where $\gamma_j$ ranges from 0 to 1. The greater the value of $\gamma_i$ is, the more relevant $x_j$ is to $x_1$. In other words, to construct the proposed model, $x_j$ can be retained for constructing the

proposed MGPM when $\gamma_j$ surpasses a threshold value $\lambda$; otherwise, $x_j$ can no longer be considered for the proposed MGPM. That means, for equation (3), both $b_j$ and $x_j^{(1)}(k+1)$ associated with $x_j$ can be retained when $\gamma_j$ is above $\lambda$; otherwise, they can be removed directly. However, $\lambda$ may not be prespecified easily beforehand.

*3.2. Construction of Multivariate Grey Prediction Models Using Genetic Algorithms.* This study aims to find the optimal solution to construct the proposed MGPM with high prediction accuracy. This problem can be thus formulated as the following single objective optimization problem by minimizing the mean absolute percentage error (MAPE):

$$\text{Minimize}\,\frac{1}{m-1}\sum_{k=2}^{m}\frac{\left|x_1^{(0)}(k)-\widehat{x}_1^{(0)}(k)\right|}{x_1^{(0)}(k)}. \tag{27}$$

Instead of OLS, a real-valued GA using MAPE as its fitness function can be developed to automatically determine the optimal values of development coefficient ($a$), the driving coefficients ($b_2, b_3, \ldots, b_N$), and the cut value ($\lambda$).

For the GA, a set of strings making up a population is generated initially. All parameters for each string are randomly generated as real numbers. Using the Genetic Algorithm and the Direct Search Toolbox in MATLAB, a real-valued GA is easily developed to automatically determine all parameters. The best chromosome with the maximum fitness value of all successive generations is the desired solution for examining the generalizability of the proposed MGPM.

*3.3. Residual Modification Using a Functional-Link Net.* $\widehat{x}_1^{(0)}(k)$ produced by the proposed MGPM can be further adjusted by residual modification to improve prediction accuracy. Let $\varepsilon_1^{(0)} = (\varepsilon_1^{(0)}(2), \varepsilon_1^{(0)}(3), \ldots, \varepsilon_1^{(0)}(m))$ denote the sequence of absolute residual values, where

$$\varepsilon_1^{(0)}(k) = \left|\varepsilon_1^{(0)}(k) - \widehat{\varepsilon}_1^{(0)}\right|, \quad k = 2, 3, \ldots, m. \tag{28}$$

Because no dependent variables are considered for $\varepsilon_1$, a residual GM (1, 1) model can be established for $\varepsilon_1^{(0)}$, where the predicted value of $\varepsilon_1^{(0)}(k)$ is

$$\widehat{\varepsilon}_1^{(0)}(k) = \left(1-e^{a_\varepsilon}\right)\left(\varepsilon_1^{(0)}(2) - \frac{b_\varepsilon}{a_\varepsilon}\right)e^{-ae(k-1)}, \quad k = 3, 4, \ldots, m, \tag{29}$$

where $a_\varepsilon$ and $b_\varepsilon$ are the developing coefficient and the control variable, respectively. $\widehat{\varepsilon}_1^{(1)}(2) = \varepsilon_1^{(0)}(2)$. The problem of constructing a residual GM (1, 1) model can be formulated as the following single objective optimization problem by minimizing the MAPE:

$$\text{Minimize}\sum_{k=3}^{m}\frac{\left|\varepsilon_1^{(0)}(k)-\widehat{\varepsilon}_1^{(0)}(k)\right|}{\varepsilon_1^{(0)}(k)}. \tag{30}$$

MAPE can be used as the fitness function of a real-valued GA that is used to determine optimal $a_\varepsilon$ and $b_\varepsilon$ instead of OLS.

Then, following a mechanism of residual modification recommended by Wang and Hu [41], the final predicted value $\widehat{x}_k^{(0)}$ of the proposed MGPM is produced by means of $\widehat{x}_1^{(0)}(k)$:

$$\widehat{x}_1^{(0)}(k) = \widehat{x}_1^{(0)}(k) + 3y_k\widehat{\varepsilon}_1^{(0)}(k), \quad k = 2, 3, \ldots, n, \tag{31}$$

where $y_k$ ranges from $-1$ to 1 and can be computed by presenting ($t_k$, $\sin(\pi t_k)$, $\cos(\pi t_k)$, $\sin(2\pi t_k)$, and $\cos(2\pi t_k)$) to a single-layer perceptron, namely, the functional-link net, with effective function approximation capability [45–48]:

$$\begin{aligned} y_k = \tanh\big(&w_1 t_k + w_2\sin(\pi t_k) + w_3\cos(\pi t_k) \\ &+ w_4\sin(2\pi t_k) + w_5\cos(2\pi t_k) + \theta\big), \end{aligned} \tag{32}$$

where $\tanh$ represents a hyperbolic tangent function, and $w_1, \ldots, w_5$ are connection weights. This means that the amount of adjusting $\widehat{x}_k^{(0)}$ could be as much as $3\widehat{\varepsilon}_k^{(0)}$ if $y_k = 1$. In contrast, the adjustable amount can be $-3\widehat{\varepsilon}_k^{(0)}$ if $y_k = -1$.

## 4. Empirical Results

Empirical studies were conducted using real datasets to compare the $CO_2$ emission forecasting ability of the proposed MGPM with the other MGPMs considered. As mentioned above, urban population (UP), GDP, and energy consumption have a dominant influence on $CO_2$ emissions. In addition to the traditional GM (1, $N$), the Autoregressive Integrated Moving Average model (ARIMA) and the aforementioned improved MGPMs with comprehensible distinctive features were considered.

*4.1. Case I.* Statistics from the IEA [3] revealed that, in 2015, the total amount of $CO_2$ emissions in China was 9,040 million tons, reaching the highest level worldwide. To devise energy plans that would effectively reduce $CO_2$ emissions while promoting green economic growth, the ability to predict $CO_2$ emissions has played a very significant role in China. Therefore, we were intrigued to examine the prediction performance of MGPMs that consider $CO_2$ emissions in China. The data on urban population (million persons) and GDP (million USD dollars) were collected from the World Bank (http://data.worldbank.org.cn), and energy consumption (million tons of oil equivalent) and $CO_2$ emissions (million tons) were collected from the IEA (http://www.iea.org).

As shown in Table 1, the historical annual data were collected from 2005 to 2015, data from 2005 to 2012 were used for the model-fitting, and data from 2013 to 2015 were used for ex-post testing. Results shown in Table 2 are summarized as follows:

(1) The MAPE of the traditional GM (1, $N$), the TNGM (1, $N$), the OGM (1, $N$), the TDVGM (1, $N$), the ARIMA, and the proposed MGPM for model-fitting were 3.34%, 0.81%, 0.02%, 3.14%, 1.98%, and 1.71%, respectively. The OGM (1, $N$) thus demonstrated its superiority in model-fitting.

Table 1: Annual carbon dioxide emissions with GDP, UP, and EC in China.

| Year | GDP | UP | EC | $CO_2$ |
|---|---|---|---|---|
| 2005 | 2285966 | 554.37 | 1184.158 | 5357.71 |
| 2006 | 2752132 | 575.12 | 1273.683 | 5911.96 |
| 2007 | 3552182 | 595.67 | 1368.733 | 6468.27 |
| 2008 | 4598206 | 616.48 | 1417.323 | 6608.14 |
| 2009 | 5109954 | 637.41 | 1480.672 | 7025.82 |
| 2010 | 6100620 | 658.50 | 1578.852 | 7706.65 |
| 2011 | 7572554 | 679.77 | 1692.063 | 8465.64 |
| 2012 | 8560547 | 700.86 | 1747.103 | 8620.58 |
| 2013 | 9607224 | 721.69 | 1816.852 | 8995.79 |
| 2014 | 10482372 | 742.30 | 1868.17 | 9036.47 |
| 2015 | 11064666 | 762.59 | 1905.679 | 9040.74 |

Table 2: MAPE (%) obtained by different MGPMs for carbon dioxide emissions in China.

| Year | Actual | GM (1, *N*) | | TNGM (1, *N*) | | OGM (1, *N*) | | TDVGM (1, *N*) | | ARIMA | | The proposed MGPM | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Predicted | APE | Predicted | APE | Predicted | APE | Predicted | APE | Predicted | APE | Predicted | APE |
| 2005 | 5357.7 | 5357.7 | 0.00 | 5357.7 | 0.00 | 5357.7 | 0.00 | 5435.9 | 1.32 | 5357.71 | 0.00 | 5357.7 | 0.00 |
| 2006 | 5912.0 | 5007.9 | 15.29 | 5869.3 | 0.72 | 5912.1 | 0.00 | 5786.9 | 2.12 | 5916.89 | 0.08 | 5919.4 | 0.13 |
| 2007 | 6468.3 | 6941.3 | 7.31 | 6478.3 | 0.15 | 6467.7 | 0.01 | 6160.3 | 4.76 | 6437.24 | 0.48 | 6850.5 | 5.91 |
| 2008 | 6608.1 | 6753.8 | 2.20 | 6495.6 | 1.70 | 6609.7 | 0.02 | 6551.9 | 0.85 | 6919.37 | 4.71 | 6739.0 | 1.98 |
| 2009 | 7025.8 | 7003.8 | 0.31 | 7083.0 | 0.81 | 7023.3 | 0.04 | 6962.3 | 0.90 | 7366.09 | 4.84 | 7026.6 | 0.01 |
| 2010 | 7706.7 | 7651.5 | 0.72 | 7703.0 | 0.05 | 7709.3 | 0.03 | 7390.9 | 4.10 | 7779.98 | 0.95 | 7556.9 | 1.94 |
| 2011 | 8465.6 | 8429.1 | 0.43 | 8255.0 | 2.49 | 8464.0 | 0.02 | 7838.3 | 7.41 | 8163.48 | 3.57 | 8186.0 | 3.30 |
| 2012 | 8620.6 | 8660.7 | 0.47 | 8665.3 | 0.52 | 8621.1 | 0.01 | 8303.9 | 3.67 | 8518.80 | 1.18 | 8582.9 | 0.44 |
| MAPE | — | — | 3.34 | — | 0.81 | — | 0.02 | — | 3.14 | — | 1.98 | — | 1.71 |
| 2013 | 8995.8 | 9039.7 | 0.49 | 9011.6 | 0.18 | 9080.8 | 0.94 | 8788.2 | 2.31 | 8848.02 | 1.64 | 8990.6 | 0.06 |
| 2014 | 9036.5 | 9248.2 | 2.34 | 9270.5 | 2.59 | 9282.2 | 2.72 | 9290.3 | 2.81 | 9153.05 | 1.29 | 9242.6 | 2.28 |
| 2015 | 9040.7 | 9342.8 | 3.34 | 9452.7 | 4.56 | 9419.1 | 4.19 | 9809.9 | 8.51 | 9435.68 | 4.37 | 9335.2 | 3.26 |
| MAPE | — | — | 2.06 | — | 2.44 | — | 2.62 | — | 4.54 | — | 2.43 | — | 1.87 |

(2) For ex-post testing, the MAPE of the traditional GM (1, *N*), the TNGM (1, *N*), the OGM (1, *N*), the TDVGM (1, *N*), the ARIMA, and the proposed MGPM were 2.06%, 2.44%, 2.62%, 4.54%, 2.43%, and 1.87%, respectively.

(3) Although the proposed MGPM is inferior to the TNGM (1, *N*) and the OGM (1, *N*) for model-fitting, it is superior to other MGPMs considered for ex-post testing.

It should be noted that, when evaluating a prediction model, more emphasis should be placed on generalization rather than on model-fitting [49]. Figure 1 also demonstrates the superiority of the generalization ability of the proposed MGPM over the other prediction models considered.

*4.2. Case II.* From statistics reported by the IEA [47], the total and average amounts of $CO_2$ emissions in Taiwan in 2015 were the 21st and the 19th highest in the world, respectively. This means that Taiwan still has room to reduce $CO_2$ emissions. The second real case involved the historical annual $CO_2$ emission data collected in Taiwan from 2005 to 2015. The data on urban population (million persons) and GDP (million USD dollars) were collected from the United Nations Conference on Trade and Development (UNCTAD) (http://unctad.org/en/Pages/statistics.aspx), and energy consumption (million tons of oil equivalent) and $CO_2$ emissions (million tons) were collected from the IEA.

As shown in Table 3, data collected from 2005 to 2012 were used for the model-fitting, and data from 2013 to 2015 were used for ex-post testing. The results obtained from the different prediction models are shown in Table 4. The results are summarized as follows:

(1) The MAPE of the traditional GM (1, *N*), the TNGM (1, *N*), the OGM (1, *N*), the TDVGM (1, *N*), the ARIMA, and the proposed MGPM for model-fitting were 3.35%, 0.23%, 0.21%, 2.59%, 2.10%, and 1.86%, respectively. The TNGM (1, *N*) and the OGM (1, N) demonstrated their superiority in model-fitting.

(2) The MAPE of the traditional GM (1, *N*), the TNGM (1, *N*), the OGM (1, *N*), the TDVGM (1, *N*), the ARIMA, and the proposed MGPM for ex-post testing were 1.28%, 4.75%, 8.43%, 4.54%, 0.86%, and 1.09%, respectively. The proposed MGPM is slightly inferior to the TDVGM (1, *N*).

(3) Although the proposed MGPM is inferior to the TNGM (1, *N*) and the OGM (1, *N*) for model-fitting, it is superior to those two MGPMs for ex-post testing.

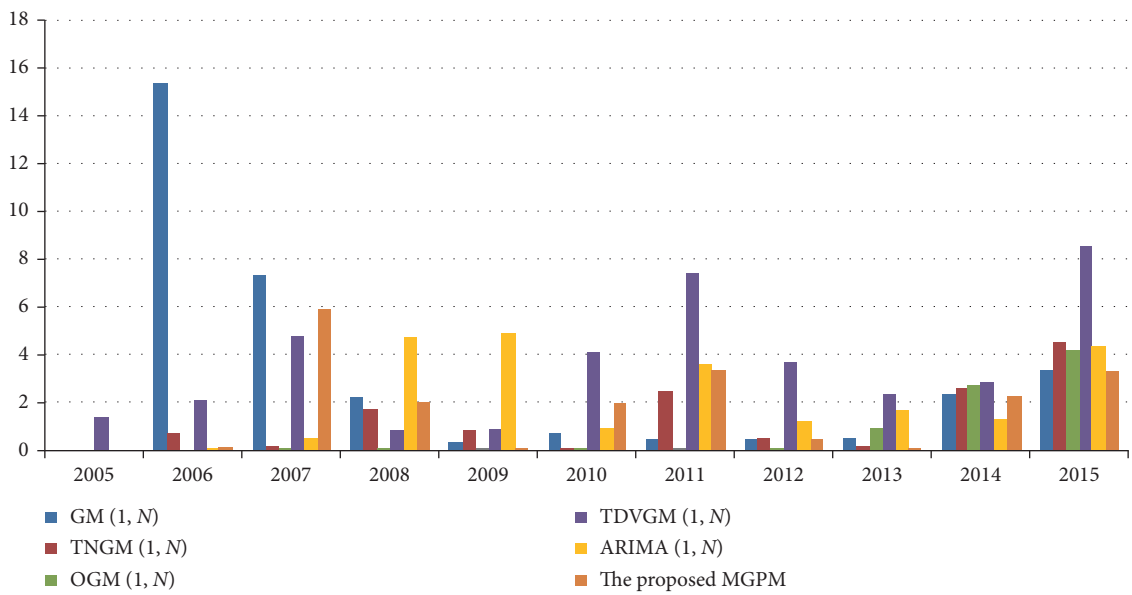



Figure 1: Comparisons between different MGPMs for carbon dioxide emissions prediction in China.

Table 3: Annual carbon dioxide emissions with GDP, UP, and EC in Taiwan.

| Year | GDP | UP | EC | $CO_2$ |
|---|---|---|---|---|
| 2005 | 375787 | 22.603 | 60.437 | 253.64 |
| 2006 | 388547 | 22.725 | 61.276 | 260.86 |
| 2007 | 408221 | 22.833 | 65.511 | 263.94 |
| 2008 | 417038 | 22.929 | 63.393 | 252.75 |
| 2009 | 392106 | 23.017 | 62.892 | 239.68 |
| 2010 | 446141 | 23.102 | 67.855 | 256.22 |
| 2011 | 485671 | 23.185 | 65.405 | 254.7 |
| 2012 | 495919 | 23.264 | 65.289 | 246.55 |
| 2013 | 511599 | 23.34 | 67.808 | 247.59 |
| 2014 | 530515 | 23.414 | 68.014 | 249.66 |
| 2015 | 525236 | 23.486 | 68.566 | 249.38 |

Table 4: MAPE (%) obtained by different MGPMs for carbon dioxide emissions in Taiwan.

| Year | Actual | GM (1, $N$) | | TNGM (1, $N$) | | OGM (1, $N$) | | TDVGM (1, $N$) | | ARIMA | | The proposed MGPM | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Predicted | APE | Predicted | APE | Predicted | APE | Predicted | APE | Predicted | APE | Predicted | APE |
| 2005 | 253.64 | 253.64 | 0.00 | 253.64 | 0.00 | 253.64 | 0.00 | 250.18 | 1.36 | 253.64 | 0.00 | 253.64 | 0.00 |
| 2006 | 260.86 | 220.24 | 15.57 | 260.33 | 0.20 | 260.38 | 0.18 | 250.79 | 3.86 | 255.91 | 1.90 | 261.36 | 0.19 |
| 2007 | 263.94 | 270.58 | 2.52 | 264.40 | 0.17 | 264.29 | 0.13 | 251.06 | 4.88 | 251.58 | 4.68 | 258.87 | 1.92 |
| 2008 | 252.75 | 253.11 | 0.14 | 253.18 | 0.17 | 252.48 | 0.11 | 251.24 | 0.60 | 252.39 | 0.14 | 249.56 | 1.26 |
| 2009 | 239.68 | 252.71 | 5.44 | 239.68 | 0.00 | 239.66 | 0.01 | 251.36 | 4.87 | 252.24 | 5.24 | 233.84 | 2.44 |
| 2010 | 256.22 | 256.63 | 0.16 | 255.82 | 0.16 | 256.81 | 0.23 | 251.43 | 1.87 | 252.26 | 1.54 | 250.40 | 2.27 |
| 2011 | 254.70 | 249.01 | 2.23 | 252.97 | 0.68 | 252.88 | 0.71 | 251.43 | 1.28 | 252.26 | 0.96 | 241.07 | 5.35 |
| 2012 | 246.55 | 248.41 | 0.75 | 247.71 | 0.47 | 247.24 | 0.28 | 251.37 | 1.95 | 252.26 | 2.32 | 242.89 | 1.48 |
| MAPE | — | — | 3.35 | — | 0.23 | — | 0.21 | — | 2.59 | — | 2.10 | — | 1.86 |
| 2013 | 253.64 | 251.97 | 1.77 | 254.77 | 2.90 | 257.98 | 4.20 | 251.24 | 1.47 | 252.26 | 1.89 | 249.91 | 0.94 |
| 2014 | 260.86 | 251.07 | 0.56 | 260.44 | 4.32 | 268.40 | 7.51 | 251.04 | 0.55 | 252.26 | 1.04 | 251.72 | 0.83 |
| 2015 | 263.94 | 253.13 | 1.50 | 266.89 | 7.02 | 283.29 | 13.60 | 250.77 | 0.56 | 252.26 | 1.15 | 253.15 | 1.51 |
| MAPE | — | — | 1.28 | — | 4.75 | — | 8.43 | — | 0.86 | — | 1.36 | — | 1.09 |

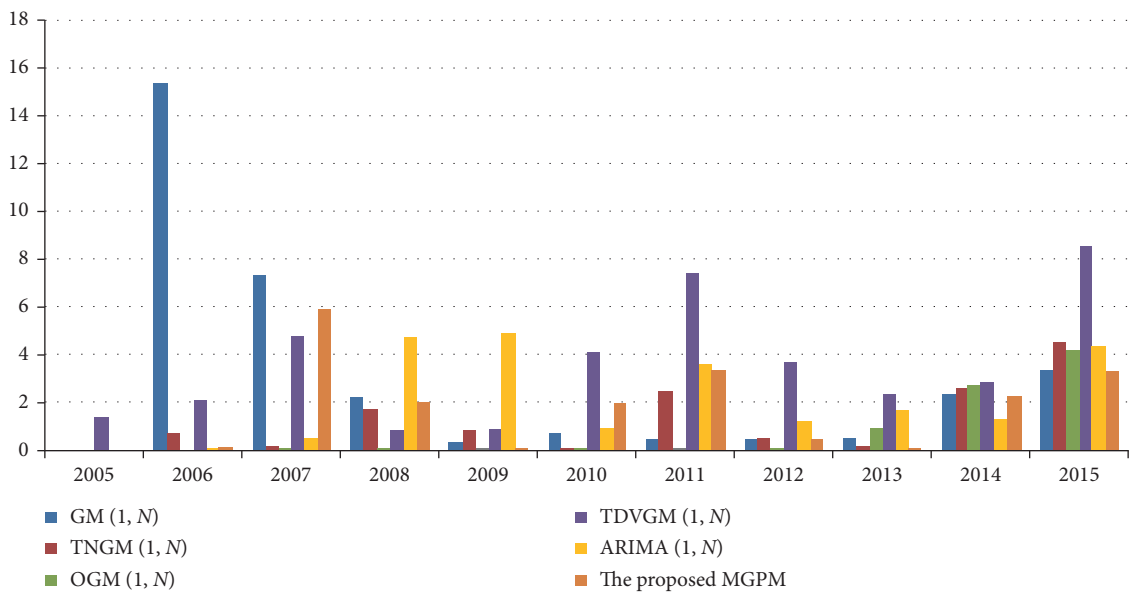



Figure 2: Comparisons between different MGPMs for carbon dioxide emissions prediction in Taiwan.

In Figure 2, we can see that the proposed MGPM performs well compared with other prediction models considered.

## 5. Discussion

What this study focuses on is the forecast rather than the projection. Compared with the projection, the forecast places value on estimating the amount of $CO_2$ emissions based on an established model such as a MGPM for which MAPE is a useful metric for measuring the prediction performance [50]. The projection can answer "what-if" kind of questions to extrapolate the development trend. In other words, it is concerned about what would happen to $CO_2$ emission based on some future scenarios, for instance, how $CO_2$ is produced and how it is influenced by what kind of factors.

As mentioned above, the forecast of $CO_2$ emissions can be regarded as a grey system problem. Furthermore, $CO_2$ emission data might well not conform to statistical assumptions. Therefore, it is reasonable to apply MGPMs to forecast the amount of $CO_2$ emissions. Compared with the other MGPMs, feature selection and residual modification are taken into account in the proposed MGPM to improve prediction accuracy. In particular, GRA and a functional-link net are employed to implement feature selection and residual modification, respectively. The empirical results reveal that feature selection and residual modification can boost the prediction performance of the proposed MGPM. It is noted that although the OGM (1, $N$) variant [25] also applied GRA to filter out irrelevant features, the cut value ($\lambda$) was arbitrarily assigned by a prespecified value. However, $\lambda$ can be optimized by the GA to improve the prediction accuracy of the proposed MGPM.

With real-world datasets, from Tables 2 and 4, we can see that the generalization ability of the proposed MGPM for $CO_2$ emissions was quite encouraging. The outcomes verified that the results obtained by the proposed MGPM are comparable to other prediction models considered. It is interesting to note that the TNGM (1, $N$) and the OGM (1, $N$) are superior to the traditional GM (1, $N$) and the proposed MGPM for model-fitting but inferior for ex-post testing. In other words, both the TNGM (1, $N$) and the OGM (1, $N$) appear to be overfitting. Experimental results show that the fitting and generalization abilities of the proposed MGPM are superior to the traditional GM (1, $N$). Thus, the prediction ability of the traditional GM (1, $N$) could be effectively improved by feature selection and residual modification indeed.

## 6. Conclusions

Undoubtedly, the reduction of greenhouse gas emissions is critical to environmental protection. For many countries, $CO_2$ is mainly produced from fuel combustion, which forms the majority of greenhouse gases. How to reduce the impact that energy consumption and economic growth have on $CO_2$ emissions has gained increasing global attention. Revelations from the IEA [3] showed that, along with global economic growth, global $CO_2$ emissions plateaued from 2014 to 2016 due to the growth of renewable electricity, the replacement of coal with natural gas, and changes to the economic structure. There appeared to be a decoupling of economic growth and environmental degradation. However, this does not mean that $CO_2$ emissions have reached a summit. To continuously inhibit carbon emissions and remain competitive, it is necessary for the authorities to make use of prediction models on carbon emissions, set up comprehensive policies on the development of new energy technologies, and increase demand for renewable energies (e.g., solar, wind, and natural gas) and environmental protection.

In addition to $CO_2$ emissions, there are several multivariate prediction problems, such as energy demand

forecasting, which need to be resolved. In fact, energy demand prediction has become increasingly important when devising development plans for a country, particularly for developing countries [51]. Meanwhile, energy demand forecasting can be regarded as a grey system problem [52] because a few factors, such as income and population, have an influence on energy demand, but how exactly these factors affect energy demand is not clear. On the basis of the conspicuous forecasting performance of the proposed MGPM for $CO_2$ emissions, it would be interesting to explore its applicability to energy demand forecasting.

## Data Availability

Statistics used in this paper are from the IEA (International Energy Agency).

## Conflicts of Interest

The authors declare that there are no conflicts of interest.

## Acknowledgments

This research was supported by the Ministry of Science and Technology, Taiwan, under Grant MOST 108-2410-H-033-038-MY2, the National Natural Science Foundation of China (no. 71772106), Humanities and Social Sciences Foundation of the Ministry of Education of China (no. 17YJCZH198), Shandong Provincial Natural Science Foundation, China (No. ZR2017MG012).

## References

[1] P. Nema, S. Nema, and P. Roy, "An overview of global climate changing in current scenario and mitigation action," *Renewable and Sustainable Energy Reviews*, vol. 16, no. 4, pp. 2329–2336, 2012.

[2] W. Wu, X. Ma, Y. Zhang, and Y. Wang, "A novel conformable fractional non-homogeneous grey model for forecasting carbon dioxide emissions of BRICS countries," *Science of The Total Environmen*, vol. 707, p. 13544, 2020.

[3] International Energy Agency, *Key World Energy Statistics 2017*, IEA/OECD, Paris, France, 2019.

[4] S. Liu and Y. Lin, *Grey Information: Theory and Practical Applications*, Springer, Berlin, Germany, 2010.

[5] S. Liu, Y. Yang, and J. Forrest, *Grey Data Analysis: Methods, Models and Applications*, Springer, Berlin, Germany, 2017.

[6] J. L. Deng, "Control problems of grey systems," *Systems and Control Letters*, vol. 1, no. 5, pp. 288–294, 1982.

[7] S. Ding, Y.-G. Dang, X.-M. Li, J.-J. Wang, and K. Zhao, "Forecasting Chinese CO2 emissions from fuel combustion using a novel grey multivariable model," *Journal of Cleaner Production*, vol. 162, pp. 1527–1538, 2017.

[8] L. Wu, S. Liu, D. Liu, Z. Fang, and H. Xu, "Modelling and forecasting CO 2 emissions in the BRICS (Brazil, Russia, India, China, and South Africa) countries using a novel multivariable grey model," *Energy*, vol. 79, pp. 489–495, 2015.

[9] O. Burcu, "The nexus between carbon emissions, energy consumption and economic growth in Middle East countries: a panel data analysis," *Energy Policy*, vol. 62, pp. 1138–1147, 2013.

[10] P. A. González and J. M. Zamarreño, "Prediction of hourly energy consumption in buildings based on a feedback artificial neural network," *Energy and Buildings*, vol. 37, no. 6, pp. 595–601, 2005.

[11] P. Lauret, E. Fock, R. N. Randrianarivony, and J.-F. Manicom-Ramsamy, "Bayesian neural network approach to short time load forecasting," *Energy Conversion and Management*, vol. 49, no. 5, pp. 1156–1166, 2008.

[12] C. Xia, J. Wang, and K. McMenemy, "Short, medium and long term load forecasting model and virtual load forecaster based on radial basis function neural networks," *International Journal of Electrical Power & Energy Systems*, vol. 32, no. 7, pp. 743–750, 2010.

[13] W. Sun and M. Liu, "Prediction and analysis of the three major industries and residential consumption CO2 emissions based on least squares support vector machine in China," *Journal of Cleaner Production*, vol. 122, pp. 144–153, 2016.

[14] O. Kaynar, I. Yilmaz, and F. Demirkoparan, "Forecasting of natural gas consumption with neural network and neuro fuzzy system," *Energy Education Science and Technology Part A: Energy Science and Research*, vol. 26, pp. 221–238, 2011.

[15] R. Li, X. Chen, T. Balezentis, D. Streimikiene, and Z. Niu, "Multi-step least squares support vector machine modeling approach for forecasting short-term electricity demand with application," *Neural Computing and Applications*, 2020.

[16] M. Meng and D. Niu, "Modeling $CO_2$ emissions from fossil fuel combustion using the logistic equation," *Energy*, vol. 36, no. 5, pp. 3355–3359, 2011.

[17] A. Azadeh, M. Khakestani, and M. Saberi, "A flexible fuzzy regression algorithm for forecasting oil consumption estimation," *Energy Policy*, vol. 37, no. 12, pp. 5567–5579, 2009.

[18] Q. Yan, W. Zhang, J. Yuan, T. Balezentis, and Y. Zhang, "How much electricity will be consumed in 2020 under the new normal economy in China?" *Transformations in Business & Economics*, vol. 18, no. 2, pp. 88–102, 2019.

[19] V. Ş. Ediger and S. Akar, "ARIMA forecasting of primary energy demand by fuel in Turkey," *Energy Policy*, vol. 35, no. 3, pp. 1701–1708, 2007.

[20] S. Tutun, C.-A. Chou, and E. Canıyılmaz, "A new forecasting framework for volatile behavior in net electricity consumption: a case study in Turkey," *Energy*, vol. 93, pp. 2406–2422, 2015.

[21] C.-H. Wang and L.-C. Hsu, "Using genetic algorithms grey theory to forecast high technology industrial output," *Applied Mathematics and Computation*, vol. 195, no. 1, pp. 256–263, 2008.

[22] K. L. Wen, *Grey Systems Modeling and Prediction*, Yang's Scientific Research Institute, Tucson, AZ, USA, 2004.

[23] H. Duan, D. Wang, X. Pang, Y. Liu, and S. Zeng, "A novel forecasting approach based on multi-kernel nonlinear multivariable grey model: a case report," *Journal of Cleaner Production*, vol. 260, 2020.

[24] M. Guo, J. Lan, J. Li, Z. Lin, and X. Sun, "Traffic flow data recovery algorithm based on gray residual GM (1, *N*) model," *Journal of Transportation Systems Engineering and Information Technology*, vol. 12, no. 1, pp. 42–47, 2012.

[25] B. Zeng, C. Luo, S. Liu, and C. Li, "A novel multi-variable grey forecasting model and its application in forecasting the amount of motor vehicles in Beijing," *Computers & Industrial Engineering*, vol. 101, pp. 479–489, 2016.

[26] Z.-X. Wang, "A GM (1, *N*)-based economic cybernetics model for the high-tech industries in China," *Kybernetes*, vol. 43, no. 5, pp. 672–685, 2014.

[27] L. L. Pei, W. M. Chen, J. H. Bai, and Z. X. Wang, "The improved GM (1, *N*) models with optimal background values: a case study of Chinese high-tech industry," *The Journal of Grey System*, vol. 27, no. 3, pp. 223–233, 2015.

[28] Z.-X. Wang and D.-J. Ye, "Forecasting Chinese carbon emissions from fossil energy consumption using non-linear grey multivariable models," *Journal of Cleaner Production*, vol. 142, pp. 600–612, 2017.

[29] Z. X. Wang and P. Hao, "An improved grey multivariable model for predicting industrial energy consumption in China," *Applied Mathematical Modelling*, vol. 40, no. 11-12, pp. 5745–5758, 2016.

[30] L. C. Hsu, "Forecasting the output of integrated circuit industry using genetic algorithm based multivariable grey optimization models," *Expert Systems with Applications*, vol. 36, no. 2, pp. 7898–7903, 2009.

[31] L.-C. Hsu and C.-H. Wang, "Forecasting integrated circuit output using multivariate grey model and grey relational analysis," *Expert Systems with Applications*, vol. 36, no. 2, pp. 1403–1409, 2009.

[32] H. Jiang and Y.-C. Hu, J.-Y. Lin and P. Jiang, "Analyzing China's OFDI using a novel multivariate grey prediction model with Fourier series" *International Journal of Intelligent Computing and Cybernetics*, vol. 12, no. 3, pp. 352–371, 2019.

[33] W. B. Wang and Y. C. Hu, "Multivariate grey prediction models for pattern classification irrespective of time series," *Journal of Grey System*, vol. 31, no. 2, pp. 135–142, 2019.

[34] Y.-C. Hu, "A multivariate grey prediction model with grey relational analysis for bankruptcy prediction problems," *Soft Computing*, vol. 24, no. 6, pp. 4259–4268, 2020.

[35] B. Zeng, C. Luo, S. Liu, Y. Bai, and C. Li, "Development of an optimization method for the GM (1, *N*) model," *Engineering Applications of Artificial Intelligence*, vol. 55, pp. 353–362, 2016.

[36] H.-T. Pao, H.-C. Fu, and C.-L. Tseng, "Forecasting of CO2 emissions, energy consumption and economic growth in China using an improved grey model," *Energy*, vol. 40, no. 1, pp. 400–409, 2012.

[37] N. Xu, S. Ding, Y. Gong, and J. Bai, "Forecasting Chinese greenhouse gas emissions from energy consumption using a novel grey rolling model," *Energy*, vol. 175, pp. 218–227, 2019.

[38] M. Xie, L. Wu, B. Li, and Z. Li, "A novel hybrid multivariate nonlinear grey model for forecasting the traffic-related emissions," *Applied Mathematical Modelling*, vol. 77, pp. 1242–1254, 2020.

[39] S. Ding, N. Xu, J. Ye, W. Zhou, and X. Zhang, "Estimating Chinese energy-related CO2 emissions by employing a novel discrete grey prediction model," *Journal of Cleaner Production*, vol. 259, Article ID 120793, 2020.

[40] Z.-X. Wang and Q. Li, "Modelling the nonlinear relationship between CO2 emissions and economic growth using a PSO algorithm-based grey Verhulst model," *Journal of Cleaner Production*, vol. 207, pp. 214–224, 2019.

[41] W. B. Wang and Y. C. Hu, "A novel grey residual modification model using neural networks," *Journal of Grey System*, vol. 30, no. 4, pp. 34–46, 2018.

[42] Y.-C. Hu, "A novel flow-based method using grey relational analysis for pattern classification," *International Journal of Information Technology & Decision Making*, vol. 12, no. 1, pp. 75–93, 2013.

[43] S. J. Feng, Y. D. Ma, Z. L. Song, and J. Ying, "Forecasting the energy consumption of China by the grey prediction model," *Energy Sources, Part B: Economics, Planning, and Policy*, vol. 7, no. 4, pp. 376–389, 2012.

[44] Y.-C. Hu, R.-S. Chen, Y.-T. Hsu, and G.-H. Tzeng, "Grey self-organizing feature maps," *Neurocomputing*, vol. 48, no. 1–4, pp. 863–877, 2002.

[45] Y. H. Pao, *Adaptive Pattern Recognition and Neural Networks*, Addison-Wesley, Boston, MA, USA, 1989.

[46] Y.-H. Pao and Y. Takefuji, "Functional-link net computing: theory, system architecture, and functionalities," *Computer*, vol. 25, no. 5, pp. 76–79, 1992.

[47] G. H. Park and Y. H. Pao, "Unconstrained word-based approach for off-line script recognition using density-based random-vector functional-link net," *Neurocomputing*, vol. 31, no. 1–4, pp. 45–65, 2000.

[48] Y.-C. Hu, "Grey prediction with residual modification using functional-link net and its application to energy demand forecasting," *Kybernetes*, vol. 46, no. 2, pp. 349–363, 2017.

[49] D. R. Luo, K. Z. Guo, and H. R. Huang, "Regional economic forecasting combination model based on RAR + SVR," *Advances in Intelligent Systems and Computing, Fuzzy Information & Engineering And Operations Research & Management*, vol. 211, pp. 329–338, 2014.

[50] S. Aminikhanghahi and D. J. Cook, "A survey of methods for time series change point detection," *Knowledge and Information Systems*, vol. 51, no. 2, pp. 339–367, 2017.

[51] D. Pi, J. Liu, and X. Qin, "A grey prediction approach to forecasting energy demand in China," *Energy Sources, Part A: Recovery, Utilization, and Environmental Effects*, vol. 32, no. 16, pp. 1517–1528, 2010.

[52] L. Suganthi and A. A. Samuel, "Energy models for demand forecasting-A review," *Renewable and Sustainable Energy Reviews*, vol. 16, no. 2, pp. 1223–1240, 2012.