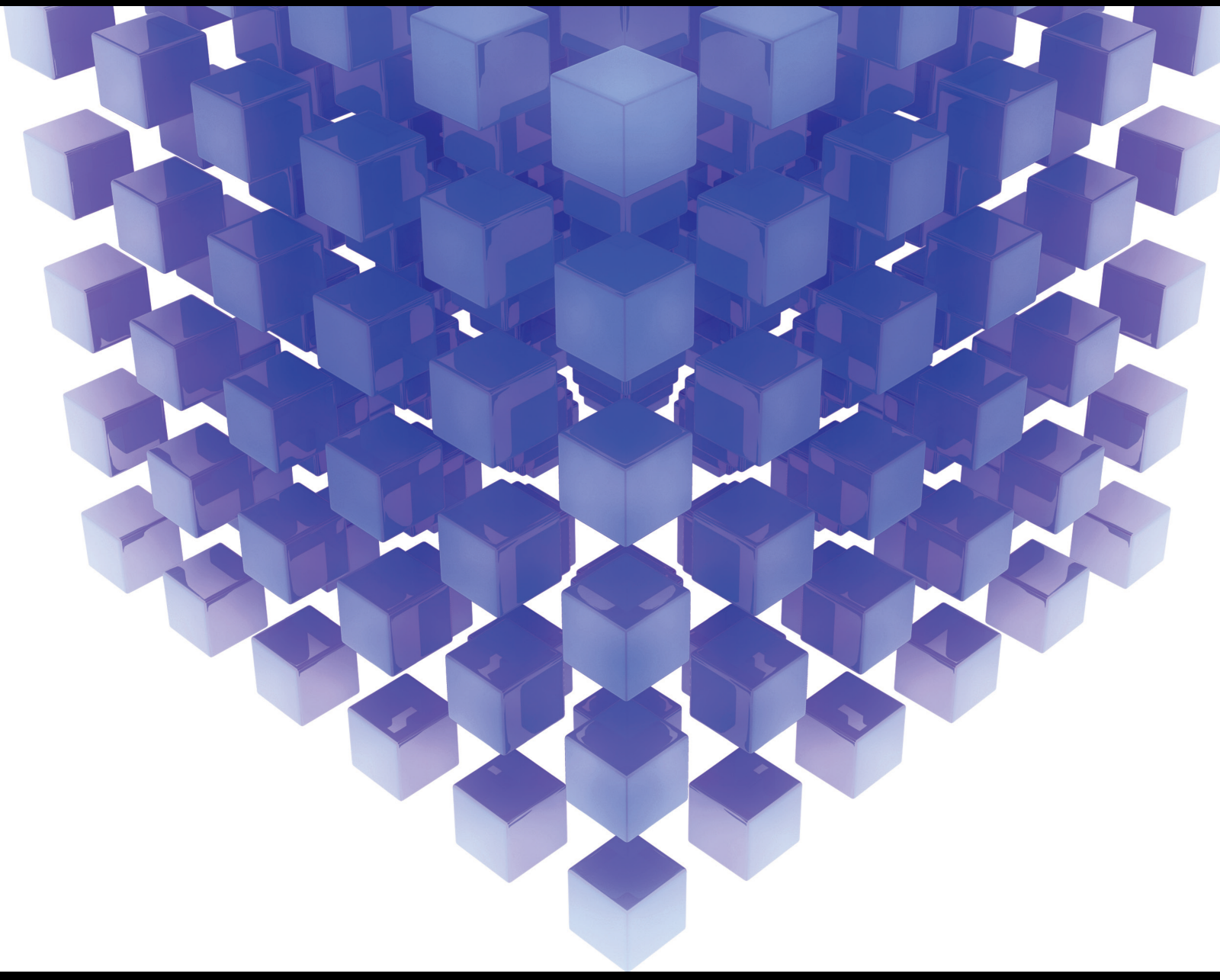
Mathematical Problems in Engineering

Structure, Dynamics, and Applications of Complex Networks in Software Engineering 2021

Lead Guest Editor: Weifeng Pan
Guest Editors: Hua Ming and Chunlai Chai



Structure, Dynamics, and Applications of Complex Networks in Software Engineering 2021

Mathematical Problems in Engineering

Structure, Dynamics, and Applications of Complex Networks in Software Engineering 2021

Lead Guest Editor: Weifeng Pan
Guest Editors: Hua Ming and Chunlai Chai

Chief Editor

Academic Editors

Contents

Hindawi
Mathematical Problems in Engineering
Volume 2022, Article ID 7416723, 18 pages
https://doi.org/10.1155/2022/7416723

Research Article

Interval-Valued q-Rung Orthopair Fuzzy Choquet Integral Operators and Their Application in Group Decision-Making

Benting Wan, Juelin Huang, Xi Chen, Youyu Cheng, and Jing Wang

School of Software and Internet of Things Engineering, Jiangxi University of Finance and Economics, Nanchang 330013, China

Correspondence should be addressed to Juelin Huang; 2202021682@stu.jxufe.edu.cn

Received 22 December 2021; Revised 18 April 2022; Accepted 6 May 2022; Published 22 August 2022

Academic Editor: Ali Ahmadian

In this paper, we develop a multiattribute group decision-making (MAGDM) method to solve problems with interactive attributes under interval-valued q-rung orthopair fuzzy set (IVq-ROFS) environment. Firstly, the interval-valued q-rung orthopair fuzzy Choquet integral average (IVq-ROFCA) operator is proposed to aggregate interval-valued q-rung orthopair fuzzy information. Then, we investigate the interval-valued q-rung orthopair fuzzy Choquet integral geometric (IVq-ROFCG) operator and offer several related properties. More importantly, for handling problems with interdependence between attributes for IVq-ROFS, a MAGDM method is developed based on the IVq-ROFCA operator. Finally, an example of the warning management system for hypertension is given to illustrate the proposed method, and parameter analysis and comparison analysis further verify the feasibility and validity of the proposed method.

1. Introduction

As a powerful and practical approach, fuzzy set (FS) [1] has been deeply used in various areas, such as medical treatment, manufacturing, and education . With the increase in people's awareness of complex and uncertain issues, fuzzy theories and methods have received great attention [2–5]. Since decision-makers need to deal with the possibilities of support, opposition, and neutrality in real life, Atanassov [6] proposed the intuitionistic fuzzy set (IFS), that is, the sum of the membership degree (u) and the nonmembership degree (v) satisfies $u + v \leq 1$. In response to the condition $u + v > 1$, Yager [7] investigated the Pythagorean fuzzy set (PFS) ($u^2 + v^2 \leq 1$). As a wider range compared with intuitionistic fuzzy, Verma and Merigo [8] defined a generalized hybrid trigonometric Pythagorean fuzzy similarity measure and developed the approach for Pythagorean fuzzy decision-making; Bakioglu and Atahan [9] addressed the prioritization of risks involved with self-driving vehicles by the proposed new hybrid MCDM method based on AHP and TOPSIS VIKOR under Pythagorean fuzzy environment; Peng and Yang [10] defined interval-valued Pythagorean fuzzy sets (IVPFSs) and developed two interval-valued Pythagorean fuzzy aggregation operators to solve MAGDM problems. Gradually, for dealing with much more complicated problems, Yager [11] proposed the q-rung orthopair fuzzy set (q-ROFS) ($u^q + v^q \leq 1, q \geq 1$) in 2016 and extended the fuzzy set to a wider range of applications based on the different values of q. Researchers have put forward numerous excellent results in recent years. Yager et al. [12, 13] further deduced the related properties and mathematical principles of q-ROFS. Liu and Wang proposed several weighted average operators of q-rung orthopair fuzzy numbers (q-ROFNs) [14]. Liu and Liu developed the q-rung orthopair fuzzy Bonferroni mean operator to deal with problems [15]. Xing et al. [16] developed the q-rung orthopair fuzzy point weighted aggregation operator and applied it to q-rung orthopair fuzzy decision-making. Garg investigated the trigonometric operation-based q-rung orthopair fuzzy aggregation operator for processing fuzzy information [17] and introduced a novel concept of the connection number-based q-rung orthopair fuzzy set (CN-q-ROFS) [18]. Hussain et al. [19] proposed q-rung orthopair fuzzy soft weighted averaging, q-rung orthopair fuzzy soft ordered weighted averaging, and q-rung orthopair fuzzy soft hybrid averaging operators. Verma [20] defined two order-α

divergence measures for q-ROFS to quantify the information of discrimination that determine completely unknown or partially known attribute weights. Aydemir and Gunduz [21] presented neutrality average and neutrality geometric aggregation operators and further designed a general score function for q-ROFS based on power aggregation operators. Liu et al. [22] constructed the normalized bidirectional projection model and generalized knowledge-based entropy measure under q-rung orthopair fuzzy environment. In particular, it is better to address strongly uncertain decision-making problems via IVq-ROFS, such that Garg [23] defined q-connection numbers (q-CNs) and gave some new q-exponential operation laws (q-EOLs) and operators over q-CNs for IVq-ROFS and presented the possibility of comparison between IVq-ROFSs [24]. Garg [25] introduced a novel concept of interval-valued q-rung orthopair fuzzy preference relations (IVq-ROFPRs) and then proposed additive consistent of IVq-ROFPR and programming model to derive the weights of alternatives.

However, attributes are not independent in decision-making, and there are more mutual influences and correlations that can be appropriately solved by the Choquet integral [26]. The Choquet integral comes with decision-making problems with independent attributes in handy [27–29], which had been conducted in an in-depth research study Tan and Chen defined the intuitionistic of the fuzzy measure based on the Choquet integral [30] and investigated the induced Choquet ordered averaging operator for aggregating expert evaluation [31]. Tan developed the generalized interval-valued intuitionistic fuzzy geometric aggregation (GIIFGA) operator and combined TOPSIS to deal with MAGDM problems [32]. Xu [33] investigated the intuitionistic fuzzy Choquet integral average operator and geometric operator, which are applied to address MAGDM problems with IFS. It is significant to mention that Grabisch [34] presented a synthesis on the application of fuzzy integrals and extended the application of fuzzy integrals in various fields. Considering the advantages of fuzzy integrals that simulate the interaction between standards in a flexible way, scholars have conducted a lot of studies on fuzzy integrals (including Choquet integral) group decision-making (GDM) methods to effectively solve decision problems. Wu et al. [35] extended some operational properties of intuitionistic fuzzy values (IFVs) and then studied the aggregation properties of the intuitionistic fuzzy-valued Choquet integral (IFCI) and the intuitionistic fuzzy-valued conjugate Choquet integral (IFCCI). Xing et al. [36] developed the Choquet integral based on q-rung orthopair fuzzy environment and proposed q-rung orthopair fuzzy decision-making methods, Keikha et al. [37] combined the Choquet integral and the TOPSIS method to process fuzzy information, and Teng and Liu [38] developed the generalized Shapley probabilistic linguistic Choquet average (GS-PLCA) operator and investigated a method to solve large group decision-making (LGDM) issues. The in-depth study of the fusion application of the Choquet integral and GDM methods is of practical significance for solving complex and uncertain problems.

Although the aforementioned studies brilliantly handle complex decision-making problems, they cannot be applied to the decision-making problem where attributes are dependent under the interval-valued q-rung orthopair fuzzy environment. Accordingly, this paper develops two interval-valued q-rung orthopair fuzzy Choquet integral operators for aggregating fuzzy information. Subsequently, a MAGDM method is constructed by employing the proposed IVq-ROFCA operator, where interaction attributes of alternatives among the MAGDM problem are taken into account. Finally, compared with existing methods, the practicability and superiority of the proposed method are demonstrated.

The remainder of this paper is constructed as follows: Section 2 reviews the concept of the IVq-ROFS and Choquet integral operator, Section 3 proposes the IVq-ROFCA operator and the IVq-ROFCG operator and extends several weighted and ordered operators, Section 4 introduces a MAGDM method based on the IVq-ROFCA operator, Section 5 gives an illustrated case and then provides parameter analysis and comparison analysis, and Section 6 concludes this paper.

2. Preliminaries

In this section, we make a brief review of the IVq-ROFS and Choquet integral.

Definition 1 (see [11]). Let $X = \{x_1, x_2, \ldots, x_n\}$ be a fixed set, $\widetilde{a} = \{x_i, t_{\widetilde{a}}(x_i), f_{\widetilde{a}}(x_i) | x_i \in X\}$ is a q-ROFS, where $t_{\widetilde{a}}: X \longrightarrow [0,1]$, $f_a: X \longrightarrow [0,1]$ and the following equation holds:

$$0 \le (t_{\widetilde{a}}(x_i))^q + (f_{\widetilde{a}}(x_i))^q \le 1, \tag{1}$$

where $q \ge 1$. For all $x_i \in X$, $t_{\widetilde{a}}(x_i)$ is the degree of membership, $f_{\widetilde{a}}(x_i)$ is the degree of nonmembership, and the degree of indeterminacy $\pi_{\widetilde{a}}(x_i)$ is shown in as follows:

$$\pi_{\widetilde{a}}(x_i) = \sqrt[q]{1 - t_{\widetilde{a}}(x_i)^q - f_{\widetilde{a}}(x_i)^q} \, (q \ge 1). \tag{2}$$

Definition 2 (see [39]). Given a fixed set $X = \{x_1, x_2, \ldots, x_n\}$, IVq-ROFS a on X is defined as

$$a = \{ < x_i, t_a(x_i), f_a(x_i) > | x_i \in X \}. \tag{3}$$

The membership is represented by interval values $t_a(x_i) = [t_a^-(x_i), t_a^+(x_i)] \subseteq [0,1]$ and the nonmembership is $f_a(x_i) = [f_a^-(x_i), f_a^+(x_i)] \subseteq [0,1]$, $0 \le (t_a^+(x_i))^q + (f_a^+(x_i))^q \le 1$, $(q \ge 1)$. The indeterminacy degree of a is shown as follows:

$$\begin{aligned} \pi_a(x_i) &= [\pi_a^-(x_i), \pi_a^+(x_i)] \\ &= \left[\sqrt[q]{1 - (t_a^+(x_i))^q - (f_a^+(x_i))^q}, \sqrt[q]{1 - (t_a^-(x_i))^q - (f_a^-(x_i))^q} \right]. \end{aligned} \tag{4}$$

In particular, IVq-ROFS extends the application of interval-valued fuzzy sets in decision-making problems. When $q = 1$, IVq-ROFS would reduce to the interval-valued

intuitionistic fuzzy set (IVIFS); when $q = 2$, IVq-ROFS would transform to IVPFS.

Definition 3 (see [39]). Let $a_1 = <[t_{a_1}^-, t_{a_1}^+], [f_{a_1}^-, f_{a_1}^+]>$ and $a_2 = <[t_{a_2}^-, t_{a_2}^+], [f_{a_2}^-, f_{a_2}^+]>$ be two interval-valued q-rung orthopair fuzzy numbers (IVq-ROFNs), $q \geq 1$. Equations (5)–(8) hold

$$a_1 \oplus a_2 = <\left[\begin{array}{c}\sqrt[q]{\left(t_{a_1}^-\right)^q + \left(t_{a_2}^-\right)^q - \left(t_{a_1}^-\right)^q \left(t_{a_2}^-\right)^q}, \\ \sqrt[q]{\left(t_{a_1}^+\right)^q + \left(t_{a_2}^+\right)^q - \left(t_{a_1}^+\right)^q \left(t_{a_2}^+\right)^q}\end{array}\right], \left[f_{a_1}^- f_{a_2}^-, f_{a_1}^+ f_{a_2}^+\right]>, \tag{5}$$

$$a_1 \otimes a_2 = <\left[t_{a_1}^- t_{a_2}^-, t_{a_1}^+ t_{a_2}^+\right], \left[\begin{array}{c}\sqrt[q]{\left(f_{a_1}^-\right)^q + \left(f_{a_2}^-\right)^q - \left(f_{a_1}^-\right)^q \left(f_{a_2}^-\right)^q}, \\ \sqrt[q]{\left(f_{a_1}^+\right)^q + \left(f_{a_2}^+\right)^q - \left(f_{a_1}^+\right)^q \left(f_{a_2}^+\right)^q}\end{array}\right]>, \tag{6}$$

$$\lambda a_1 = <\left[\sqrt[q]{1 - \left(1 - \left(t_{a_1}^-\right)^q\right)^\lambda}, \sqrt[q]{1 - \left(1 - \left(t_{a_1}^+\right)^q\right)^\lambda}\right], \left[\left(f_{a_1}^-\right)^\lambda, \left(f_{a_1}^+\right)^\lambda\right]>, \tag{7}$$

$$a_1^\lambda = <\left[\left(t_{a_1}^-\right)^\lambda, \left(t_{a_1}^+\right)^\lambda\right], \left[\sqrt[q]{1 - \left(1 - \left(f_{a_1}^-\right)^q\right)^\lambda}, \sqrt[q]{1 - \left(1 - \left(f_{a_1}^+\right)^q\right)^\lambda}\right]>. \tag{8}$$

Definition 4 (see [39]). For the IVq-ROFN $a = <[t_a^-, t_a^+], [f_a^-, f_a^+]>$, the score function is defined as

$$S(a) = \frac{1}{2}\left[\left(t_a^-\right)^q + \left(t_a^+\right)^q - \left(f_a^-\right)^q - \left(f_a^+\right)^q\right], (q \geq 1). \tag{9}$$

Definition 5 (see [39]). For the IVq-ROFN $a = <[t_a^-, t_a^+], [f_a^-, f_a^+]>$, the accuracy function is defined as

$$H(a) = \frac{1}{2}\left[\left(t_a^-\right)^q + \left(t_a^+\right)^q + \left(f_a^-\right)^q + \left(f_a^+\right)^q\right], (q \geq 1). \tag{10}$$

Definition 6 (see [39]). For two IVq-ROFNs a_1 and a_2, the comparison method is defined as follows:

(1) If $S(a_1) > S(a_2)$, then $a_1 > a_2$

(2) If $S(a_1) < S(a_2)$, then $a_1 < a_2$

(3) If $S(a_1) = S(a_2)$, and if $H(a_1) > H(a_2)$, then $a_1 > a_2$; if $H(a_1) < H(a_2)$, then $a_1 < a_2$; if $H(a_1) = H(a_2)$, then $a_1 = a_2$

Definition 7 (see [26]). Let $X = \{x_1, x_2, \ldots, x_n\}$ be a universe of discourse, f be a positive real-valued function, and μ be the fuzzy measure on X. Then, the discrete Choquet integral of f on fuzzy measure μ is defined as

$$\int f \mathrm{d}\mu = \sum_{i=1}^{n} f\left(x_{\sigma(i)}\right)\left[\mu\left(B_{\sigma(i)}\right) - \mu\left(B_{\sigma(i-1)}\right)\right], \tag{11}$$

where $(\sigma(1), \sigma(2), \ldots, \sigma(n))$ is a permutation of $(1, 2, \ldots, n)$ that satisfies $f(x_{\sigma(1)}) \geq f(x_{\sigma(2)}) \geq \cdots \geq f(x_{\sigma(n)})$, $B_{\sigma(I)} = \left\{x_{\sigma(1)}, x_{\sigma(2)}, \ldots, x_{\sigma(i)}\right\}$ $(i = 1, 2, \ldots, n, B_{\sigma(0)} = \varnothing)$.

3. Several Interval-Valued q-Rung Orthopair Fuzzy Choquet Integral Operators

In this section, we investigate IVq-ROFCA and IVq-ROFCG operators, discuss their properties, and extend their weighted operators.

3.1. IVq-ROFCA

Definition 8. Let μ be the fuzzy measure on the nonempty finite set $X = \{x_1, x_2, \ldots, x_n\}$ $(\mu(\varnothing) = 0)$ and $a(x_i) = <[t_a^-(x_i), t_a^+(x_i)], [f_a^-(x_i), f_a^+(x_i)]>$ $(i = 1, 2, \ldots, n)$ are IVq-ROFNs. The IVq-ROFCA operator is defined as

$$\begin{aligned}(C_1)\int a\mathrm{d}\mu &= IVq - ROFCA\left(a\left(x_1\right), a\left(x_2\right), \ldots, a\left(x_n\right)\right) \\ &= \sum_{i=1}^{n}\left[\mu\left(B_{\sigma(i)}\right) - \mu\left(B_{\sigma(i-1)}\right)\right]a\left(x_{\sigma(i)}\right),\end{aligned} \tag{12}$$

where $(C_1)\int \alpha \mathrm{d}\mu$ indicates the Choquet integral, and $(\sigma(1), \sigma(2), \ldots, \sigma(n))$ denotes a permutation of $(1, 2, \ldots, n)$ that satisfies $a(x_{\sigma(1)}) \geq a(x_{\sigma(2)}) \geq \cdots \geq a(x_{\sigma(n)}), B_{\sigma(I)} = \left\{x_{\sigma(1)}, x_{\sigma(2)}, \ldots, x_{\sigma(i)}\right\}(i = 1, 2, \ldots, n, B_{\sigma(0)} = \varnothing)$. The IVq-

ROFCA operator aggregates information according to the fuzzy measures between attributes, and it is easy to find that the aggregation results obtained are still IVq-ROFNs.

Theorem 1. *For IVq-ROFNs* $a(x_i) = <[t_a^-(x_i), t_a^+(x_i)], [f_a^-(x_i), f_a^+(x_i)]>$ $(i = 1, 2, \ldots, n)$, $\mu(\mu(\varnothing) = 0)$ *is the fuzzy measure on the nonempty finite set* $X = \{x_1, x_2, \ldots, x_n\}$, *and the IVq-ROFCA operator can be expressed as*

$$IVq - ROFCA(a(x_1), a(x_2), \ldots, a(x_n)) = <\left[\begin{array}{c}\sqrt[q]{1 - \prod_{i=1}^{n}\left(1 - t_a^-\left(x_{\sigma(i)}\right)^q\right)^{\mu\left(B_{\sigma(i)}\right) - \mu\left(B_{\sigma(i-1)}\right)}}, \\ \sqrt[q]{1 - \prod_{i=1}^{n}\left(1 - t_a^+\left(x_{\sigma(i)}\right)^q\right)^{\mu\left(B_{\sigma(i)}\right) - \mu\left(B_{\sigma(i-1)}\right)}}\end{array}\right], \left[\begin{array}{c}\prod_{i=1}^{n}\left(f_a^-\left(x_{\sigma(i)}\right)\right)^{\mu\left(B_{\sigma(i)}\right) - \mu\left(B_{\sigma(i-1)}\right)}, \\ \prod_{i=1}^{n}\left(f_a^+\left(x_{\sigma(i)}\right)\right)^{\mu\left(B_{\sigma(i)}\right) - \mu\left(B_{\sigma(i-1)}\right)}\end{array}\right]>. \tag{13}$$

Proof. Based on Definition 3, we prove equation (13) by mathematical induction.

If $n = 2$,

$$IVq - ROFCA(a(x_1), a(x_2)) = \prod_{i=1}^{2}\left[\mu\left(B_{\sigma(i)}\right) - \mu\left(B_{\sigma(i-1)}\right)\right]a\left(x_{\sigma(i)}\right) = <\left[\begin{array}{c}\sqrt[q]{1 - \prod_{i=1}^{2}\left(1 - t_a^-\left(x_{\sigma(i)}\right)^q\right)^{\mu\left(B_{\sigma(i)}\right) - \mu\left(B_{\sigma(i-1)}\right)}}, \\ \sqrt[q]{1 - \prod_{i=1}^{2}\left(1 - t_a^+\left(x_{\sigma(i)}\right)^q\right)^{\mu\left(B_{\sigma(i)}\right) - \mu\left(B_{\sigma(i-1)}\right)}}\end{array}\right], \left[\begin{array}{c}\prod_{i=1}^{2}\left(f_a^-\left(x_{\sigma(i)}\right)\right)^{\mu\left(B_{\sigma(i)}\right) - \mu\left(B_{\sigma(i-1)}\right)}, \\ \prod_{i=1}^{2}\left(f_a^+\left(x_{\sigma(i)}\right)\right)^{\mu\left(B_{\sigma(i)}\right) - \mu\left(B_{\sigma(i-1)}\right)}\end{array}\right]>. \tag{14}$$

If $n = k$,

$$IVq - ROFCA(a(x_1), a(x_2), \ldots, a(x_k)) = <\left[\begin{array}{c}\sqrt[q]{1 - \prod_{i=1}^{k}\left(1 - t_a^-\left(x_{\sigma(i)}\right)^q\right)^{\mu\left(B_{\sigma(i)}\right) - \mu\left(B_{\sigma(i-1)}\right)}}, \\ \sqrt[q]{1 - \prod_{i=1}^{k}\left(1 - t_a^+\left(x_{\sigma(i)}\right)^q\right)^{\mu\left(B_{\sigma(i)}\right) - \mu\left(B_{\sigma(i-1)}\right)}}\end{array}\right], \left[\begin{array}{c}\prod_{i=1}^{k}\left(f_a^-\left(x_{\sigma(i)}\right)\right)^{\mu\left(B_{\sigma(i)}\right) - \mu\left(B_{\sigma(i-1)}\right)}, \\ \prod_{i=1}^{k}\left(f_a^+\left(x_{\sigma(i)}\right)\right)^{\mu\left(B_{\sigma(i)}\right) - \mu\left(B_{\sigma(i-1)}\right)}\end{array}\right]>. \tag{15}$$

If $n = k + 1$, the results of IVq-ROFCA are as follows:

$$
\begin{aligned}
&IVq-ROFCA\left(a\left(x_1\right),a\left(x_2\right),\ldots,a\left(x_{k+1}\right)\right)=\prod_{i=1}^{k+1}\left[\mu\left(B_{\sigma(i)}\right)-\mu\left(B_{\sigma(i-1)}\right)\right]a\left(x_{\sigma(i)}\right)\\
&=\prod_{i=1}^{k}\left[\mu\left(B_{\sigma(i)}\right)-\mu\left(B_{\sigma(i-1)}\right)\right]a\left(x_{\sigma(i)}\right)\oplus\left[\mu\left(B_{\sigma(k+1)}\right)-\mu\left(B_{\sigma(k)}\right)\right]a\left(x_{\sigma(k+1)}\right)\\
&=<\begin{bmatrix}\sqrt[q]{1-\prod_{i=1}^{k}\left(1-t_a^-\left(x_{\sigma(i)}\right)^q\right)^{\mu\left(B_{\sigma(i)}\right)-\mu\left(B_{\sigma(i-1)}\right)}},\\ \sqrt[q]{1-\prod_{i=1}^{k}\left(1-t_a^+\left(x_{\sigma(i)}\right)^q\right)^{\mu\left(B_{\sigma(i)}\right)-\mu\left(B_{\sigma(i-1)}\right)}}\end{bmatrix},\begin{bmatrix}\prod_{i=1}^{k}\left(f_a^-\left(x_{\sigma(i)}\right)\right)^{\mu\left(B_{\sigma(i)}\right)-\mu\left(B_{\sigma(i-1)}\right)},\\ \prod_{i=1}^{k}\left(f_a^+\left(x_{\sigma(i)}\right)\right)^{\mu\left(B_{\sigma(i)}\right)-\mu\left(B_{\sigma(i-1)}\right)}\end{bmatrix}>\\
&\oplus<\begin{bmatrix}\sqrt[q]{1-\left(1-t_a^-\left(x_{\sigma(k+1)}\right)^q\right)^{\mu\left(B_{\sigma(k+1)}\right)-\mu\left(B_{\sigma(k)}\right)}},\\ \sqrt[q]{1-\left(1-t_a^+\left(x_{\sigma(k+1)}\right)^q\right)^{\mu\left(B_{\sigma(k+1)}\right)-\mu\left(B_{\sigma(k)}\right)}}\end{bmatrix},\begin{bmatrix}\left(f_a^-\left(x_{\sigma(k+1)}\right)\right)^{\mu\left(B_{\sigma(k+1)}\right)-\mu\left(B_{\sigma(k)}\right)},\\ \left(f_a^+\left(x_{\sigma(k+1)}\right)\right)^{\mu\left(B_{\sigma(k+1)}\right)-\mu\left(B_{\sigma(k)}\right)}\end{bmatrix}>\\
&=<\begin{bmatrix}\sqrt[q]{1-\prod_{i=1}^{k+1}\left(1-t_a^-\left(x_{\sigma(i)}\right)^q\right)^{\mu\left(B_{\sigma(i)}\right)-\mu\left(B_{\sigma(i-1)}\right)}},\\ \sqrt[q]{1-\prod_{i=1}^{k+1}\left(1-t_a^+\left(x_{\sigma(i)}\right)^q\right)^{\mu\left(B_{\sigma(i)}\right)-\mu\left(B_{\sigma(i-1)}\right)}}\end{bmatrix},\begin{bmatrix}\prod_{i=1}^{k+1}\left(f_a^-\left(x_{\sigma(i)}\right)\right)^{\mu\left(B_{\sigma(i)}\right)-\mu\left(B_{\sigma(i-1)}\right)},\\ \prod_{i=1}^{k+1}\left(f_a^+\left(x_{\sigma(i)}\right)\right)^{\mu\left(B_{\sigma(i)}\right)-\mu\left(B_{\sigma(i-1)}\right)}\end{bmatrix}>.
\end{aligned}
\tag{16}
$$

Equation (13) holds, and Theorem 1 holds. □

Example 1. Suppose there are three suppliers $\{x_1, x_2, x_3\}$ to be chosen. The core competitiveness of suppliers can be evaluated by three criteria $\{C_1, C_2, C_3\}$: C_1 denotes the level of technological innovation, C_2 denotes the ability of circulation control, and C_3 denotes the capability of management. The decision matrix $A = (a_{ij})_{3\times3}$ is generated by evaluation of experts, in which the IVq-ROFN evaluation values $a_{ij} = <[t_{a_{ij}}^-, t_{a_{ij}}^+], [f_{a_{ij}}^-, f_{a_{ij}}^+]>$ $(i, j = 1, 2, 3)$ are represented in Table 1. Now, it needs to evaluate the core competitiveness of $\{x_1, x_2, x_3\}$ according to the decision matrix $A = (a_{ij})_{3\times3}$, which provides the reference for the manufacturing enterprise to select the best solution. The fuzzy measure μ between attributes are set as follows:

$$
\begin{aligned}
\mu(\varnothing) &= 0,\\
\mu(\{C_1, C_2, C_3\}) &= 1,\\
\mu(\{C_1\}) &= \mu(\{C_2\})\\
&= 0.4,\\
\mu(\{C_3\}) &= 0.3,\\
\mu(\{C_1, C_2\}) &= 0.5,\\
\mu(\{C_1, C_3\}) &= \mu(\{C_2, C_3\})\\
&= 0.8.
\end{aligned}
\tag{17}
$$

Let $q = 3$, all the IVq-ROFNs in A satisfy $t_{a_{ij}}^+ + f_{a_{ij}}^+ \le 1$, and the results are obtained as follows:

$$
\begin{aligned}
r_1 &= IVq-ROFCA\left(x_1\left(C_1\right), x_1\left(C_2\right), x_1\left(C_3\right)\right)\\
&= <[0.57, 0.77], [0.39, 0.54]>,\\
r_2 &= IVq-ROFCA\left(x_2\left(C_1\right), x_2\left(C_2\right), x_2\left(C_3\right)\right)\\
&= <[0.49, 0.70], [0.41, 0.57]>,\\
r_3 &= IVq-ROFCA\left(x_3\left(C_1\right), x_3\left(C_2\right), x_3\left(C_3\right)\right)\\
&= <[0.52, 0.72], [0.34, 0.46]>.
\end{aligned}
\tag{18}
$$

It is easy to find r_1, r_2, and r_3 are IVq-ROFNs. According to equation (9), $S(r_1) > S(r_3) > S(r_2)$ can be derived, which indicates that the supplier x_1 is better than x_2 and x_3.

Theorem 2. *Suppose* $a_i = <[t_{a_i}^-, t_{a_i}^+], [f_{a_i}^-, f_{a_i}^+]>$ $(i = 1, 2, \ldots, n)$ *are IVq-ROFNs, μ is the fuzzy measure on a nonempty finite set* $X = \{x_1, x_2, \ldots, x_n\}$, *and then, the following four properties of the IVq-ROFCA operator hold*

(1) *Idempotency: for an IVq-ROFN* $a = <[t_a^-, t_a^+], [f_a^-, f_a^+]>$, *if* $a_i = a$ $(i = 1, 2, \ldots, n)$, *then* $IVq-ROFCA(a_1, a_2, \ldots, a_n) = a$

(2) *Boundedness: if* $a_{min} = <[min(t_{a_i}^-), min(t_{a_i}^+)], [max(f_{a_i}^-), max(f_{a_i}^+)]>$ *and* $a_{max} = <[max(t_{a_i}^-), max(t_{a_i}^+)], [min(f_{a_i}^-), min(f_{a_i}^+)]>$, *then* $a_{\min} \le IVq-ROFCA(a_1, a_2, \ldots, a_n) \le a_{max}$

(3) *Commutativity: suppose* $(a_1', a_2', \ldots, a_n')$ *is a permutation of* $(a_1, a_2, \ldots, a_n)$, *then* $IVq-ROFCA(a_1, a_2, \ldots, a_n) = IVq-ROFCA(a_1', a_2', \ldots, a_n')$

(4) *Monotonicity: if* $\beta_i = <[t_{\beta_i}^-, t_{\beta_i}^+], [f_{\beta_i}^-, f_{\beta_i}^+]>$ $(i = 1, 2, \ldots, n)$ *are IVq-ROFNs and* $t_{a_i}^- \le t_{\beta_i}^-, t_{a_i}^+ \le t_{\beta_i}^+, f_{a_i}^- \ge$

Table 1: Decision matrix $A = (a_{ij})_{3\times 3}$.

	$\mathbf{C}_1$	$\mathbf{C}_2$	$\mathbf{C}_3$
$\mathbf{x}_1$	<[0.7, 0.9], [0.3, 0.5]>	<[0.3, 0.4], [0.5, 0.6]>	<[0.5, 0.6], [0.4, 0.5]>
$\mathbf{x}_2$	<[0.6, 0.8], [0.4, 0.5]>	<[0.3, 0.5], [0.5, 0.7]>	<[0.4, 0.7], [0.3, 0.5]>
$\mathbf{x}_3$	<[0.6, 0.8], [0.3, 0.4]>	<[0.4, 0.6], [0.4, 0.5]>	<[0.5, 0.7], [0.3, 0.5]>

$f_{\beta_i}^{-}, f_{a_i}^{+} \geq f_{\beta_i}^{+}$, *then* $IVq - ROFCA(a_1, a_2, \ldots, a_n) \leq IVq - ROFCA(\beta_1, \beta_2, \ldots, \beta_n)$

The proof of Theorem 2 is given as follows.

Proof. (1) Because all elements in the a_i are equal and $a_i = a$,

$$\mu(x_i) = \frac{1}{n} \text{ and } \mu\left(B_{\sigma(i)}\right) - \mu\left(B_{\sigma(i-1)}\right) = \frac{1}{n},$$

$$IVq - ROFCA(a_1, a_2, \ldots, a_n) = < \begin{bmatrix} \sqrt[q]{1 - \prod_{i=1}^{n} \left(1 - t_{a_{\sigma(i)}}^{-q}\right)^{\mu\left(B_{\sigma(i)}\right) - \mu\left(B_{\sigma(i-1)}\right)}}, \\ \sqrt[q]{1 - \prod_{i=1}^{n} \left(1 - t_{a_{\sigma(i)}}^{+q}\right)^{\mu\left(B_{\sigma(i)}\right) - \mu\left(B_{\sigma(i-1)}\right)}} \end{bmatrix}, \begin{bmatrix} \prod_{i=1}^{n} \left(f_{a_{\sigma(i)}}^{-}\right)^{\mu\left(B_{\sigma(i)}\right) - \mu\left(B_{\sigma(i-1)}\right)}, \\ \prod_{i=1}^{n} \left(f_{a_{\sigma(i)}}^{+}\right)^{\mu\left(B_{\sigma(i)}\right) - \mu\left(B_{\sigma(i-1)}\right)} \end{bmatrix} >$$

$$= < \begin{bmatrix} \sqrt[q]{1 - \prod_{i=1}^{n} \left(1 - t_a^{-q}\right)^{(1/n)}}, \\ \sqrt[q]{1 - \prod_{i=1}^{n} \left(1 - t_a^{+q}\right)^{(1/n)}} \end{bmatrix}, \begin{bmatrix} \prod_{i=1}^{n} \left(f_a^{-}\right)^{(1/n)}, \\ \prod_{i=1}^{n} \left(f_a^{+}\right)^{(1/n)} \end{bmatrix} > \tag{19}$$

$$= < \begin{bmatrix} \sqrt[q]{1 - \left(1 - t_a^{-q}\right)}, \\ \sqrt[q]{1 - \left(1 - t_a^{+q}\right)} \end{bmatrix}, \begin{bmatrix} f_a^{-}, \\ f_a^{+} \end{bmatrix} >$$

$$= < \begin{bmatrix} \sqrt[q]{t_a^{-q}}, \\ \sqrt[q]{t_a^{+q}} \end{bmatrix}, \begin{bmatrix} f_a^{-}, \\ f_a^{+} \end{bmatrix} >$$

$$= a.$$

(2) According to equation (9), it can be seen that

$$S(a_{\min}) = \frac{1}{2}\left[\left(\min\left(t_{a_i}^{-}\right)\right)^q + \left(\min\left(t_{a_i}^{+}\right)\right)^q - \left(\max\left(f_{a_i}^{-}\right)\right)^q - \left(\max\left(f_{a_i}^{+}\right)\right)^q\right],$$
$$S(a_{\max}) = \frac{1}{2}\left[\left(\max\left(t_{a_i}^{-}\right)\right)^q + \left(\max\left(t_{a_i}^{+}\right)\right)^q - \left(\min\left(f_{a_i}^{-}\right)\right)^q - \left(\min\left(f_{a_i}^{+}\right)\right)^q\right]. \tag{20}$$

For any $t_{a_i}^{-}$, it satisfies

$$\begin{aligned}
&\left(\min\left(t_{a_i}^{-}\right)\right)^q \le \left(t_{a_i}^{-}\right)^q \\
&\le \left(\max\left(t_{a_i}^{-}\right)\right)^q, 1-\left(\min\left(t_{a_i}^{-}\right)\right)^q \\
&\ge 1-\left(t_{a_i}^{-}\right)^q \\
&\ge 1-\left(\max\left(t_{a_i}^{-}\right)\right)^q; \\
&\prod_{i=1}^{n}\left(1-\left(\min\left(t_{a_i}^{-}\right)\right)^q\right)^{\mu\left(B_{\sigma(i)}\right)-\mu\left(B_{\sigma(i-1)}\right)} \\
&\ge \prod_{i=1}^{n}\left(1-\left(t_{a_i}^{-}\right)^q\right)^{\mu\left(B_{\sigma(i)}\right)-\mu\left(B_{\sigma(i-1)}\right)} \\
&\ge \prod_{i=1}^{n}\left(1-\left(\max\left(t_{a_i}^{-}\right)\right)^q\right)^{\mu\left(B_{\sigma(i)}\right)-\mu\left(B_{\sigma(i-1)}\right)} \\
&1-\prod_{i=1}^{n}\left(1-\left(\min\left(t_{a_i}^{-}\right)\right)^q\right)^{\mu\left(B_{\sigma(i)}\right)-\mu\left(B_{\sigma(i-1)}\right)} \\
&\le 1-\prod_{i=1}^{n}\left(1-\left(t_{a_i}^{-}\right)\right)^{\mu\left(B_{\sigma(i)}\right)-\mu\left(B_{\sigma(i-1)}\right)} \\
&\le 1-\prod_{i=1}^{n}\left(1-\left(\max\left(t_{a_i}^{-}\right)\right)^q\right)^{\mu\left(B_{\sigma(i)}\right)-\mu\left(B_{\sigma(i-1)}\right)}.
\end{aligned} \tag{21}$$

Namely,

$$\left(\min\left(t_{a_i}^{-}\right)\right)^q \le 1-\prod_{i=1}^{n}\left(1-\left(t_{a_i}^{-}\right)^q\right)^{\mu\left(B_{\sigma(i)}\right)-\mu\left(B_{\sigma(i-1)}\right)} \le \left(\max\left(t_{a_i}^{-}\right)\right)^q. \tag{22}$$

Similarly,

$$\begin{aligned}
&\left(\min\left(t_{a_i}^{+}\right)^q\right) \le -\prod_{i=1}^{n}\left(1-\left(t_{a_i}^{+}\right)^q\right)^{\mu\left(B_{\sigma(i)}\right)-\mu\left(B_{\sigma(i-1)}\right)} \le \left(\max\left(t_{a_i}^{+}\right)\right)^q, \\
&\left(\min\left(f_{a_i}^{-}\right)\right)^q \le \prod_{i=1}^{n}\left(\left(f_{a_i}^{-}\right)^q\right)^{\mu\left(B_{\sigma(i)}\right)-\mu\left(B_{\sigma(i-1)}\right)} \le \left(\max\left(f_{a_i}^{-}\right)\right)^q, \\
&\left(\min\left(f_{a_i}^{+}\right)\right)^q \le \prod_{i=1}^{n}\left(\left(f_{a_i}^{-}\right)^q\right)^{\mu\left(B_{\sigma(i)}\right)-\mu\left(B_{\sigma(i-1)}\right)} \le \left(\max\left(f_{a_i}^{+}\right)\right)^q,
\end{aligned} \tag{23}$$

sorted out

$$\begin{aligned}
S\left(a_{\min}\right) &\le S\left(IVq-ROFCA\left(a_1,a_2,\ldots,a_n\right)\right) \\
&\le S\left(a_{\max}\right), \\
a_{\min} &\le IVq-ROFCA\left(a_1,a_2,\ldots,a_n\right) \le a_{\max}.
\end{aligned} \tag{24}$$

(3) According to Definition 8, it can be easily derived that

$$\begin{aligned}
&IVq-ROFCA\left(a_1,a_2,\ldots,a_n\right) = \\
&< \left[\sqrt[q]{1-\prod_{i=1}^{n}\left(1-t_{a_{\sigma(i)}}^{-q}\right)^{\mu\left(B_{\sigma(i)}\right)-\mu\left(B_{\sigma(i-1)}\right)}}, \sqrt[q]{1-\prod_{i=1}^{n}\left(1-t_{a_{\sigma(i)}}^{+q}\right)^{\mu\left(B_{\sigma(i)}\right)-\mu\left(B_{\sigma(i-1)}\right)}}\right], \left[\prod_{i=1}^{n}\left(f_{a_{\sigma(i)}}^{-}\right)^{\mu\left(B_{\sigma(i)}\right)-\mu\left(B_{\sigma(i-1)}\right)}, \prod_{i=1}^{n}\left(f_{a_{\sigma(i)}}^{+}\right)^{\mu\left(B_{\sigma(i)}\right)-\mu\left(B_{\sigma(i-1)}\right)}\right] > \\
&= IVq-ROFCA\left(a_1',a_2',\ldots,a_n'\right)\square.
\end{aligned} \tag{25}$$

(4) Based on $t_{a_i}^{-} \le t_{\beta_i}^{-}$ and Definition 8, it can be found that

$$\begin{aligned}
t_{a_{\sigma(i)}}^{-q} &\le t_{\beta_{\sigma(i)}}^{-q}, \\
1-t_{a_{\sigma(i)}}^{-q} &\ge 1-t_{\beta_{\sigma(i)}}^{-q}.
\end{aligned} \tag{26}$$

Thus,

$$\begin{aligned}
&\prod_{i=1}^{n}\left(1-t_{a_{\sigma(i)}}^{-q}\right)^{\mu\left(B_{\sigma(i)}\right)-\mu\left(B_{\sigma(i-1)}\right)} \ge \prod_{i=1}^{n}\left(1-t_{\beta_{\sigma(i)}}^{-q}\right)^{\mu\left(B_{\sigma(i)}\right)-\mu\left(B_{\sigma(i-1)}\right)}, \\
&\sqrt[q]{1-\prod_{i=1}^{n}\left(1-t_{a_{\sigma(i)}}^{-q}\right)^{\mu\left(B_{\sigma(i)}\right)-\mu\left(B_{\sigma(i-1)}\right)}} \le \sqrt[q]{1-\prod_{i=1}^{n}\left(1-t_{\beta_{\sigma(i)}}^{-q}\right)^{\mu\left(B_{\sigma(i)}\right)-\mu\left(B_{\sigma(i-1)}\right)}}.
\end{aligned} \tag{27}$$

Similarly,

$$\sqrt[q]{1-\prod_{i=1}^{n}\left(1-t_{a_{\sigma(i)}}^{+q}\right)^{\mu\left(B_{\sigma(i)}\right)-\mu\left(B_{\sigma(i-1)}\right)}} \le \sqrt[q]{1-\prod_{i=1}^{n}\left(1-t_{\beta_{\sigma(i)}}^{+q}\right)^{\mu\left(B_{\sigma(i)}\right)-\mu\left(B_{\sigma(i-1)}\right)}},$$
$$\prod_{i=1}^{n}\left(f_{a_{\sigma(i)}}^{-}\right)^{\mu\left(B_{\sigma(i)}\right)-\mu\left(B_{\sigma(i-1)}\right)} \ge \prod_{i=1}^{n}\left(f_{\beta_{\sigma(i)}}^{-}\right)^{\mu\left(B_{\sigma(i)}\right)-\mu\left(B_{\sigma(i-1)}\right)}, \tag{28}$$
$$\prod_{i=1}^{n}\left(f_{a_{\sigma(i)}}^{+}\right)^{\mu\left(B_{\sigma(i)}\right)-\mu\left(B_{\sigma(i-1)}\right)} \ge \prod_{i=1}^{n}\left(f_{\beta_{\sigma(i)}}^{+}\right)^{\mu\left(B_{\sigma(i)}\right)-\mu\left(B_{\sigma(i-1)}\right)}.$$

According to Definition 6, it is obviously discovered that $IVq-ROFCA(a_1,a_2,\ldots,a_n) \le IVq-ROFCA(\beta_1,\beta_2,\ldots,\beta_n)$ □.

For some special relationships that exist between evaluation values in decision-making, we then develop the following operators to facilitate calculation. □

Definition 9. Let μ be the fuzzy measure on the nonempty finite set $X=\{x_1,x_2,\ldots,x_n\}$ $(\mu(\varnothing)=0)$ and $a(x_i)=<[t_a^-(x_i),t_a^+(x_i)],[f_a^-(x_i),f_a^+(x_i)]>$ $(i=1,2,\ldots,n)$ be IVq-ROFNs.

(1) If $\mu(B\cup C)=\mu(B)+\mu(C)$ for all $B,C\subseteq X$, $B\cap C=\varnothing$, and it is independent for any elements in X that means $\mu(B)=\sum_{x_i\in B}\mu(\{x_i\})$. The IVq-ROFCA operator transforms to the interval-valued q-rung orthopair fuzzy weighted Choquet average (IVq-ROFWCA) operator, which is expressed as

$$IVq-ROFWCA(a(x_1),a(x_2),\ldots,a(x_n)) = <\left[\begin{array}{l}\sqrt[q]{1-\prod_{i=1}^{n}\left(1-t_a^-\left(x_{\sigma(i)}\right)^q\right)^{\mu(\{x_i\})}},\\ \sqrt[q]{1-\prod_{i=1}^{n}\left(1-t_a^+\left(x_{\sigma(i)}\right)^q\right)^{\mu(\{x_i\})}}\end{array}\right],\left[\begin{array}{l}\prod_{i=1}^{n}\left(f_a^-\left(x_{\sigma(i)}\right)\right)^{\mu(\{x_i\})},\\ \prod_{i=1}^{n}\left(f_a^+\left(x_{\sigma(i)}\right)\right)^{\mu(\{x_i\})}\end{array}\right]>. \tag{29}$$

(2) If $\mu(B)=\sum_{i=1}^{|B|}\lambda_i$ for all $B\subseteq X$, we have $\lambda_i=\mu(B_{\sigma(i)})-\mu(B_{\sigma(i-1)})$ $(i=1,2,\ldots,n)$, where $\lambda=(\lambda_1,\lambda_2,\ldots,\lambda_n)^T$ and $\sum_{i=1}^{n}\lambda_i=1$ $(\lambda_i\ge 0)$. Furthermore, the IVq-ROFCA operator reduces to the interval-valued q-rung orthopair fuzzy order Choquet average (IVq-ROFOCA) operator that is represented as

$$IVq-ROFOCA(a(x_1),a(x_2),\ldots,a(x_n)) = <\left[\begin{array}{l}\sqrt[q]{1-\prod_{i=1}^{n}\left(1-t_a^-\left(x_{\sigma(i)}\right)^q\right)^{\lambda_i}},\\ \sqrt[q]{1-\prod_{i=1}^{n}\left(1-t_a^+\left(x_{\sigma(i)}\right)^q\right)^{\lambda_i}}\end{array}\right],\left[\begin{array}{l}\prod_{i=1}^{n}\left(f_a^-\left(x_{\sigma(i)}\right)\right)^{\lambda_i},\\ \prod_{i=1}^{n}\left(f_a^+\left(x_{\sigma(i)}\right)\right)^{\lambda_i}\end{array}\right]>. \tag{30}$$

(3) If $\mu(B)=Q(\sum_{x_i\in B}(\mu\{x_i\}))$ for all $B\subseteq X$, where $Q$is a basic unit-interval monotonic function, satisfies monotonicity in [0,1] and follows properties: $(i)Q(0)=0$; $(ii)Q(1)=1$; $(iii)$for $x>y$, $Q(x)\ge Q(y)$. For $w_i=\mu(B_{\sigma}(i))-\mu(B_{\sigma(i-1)})=Q(\sum_{j\le i}\mu(\{x_{\sigma(i)}\}))-Q(\sum_{j<i}\mu(\{x_{\sigma(i)}\}))$ $(i=1,2,\ldots,n)$, where $w=(w_1,w_2,\ldots,w_n)^T$ and $\sum_{i=1}^{n}w_i=1$ $(w_i\ge 0)$, the IVq-ROFCA operator changes to the interval-valued q-rung orthopair fuzzy order weighted Choquet average (IVq-ROFOWCA) operator that is proposed as

$$IVq-ROFOWCA(a(x_1),a(x_2),\ldots,a(x_n) = <\left[\begin{array}{l}\sqrt[q]{1-\prod_{i=1}^{n}\left(1-t_a^-\left(x_{\sigma(i)}\right)^q\right)^{w_i}},\\ \sqrt[q]{1-\prod_{i=1}^{n}\left(1-t_a^+\left(x_{\sigma(i)}\right)^q\right)^{w_i}}\end{array}\right], \left[\prod_{i=1}^{n}\left(f_a^-\left(x_{\sigma(i)}\right)\right)^{w_i},\prod_{i=1}^{n}\left(f_a^+\left(x_{\sigma(i)}\right)\right)^{w_i}\right]>. \tag{31}$$

Especially, if $\mu(\{x_i\})=(1/n)(i=1,2,\ldots,n)$, then the IVq-ROFOWCA operator reduces to the IVq-ROFOCA operator. In addition, we develop the IVq-ROFCG operator in Section 3.2.

3.2. IVq-ROFCG

Definition 10. Let μ be the fuzzy measure on the nonempty finite set $X=\{x_1,x_2,\ldots,x_n\}$ $(\mu(\varnothing)=0)$ and $a(x_i)=<[t_a^-(x_i),t_a^+(x_i)],[f_a^-(x_i),f_a^+(x_i)]>$ $(i=1,2,\ldots,n)$ be IVq-ROFNs. Then, the IVq-ROFCG operator is defined as

$$(C_2)\int a\,d\mu = IVq-ROFCG(a(x_1),a(x_2),\ldots,a(x_n)) = \prod_{i=1}^{n}\left[\mu\left(B_{\sigma(i)}\right)-\mu\left(B_{\sigma(i-1)}\right)\right]a\left(x_{\sigma(i)}\right), \tag{32}$$

where $(C_2)\int a\,d\mu$ indicates the Choquet integral, $(\sigma(1),\sigma(2),\ldots,\sigma(n))$ is a permutation of $(1,2,\ldots n)$, which satisfies $a(x_{\sigma(1)})\ge a(x_{\sigma(2)})\ge\cdots\ge a(x_{\sigma(n)})$, $B_{\sigma(i)}=\{x_{\sigma(1)},x_{\sigma(2)},\ldots,x_{\sigma(i)}\}$ $(i=1,2,\ldots,n, B_{\sigma(0)}=\varnothing)$.

Theorem 3. *If μ is the fuzzy measure on the nonempty finite set $X = \{x_1, x_2, \ldots, x_n\}$ and $a(x_i) = <[t_a^-(x_i), t_a^+(x_i)], [f_a^-(x_i), f_a^+(x_i)]> (i = 1, 2, \ldots n)$ are IVq-ROFNs, $\mu(\varnothing) = 0$. Then, the IVq-ROFCG operator is represented by*

$$IVq - ROFCG(a(x_1), a(x_2), \ldots, a(x_n)) = < \left[\prod_{i=1}^{n} \left(t_a^-\left(x_{\sigma(i)}\right)\right)^{\mu\left(B_{\sigma(i)}\right) - \mu\left(B_{\sigma(i-1)}\right)}, \prod_{i=1}^{n} \left(t_a^+\left(x_{\sigma(i)}\right)\right)^{\mu\left(B_{\sigma(i)}\right) - \mu\left(B_{\sigma(i-1)}\right)} \right], \left[\sqrt[q]{1 - \prod_{i=1}^{n} \left(1 - f_a^-\left(x_{\sigma(i)}\right)^q\right)^{\mu\left(B_{\sigma(i)}\right) - \mu\left(B_{\sigma(i-1)}\right)}}, \sqrt[q]{1 - \prod_{i=1}^{n} \left(1 - f_a^+\left(x_{\sigma(i)}\right)^q\right)^{\mu\left(B_{\sigma(i)}\right) - \mu\left(B_{\sigma(i-1)}\right)}} \right] > . \tag{33}$$

The proof of Theorem 3 is similar to Theorem 1, and the proof is omitted.

Theorem 4. *If $a_i = <[t_{a_i}^-, t_{a_i}^+], [f_{a_i}^-, f_{a_i}^+]> (i = 1, 2, \ldots, n)$ is a set of IVq-ROFNs, μ is the fuzzy measure on the nonempty finite set $X = \{x_1, x_2, \ldots, x_n\}$, then the following properties of the IVq-ROFC G operator hold:*

(1) Idempotency: for an IVq-ROFN $a = <[t_a^-, t_a^+], [f_a^-, f_a^+]>$, if $a_i = a (i = 1, 2, \ldots, n)$, then $IVq - ROFCG(a_1, a_2, \ldots, a_n) = a$

(2) Boundedness: if $a_{min} = <[min(t_{a_i}^-), min(t_{a_i}^+)], [max(f_{a_i}^-), max(f_{a_i}^+)]>$ and $a_{max} = <[max(t_{a_i}^-), max(t_{a_i}^+)], [min(f_{a_i}^-), min(f_{a_i}^+)]>$, then $a_{\min} \le IVq - ROFCG(a_1, a_2, \ldots, a_n) \le a_{max}$

(3) Commutativity: suppose that $(a_1', a_2', \ldots, a_n')$ is a permutation of $(a_1, a_2, \ldots, a_n)$, then $IVq - ROFCG(a_1, a_2, \ldots, a_n) = IVq - ROFCG(a_1', a_2', \ldots, a_n')$

(4) Monotonicity: if $\beta_i = <[t_{\beta_i}^-, t_{\beta_i}^+], [f_{\beta_i}^-, f_{\beta_i}^+]> (i = 1, 2, \ldots, n)$ is a set of IVq-ROFNs and $t_{a_i}^- \le t_{\beta_i}^-, t_{a_i}^+ \le t_{\beta_i}^+, f_{a_i}^- \ge f_{\beta_i}^-, f_{a_i}^+ \ge f_{\beta_i}^+$, then $IVq - ROFCG(a_1, a_2, \ldots, a_n) \le IVq - ROFCG(\beta_1, \beta_2, \ldots, \beta_n)$

The proof of Theorem 4 is similar to Theorem 2, and the proof is omitted.

Definition 11. Let μ be the fuzzy measure on the nonempty finite set $X = \{x_1, x_2, \ldots, x_n\}$ $(\mu(\varnothing) = 0)$ and $a(x_i) = <[t_a^-(x_i), t_a^+(x_i)], [f_a^-(x_i), f_a^+(x_i)]> (i = 1, 2, \ldots, n)$ be n IVq-ROFNs. Then,

(1) If $\mu(B \cup C) = \mu(B) + \mu(C)$ for all $B, C \subseteq X, B \cap C = \varnothing$, and it is independent for any elements in X that means $\mu(B) = \sum_{x_i \in B} \mu(\{x_i\})$. Then, the IVq-ROFCG operator transforms to the interval-valued q-rung orthopair fuzzy weighted Choquet geometric (IVq-ROFWCG) operator that is represented by

$$IVq - ROFWCG(a(x_1), a(x_2), \ldots, a(x_n)) = < \left[\prod_{i=1}^{n} \left(t_a^-\left(x_{\sigma(i)}\right)\right)^{\mu(\{x_i\})}, \prod_{i=1}^{n} \left(t_a^+\left(x_{\sigma(i)}\right)\right)^{\mu(\{x_i\})} \right], \left[\sqrt[q]{1 - \prod_{i=1}^{n} \left(1 - f_a^-\left(x_{\sigma(i)}\right)^q\right)^{\mu(\{x_i\})}}, \sqrt[q]{1 - \prod_{i=1}^{n} \left(1 - f_a^+\left(x_{\sigma(i)}\right)^q\right)^{\mu(\{x_i\})}} \right] > . \tag{34}$$

(2) If $\mu(B) = \sum_{i=1}^{|B|} \lambda_i$ for all $B \subseteq X$, we have $\lambda_i = \mu(B_{\sigma(i)}) - \mu(B_{\sigma(i-1)}) (i = 1, 2, \ldots, n)$, where $\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_n)^T$ and $\sum_{i=1}^{n} \lambda_i = 1 (\lambda_i \ge 0)$, the IVq-ROFCG operator reduces to the interval-valued q-rung orthopair fuzzy order Choquet geometric (IVq-ROFOCG) operator expressed by

$$IVq - ROFOCG(a(x_1), a(x_2), \ldots, a(x_n)) = < \left[\prod_{i=1}^{n} \left(t_a^-\left(x_{\sigma(i)}\right)\right)^{\lambda_i}, \prod_{i=1}^{n} \left(t_a^+\left(x_{\sigma(i)}\right)\right)^{\lambda_i} \right], \left[\sqrt[q]{1 - \prod_{i=1}^{n} \left(1 - f_a^-\left(x_{\sigma(i)}\right)^q\right)^{\lambda_i}}, \sqrt[q]{1 - \prod_{i=1}^{n} \left(1 - f_a^+\left(x_{\sigma(i)}\right)^q\right)^{\lambda_i}} \right] > . \tag{35}$$

(3) If $\mu(B) = Q(\sum_{x_i \in B} \mu(\{x_i\}))$ for all $B \subseteq X$, where Q is a basic unit-interval monotonic function, satisfies monotonicity in [0,1] and follows properties: $(i)Q(0) = 0$; $(ii)Q(1) = 1$; $(iii)Q(x) \geq Q(y)$ for $x > y$. Then, we let $w_i = \mu(B_{\sigma(i)}) - \mu(B_{\sigma(i-1)}) = Q(\sum_{j \leq i} \mu(\{x_{\sigma(i)}\})) - Q(\sum_{j<i} \mu(\{x_{\sigma(i)}\}))$ $(i = 1, 2, \ldots, n)$, where $w = (w_1, w_2, \ldots, w_n)^T$ and $\sum_{i=1}^{n} w_i = 1 (w_i \geq 0)$. Then, the IVq-ROFCG operator changes to the interval-valued q-rung orthopair fuzzy order weighted Choquet geometric (IVq-ROFOWCG) operator that is shown as

$$IVq - ROFOWCG(a(x_1), a(x_2), \ldots, a(x_n)) = \left\langle \left[\prod_{i=1}^{n} \left(t_a^-(x_{\sigma(i)})\right)^{w_i}, \prod_{i=1}^{n} \left(t_a^+(x_{\sigma(i)})\right)^{w_i} \right], \left[\sqrt{1 - \prod_{i=1}^{n} \left(1 - f_a^-(x_{\sigma(i)})^q\right)^{w_i}}, \sqrt{1 - \prod_{i=1}^{n} \left(1 - f_a^+(x_{\sigma(i)})^q\right)^{w_i}} \right] \right\rangle. \tag{36}$$

Especially, if $\mu(\{x_i\}) = (1/n)(i = 1, 2, \ldots, n)$, it is obvious to find that the IVq-ROFOWCG operator reduces to the IVq-ROFOCG operator.

4. A MAGDM Method Based on the IVq-ROFCA Operator

In this section, we develop a MAGDM method based on the IVq-ROFCA operator. For a MAGDM problem, $X = \{x_1, x_2, \ldots, x_m\}$ is the set of alternatives, $C = \{C_1, C_2, \ldots, C_n\}$ is the set of interaction between attributes, and $\mu(C_j)(j = 1, 2, \ldots, n)$ is the fuzzy measure of attributes, $e = \{e_1, e_2, \ldots, e_t\}$ indicates the set of experts, and $\mu(e_k)(k = 1, 2, \ldots, t)$ is the fuzzy measure of experts. For each expert, a decision matrix of different alternatives is $A = (a_{ij})_{m \times n} = \langle [t^-_{a_{ij}}, t^+_{a_{ij}}], [f^-_{a_{ij}}, f^+_{a_{ij}}] \rangle$, where a_{ij} are IVq-ROFNs that satisfy $(t^+_{a_{ij}})^q + (f^+_{a_{ij}})^q \leq 1 (q \geq 1)$. For the k th expert e_k, the decision matrix is expressed as $A^{(k)} = (a_{ij}^{(k)})_{m \times n}$, $a_{ij}^{(k)}$ denotes the kth expert's evaluation of alternative x_i with the attribute C_j. Because of the differences between physical dimensions of attributes, the decision matrix needs to be standardized. Correspondingly, Ω_1 indicates the benefit type and Ω_2 indicates the cost type, and the standardization processing is shown in equation (37) and the complement operation of fuzzy numbers is shown in equation (38):

$$r_{ij}^{(k)} = \begin{cases} a_{ij}^{(k)}, & a_{ij}^{(k)} \in \Omega_1, \\ \left(a_{ij}^{(k)}\right)^c, & a_{ij}^{(k)} \in \Omega_2, \end{cases} (i = 1, 2, \ldots, m, j = 1, 2, \ldots, n), \tag{37}$$

$$\left(a_{ij}^{(k)}\right)^c = \left\langle \left[f^-_{a_{ij}^{(k)}}, f^+_{a_{ij}^{(k)}}\right], \left[t^-_{a_{ij}^{(k)}}, t^+_{a_{ij}^{(k)}}\right] \right\rangle (i = 1, 2, \ldots, m, j = 1, 2, \ldots, n). \tag{38}$$

The steps of the MAGDM method based on the IVq-ROFCA operator are as follows:

(1) According to the evaluation matrix given by experts, the appropriate value of q is chosen. When the amount of data is small, the value of q can be determined by observation. If the amount of data is large, q can be determined by the traversal method utilizing $(t^+_{a_{ij}^{(k)}})^q + (f^+_{a_{ij}^{(k)}})^q \leq 1 (q \geq 1)$.

(2) Using equations (37) and (38), the decision matrix $A^{(k)}$ is converted into the standardize matrix $A'^{(k)}$.

(3) $A'^{(k)} = (a_{ij}^{'(k)})_{m \times n}$ is aggregated into $R = (r_{ij})_{m \times n}$ employing the IVq-ROFCA operator, and the collection matrix is derived by

$$r_{ij} = IVq - ROFCA\left(a_{ij}^{'(1)}, a_{ij}^{'(2)}, \ldots, a_{ij}^{'(t)}\right) (i = 1, 2, \ldots, m, j = 1, 2, \ldots, n), \tag{39}$$

where $a_{ij}^{'(t)}$ represents the IVq-ROFNs in the decision matrix given by the t-th expert.

(4) Then, the collective value r_i of x_i is derived by equation (40), which aggregates the fuzzy information corresponding to attributes:

$$r_i = IVq - ROFCA(r_{i1}, r_{i2}, \ldots, r_{in}) (i = 1, 2, \ldots, m). \tag{40}$$

(5) The score value and the accuracy value of r_i are obtained according to equations (9) and (10), which determine the ranking of alternatives.

5. Case of the Warning Management System for Hypertension

5.1. Decision Result by the Proposed MAGDM Method. In order to improve the management efficiency of doctors, we

Table 2: The fuzzy measures of attributes.

Fuzzy measures of $\{\mathbf{C}_1, \mathbf{C}_2, \mathbf{C}_3, \mathbf{C}_4\}$	
$\mu(\{\varnothing\}) = 0$	$\mu(\{C_2, C_3\}) = 0.45$
$\mu(\{C_1\}) = 0.2$	$\mu(\{C_2, C_4\}) = 0.5$
$\mu(\{C_2\}) = 0.2$	$\mu(\{C_3, C_4\}) = 0.57$
$\mu(\{C_3\}) = 0.2\ 5$	$\mu(\{C_1, C_2, C_3\}) = 0.65$
$\mu(\{C_4\}) = 0.35$	$\mu(\{C_1, C_2, C_4\}) = 0.77$
$\mu(\{C_1, C_2\}) = 0.42$	$\mu(\{C_1, C_3, C_4\}) = 0.8$
$\mu(\{C_1, C_3\}) = 0.45$	$\mu(\{C_2, C_3, C_4\}) = 0.8$
$\mu(\{C_1, C_4\}) = 0.5$	$\mu(\{C_1, C_2, C_3, C_4\}) = 1.0$

plan to develop a daily follow-up warning management system for hypertension. The warning degrees are divided into five levels that are represented in five colors denoted by $X = \{x_1, x_2, x_3, x_4, x_5\}$. x_1 indicates that the resident shall seek treatment immediately, showing red; x_2 indicates the resident needs treatment in time, showing orange; x_3 indicates the resident should be treated for the timely follow-up to promote management, showing yellow; x_4 indicates that blood pressure management may be required in the future, showing blue; and x_5 indicates that the blood pressure of the resident is normal, and the color is green. The warning levels are related to numerous factors of hypertension patients, among which the blood pressure measurements, related diseases, related risk factors, and follow-up time intervals are represented as the attributes denoted by $C = \{C_1, C_2, C_3, C_4\}$. From a managerial perspective, C_1, C_2, and C_3 are benefit types, and C_4 is the cost type. With different importance, the fuzzy measures of attributes are listed in Table 2.

Without taking medicine, a 60-year-old resident has a measured value and related factors as follows: systolic blood pressure is 153 mmHg, diastolic blood pressure is 98 mmHg, there is no relevant disease record, obesity and family genetic history, and the doctor had no follow-up record of him. Three experts $e = \{e_1, e_2, e_3\}$ are invited to evaluate the warning degree of this resident using IVq-ROFNs, and the fuzzy measures of them are $\mu\{e_1\} = \mu\{e_2\} = \mu\{e_3\} = 0.4$, $\mu\{e_1, e_2\} = \mu\{e_2, e_3\} = \mu\{e_1, e_3\} = 0.7$, $\mu\{e_1, e_2, e_3\} = 1$. Tables 3–5 are the evaluation values of the three experts after standardization. Subsequently, the decision result is determined utilizing the proposed MAGDM method.

We set $q = 3$ via observation, and then, the collective matrix R is obtained by equation (39) which is presented in Table 6.

The comprehensive attribute values of each alternative are obtained through equation (40), and the results are obtained as follows:

$$\begin{aligned} r_1 &= <[0.63, 0.74], [0.32, 0.46]>, \\ r_2 &= <[0.67, 0.75], [0.27, 0.38]>, \\ r_3 &= <[0.71, 0.78], [0.32, 0.42]>, \\ r_4 &= <[0.56, 0.64], [0.40, 0.50]>, \\ r_5 &= <[0.47, 0.59], [0.57, 0.68]>. \end{aligned} \tag{41}$$

The score of each alternative is determined by equation (9):

$$\begin{aligned} S(r_1) &= 0.53, \\ S(r_2) &= 0.65, \\ S(r_3) &= 0.73, \\ S(r_4) &= 0.25, \\ S(r_5) &= -0.19. \end{aligned} \tag{42}$$

The ranking result of alternatives' scores is derived as follows:

$$S(r_3) > S(r_2) > S(r_1) > S(r_4) > S(r_5). \tag{43}$$

Therefore, the ranking of alternatives is $x_3 > x_2 > x_1 > x_4 > x_5$, and x_3 is the optimal selection, which means the doctor is required for timely follow-up to promote management of blood pressure of this patient who receives yellow warning.

5.2. Parameter Analysis. For further studying the influence of q, we analyze different values of q in this subsection, which are 2, 3, 4, and 5 in this case, and the results are listed in Table 7 and Figure 1.

It can be seen from Table 7 that the score rankings are always $S(r_3) > S(r_2) > S(r_1) > S(r_4) > S(r_5)$, and the early-warning results are all yellow warnings, which are consistent

Table 3: Decision matrix $A'^{(1)}$ by e_1.

	$\mathbf{C}_1$	$\mathbf{C}_2$	$\mathbf{C}_3$	$\mathbf{C}_4$
$\mathbf{x}_1$	< [0.7,0.8], [0.2,0.3] >	< [0.5,0.6], [0.6,0.7] >	< [0.8,0.9], [0.1,0.2] >	<[0.1,0.2], [0.8,0.95] >
$\mathbf{x}_2$	< [0.8,0.9], [0.1,0.2] >	<[0.5,0.55], [0.4,0.5] >	<[0.8,0.9], [0.1,0.15] >	<[0.1,0.2], [0.8,0.95] >
$\mathbf{x}_3$	<[0.9,0.95], [0.1,0.2] >	< [0.5,0.6], [0.3,0.4] >	<[0.8,0.85], [0.2,0.3] >	< [0.1,0.2], [0.8,0.9] >
$\mathbf{x}_4$	< [0.6,0.7], [0.4,0.5] >	< [0.4,0.5], [0.3,0.4] >	<[0.75,0.8], [0.2,0.3] >	< [0.2,0.3], [0.7,0.8] >
$\mathbf{x}_5$	< [0.1,0.2], [0.8,0.9] >	< [0.2,0.3], [0.5,0.6] >	<[0.7,0.85], [0.3,0.4] >	< [0.2,0.3], [0.6,0.7] >

Table 4: Decision matrix $A'^{(2)}$ by e_2.

	$\mathbf{C}_1$	$\mathbf{C}_2$	$\mathbf{C}_3$	$\mathbf{C}_4$
$\mathbf{x}_1$	<[0.6,0.7],[0.2,0.3]>	<[0.6,0.65],[0.4,0.5]>	<[0.8,0.9], [0.1,0.2]>	<[0.1,0.2],[0.8,0.95]>
$\mathbf{x}_2$	<[0.8,0.85],[0.1,0.15]>	<[0.5,0.55],[0.4,0.5]>	<[0.8,0.85],[0.1,0.15]>	<[0.1,0.2],[0.8,0.95]>
$\mathbf{x}_3$	<[0.85,0.9],[0.1,0.15]>	<[0.5,0.6], [0.3,0.4]>	<[0.8,0.85], [0.2,0.3]>	<[0.1,0.2],[0.85,0.9]>
$\mathbf{x}_4$	<[0.4,0.5],[0.3,0.4]>	<[0.4,0.5], [0.3,0.4]>	<[0.75,0.8], [0.2,0.3]>	<[0.2,0.3],[0.85,0.9]>
$\mathbf{x}_5$	<[0.1,0.2],[0.8,0.95]>	<[0.2,0.3], [0.5,0.6]>	<[0.7,0.85], [0.3,0.4]>	<[0.2,0.3],[0.7,0.85]>

Table 5: Decision matrix $A'^{(3)}$ by e_3.

	$\mathbf{C}_1$	$\mathbf{C}_2$	$\mathbf{C}_3$	$\mathbf{C}_4$
$\mathbf{x}_1$	< [0.7,0.8],[0.2,0.3] >	<[0.5,0.6],[0.4,0.5] >	<[0.8,0.9], [0.1,0.2]>	< [0.1,0.2],[0.8,0.95] >
$\mathbf{x}_2$	< [0.85,0.9],[0.1,0.2] >	<[0.5,0.55],[0.4,0.5] >	<[0.8,0.85],[0.1,0.15]>	< [0.1,0.2],[0.8,0.95] >
$\mathbf{x}_3$	<[0.9,0.95],[0.15,0.2] >	< [0.6,0.7], [0.3,0.4] >	<[0.8,0.85],[0.2,0.3]>	< [0.1,0.2],[0.85,0.9] >
$\mathbf{x}_4$	< [0.6,0.7],[0.4,0.5] >	< [0.5,0.6], [0.3,0.4] >	<[0.75,0.85],[0.2,0.3] >	< [0.2,0.3],[0.85,0.9] >
$\mathbf{x}_5$	< [0.1,0.2],[0.8,0.9] >	< [0.4,0.5], [0.7,0.8] >	< [0.75,0.8], [0.3,0.4] >	<[0.2,0.3],[0.75,0.85] >

Table 6: Collective matrix R.

	$\mathbf{C}_1$	$\mathbf{C}_2$	$\mathbf{C}_3$	$\mathbf{C}_4$
$\mathbf{r}_1$	<[0.67,0.78], [0.20,0.30]>	<[0.54,0.62], [0.47,0.57]>	<[0.80,0.90], [0.10,0.20]>	<[0.10,0.20], [0.80,0.95]>
$\mathbf{r}_2$	<[0.82,0.89], [0.10,0.18]>	<[0.50,0.55], [0.40,0.50]>	<[0.80,0.87], [0.10,0.15]>	<[0.10,0.20], [0.80,0.95]>
$\mathbf{r}_3$	<[0.89,0.94], [0.11,0.18]>	<[0.54,0.64], [0.30,0.40]>	<[0.80,0.85], [0.20,0.30]>	<[0.10,0.20], [0.83,0.90] >
$\mathbf{r}_4$	<[0.56,0.66], [0.37,0.47]>	<[0.44,0.54], [0.30,0.40]>	<[0.75,0.82], [0.20,0.30]>	<[0.20,0.30], [0.79,0.86]>
$\mathbf{r}_5$	<[0.10,0.20], [0.80,0.91]>	<[0.29,0.39], [0.55,0.65]>	<[0.72,0.84], [0.30,0.40]>	<[0.20,0.30], [0.67,0.79]>

with the results obtained in Table 7. Hence, the decision result does not change due to the variation in the value of q. However, with the increase in the q value, it can be discovered that scores of the alternatives are declining completely but show different trends of downward. Considering the practical application, there is not much practical significance for IVq-ROFNs with the large value of q, which often does not exceed 5. Therefore, the value of q will not have an impact on the final decision result.

5.3. Comparison Analysis. In this subsection, the proposed method is used to solve the case in [32], which is compared with the method given in [32, 40]. There are 5 alternatives with 4 attributes C_1, C_2, C_3, and C_4 for 3 experts e_1, e_2, and e_3 to make a decision, and the evaluation matrices $A^{(1)}$, $A^{(2)}$, and $A^{(3)}$ of interval-valued intuitionistic fuzzy numbers (IVIFNs) given are listed in Tables 8–10. Besides, the fuzzy measures of experts are $\mu\{e_1\} = \mu\{e_2\} = \mu\{e_3\} = 0.4$, $\mu\{e_1, e_2\} = \mu\{e_2, e_3\} = \mu\{e_1, e_3\} = 0.73$, and $\mu\{e_1, e_2, e_3\} = 1$, and the fuzzy measures of attributes are shown in Table 11.

If μ is the fuzzy measure on the nonempty finite set $X = \{x_1, x_2, \ldots, x_n\}$ $(\mu(\varnothing) = 0)$, for an IVIFS $\widetilde{a}(x_i) = < [t^-_{\widetilde{a}}(x_i), t^+_{\widetilde{a}}(x_i)], [f^-_{\widetilde{a}}(x_i), f^+_{\widetilde{a}}(x_i)] > (i = 1, 2, \ldots, n)$, the GIIFGA operator and the interval-valued

Table 7: Decision result with $q = 2,3,4,$ and5.

q	Results of IVq-ROFCA	Score of alternatives	Ranking of alternatives
2	$r_1 = <[0.60, 0.71], [0.32, 0.46]> r_2 = <[0.64, 0.73], [0.27, 0.38]> r_3 = <[0.69, 0.76], [0.32, 0.42]> r_4 = <[0.53, 0.62], [0.40, 0.50]> r_5 = <[0.42, 0.54], [0.57, 0.68]>$	$S(r_1) = 0.55 S(r_2) = 0.73 S(r_3) = 0.77 S(r_4) = 0.26 S(r_5) = -0.32$	$x_3 > x_2 > x_1 > x_4 > x_5$
3	$r_1 = <[0.63, 0.74], [0.32, 0.46]> r_2 = <[0.67, 0.75], [0.27, 0.38]> r_3 = <[0.71, 0.78], [0.32, 0.42]> r_4 = <[0.56, 0.64], [0.40, 0.50]> r_5 = <[0.47, 0.59], [0.57, 0.68]>$	$S(r_1) = 0.53 S(r_2) = 0.65 S(r_3) = 0.73 S(r_4) = 0.25 S(r_5) = -0.19$	$x_3 > x_2 > x_1 > x_4 > x_5$
4	$r_1 = <[0.65, 0.75], [0.32, 0.46]> r_2 = <[0.69, 0.77], [0.27, 0.38]> r_3 = <[0.73, 0.80], [0.32, 0.42]> r_4 = <[0.58, 0.66], [0.40, 0.50]> r_5 = <[0.51, 0.62], [0.57, 0.68]>$	$S(r_1) = 0.44 S(r_2) = 0.55 S(r_3) = 0.65 S(r_4) = 0.21 S(r_5) = -0.10$	$x_3 > x_2 > x_1 > x_4 > x_5$
5	$r_1 = <[0.67, 0.77], [0.32, 0.46]> r_2 = <[0.71, 0.78], [0.27, 0.38]> r_3 = <[0.75, 0.81], [0.32, 0.42]> r_4 = <[0.60, 0.68], [0.40, 0.50]> r_5 = <[0.54, 0.65], [0.57, 0.68]>$	$S(r_1) = 0.37 S(r_2) = 0.46 S(r_3) = 0.57 S(r_4) = 0.18 S(r_5) = -0.04$	$x_3 > x_2 > x_1 > x_4 > x_5$

Table 8: Decision matrix $A^{(1)}$.

	$\mathbf{C}_1$	$\mathbf{C}_2$	$\mathbf{C}_3$	$\mathbf{C}_4$
$\mathbf{x}_1$	< [0.4,0.5],[0.3,0.4] >	< [0.4,0.6],[0.2,0.4] >	< [0.1,0.3],[0.5,0.6] >	< [0.3,0.4],[0.3,0.5] >
$\mathbf{x}_2$	< [0.6,0.7],[0.2,0.3] >	< [0.6,0.7],[0.2,0.3] >	< [0.4,0.7],[0.1,0.2] >	< [0.5,0.6],[0.1,0.3] >
$\mathbf{x}_3$	< [0.6,0.7],[0.1,0.2] >	< [0.5,0.6],[0.3,0.4] >	< [0.5,0.6],[0.1,0.3] >	< [0.4,0.5],[0.2,0.4] >
$\mathbf{x}_4$	< [0.3,0.4],[0.2,0.3] >	< [0.6,0.7],[0.1,0.3] >	< [0.3,0.4],[0.1,0.2] >	< [0.3,0.7],[0.1,0.2] >
$\mathbf{x}_5$	< [0.7,0.8],[0.1,0.2] >	< [0.3,0.5],[0.1,0.3] >	< [0.5,0.6],[0.2,0.3] >	< [0.3,0.4],[0.5,0.6] >

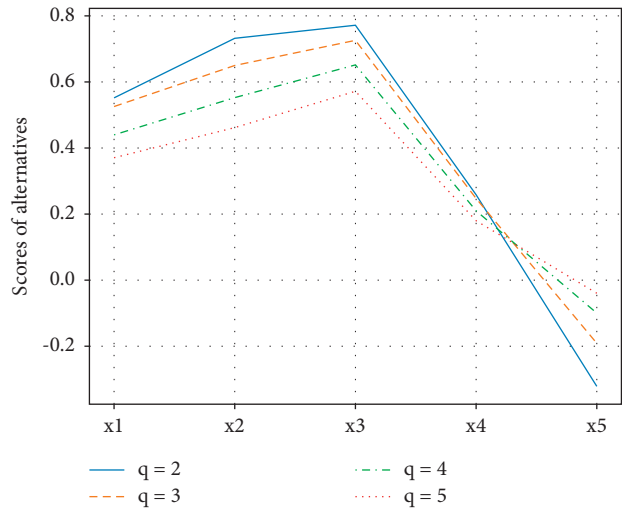


Figure 1: Decision result with $q = 2,3,4$, and 5.

Table 9: Decision matrix $A^{(2)}$.

	$\mathbf{C}_1$	$\mathbf{C}_2$	$\mathbf{C}_3$	$\mathbf{C}_4$
$\mathbf{x}_1$	< [0.3,0.4],[0.4,0.5] >	< [0.5,0.6],[0.1,0.3] >	< [0.4,0.5],[0.3,0.4] >	< [0.4,0.6],[0.2,0.4] >
$\mathbf{x}_2$	< [0.3,0.6],[0.3,0.4] >	< [0.4,0.7],[0.1,0.2] >	< [0.5,0.6],[0.2,0.3] >	< [0.6,0.7],[0.2,0.3] >
$\mathbf{x}_3$	< [0.6,0.8],[0.1,0.2] >	< [0.5,0.6],[0.1,0.2] >	< [0.5,0.7],[0.2,0.3] >	< [0.1,0.3],[0.5,0.6] >
$\mathbf{x}_4$	< [0.4,0.5],[0.3,0.5] >	< [0.5,0.8],[0.1,0.2] >	< [0.2,0.5],[0.3,0.4] >	< [0.4,0.7],[0.1,0.2] >
$\mathbf{x}_5$	< [0.6,0.7],[0.2,0.3] >	< [0.6,0.7],[0.1,0.2] >	< [0.5,0.7],[0.2,0.3] >	< [0.6,0.7],[0.1,0.3] >

Table 10: Decision matrix $A^{(3)}$.

	$\mathbf{C}_1$	$\mathbf{C}_2$	$\mathbf{C}_3$	$\mathbf{C}_4$
$\mathbf{x}_1$	< [0.2,0.5],[0.3,0.4] >	< [0.4,0.5],[0.1,0.2] >	< [0.3,0.6],[0.2,0.3] >	< [0.3,0.7],[0.1,0.3] >
$\mathbf{x}_2$	< [0.2,0.7],[0.2,0.3] >	< [0.3,0.6],[0.2,0.4] >	< [0.4,0.7],[0.1,0.2] >	< [0.5,0.8],[0.1,0.2] >
$\mathbf{x}_3$	< [0.5,0.6],[0.3,0.4] >	< [0.7,0.8],[0.1,0.2] >	< [0.5,0.6],[0.2,0.3] >	< [0.4,0.5],[0.3,0.4] >
$\mathbf{x}_4$	< [0.3,0.6],[0.2,0.4] >	< [0.4,0.6],[0.2,0.3] >	< [0.1,0.4],[0.3,0.6] >	< [0.3,0.7],[0.1,0.2] >
$\mathbf{x}_5$	< [0.6,0.7],[0.1,0.3] >	< [0.5,0.6],[0.3,0.4] >	< [0.5,0.6],[0.2,0.3] >	< [0.5,0.6],[0.2,0.4] >

Table 11: Fuzzy measures of attributes.

Fuzzy measures of attributes	
$\mu\{C_1\} = 0.4$	$\mu\{C_2, C_4\} = 0.43$
$\mu\{C_2\} = 0.25$	$\mu\{C_3, C_4\} = 0.54$
$\mu\{C_3\} = 0.37$	$\mu\{C_1, C_2, C_3\} = 0.88$
$\mu\{C_4\} = 0.2$	$\mu\{C_1, C_2, C_4\} = 0.75$
$\mu\{C_1, C_2\} = 0.6$	$\mu\{C_1, C_3, C_4\} = 0.84$
$\mu\{C_1, C_3\} = 0.7$	$\mu\{C_2, C_3, C_4\} = 0.73$
$\mu\{C_1, C_4\} = 0.56$	$\mu\{C_1, C_2, C_3, C_4\} = 1$
$\mu\{C_2, C_3\} = 0.68$	

intuitionistic fuzzy Einstein geometric Choquet integral (IVIFEGC) operator are defined as follows:

$$GIIFGA(\widehat{a}(x_1), \widehat{a}(x_2), \ldots, \widehat{a}(x_n)) =$$

$$< \left[\prod_{i=1}^{n} \left(t_{\widehat{a}}^{-}\left(x_{\sigma(i)}\right) \right)^{\mu\left(B_{\sigma(i)}\right)-\mu\left(B_{\sigma(i+1)}\right)}, \prod_{i=1}^{n} \left(t_{\widehat{a}}^{+}\left(x_{\sigma(i)}\right) \right)^{\mu\left(B_{\sigma(i)}\right)-\mu\left(B_{\sigma(i+1)}\right)} \right], \left[1 - \prod_{i=1}^{n} \left(1 - f_{\widehat{a}}^{-}\left(x_{\sigma(i)}\right) \right)^{\mu\left(B_{\sigma(i)}\right)-\mu\left(B_{\sigma(i+1)}\right)}, 1 - \prod_{i=1}^{n} \left(1 - f_{\widehat{a}}^{+}\left(x_{\sigma(i)}\right) \right)^{\mu\left(B_{\sigma(i)}\right)-\mu\left(B_{\sigma(i+1)}\right)} \right] >,$$

$$IVIFEGC(\widehat{a}(x_1), \widehat{a}(x_2), \ldots, \widehat{a}(x_n)) =$$

$$< \left[\frac{2\prod_{j=1}^{n} t_{\widehat{a}}^{-}\left(x_{\sigma(i)}\right)_{\sigma(j)}^{\mu\left(B_{\sigma(i)}\right)-\mu\left(B_{\sigma(i+1)}\right)}}{\prod_{j=1}^{n}\left(2 - t_{\widehat{a}}^{-}\left(x_{\sigma(i)}\right)_{\sigma(j)}\right)^{\mu\left(B_{\sigma(i)}\right)-\mu\left(B_{\sigma(i+1)}\right)} + \prod_{j=1}^{n} t_{\widehat{a}}^{-}\left(x_{\sigma(i)}\right)_{\sigma(j)}^{\mu\left(B_{\sigma(i)}\right)-\mu\left(B_{\sigma(i+1)}\right)}}, \frac{2\prod_{j=1}^{n} t_{\widehat{a}}^{+}\left(x_i\right)_{\sigma(j)}^{\mu\left(B_{\sigma(i)}\right)-\mu\left(B_{\sigma(i+1)}\right)}}{\prod_{j=1}^{n}\left(2 - t_{\widehat{a}}^{+}\left(x_i\right)_{\sigma(j)}\right)^{\mu\left(B_{\sigma(i)}\right)-\mu\left(B_{\sigma(i+1)}\right)} + \prod_{j=1}^{n} t_{\widehat{a}}^{+}\left(x_i\right)_{\sigma(j)}^{\mu\left(B_{\sigma(i)}\right)-\mu\left(B_{\sigma(i+1)}\right)}} \right],$$

$$\left[\frac{\prod_{j=1}^{n}\left(1 + f_{\widehat{a}}^{-}\left(x_i\right)\right)^{\mu\left(B_{\sigma(i)}\right)-\mu\left(B_{\sigma(i+1)}\right)} - \prod_{j=1}^{n}\left(1 - f_{\widehat{a}}^{-}\left(x_i\right)\right)^{\mu\left(B_{\sigma(i)}\right)-\mu\left(B_{\sigma(i+1)}\right)}}{\prod_{j=1}^{n}\left(1 + f_{\widehat{a}}^{-}\left(x_i\right)\right)^{\mu\left(B_{\sigma(i)}\right)-\mu\left(B_{\sigma(i+1)}\right)} + \prod_{j=1}^{n}\left(1 - f_{\widehat{a}}^{-}\left(x_i\right)\right)^{\mu\left(B_{\sigma(i)}\right)-\mu\left(B_{\sigma(i+1)}\right)}}, \frac{\prod_{j=1}^{n}\left(1 + f_{\widehat{a}}^{+}\left(x_i\right)\right)^{\mu\left(B_{\sigma(i)}\right)-\mu\left(B_{\sigma(i+1)}\right)} - \prod_{j=1}^{n}\left(1 - f_{\widehat{a}}^{+}\left(x_i\right)\right)^{\mu\left(B_{\sigma(i)}\right)-\mu\left(B_{\sigma(i+1)}\right)}}{\prod_{j=1}^{n}\left(1 + f_{\widehat{a}}^{+}\left(x_i\right)\right)^{\mu\left(B_{\sigma(i)}\right)-\mu\left(B_{\sigma(i+1)}\right)} + \prod_{j=1}^{n}\left(1 - f_{\widehat{a}}^{+}\left(x_i\right)\right)^{\mu\left(B_{\sigma(i)}\right)-\mu\left(B_{\sigma(i+1)}\right)}} \right] >, \tag{44}$$

where $(\sigma(1), \sigma(2), \ldots, \sigma(n))$ is a permutation of $(1, 2, \ldots n)$ that satisfies $\widehat{a}(x_{\sigma(1)}) \leq \widehat{a}(x_{\sigma(2)}) \leq \cdots \leq \widehat{a}(x_{\sigma(n)}), B_{\sigma(i)} = \left\{x_{\sigma(i)}, x_{\sigma(i+1)}, \ldots, x_{\sigma(n)}\right\} (i = 1, 2, \ldots, n, B_{\sigma(n+1)} = \varnothing)$.

In this case, we set $q = 1$, IVq-ROFNs are equal to IVIFNs, which are applied to the comparison with existing methods. The obtained results of different methods are shown in Figure 2 and Table 12.

In this case, compared with the method proposed in [40], the same best and worst solutions are obtained by the method proposed in this paper, and obvious differences in the scores of alternatives are discovered. However, there are differences between x_2 and x_3, and the small deviation is derived between them. From the method proposed in [32], the decision result is completely different from the other two methods, and the closeness of each alternative has a few differences. We analyze the reason as follows.

In this case, the membership degrees of IVIFNs are generally greater than nonmembership degrees. From the method proposed in [32], multiple decision matrices are aggregated by CITOPSIS, which reduces the difference in the influence of the membership and nonmembership of each alternative on the results. Thus, the deviations of alternatives are not obvious. Contrarily, the method proposed in this paper employs IVq-ROFCA to process fuzzy information, which further highlights the degree of expert support for the alternatives. The collective matrices of GIIFGA and IVq-ROFCA of Tables 8–10 are listed in Tables 13 and 14, respectively. According to equation (9), the scores of these two matrices are presented in Figure 3.

From Figure 3, the abscissa in the figure represents the alternatives with different attributes and the ordinate indicates their scores. It is easy to find there are only few differences between membership and nonmembership, except for the evaluation of x_2 with C_1. Correspondingly, the span of the interval in membership with a wide range can be found in the experts' evaluation of x_2 with C_1 from Tables 8–10, such as [0.2, 0.7]. Considering the above, there are differences in the selection of x_2 between the method proposed in this paper and the proposed method in [32]. On the other hand, it reflects that more advantages of the method proposed in this paper will be discovered, while the

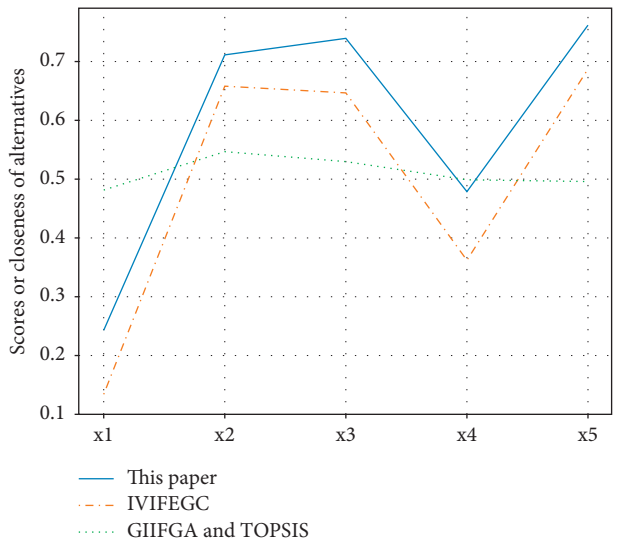


Figure 2: Ranking of different methods.

Table 12: Results by different methods.

Operators	Score or closeness	Alternative ranking
Proposed method	Score: $S(x_1) = 0.16 S(x_2) = 0.71 S(x_3) = 0.74 S(x_4) = 0.38 S(x_5) = 0.77$	$x_5 > x_3 > x_2 > x_4 > x_1$
IVIFEGC [40]	Score: $S(x_1) = 0.05 S(x_2) = 0.66 S(x_3) = 0.65 S(x_4) = 0.26 S(x_5) = 0.70$	$x_5 > x_2 > x_3 > x_4 > x_1$
Choquet integral-based TOPSIS (CITOPSIS) [32]	Closeness: $r(x_1) = 0.4817 r(x_2) = 0.5465 r(x_3) = 0.5297 r(x_4) = 0.4990 r(x_5) = 0.4958$	$x_2 > x_3 > x_4 > x_5 > x_1$

Table 13: Collective matrix by GIIFGA.

	$\mathbf{C}_1$	$\mathbf{C}_2$	$\mathbf{C}_3$	$\mathbf{C}_4$
x_1	<[0.3017,0.4645], [0.2685,0.3687]>	<[0.4373,0.5650], [0.1282,0.2983]>	<[0.2452,0.4685], [0.3257,0.4280]>	<[0.3299,0.5720], [0.1911,0.3925]>
x_2	<[0.3463,0.5386], [0.2917,0.3925]>	<[0.4353,0.6715], [0.1683,0.2983]>	<[0.4248,0.6715], [0.1282,0.2283]>	<[0.5310,0.7083], [0.1343,0.2616]>
x_3	<[0.5712,0.7083], [0.1590,0.2598]>	<[0.5720,0.6732], [0.1590,0.2598]>	<[0.5000,0.6382], [0.1683,0.3000]>	<[0.2751,0.4356], [0.3257,0.4622] >
x_4	<[0.3242,0.4996], [0.2283,0.3990]>	<[0.5000,0.7083], [0.1282,0.2616]>	<[0.1951,0.4306], [0.2260,0.3966]>	<[0.3366,0.7000], [0.1000,0.2000]>
x_5	<[0.6382,0.7384], [0.1282,0.2616]>	<[0.4685,0.6075], [0.1716,0.2982]>	<[0.5000,0.6382], [0.2000,0.3000]>	<[0.4685,0.5720], [0.2614,0.4280]>

Table 14: Collective matrix by IVq-ROFCA.

	$\mathbf{C}_1$	$\mathbf{C}_2$	$\mathbf{C}_3$	$\mathbf{C}_4$
$\mathbf{x}_1$	<[0.3122,0.4748], [0.3242,0.4248]>	<[0.4288,0.5694], [0.1320,0.2944]>	<[0.2575,0.4686], [0.3219,0.4278]>	<[0.3285,0.5722], [0.1871,0.3977]>
x_2	<[0.4152,0.6758], [0.2231,0.3242]>	<[0.4632,0.6701], [0.1659,0.2957]>	<[0.4288,0.6758], [0.1206,0.2231]>	<[0.5292,0.7056], [0.1206,0.2624]>
x_3	<[0.5694,0.7043], [0.1437,0.2514]>	<[0.5776,0.6818], [0.1552,0.2639]>	<[0.5000,0.6299], [0.1516,0.3000]>	<[0.3306,0.4524], [0.2928,0.4563]>
x_4	<[0.3285,0.5004], [0.2231,0.3787]>	<[0.5143,0.7043], [0.1257,0.2689]>	<[0.2116,0.4288], [0.1933,0.3456]>	<[0.3285,0.7000], [0.1000,0.2000]>
x_5	<[0.6435,0.7449], [0.1206,0.2551]>	<[0.4614,0.5953], [0.1437,0.2957]>	<[0.5000,0.6299], [0.2000,0.3000]>	<[0.4614,0.5647], [0.2393,0.4353]>

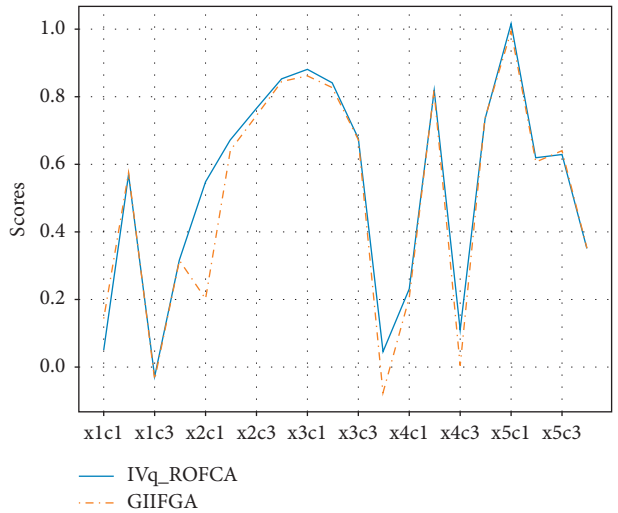


Figure 3: The scores of different matrices obtained by GIFFGA and IVq-ROFCA.

degree of support of experts for the alternatives is significantly higher than that of the opposition.

Above all, the proposed method employs the Choquet integral to aggregate interval-valued q-rung orthopair fuzzy information in this paper, which is a powerful tool to deal with MAGDM problems with dependent attributes. Moreover, the proposed method will highlight the support of experts for the alternatives and better reflect the superiority of the optimal alternative. Nevertheless, the proposed operators and the MAGDM method are used to handle decision problems for IVq-ROFNs, which are not able to be applied to different fuzzy environments. Besides, as a great impact on decision results remains to be studied, the degree of consensus among experts needs to be further studied in the proposed method.

6. Conclusion

The Choquet integral is an efficient tool to solve decision problems with interaction between attributes. For addressing complex MAGDM problems under interval-valued q-rung orthopair fuzzy information, we develop the IVq-ROFCA operator and the IVq-ROFCG operator and discuss some properties of them, including idempotency, commutativity, monotonicity, and boundedness. Particularly, we further design Choquet integral weighted and ordered operators for IVq-ROFS. Subsequently, a novel method is devised to process fuzzy information employing the IVq-ROFCA operator. Finally, a case of early-warning in the daily management of hypertension is given to illustrate the proposed method, and the results obtained are consistent with the actual situation provided by medical experts. The feasibility and effectiveness of the proposed method are proved by sensitivity analysis and comparative analysis additionally. Ulteriorly, our future research will focus on consensus models of IVq-ROFS, for instance, the MAGDM method with the Choquet integral based on consistency and consensus of experts. Moreover, considering the broad development prospect of fuzzy theory in the era of big data [41], the integration of the decision-making method, machine learning and big data are also one of our research directions in the future.

Data Availability

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

Disclosure

The innovation of this paper has been preprinted in arXiv: "Interval-Valued q-Rung Orthopair Fuzzy Choquet Integral Operators and Its Application in Group Decision Making" [42] by the authors of this paper.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported in part by the Department of Education of Jiangxi Province of China under Grant GJJ180270, National Natural Science Foundation of China grant no. 61866014, the Department of Shenzhen Local Science and Technology Development under Grant 2021Szvup052, and Jiangxi Provincial Natural Science Foundation (20202BAB202006).

References

[1] L. A. Zadeh, "Fuzzy sets," *Information and Control*, vol. 8, no. 3, pp. 338–353, 1965.

[2] R. R. Yager, "Pythagorean membership grades in multicriteria decision making," *IEEE Transactions on Fuzzy Systems*, vol. 22, no. 4, pp. 958–965, 2014.

[3] X. Zhang and Z. Xu, "Extension of TOPSIS to multiple criteria decision making with pythagorean fuzzy sets," *International Journal of Intelligent Systems*, vol. 29, no. 12, pp. 1061–1078, 2014.

[4] X. Zhang, "Multicriteria Pythagorean fuzzy decision analysis: a hierarchical QUALIFLEX approach with the closeness index-based ranking methods," *Information Sciences*, vol. 330, pp. 104–124, 2016.

[5] X. Peng and Y. Yang, "Some results for Pythagorean fuzzy sets," *International Journal of Intelligent Systems*, vol. 30, no. 11, pp. 1133–1160, 2015.

[6] K. T. Atanassov, "Intuitionistic fuzzy sets," *Fuzzy Sets and Systems*, vol. 20, no. 1, pp. 87–96, 1986.

[7] R. R. Yager, "Pythagorean Fuzzy Subsets," in *Proceedings of the 2013 Joint IFSA World Congress and NAFIPS Annual Meeting (IFSA/NAFIPS)*, pp. 57–61, Edmonton, AB, Canada, June 2013.

[8] R. Verma and J. M. Merigo, "On generalized similarity measures for Pythagorean fuzzy sets and their applications to multiple attribute decision-making," *International Journal of Intelligent Systems*, vol. 34, no. 10, pp. 2556–2583, 2019.

[9] G. Bakioglu and A. O. Atahan, "AHP integrated TOPSIS and VIKOR methods with Pythagorean fuzzy sets to prioritize risks in self-driving vehicles," *Applied Soft Computing*, vol. 99, Article ID 106948, 2021.

[10] X. Peng and Y. Yang, "Fundamental properties of interval-valued pythagorean fuzzy aggregation operators," *International Journal of Intelligent Systems*, vol. 31, no. 5, pp. 444–487, 2016.

[11] R. R. Yager, "Generalized orthopair fuzzy sets," *IEEE Transactions on Fuzzy Systems*, vol. 25, no. 5, pp. 1222–1230, 2017.

[12] R. R. Yager, N. Alajlan, and Y. Bazi, "Aspects of generalized orthopair fuzzy sets," *International Journal of Intelligent Systems*, vol. 33, no. 11, pp. 2154–2174, 2018.

[13] R. R. Yager and N. Alajlan, "Approximate reasoning with generalized orthopair fuzzy sets," *Information Fusion*, vol. 38, pp. 65–73, 2017.

[14] 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.

[15] P. Liu and J. Liu, "Someq-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.

[16] Y. Xing, R. Zhang, Z. Zhou, and J. Wang, "Some q-rung orthopair fuzzy point weighted aggregation operators for multi-attribute decision making," *Soft Computing*, vol. 23, no. 22, Article ID 11627, 2019.

[17] H. Garg, "A novel trigonometric operation-based q-rung orthopair fuzzy aggregation operator and its fundamental properties," *Neural Computing & Applications*, vol. 32, no. 18, 2020.

[18] H. Garg, "CN-q-ROFS: connection number-based q -rung orthopair fuzzy set and their application to decision-making process," *International Journal of Intelligent Systems*, vol. 36, no. 7, pp. 3106–3143, 2021.

[19] 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.

[20] R. Verma, "Multiple attribute group decision-making based on order-α divergence and entropy measures under q-rung orthopair fuzzy environment," *International Journal of Intelligent Systems*, vol. 35, no. 4, pp. 718–750, 2020.

[21] S. B. Aydemir and S. Y. Gunduz, "A novel approach to multi-attribute group decision making based on power neutrality aggregation operator for q -rung orthopair fuzzy sets," *International Journal of Intelligent Systems*, vol. 36, no. 3, pp. 1454–1481, 2021.

[22] Z. Liu, X. Wang, L. Li, X. Zhao, and P. Liu, "Q-rung orthopair fuzzy multiple attribute group decision-making method based on normalized bidirectional projection model and generalized knowledge-based entropy measure," *Journal of Ambient Intelligence and Humanized Computing*, vol. 12, no. 2, pp. 2715–2730, 2021.

[23] Z. Yang, L. Zhang, and T. Li, "Group decision making with incomplete interval-valued q-rung orthopair fuzzy preference relations," *International Journal of Intelligent Systems*, vol. 36, no. 12, pp. 7274–7308, 2021.

[24] H. Garg, "New exponential operation laws and operators for interval-valued q-rung orthopair fuzzy sets in group decision making process," *Neural Computing & Applications*, vol. 33, no. 20, Article ID 13937, 2021.

[25] H. Garg, "A new possibility degree measure for interval-valued q-rung orthopair fuzzy sets in decision-making," *International Journal of Intelligent Systems*, vol. 36, no. 1, pp. 526–557, 2021.

[26] G. Choquet, "Theory of capacities," *Annales de l'Institut Fourier*, vol. 5, pp. 131–295, 1954.

[27] J. L. Marichal, "An axiomatic approach of the discrete Choquet integral as a tool to aggregate interacting criteria," *IEEE Transactions on Fuzzy Systems*, vol. 8, no. 6, pp. 800–807, 2000.

[28] M. Sugeno, *Theory of Fuzzy Integral and its Application*, Tokyo Institute of Technology, Tokyo, Japan, 1974.

[29] T. Murofushi and M. Sugeno, "An interpretation of fuzzy measures and the Choquet integral as an integral with respect to a fuzzy measure," *Fuzzy Sets and Systems*, vol. 29, no. 2, pp. 201–227, 1989.

[30] C. Tan and X. Chen, "Intuitionistic fuzzy Choquet integral operator for multi-criteria decision making☆," *Expert Systems with Applications*, vol. 37, no. 1, pp. 149–157, 2010.

[31] C. Tan and X. Chen, "Induced Choquet ordered averaging operator and its application to group decision making," *International Journal of Intelligent Systems*, vol. 25, no. 1, pp. 59–82, 2010.

[32] C. Tan, "A multi-criteria interval-valued intuitionistic fuzzy group decision making with Choquet integral-based TOPSIS," *Expert Systems with Applications*, vol. 38, no. 4, pp. 3023–3033, 2011.

[33] Z. Xu, "Choquet integrals of weighted intuitionistic fuzzy information," *Information Sciences*, vol. 180, no. 5, pp. 726–736, 2010.

[34] M. Grabisch, "The application of fuzzy integrals in multi-criteria decision making," *European Journal of Operational Research*, vol. 89, no. 3, pp. 445–456, 1996.

[35] J. Wu, F. Chen, C. Nie, and Q. Zhang, "Intuitionistic fuzzy-valued Choquet integral and its application in multicriteria decision making," *Information Sciences*, vol. 222, pp. 509–527, 2013.

[36] Y. Xing, R. Zhang, X. Zhu, and K. Bai, "q-Rung orthopair fuzzy uncertain linguistic choquet integral operators and their application to multi-attribute decision making," *Journal of Intelligent and Fuzzy Systems*, vol. 37, no. 1, pp. 1123–1139, 2019.

[37] A. Keikha, H. Garg, and H. Mishmast Nehi, "An approach based on combining Choquet integral and TOPSIS methods to uncertain MAGDM problems," *Soft Computing*, vol. 25, no. 10, pp. 7181–7195, 2021.

[38] F. Teng and P. Liu, "A large group decision-making method based on a generalized Shapley probabilistic linguistic Choquet average operator and the TODIM method," *Computers & Industrial Engineering*, vol. 151, no. 1, Article ID 106971, 2021.

[39] B. P. Joshi, A. Singh, P. K. Bhatt, and K. S. Vaisla, "Interval valued q-rung orthopair fuzzy sets and their properties," *Journal of Intelligent and Fuzzy Systems*, vol. 35, no. 5, pp. 5225–5230, 2018.

[40] Q. Wang and H. Sun, "Interval-valued intuitionistic fuzzy Einstein geometric choquet integral operator and its application to multiattribute group decision-making," *Mathematical Problems in Engineering*, vol. 2018, pp. 1–11, 2018.

[41] G. Srivastava, J. C. W. Lin, D. Pamucar, and S. Kotsiantis, "Editorial: applications of fuzzy systems in data science and big data," *IEEE Transactions on Fuzzy Systems*, vol. 29, no. 1, pp. 1–3, 2021.

[42] B. T. Wan, J. L. Huang, and X. Chen, "Interval-valued Q-Rung Orthopair Fuzzy Choquet Integral Operators and its Application in Group Decision Making," 2021, https://arxiv.org/abs/2111.15108.

Hindawi
Mathematical Problems in Engineering
Volume 2022, Article ID 2617640, 12 pages
https://doi.org/10.1155/2022/2617640

Research Article

Dynamic Design Quality Evaluation of Power Enterprise Digital System Based on Fuzzy Information Axiom

Xinping Wu,[1] Xinzhou Geng,[1] Zhiyi Chen,[1] Aidi Dong,[2] and Jinchao Li[3]

[1]*State Grid Economic and Technology Research Institute Co., Ltd., Beijing 102209, China*
[2]*State Grid Jilin Electric Power Co., Ltd., Changchun 130012, China*
[3]*School of Economics and Management, North China Electric Power University, Beijing 102206, China*

Correspondence should be addressed to Jinchao Li; lijc@ncepu.edu.cn

Received 1 March 2022; Revised 21 June 2022; Accepted 12 July 2022; Published 18 August 2022

Academic Editor: Weifeng Pan

With the deepening of digital transformation and upgrading of power grid enterprises, the digital system evaluation method of power grid enterprises based on experts' subjective experience has been unable to meet the management needs of modern enterprises. In this paper, a method based on fuzzy information axiom for dynamic design quality evaluation of digital system in electric power enterprises is proposed. Firstly, the electric power enterprise digital system dynamic design quality comprehensive evaluation index system is set up from three aspects, which are achievement degree of target business function, logical relation rationality, and technical economy of physical model. Secondly, the quantitative and qualitative index values are processed by using the information calculation formula of minimum information axiom and fuzzy membership function. And then best-worst method and antientropy weight method are used to form the comprehensive evaluation model. Finally, the feasibility and effectiveness of the design scheme are verified by an example of dynamic design of digital system in power enterprise.

1. Introduction

With the vigorous development of the digital economy and our country's vigorous promotion of the "Internet +" and green energy revolution development strategies, as well as the change of global market players, power companies are making full use of cloud computing, big data, mobile Internet, artificial intelligence, Internet of Things, and other information technologies, actively exploring the energy Internet, green energy substitution, and assisting the development the model of the digital strategic transformation of power companies. In 2020, China's power grid enterprises have increased their investment in the direction of digital transformation, and increased their investment in power grid digital platform, energy big data center, power big data application, power Internet of things, etc. in 2020, China's digital power grid investment exceeded 110 billion yuan, and the investment scale is expected to reach 158 billion yuan by 2025 [1]. At present, facing the technical and economic evaluation needs of multiscene and complex application, multidimensional comprehensive evaluation and quantitative control dynamic evaluation of digital construction, the traditional postevaluation method based on expert subjective evaluation, and manual method cannot meet the management requirements of modern enterprises and meet the needs of digital development under the new situation. Based on this background, this paper establishes the electric power enterprise digital system dynamic design quality comprehensive evaluation index system, and evaluates the digital design schemes based on comprehensive evaluation model of subjective and objective combination.

The paper is organized as follows. The second part serves as a literature review. The third part introduces the establishment of the index system. The fourth part sets up the evaluation model of digital system dynamic design quality. The fifth part analyzes the examples. Finally, the eighth part offers conclusion.

Table 1: Research on digital system design quality.

Author	Evaluation content	Indexes/dimensions	Method/model
Fang [1]	Design quality of software systems	Control area, coupling degree, degree of condensation	AHP method
Ma et al. [2]	Software quality	Operability, modifiability, and adaptability	Fuzzy theory
Enríquez et al. [3]	Software quality	Concepts, design, production, support/use, general	QuEF methodology
Corbin et al. [4]	Software design	Systems engineering, software development, test, quality assurance, configuration management, data management, process group	Empirical analysis
Jing et al. [5]	Software quality	Metering, communication, freezing, event recording, load curve, reliability	Analytic hierarchy process
Liu [6]	Software quality	Software operation, software modification, software transfer	Grey fixed weight clustering
Li et al. [7]	Software quality	Response time, database size, accuracy, language number, special clicks, dead links, update time and format	Fuzzy triangular number fuzzy neural network
Yue and Zhang [8]	Software quality	Maintainability, reliability, reusability	Improved TOPSIS method
Jianli et al. [9]	Software quality	Functionality, reliability, ease of use and portability	Hesitant fuzzy sets and multiattribute decision making
Yu et al. [10]	Software quality	Functionality, reliability, ease of use, efficiency, maintainability, portability	Generalized intuitionistic fuzzy hybrid weighted averaging
Zhou et al. [11]	Software quality	Functionality, reliability, ease of use, efficiency, maintainability, portability	Improved vague set method
Yue [12]	Software quality	Efficiency, reliability, functionality, maintainability	Entropy
Bao and Liu [13]	Software quality	General attributes, domain attributes, application attributes	Expert method
Akay et al. [14]	Conceptual design evaluation of adhesive tape dispenser	Not about software or digital system	Fuzzy information axiom

2. Literature Review

The quality of digital system design is the key to realizing the effective integration of the underlying driving energy technology and digital technology, prompting a comprehensive transformation of traditional power generation methods to new power systems, promoting the flexible configuration of energy supply and demand, and driving the steady growth of digital emerging strategic industries. To ensure the quality of digital system design, many scholars have carried out relevant research on design quality evaluation. The specific literature is shown in Table 1.

From Table 1 we can find that there are few studies on digital system design quality evaluation of power grid. Therefore, we summarize the software quality evaluation methods related to digital design. We find that AHP, Entropy, and Fuzzy evaluation are the main methods of comprehensive evaluation. The evaluation indexes are mainly set up from functions, defects, process, and performance. The above quality evaluation system cannot fully reflect the dynamic formation of design quality. Especially, the rapid iteration of the current power production mode, organization mode, power dispatch form, function positioning, etc., makes the dynamic characteristics of the related power digital system in the design link more significant, that is, while the power enterprise digital system is in the design process, based on the ever-changing internal and external demand information, the conceptual design, logical design, and functional design are dynamically adjusted, supplemented, and improved. For example, based on the software engineering system life cycle, the literature [15–17] defines the life cycle process of digital system construction, clarifies the phase characteristics and internal and external influence factors of the system construction process, and analyzes the impact of standardized design on digital systems, the importance of the construction stage, and its design quality evaluation must run through the whole process of dynamic design. The literature [18–21] stipulated and standardized the information architecture of electric power enterprises and strengthened the overall design of data element model (conceptual model, logical model, physical model), technical process (development activity), and visualization design requirements of the project process (enable process). It clarified the relationship between subject domains (business domain, application domain, data domain, and technology domain) and the dynamic path of data cross-domain reuse. However, the existing design quality evaluation methods fail to design the evaluation index system in the iterative process of dynamic design. Its specificity, the reliability, and practicability need to be further improved.

In summary, based on the software system engineering theory and enterprise architecture (TOGAF) [22] standard perspective, this paper takes data as the core evaluation element of the design quality of digital systems, takes the characteristics of the agile development cycle of digital systems as the dynamic evaluation mechanism, and introduces the information axiom [23, 24] in the modernized design theory, provides a new solution path for the multi-attribute decision-making problem of complex systems by

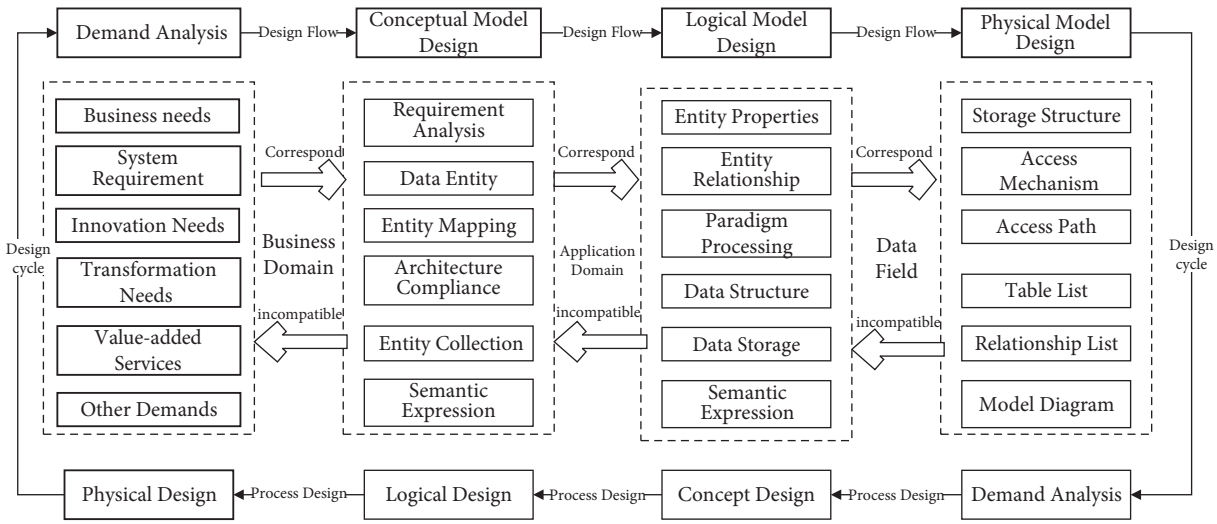


Figure 1: Characteristics of digital system dynamic design flow.

integrating the size of the information of each index, and completes the overall evaluation of the quality of the dynamic design of the digital system of the electric power enterprise. At the same time, due to the network characteristics, scale characteristics, functional diversity, and other characteristics of the digital system of electric power enterprises, the system has complexity and uncertainty in the design process, which causes a certain degree of human interference and inaccuracy in the design index evaluation information. In the process of digital project design quality evaluation, some indicators are difficult to quantify. Fuzzy evaluation is an effective way to solve this problem. The fuzzy evaluation method was founded by Professor Zadeh, an American scientist, in the 1960s. It is an evaluation model and method designed for the fuzziness of a large number of economic phenomena in reality. It has been constantly evolving by relevant experts in application practice. And it has been tested that which has the characteristics of clear results and strong systematicness. It can better solve the fuzzy and difficult to quantify problems and is suitable for solving all kinds of uncertain problems [25]. In this paper, how to eliminate the subjectivity of evaluation and give full play to the objective advantages of the fuzzy information axiom method is also the focus of this article.

3. Dynamic Design Quality Index System of Power Enterprise Digital System

3.1. The Dynamic Design Characteristics of the Digital System of the Electric Power Enterprise. The dynamic design of power enterprise digital system is based on the characteristics of digital system life cycle, and gives full play to the role of data as the core production factor. According to the strategic direction of enterprise digital development and the new operation and management mode, unify and solidify the entity, attribute and their relationship, introduce the standardized enterprise data model, strengthen the design of digital system data model, through periodic iteration, improve all business concepts and logical rules involved in the process of enterprise operation and management, and build an enterprise level digital system under the new power system. The dynamic design of the digital system of the electric power enterprise specifically includes three levels: conceptual data model design, logical data model design, and physical data model design. They rely on and interact with each other, with design modularity, cross-domain closed-loop reuse of data, and multilevel design collaboration features. Characteristics of dynamic design process of digital system are shown in Figure 1.

3.2. Dynamic Design Quality Evaluation Index System. According to the characteristics of dynamic design process in each stage of digital system, this paper establishes the quality evaluation index system of digital system dynamic design of power grid enterprises, as shown in Figure 2.

3.2.1. Quality Indexes and Measurement Method of Conceptual Design. Conceptual design is mainly used to describe the conceptual structure of things, including subject domain, entity, object, class, domain, generalization, aggregation, combination, dependency, and other specific design contents. Conceptual design is the communication bridge between requirements analysts and database designers. The evaluation indicators of conceptual design quality mainly involve business compliance, business architecture compliance, CIM model compliance, and data entity accuracy. The specific indicators are as follows.

(1) Compliance with Business Requirements. Mapping business requirements to business purposes is one task of conceptual design, which formulates business goals to meet certain user behaviors. This indicator is used to measure the degree to which the conceptual data model satisfies the business information required by the user's behavior. The calculation formula for the compliance degree of the business requirements is as follows:

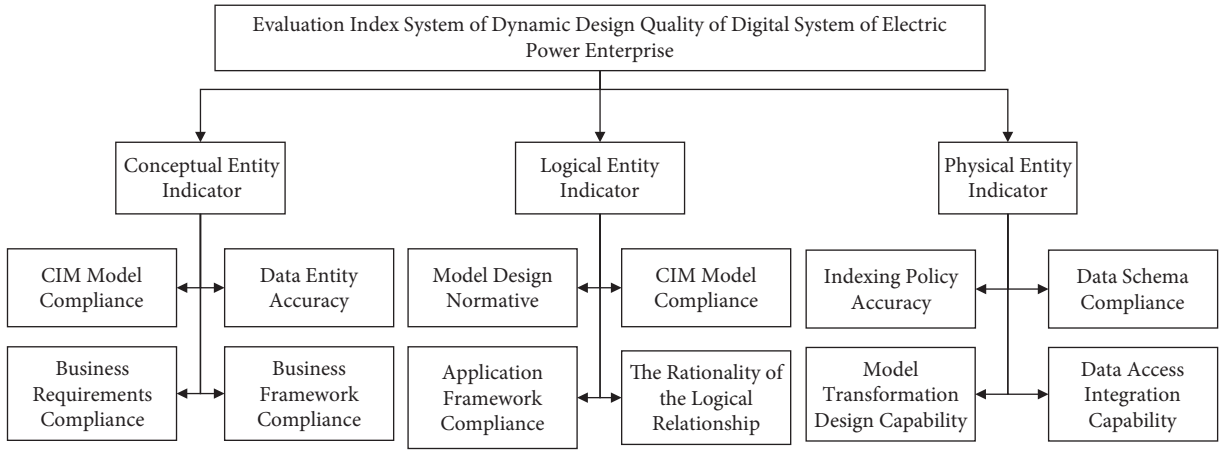


Figure 2: Digital system dynamic design quality evaluation index system.

$$C_1 = \left(\frac{1}{N_1} \sum_{i=1}^{N_1} C_i \right) \times 100\%, \tag{1}$$

where C_1 is the degree of compliance with business requirements; N_1 is the number of business requirements in the digital system; C_i is the state of demand satisfaction. If the demand is met in the digital system, $C_i = 1$, if partially met, $C_i < 1$, if not met, $C_i = 0$.

(2) Business Architecture Compliance. Measure whether the conceptual description of the name, attributes, and model relationships of the conceptual data model is clear and readable, that is, concisely reflect the semantics of the object, reflect the relationship and position of the object in the system, and facilitate the comparison with the business architecture design standards, so as to judge the compliance of the standardized design module. The calculation formula is as follows:

$$C_2 = \left(\frac{1}{N_{F2}} \sum_{i=1}^{N_{F2}} C_{Fi} + \frac{1}{N_{D2}} \sum_{i=1}^{N_{D2}} C_{Di} \right) / 2 \times 100\%, \tag{2}$$

where C_2 is the concept expressing business architecture compliance; N_{F2} is the number of key business activities; C_{Fi} is the business capability comment status, if there is a comment, then $C_{Fi} = 1$, otherwise $C_{Fi} = 0$; N_{D2} represents the number of business processes, C_{Di} represents the legibility of object naming, if the legibility is met, then $C_{Di} = 1$, otherwise $C_{Di} = 0$.

(3) CIM Model Compliance. Measure the percentage of the conceptual data model contained in a digital system that belongs to the unified data model of the enterprise. The calculation formula is as follows:

$$C_3 = \left(\frac{1}{N_3} \sum_{i=1}^{N_3} C_{3i} \right) \times 100\%, \tag{3}$$

where C_3 is the standardized module adoption rate; N_3 is the number of modules in the digital system; C_{3i} is the adoption of standardized modules, if standardized modules are used then $C_{3i} = 1$, otherwise $C_{3i} = 0$.

(4) Accuracy of Data Entities. Calculate the accuracy of business subject analysis of power enterprise digital system in the conceptual design stage. The calculation formula is as follows:

$$C_4 = \left(\frac{1}{N_4} \sum_{i=1}^{N_4} C_{4i} \right) \times 100\%, \tag{4}$$

where C_4 is the accuracy of the data entity; N_4 is the number of topics; C_i is the accuracy of the topic domain, if the topic domain is accurate then $C_{4i} = 1$, otherwise $C_{4i} = 0$.

3.2.2. Logical Design Quality Indicators and Measurement Methods. Logic design is the further decomposition and refinement of conceptual design, describing entities, attributes and entity relationships, and mainly solving detailed business problems. The design generally follows the "third paradigm," which makes the logic of the digital system clearer, enhances compatibility, and reduces partial iterative update difficulties. The specific evaluation indicators are as follows.

(1) Normative Design Model. The standardization of the logic structure of power enterprise digital system is to overcome the problems of redundancy and abnormality in the logic structure. It is evaluated in terms of the number of redundant points, the number of abnormal points, the degree of satisfaction of the "third normal form," and the degree of compliance with the unified model of the digital system. The calculation formula is as follows:

$$L_1 = \left(1 - \left(\frac{1}{M_1} \sum_{i=1}^{M_1} r_i + \frac{1}{M_1} \sum_{i=1}^{M_1} A_i + \frac{1}{M_1} \sum_{i=1}^{M_1} T_i + \frac{1}{M_1} \sum_{i=1}^{M_1} u_i \right) / 4 \right) \times 100\%, \tag{5}$$

where M_1 represents the number of logical structure points, r_i represents the number of redundancy points, A_i represents the number of anomaly points, T_i represents the degree to which the "third paradigm" is satisfied, and u_i represents the unified model of the digital system.

(2) Application Architecture Compliance. The object-oriented design method is adopted to measure the accuracy of the entities, attributes and entity relationships of the logical data model in complying with the application architecture design standards, and to assess whether the semantics of the object and the description of the relationship and position of the object in the system are clear and readable, so as to judge the compliance of the standardized design module. The calculation formula is as follows:

$$L_2 = \left(\frac{1}{N_{F2}} \sum_{i=1}^{N_{F2}} C_{Fi} + \frac{1}{N_{D2}} \sum_{i=1}^{N_{D2}} C_{Di} \right) / 2 \times 100\%, \quad (6)$$

where L_2 is the logical expression of application architecture compliance; N_{F2} is the number of key applications; C_{Fi} is the comment status of the application, if there is a comment $C_{Fi} = 1$, otherwise $C_{Fi} = 0$; N_{D2} represents the number of associated applications, C_{Di} means the legibility of the object name, if the legibility is met, then $C_{Di} = 1$, otherwise $C_{Di} = 0$.

(3) Adoption Rate of Standardized Modules. Measure the percentage of the logical data model contained in a digital system that belongs to the unified data model of the enterprise. The calculation formula is as follows:

$$L_3 = \left(\frac{1}{N_{L3}} \sum_{i=1}^{N_{L3}} C_{L3i} \right) \times 100\%, \quad (7)$$

where L_3 is the adoption rate of standardized modules; N_{L3} is the number of digital system modules; C_{L3i} is the case of using standardized modules, if standardized modules are used then $C_{L3i} = 1$, otherwise $C_{L3i} = 0$.

(4) Rationality of Logical Relationship. The rationality of business logic relationship refers to the clear hierarchical structure and smooth transmission path between digital system businesses. The calculation formula is as follows:

$$L_4 = \left(\frac{1}{M_H} \sum_{i=1}^{M_H} H_i + \frac{1}{M_R} \sum_{i=1}^{M_R} R_i \right) / 2 \times 100\%, \quad (8)$$

where M_H represents the number of levels, H_i means the degree of clarity of each level, if it is clear, $H_i = 1$, otherwise $H_i = 0$; M_R means the number of paths, R_i means the smoothness of the path, if there is no obstruction, $R_i = 1$, otherwise $R_i = 0$.

3.2.3. Physical Design Quality Indicators and Measurement Methods. Physical design is based on logical design, taking into account various specific technical realization factors, and designing digital system structure to provide the most detailed design for the digital system development. The quality of the design at this stage is determined by the system transformation capability, data processing capability, data application analysis capability, model stability, and data reuse degree.

(1) Model Conversion Design Ability. To measure the ability to be automatically transformed into a physical data model through a logical data model, this indicator is measured by the development experience of each logical module.

$$P_1 = \frac{1}{K_1} \sum_{i}^{K_1} TA_i, \quad (9)$$

where P_1 represents the model conversion design capability, K_1 represents the number of function points of the physical data model, and TA_i means the automatic function point conversion capability. If there is similar development experience, then $TA_i = 1$, otherwise $TA_i = 0$.

(2) Data Structure Compliance. According to the basic elements required for the initial design and the relationship between related elements, measure the accuracy of the storage structure, record the sequence and access mechanism of the physical data model in compliance with the data architecture design standards, so as to determine the compliance of the standardized design module, and the formula is as follows:

$$P_2 = \left(\frac{1}{N_{P2}} \sum_{i=1}^{N_{P2}} C_{Pi} + \frac{1}{N_{S2}} \sum_{i=1}^{N_{S2}} C_{Si} \right) / 2 \times 100\%, \quad (10)$$

where P_2 is the compliance degree of the physical representation data structure; N_{P2} is the number of physical data models; C_{Pi} is the comment situation, if there is comment $C_{Pi} = 1$; otherwise $C_{Pi} = 0$; N_{S2} means the number of standard physical data models used, C_{Si} means the integrity of the physical data model used, if it satisfies $C_{Si} = 1$, otherwise $C_{Si} = 0$.

(3) Data Access Integration Capability. This indicator reflects the ability of digital systems to integrate different data sources, supporting both traditional data and big data platforms and supporting both structured data and unstructured data access. The specific calculation formula is as follows:

$$P_3 = \frac{1}{K_3} \sum_{i}^{K_3} DA_i, \quad (11)$$

where P_3 represents the data access integration capability, K_3 represents the type of data, DA_i represents the data access integration capability. If it can be accessed effectively, $DA_i = 1$, otherwise $DA_i = 0$.

(4) Accuracy of Indexing Strategy. Measure the physical design stage to improve the data access speed of database tables by creating indexes, and improve the accuracy of queries through strategy settings and algorithms.

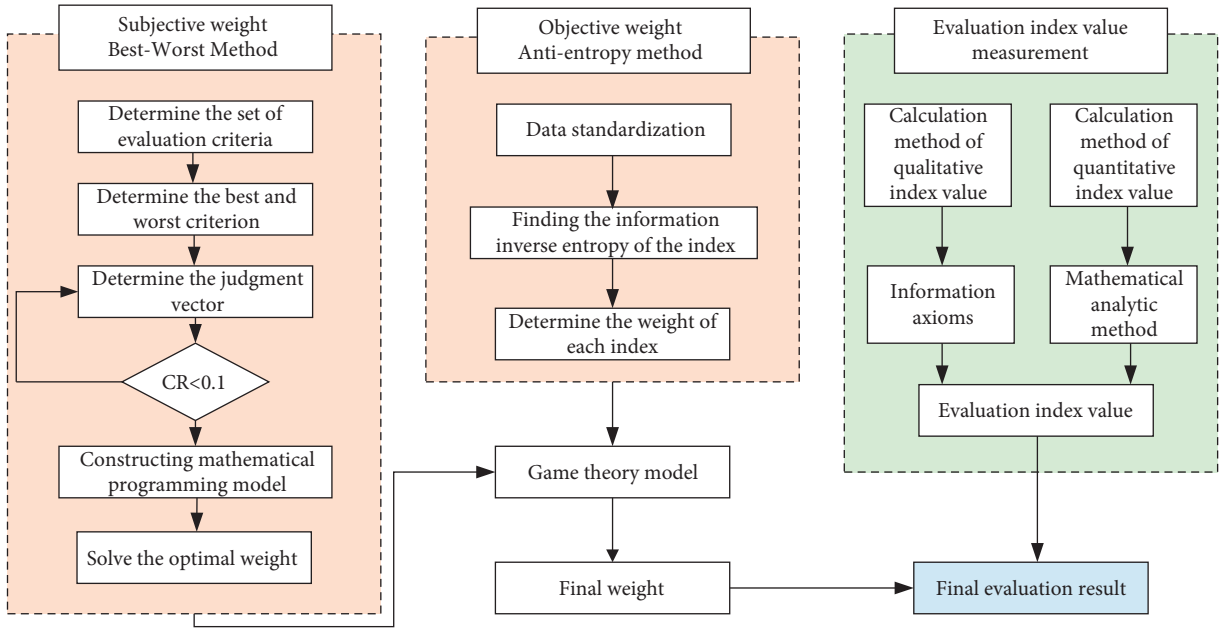


Figure 3: Digital system dynamic design quality evaluation model roadmap.

4. Dynamic Design Quality Evaluation Model of Power Enterprise Digital System

In the dynamic design process of digital system of power grid enterprises, digital experts play an important role in determining the index weight and estimating the index value. In this paper, BWM method is introduced to determine the subjective weight of evaluation indexes, and the antientropy weight method is selected to overcome the influence of inconsistent opinions of evaluation experts on the evaluation results of design quality. Information axiom method can more systematically and scientifically obtain the evaluation value of evaluation experts on indicators. The specific technical route of the evaluation model is shown in Figure 3.

4.1. Calculation Method of Index Weight

4.1.1. BWM Method. The BWM method was proposed by Rezaei in 2015 [26], according to the experience and actual needs of the project, the expert (decision-maker) selects the criteria other than the best (most important) and the worst (least important), and then compares the optimal criteria with other indicators in turn, and the other indicators with the worst criteria in turn. After comparing each indicator with the best (worst) indicator, an integer value of 1~9 reflecting the relative advantages and disadvantages will be formed. Finally, the BWM solution can be transformed into a mathematical programming problem, and the result calculation can be realized by lingo software. The specific operation steps are as follows [27–30].

(1) Determine the set of evaluation criteria

Experts (decision makers) discuss and determine the influencing factors of multicriteria decision-making problems to be studied, and then determine the set of evaluation criteria $\{c_1, c_2, \cdots, c_n\}$.

(2) Determine the best criterion and the worst criterion

In the criterion set $\{c_1, c_2, \cdots, c_n\}$, the optimal criterion C_B and the worst criterion C_W are determined. The optimal criterion is the relatively most important criterion determined by experts (decision makers) according to their experience, cognition, and actual needs of engineering, which has the most prominent impact on the decision-making results; Similarly, the worst criterion is the criterion that is relatively least important and has the least impact on the decision-making results. If experts (decision makers) believe that there is more than one optimal (worst) criterion, they can choose one of these optimal (worst) indexes without affecting the calculation results.

(3) Compare the preference between the optimal criteria and all criteria, and construct the judgment vector $A_B = (a_{B1}, a_{B2}, \cdots, a_{Bn})$.

The expert (decision maker) compares the optimal criteria with other criteria one by one, determines the preference degree of the optimal criteria relative to other criteria, scores its preference degree with 1–9, and successively constructs the comparison vector $A_B = (a_{B1}, a_{B2}, \cdots, a_{Bn})$ based on the optimal criteria, where a_{Bi} represents the preference degree of the optimal criteria B compared with the criteria i, and it is easy to know that $A_{BB} = 1$. It should be noted that the scale scoring method of 1–9 actually has the same meaning as that of AHP. The degree and meaning of each scale are very similar. The biggest difference is that the comparison vector set constructed by BWM with 1–9 scale is an integer, while the judgment matrix constructed by AHP with 1–9 scale is

composed of fraction. The preference scoring table of experts (decision makers) based on the optimal criteria is shown in Table 2.

(4) Compare the preference between all criteria and the worst criterion, and construct the judgment vector $A_W = (a_{1W}, a_{2W}, \cdots, a_{nW})$.

The expert (decision maker) compares the worst criteria with other criteria one by one, determines the preference degree of the worst criteria relative to other criteria, scores its preference degree with 1–9, and successively constructs a comparison vector $\mathrm{A_W} = (\mathrm{a_{1W}}, \mathrm{a_{2W}}, \cdots, \mathrm{a_{nW}})^T$ based on the worst criteria, where a represents the preference degree of criterion i compared with the worst criterion W, and it is easy to know that $A_{WW} = 1$. The preference scoring table of experts (decision makers) based on the worst criteria is shown in Table 3.

(5) Constructing mathematical programming problem, solve the optimal weight $(w_1^*, w_2^*, \cdots, w_n^*)$.

Criterion preference comparison is the comparison of criterion weights, so the optimal weight should meet the following conditions: the weight W of any criterion j is $\frac{j}{w}$.

$$\frac{w_B}{w_j} = a_{Bj}, \frac{w_j}{w_w} = a_{jw}. \tag{12}$$

Therefore, in order to determine the optimal weight, the following mathematical programming problem can be constructed:

$$\begin{aligned} &\min \max_j \left\{ \left| \frac{w_B}{w_j} - a_{Bj} \right|, \left| \frac{w_j}{w_w} - a_{jw} \right| \right\} \\ &\text{s.t.} \\ &\sum_j w_j = 1 \\ &w_j \geq 0. \end{aligned} \tag{13}$$

where the objective function is to minimize the largest one of $|W_B/W_j - a_{Bj}|$ and $|W_j/W_w - a_{jw}|$ among all j.

For the convenience of solution, mathematical programming can be transformed into the following problems:

$$\begin{aligned} &\min \xi \\ &\text{s.t.} \\ &\left| \frac{w_B}{w_j} - a_{Bj} \right| \leq \xi, \text{for all } j \\ &\left| \frac{w_j}{w_w} - a_{jw} \right| \leq \xi, \text{for all } j \\ &\sum_j w_j = 1 \\ &w_j \geq 0, \text{for all } j. \end{aligned} \tag{14}$$

Table 2: Optimal preference scoring criteria.

	c_1	c_2	$\cdots$	c_n
C_B	a_{B1}	a_{B2}	$\cdots$	a_{Bn}

Table 3: Worst preference scoring criteria.

	c_1	c_2	$\cdots$	c_n
C_W	a_{1W}	a_{2W}	$\cdots$	a_{nW}

The optimal weight $W_1(w_1^*, w_2^*, \cdots, w_n^*)$ based on BWM can be obtained by solving the mathematical programming.

4.1.2. Antientropy Method. According to the basic principle of information theory, information is a measure of the degree of system order, and entropy is a measure of the degree of system disorder. If the system may be in many different states, and the probability of each state is $p_i (i = 1, 2, \ldots, m)$, the entropy of the system is defined as

$$E = -\sum_{i=1}^{n} P_i ln P_i, \tag{15}$$

where $0 \leq P_i \leq 1$; $\sum_{i=1}^{n} P_i = 1$.

In the digital project evaluation of power grid enterprises, the closer the evaluation experts score on a certain index, it shows that the score consistency is strong, and the index should be given a high weight, otherwise a low weight. Therefore, the antientropy weight method is used to calculate the index weight. This method is based on the idea that the greater the difference of the index, the greater the antientropy. The antientropy calculation formulas are constructed, as shown in the following.

$$\begin{aligned} h_i' &= -\sum_{i=1}^{n} P_i ln(1 - P_i) \\ w_i' &= h_i' / \sum_{i=1}^{n} h_i', \end{aligned} \tag{16}$$

where h_i' express the antientropy of the index i, w_i' express the weight of the index i, and so we can get the weights W_2 of the indexes system.

4.1.3. Combined Weighting Method Based on Game Theory. The idea of game theory is used to seek agreement or compromise between weights W_1 and W_2 in order to obtain the most satisfactory weight. Let the comprehensive weight value composed of any linear combination of W_1 and W_2 be W. The calculation formula is as follows:

$$W = \lambda_1 W_1^T + \lambda_2 W_2^T, \tag{17}$$

where λ_1 and λ_2 are the linear combination coefficients, which all over zero.

(1) Determine the objective function

Through the above formula, the problem of combined weighting is transformed into the change of coefficients λ, so as to minimize the range between the comprehensive weight W and the weights W_1 and W_2. The calculation formula is as follows:

$$\min\left(\left\|W^T - W_1^T\right\|^2 + \left\|W^T - W_2^T\right\|^2\right) = \min\left(\left\|\lambda_1 W_1^T + \lambda_2 W_2^T - W_1^T\right\|^2 + \left\|\lambda_1 W_1^T + \lambda_2 W_2^T - W_2^T\right\|^2\right). \tag{18}$$

(2) Solving objective function

According to the properties of differential matrix, the first derivative condition of formula (18) optimization is

$$\begin{aligned} \lambda_1 W_1 W_1^T + \lambda_2 W_1 W_2^T &= W_1 W_1^T, \\ \lambda_1 W_2 W_1^T + \lambda_2 W_2 W_2^T &= W_2 W_2^T. \end{aligned} \tag{19}$$

(3) Calculate weight combination coefficient

The linear coefficient obtained after normalization is processed as follows:

$$\lambda_k^* = \frac{\lambda_k}{\lambda_1 + \lambda_2}, \quad k = 1, 2. \tag{20}$$

Final combination weights of the indexes system are $\overset{*}{W}$, which are calculated by the following formula:

$$W^* = \lambda_1^* W_1^T + \lambda_2^* W_2^T. \tag{21}$$

4.2. Index Value Measurement Method

4.2.1. The Establishment of Information Axioms. The information axiom method can obtain the evaluation value of the evaluation experts in a more systematic and scientific manner. After the axioms of information are proposed, they are widely used in the fields of design plan evaluation, advanced manufacturing system selection, and control decision-making. The basic idea of this method is that the overall amount of information is the smallest is the best.

(1) Conversion of Language Forms. Use fuzzy mathematics to convert qualitative language phrase descriptions into quantitative values. The conversion formula is shown as follows:

$$\widetilde{l}_q = \left(l_p^1, l_p^2, l_p^3\right) = \left[\max\left\{\frac{q-1}{t}, 0\right\}, \frac{q}{t}, \min\left\{\frac{q+1}{t}, 1\right\}\right], \quad q \in \{0, 1, \cdots, t\}. \tag{22}$$

In the formula, $\widetilde{l}_q$ represents the triangular fuzzy number, l_p^1, l_p^2, l_p^3 means the language phrase description, q means the number of reviews, and t means the number of reviews.

In this article, n experts will be organized to evaluate the qualitative index i, and the evaluation index will be evaluated according to the comment set [Very good (VH), good (H), fair (M), poor (L), very poor (VL)]. For evaluation, n experts obtain the evaluation result sequence $[v_{i,1}, v_{i,2}, \cdots, v_{i,n}]$ after completing the evaluation, and then use the triangular membership function (as shown in Figure 4) to transform the fuzzy evaluation result into a numerical sequence $[d_{i,1}, d_{i,2}, \cdots, d_{i,n}]$.

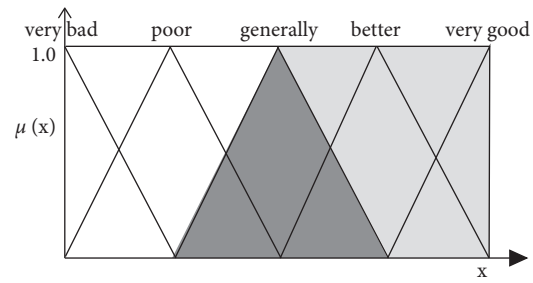


Figure 4: Range diagram of triangle membership function.

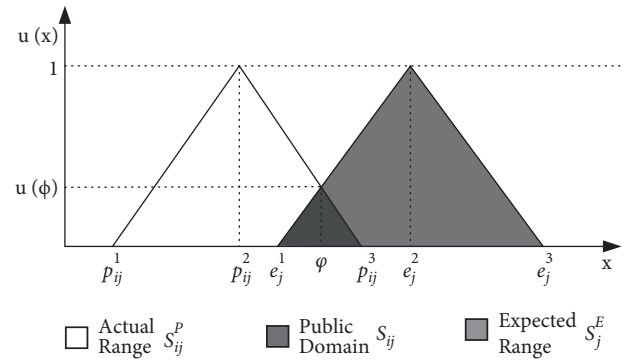


Figure 5: Extraction diagram of linguistic form index information.

According to formula (22), the membership degree for converting expert comment information into triangular fuzzy numbers $\widetilde{s}_j = (s_j^1, s_j^2, s_j^3)$ is defined as

$$u_j(x) = \begin{cases} \dfrac{x - s_j^1}{s_j^2 - s_j^1}, & s_j^1 \le x \le s_j^2, \\ \dfrac{s_j^3 - x}{s_j^3 - s_j^2}, & s_j^2 \le x \le s_j^3, \\ 0, & \text{else.} \end{cases} \tag{23}$$

The information value in the language form index is extracted through the intersection of the comments of the review experts, and the specific formula is as follows:

The S_j^E in Figure 5 shows the expected range, that is, the area enclosed by the expected information of the index $\widetilde{e}_j$ and the coordinate axis after the conversion. The expected range is expressed by the standard value of each index level.

$$S_j^E = \int_{-\infty}^{+\infty} u_j^E(x)dx = \int_{e_j^1}^{e_j^2} \frac{x - e_j^1}{e_j^2 - e_j^1}dx + \int_{e_j^2}^{e_j^3} \frac{e_j^3 - x}{e_j^3 - e_j^2}dx = \frac{e_j^3 - e_j^1}{2}. \tag{24}$$

In the same way, S_{ij}^{P} is the actual range, that is, the area enclosed by the converted evaluation information $\widetilde{p}_{ij}$ and the coordinate axis

$$S_{ij}^{P}=\int_{-\infty}^{+\infty}u_{ij}^{P}(x)dx=\int_{p_{ij}^{1}}^{p_{ij}^{2}}\frac{x-p_{ij}^{1}}{p_{ij}^{2}-p_{ij}^{1}}dx+\int_{p_{ij}^{2}}^{p_{ij}^{3}}\frac{p_{ij}^{3}-x}{p_{ij}^{3}-p_{ij}^{2}}dx=\frac{p_{ij}^{3}-p_{ij}^{1}}{2}. \tag{25}$$

where $i \in M$, $j \in N$, S_{ij} represents the common range, that is, the area S_j^E enclosed by the intersection of the expected range and the actual range S_{ij}^P, Here, when the index evaluation is higher than the decision maker's expectation, the public range is considered to be equal to the expected range. The calculation formula for the public range is as follows:

$$S_{ij}=\begin{cases}0, & p_{ij}^{3}\le e_j^{1},\\ \int_{e_j^{1}}^{\phi}\frac{x-e_j^{1}}{e_j^{2}-e_j^{1}}dx+\int_{\phi}^{p_{ij}^{3}}\frac{p_{ij}^{3}-x}{p_{ij}^{3}-p_{ij}^{2}}dx, & e_j^{1}\le p_{ij}^{3}\le e_j^{3},\\ S_j^{E}, & p_{ij}^{3}>e_j^{3},\end{cases} \tag{26}$$

where ϕ represents the mapping on the x axis of the intersection points on the boundary between the expected range S_j^E and the actual range S_{ij}^P

$$\phi=\frac{e_j^{2}p_{ij}^{3}-e_j^{1}p_{ij}^{2}}{e_j^{2}-e_j^{1}+p_{ij}^{3}-p_{ij}^{2}},\quad i\in M, j\in N. \tag{27}$$

Finally, j calculate the amount of information I_j contained in the indicator, which is the degree to which the indicator does not meet the expectations, and the calculation formula is as follows:

$$I_j=\begin{cases}\infty, & p_{ij}^{3}\le e_{ij}^{1},\\ \log_2\left(\frac{S_{ij}^{P}}{S_{ij}}\right), & e_{ij}^{1}<p_{ij}^{3}\le e_{ij}^{3},\\ 0, & e_{ij}^{3}<p_{ij}^{3},\end{cases} \tag{28}$$

According to formula (28), $I_j \in [0\ ,\ \infty)$ can be known. When $I_j = 0$, it means that the indicator is completely consistent with expectations. The larger the value, the greater the difference the experts expect. When $I_j = \infty$, the indicator has not reached expectations at all, and adjustments should be made.

(2) Index Information Extraction in the Form of Numerical Statistics. The amount of information (I) of indicators in the form of numerical statistics is calculated by the probability of meeting requirements and design standards. The specific formula is as follows:

Table 4: Design quality grade value of digital system.

Grade	Level 1	Level 2	Level 3
C1	0.7	0.8	0.95
C2	0.7	0.8	0.95
C3	0.7	0.8	0.95
C4	0.7	0.8	0.95
L1	0.7	0.8	0.95
L2	0.7	0.8	0.95
L3	0.7	0.8	0.95
L4	0.7	0.8	0.95
P1	0.7	0.8	0.95
P2	0.7	0.8	0.95
P3	0.7	0.8	0.95
P4	M	H	VH

$$I=-\log_2 p. \tag{29}$$

In the formula, p represents the probability of meeting system requirements and design standards. It is obtained by evaluation experts or statistical data processed in different ways.

4.2.2. Index Classification. With reference to the CMMI software quality management theory, combined with the life cycle characteristics of the software development process quality of power companies, according to the relevant design standards and engineering practices of the power company software engineering in the design phase, three quality levels and expectations are set for the software design quality as follows.

Level 1: The designed product has a basic standardized design process and quality control mechanism, and is based on and complies with a certain standardized model. Changes that occur can be tracked, and the accumulated experience can be used for the development of new projects.

Level 2: Design finished products based on and comply with enterprise-level information, architecture design specifications, and standardized models, which can effectively achieve quantitative quality control and can perform statistics and analysis on the accumulated data, but further refinement of the design is required to improve quality stability.

Level 3: On the basis of Level 2, the design product uses advanced theories and concepts and uses new technologies to continuously improve the design process. It can flexibly deploy and operate and maintain for the realization of business needs and subsequent changes that may occur to meet the needs of the enterprise and the overall layout requirements of the level information architecture.

The index values of each grade obtained through the expert scoring method are shown in Table 4.

4.3. Example Analysis. The digital system of electric power enterprises is the use of 5G, Internet of Things, cloud computing, big data analysis, artificial intelligence, and other emerging technologies to upgrade and transform traditional electric power enterprises. It is an effort to empower the

Table 5: Score table of digital system dynamic design quality evaluation index of provincial electric power enterprises.

Expert	C1	C2	C3	C4	L1	L2	L3	L4	P1	P2	P3	P4
E1	0.95	0.85	0.86	0.81	0.78	0.84	0.86	0.81	0.5	0.76	0.91	VH
E2	0.88	0.93	0.84	0.74	0.71	0.65	0.87	0.82	0.55	0.76	0.85	H
E3	0.99	0.93	0.78	0.67	0.76	0.84	0.8	0.81	0.51	0.78	0.89	M
E4	0.86	0.89	0.89	0.7	0.68	0.85	0.82	0.88	0.56	0.85	0.88	VH
E5	0.8	0.9	0.94	0.68	0.73	0.78	0.8	0.91	0.66	0.78	0.9	H
E6	0.87	0.86	0.85	0.65	0.67	0.85	0.75	0.93	0.66	0.81	0.88	VH
E7	0.9	0.86	0.8	0.7	0.68	0.73	0.89	0.87	0.61	0.85	0.93	H
E8	0.85	0.89	0.8	0.79	0.74	0.66	0.83	0.86	0.53	0.76	0.82	M
E9	0.85	0.93	0.84	0.74	0.73	0.8	0.8	0.82	0.59	0.76	0.89	VH
E10	0.91	0.9	0.87	0.65	0.63	0.7	0.87	0.83	0.65	0.86	0.89	VH
E11	0.8	0.89	0.78	0.71	0.72	0.84	0.76	0.9	0.66	0.77	0.95	H
E12	0.8	0.95	0.88	0.77	0.77	0.68	0.81	0.82	0.67	0.82	0.91	H
E13	0.93	0.9	0.86	0.84	0.64	0.8	0.89	0.83	0.58	0.85	0.83	VH
E14	1	0.94	0.88	0.69	0.66	0.78	0.84	0.86	0.6	0.85	0.81	VH
E15	0.89	0.91	0.85	0.85	0.75	0.84	0.82	0.84	0.51	0.79	0.89	VH
E16	0.83	0.89	0.93	0.78	0.76	0.82	0.88	0.88	0.7	0.79	0.83	H
E17	0.99	0.94	0.9	0.7	0.8	0.79	0.85	0.92	0.68	0.77	0.81	H
E18	0.95	0.91	0.85	0.69	0.63	0.66	0.88	0.87	0.55	0.85	0.87	M
E19	0.92	0.9	0.88	0.76	0.75	0.65	0.84	0.9	0.57	0.82	0.92	VH
E20	0.82	0.9	0.87	0.72	0.71	0.7	0.76	0.94	0.62	0.79	0.84	VH

Table 6: Information values of digital statistics indexes at different design quality levels.

Grade	First level	Level 2	Level 3
C1	0.000	0.000	2.000
C2	0.000	0.000	4.322
C3	0.000	0.152	∞
C4	0.515	2.737	∞
L1	0.621	4.322	∞
L2	0.415	1.152	∞
L3	0.000	0.234	∞
L4	0.000	0.000	∞
P1	0.322	1.737	∞
P2	0.000	1.152	∞
P3	0.000	0.000	4.322

Table 7: Transformation of language phrases and trigonometric fuzzy numbers.

Language phrase	Triangular fuzzy number
VL	[0.00, 0.00, 0.25]
L	[0.00, 0.25, 0.50]
M	[0.25, 0.50, 0.75]
H	[0.50, 0.75, 1.00]
VH	[0.75, 1.00, 1.00]

construction of Digital China and build an internationally leading energy Internet with Chinese characteristics. The digital engine built by the enterprise. This article takes the dynamic design of the digital system of the provincial power enterprise as the analysis object, and hires 20 industry experts to statistically score the 12 indicators of the dynamic design. The specific scoring results are shown in Table 5.

First, according to formula (19), the information value of each numerical statistical form index under different design quality grade standards is calculated, as shown in Table 6.

Table 8: Digital system dynamic design quality evaluation information based on triangular fuzzy numbers.

Expert	P4
E1	[0.75, 1.00, 1.00]
E2	[0.50, 0.75, 1.00]
E3	[0.25, 0.50, 0.75]
E4	[0.75, 1.00, 1.00]
E5	[0.50, 0.75, 1.00]
E6	[0.75, 1.00, 1.00]
E7	[0.50, 0.75, 1.00]
E8	[0.25, 0.50, 0.75]
E9	[0.75, 1.00, 1.00]
E10	[0.75, 1.00, 1.00]
E11	[0.50, 0.75, 1.00]
E12	[0.50, 0.75, 1.00]
E13	[0.75, 1.00, 1.00]
E14	[0.75, 1.00, 1.00]
E15	[0.75, 1.00, 1.00]
E16	[0.50, 0.75, 1.00]
E17	[0.50, 0.75, 1.00]
E18	[0.25, 0.50, 0.75]
E19	[0.75, 1.00, 1.00]
E20	[0.75, 1.00, 1.00]

The five-level language form is transformed into triangular fuzzy numbers and the results are shown in Table 7.

The comment of index P4 is converted based on the triangular fuzzy number and the conversion result is shown in Table 8.

According to the design quality grade division value, the expected values of different grades of the index are calculated using formulas (14)–(18). The specific results are shown in Table 9.

The average value of the information value of the index at each level is calculated and the specific results are shown in Table 10.

Table 9: Evaluation index information value.

Expert	First level	Level 2	Level 3
E1	0	0	0
E2	0	0	2
E3	0	2	∞
E4	0	0	0
E5	0	0	2
E6	0	0	0
E7	0	0	2
E8	0	2	∞
E9	0	0	0
E10	0	0	0
E11	0	0	2
E12	0	0	2
E13	0	0	0
E14	0	0	0
E15	0	0	0
E16	0	0	2
E17	0	0	2
E18	0	2	∞
E19	0	0	0
E20	0	0	0

Table 10: Evaluation index P4' information values at different levels.

	P4 under 1st level	P4 under 2nd level	P4 under 3rd level
Information value	0	0.3	∞

Table 11: Index weight values.

Index	BWM	Antientropy	Combined weights
C1	0.363	0.083	0.223
C2	0.141	0.084	0.113
C3	0.106	0.084	0.095
C4	0.050	0.083	0.067
L1	0.041	0.082	0.061
L2	0.019	0.082	0.050
L3	0.033	0.084	0.058
L4	0.148	0.084	0.116
P1	0.017	0.084	0.051
P2	0.049	0.084	0.067
P3	0.010	0.083	0.047
P4	0.023	0.083	0.053

Using the BWM, antientropy method, and Game Theory to calculate the weight of each evaluation index is shown in Table 11.

Through calculation, the dynamic design quality of the power company's digital system is calculated to be 0.1096, 0.7143, and ∞, respectively. The probability of the system design quality reaching the first level is 92.69%, and the probability of reaching the second level is 60.95%. It does not meet the third level at all. Based on the comprehensive judgment, the design system is still at the first-level design level, and some design indicators have reached the second-level level.

5. Conclusion

Digital transformation is a new system engineering faced by enterprises. The digital construction of power grid enterprises is still in the stage of exploration and running in with the actual operation of enterprise organizational structure, digital system construction process, and technological innovation capability. From the perspective of software system engineering and power enterprise architecture standards, this paper integrates interdisciplinary, interdisciplinary and interdisciplinary technology and management knowledge theory systems, makes a structural analysis on the quality of digital cycle characteristics, model construction, design process and dynamic design in the process of power enterprise digital transformation, and puts forward a quality evaluation method of power enterprise digital system dynamic design based on fuzzy information axiom.

The index system in the evaluation method covers three aspects: business function achievement degree, logical relationship rationality, and physical model technical economy, with a total of twelve specific indicators. The index value measurement methods are mainly analytical method and fuzzy information axiom. The evaluation index division standard is set by the digital evaluation experts of power grid enterprises according to their work experience.

The weight calculation method in the evaluation method is obtained by using the game theory to deal with the subjective weight obtained by BWM and the objective weight obtained by the antientropy weight method. The combined weight calculation method realizes the principle of combining subjective and objective and helps to improve the reliability of the evaluation results.

Combined with the characteristics and current situation of the digital transformation of power grid enterprises, this paper carries out a numerical example analysis of the design quality evaluation of the digital system of power grid enterprises and proves the feasibility of the model.

In the future, with the further deepening of the transformation and development of power grid enterprises to energy Internet enterprises, the evaluation indicators will be further deepened. The indicator system will fully reflect the digital development trend of power grid enterprises and truly help the steady and healthy development of new power systems under the energy Internet.

Data Availability

This article study the dynamic design of the digital system of the provincial power enterprise as the analysis object and hires 20 industry experts to statistically score the 12 indicators of the dynamic design.

Conflicts of Interest

The authors declare no conflicts of interest.

Authors' Contributions

Xinping Wu contributed to the conception and design. Xinzhou Geng and Zhiyi Chen contributed to the index

system design. Aidi Dong collected and interpreted the data. Jinchao Li contributed to the evaluation method design and computation. All authors drafted and revised the manuscript together and approved its final publication.

Acknowledgments

This work was supported by the State Grid Corporation Headquarters Science and Technology Project "Research on Enterprise Digital Technology and Economic Evaluation System and Technical Tools" (5700-202129180A-0-0-00). The formulation of the calculation example in this paper and the measurement and recording of the verification data were completed with the support of the staff of State Grid Jilin Electric Power Company and State Grid Electric Power Academy Co., Ltd., and I would like to express my heartfelt thanks to them.

References

[1] M. Fang, "Application of analytic hierarchy process in software system design quality evaluation [J]," *Mathematical Statistics and Management*, vol. 1997, no. 06, pp. 11–14, 1997.

[2] S. Ma, H. Wang, and Y. Sun, "Study and application of software quality evaluation based on fuzzy theory," *Journal of Dalian Jiaotong University*, vol. 2007, no. 04, pp. 18–21+36, 2007.

[3] J. G. Enríquez, J. M. Sánchez-Begines, F. J. Domínguez-Mayo, J. A. García-García, and M. J. Escalona, "An approach to characterize and evaluate the quality of product lifecycle management software systems," *Computer Standards & Interfaces*, vol. 61, pp. 77–88, 2019.

[4] R. D. Corbin, C. B. Dunbar, and Q. Zhu, "A three-tier knowledge management scheme for software engineering support and innovation," *Journal of Systems and Software*, vol. 80, no. 9, pp. 1494–1505, 2007.

[5] J. Jing, X. Hou, H. Chen, X. Ou, F. Zhao, and K. Sun, "Software quality evaluation of smart electric energy meters based on analytic hierarchy process," *Electrical Measurement and Instrumentation*, vol. 52, no. 08, pp. 5–9, 2015.

[6] Y. Liu, "Research on software quality evaluation model based on grey fixed weight clustering[J]," *Journal of Southwest University (Natural Science Edition)*, vol. 2008, no. 05, pp. 177–180, 2008.

[7] K. Li, Y. Zhang, J. Ma, and H. Liu, "Software quality evaluation method based on fuzzy triangular number fuzzy neural network," *Computer Engineering and Science*, vol. 36, no. 07, pp. 1301–1306, 2014.

[8] C. Yue and J. Zhang, "Software quality evaluation model based on group decision and projection measurement," *Computer Applications*, vol. 40, no. 01, pp. 218–226, 2020.

[9] Y. Jianli, J. Lu, and H. Chenn, "Software quality evaluation model based on hesitant fuzzy sets," *Computer Engineering and Science*, vol. 42, no. 05, pp. 819–824, 2020.

[10] J. Yu, J. Lu, and H. Chen, "Software quality evaluation model based on intuitionistic fuzzy sets," *Science Technology and Engineering*, vol. 20, no. 27, pp. 11180–11184, 2020.

[11] L. Zhou, J. Xiao, and G. Wang, "Research on software quality evaluation of an improved vague set method," *Mathematics in Practice and Knowledge*, vol. 47, no. 22, pp. 36–45, 2017.

[12] C. Yue, "Projection-based approach to group decision-making with hybrid information representations and application to software quality evaluation," *Computers & Industrial Engineering*, vol. 132, pp. 98–113, 2019.

[13] T. Bao and S. Liu, "Quality evaluation and analysis for domain software: application to management information system of power plant," *Information and Software Technology*, vol. 78, pp. 53–65, 2016.

[14] D. Akay, O. Kulak, and B. Henson, "Conceptual design evaluation using interval type-2 fuzzy information axiom," *Computers in Industry*, vol. 62, no. 2, pp. 138–146, 2011.

[15] International System Engineering Association, *System Engineering Handbook-System Life Cycle Process and Activity Guide*, Vol. 04, Mechanical Industry Press, , Beijing, 2017.

[16] Gb/T 22032-2021, *System Engineering, System Life Cycle Process*, 2021.

[17] *Systmes and Software Engineering--System Life Cycle Processes ISO/IEC 15288:2015*, 2015.

[18] *Information Architecture of Electric Power Enterprises DL/T 2075—2019*, 2019.

[19] *The Overall Architecture and Technical Requirements of a Unified Data center for All Services QGDW/11816—2018*, 2018.

[20] *Information Architecture of State Grid Corporation of China (SG-EA) Q/GDW 11209—2018*, 2018.

[21] *In-depth Regulations on the Outline Design Content of Informatization Projects Q/GDW 11831—2018*, 2018.

[22] *TOGF Standard Version 9.1*, Available at: https://pubs.opengroup.org/architecture/togaf91-doc/arch/index.html, 2018.

[23] L. Wang, J. Chen, J. Pu, and Y. Yang, "Design evaluation method based on normal cloud model and information axioms," *Journal of Logistics Engineering Institute*, vol. 30, no. 02, pp. 65–70, 2014.

[24] J. Zhao, X. Wang, W. Guan, L. Yin, and Y. Zhou, "Virtual enterprise partner selection based on fuzzy information axioms and cloud model," *Operations Research and Management*, vol. 29, no. 01, pp. 202–208, 2020.

[25] G. . Büyüközkan and F. Göçer, "Evaluation of software development projects based on integrated Pythagorean fuzzy methodology," *Expert Systems with Applications*, vol. 183, p. 115355, 2021.

[26] J. Rezaei, "Best-worst multi-criteria decision-making method," *Omega*, vol. 53, no. 6, pp. 49–57, 2015.

[27] S.-P. Wan, "Jiuying. A novel extension of best-worst method with intuitionistic fuzzy reference comparisons [J]," *IEEE Transactions on Fuzzy Systems*, vol. 3, pp. 1–15, 2021.

[28] S. Wan, J. Dong, and S.-M. Chen, "Fuzzy best-worst method based on generalized interval-valued trapezoidal fuzzy numbers for multi-criteria decision-making," *Information Sciences*, vol. 573, no. 3, pp. 493–518, 2021.

[29] J. Dong, S. Wan, and S.-M. Chen, "Fuzzy best-worst method based on triangular fuzzy numbers for multi-criteria decision-making," *Information Sciences*, vol. 547, no. 2, pp. 1080–1104, 2021.

[30] S.-ping Wan, Ze-hui Chen, and J.-ying Dong, "An integrated interval type-2 fuzzy technique for democratic–autocratic-multi-criteria decision making," *Knowledge-Based Systems*, vol. 214, p. 106735, 2021.

Hindawi
Mathematical Problems in Engineering
Volume 2022, Article ID 4283995, 11 pages
https://doi.org/10.1155/2022/4283995

Research Article

Topological Structure Analysis of Software Using Complex Network Theory

Xinxin Xu,[1] **Zengyou Zhang**,[1] **Yan Liang**,[2] **and Liya Wang**[1]

[1]*College of Artificial Intelligence, Zhejiang Industry & Trade Vocational College, Zhejiang 325003, China*
[2]*School of Business, Shanghai Jianqiao University, Shanghai 201306, China*

Correspondence should be addressed to Yan Liang; irisly1020@163.com

Received 24 January 2022; Revised 12 April 2022; Accepted 13 May 2022; Published 27 May 2022

Academic Editor: Chunlai Chai

Due to people's increasing dependence on software, the emergence of software defects will lead to serious consequences. And the essential cause of software defects is the increasing complexity of software. The premise of reducing software defects is to understand the software topology to ensure software quality. The software topology refers to the connection between the internal elements of the software, and it has become an important factor affecting the quality of the software. In this paper, we use complex network theory as a tool to analyze the software topology. Firstly, we extract the software structural information from the source code of the software system and abstract the extracted software structural information with software network theory. Secondly, the metrics widely used in complex networks are introduced to analyze the built software network. When tracking the values of these metrics in the software system, we have a deeper understanding of the software topology. These results provide a different dimension to understanding the software topology, which has important guiding significance for the subsequent understanding of the software and is also very useful for reducing software defects and ensuring software quality.

1. Introduction

Large-scale software systems are composed of countless small elements (class, process, method, etc.), and every tiny error may lead to catastrophic consequences, especially for projects with extremely high software reliability requirements. When the software system becomes more and more complex, how to recognize and measure the software system has become a matter of constant concern and urgent solution.

Some researchers have proposed to study existing software systems from the perspective of software structure. Software structure analysis [1–4] can help us understand the specific situation of the software system to carry out corresponding maintenance and upgrades according to the characteristics. At present, many achievements in the field of software structure analysis have been published. The main software structure analysis approaches are divided into traditional software structure measurement approaches and software structure measurement approaches based on complex networks.

Traditional software structure analysis focuses on analysis from a single module. For example, the McCabe method [5], the Halstead method, the C&K metrics proposed by Chidanber and Kemerer and the MOOD method proposed by Brito all describe the complexity of the software structure from different aspects but focus on the analysis of the local structure and properties of functional individuals of the software system. Therefore, traditional analysis approaches lack the measurement of the overall software structure. As a kind of complex system, the overall structure of the computer software system has a huge impact on its function, performance, and quality [6].

However, some researchers have introduced the theory of complex networks into software research [7–10]. By constructing software networks from software source code, they can use complex networks to understand, analyze, and control the system from a global perspective, rather than from a local perspective. Complex network theory provides us with a new way to understand the structure of software systems.

Based on a weighted network, Wang and Xiao [11] used the theory and technology of complex networks to explore the execution process of Linux. Trindade et al. [3] represented class-level software as Little House.

At present, there is still little research on software networks, and the existing research still has the following shortcomings: (1) Existing research is not accurate enough for the construction of a software network model. (2) In the existing research, the metrics used in the software network analysis and the data sets used in the experiments are not comprehensive enough. In this paper, we adopt the idea of interdisciplinary and propose the software topology analysis approach. Firstly, we extract the software structural information from the source code of the software system and abstract the extracted software structural information with software network theory. Secondly, the metrics widely used in complex networks are introduced to analyze the built software network. By analyzing a series of metrics widely used in complex networks, we can discover the underlying laws that, in the software system, only a few classes contain more important information and have a strong influence, while most other nodes have little influence. That is, developers can quickly locate software defects by finding key classes in the software system.

The rest of this paper is organized as follows. Section 2 introduces some preliminary knowledge with a focus on the framework of the proposed software topology analysis approach and the formal definition of the software network model. Section 3 illustrates our approach by analyzing the experimental results of 6 subject software systems. And we conclude this paper in Section 4.

2. Related Work

Many results of software system research have been reported in the past few years. These studies can be roughly classified into two groups, that is, approaches based on traditional software metrics and approaches based on complex networks. For the approaches based on traditional software metrics, they pay more attention to analyzing software from a single module. There are mainly the following approaches. The McCabe method [5] is mainly based on graph theory and program structure control theory and uses directed graphs to represent program control flow, thereby representing the complexity of the network according to the cyclic complexity in the graph. The programming complexity measured using McCabe's method mainly depends on the complexity of the structural control flow. The Halstead method measures the complexity of the software system by counting the number of operators and operands in the program. However, this method only considers the program data flow but does not consider the control flow, so it can not reflect the complexity of the program fundamentally. The C&K metric proposed by Chidanber and Kemerer is based on object-oriented metric theory, including six metrics: (1) the number of subclasses (the number of direct subclasses of a class); the number of weighted methods of the class; (2) the depth of the inheritance tree (if it is multiple inheritances, calculate the maximum depth from the node to the root of the tree); (3) the number of weighted methods of the class; (4) the degree of coupling between objects (when a class uses member variables or methods of other classes, the two classes are said to be coupled); (5) the number of responses of the class (the total number of out-of-class methods called by all methods in the class); (6) the lack of cohesion in the class method. The MOOD method proposed by Brito indirectly measures the inheritance, encapsulation, polymorphism, and coupling of object-oriented software systems. The traditional software structure measurement method describes the complexity of the software structure from different aspects, but it focuses on analyzing the local structure and properties of functional individuals (classes, procedures, methods, etc.) in the software system. Therefore, the traditional analysis methods lack the overall software structure measurement. However, the measurement method based on a single module cannot understand the software system from the perspective of the overall structure.

As a kind of complex software system, the overall structure of the system has a great impact on its function, performance, and quality. Therefore, compared with approaches based on traditional software metrics, approaches based on complex networks have great application potential. In this work, we mainly discuss research based on complex network analysis. Based on a weighted network, Wang and Xiao [11] used the theory and technology of complex networks to explore the execution process of Linux. They found that the weight distribution obeys the power-law distribution, and the process management component of Linux plays the most important role. Trindade et al. [3] represented class-level software as Little House. Based on Little House, they analyzed 81 versions of 6 software systems and found some software evolution patterns. Šubelj and Bajec [12] used an Associative Software Graph (ASG) to represent a class-level software system, where nodes represent classes and edges represent "inheritance," "composition," and "dependency" relationships between classes. Based on ASG, they calculated the number of communities, the modularity of the software network, and other network metrics such as clustering coefficient, average path length, and average degree. They then analyzed the correlation between these indicators and the number of defects in the software. They found that medium-sized systems with a community structure tended to have a greater probability of defects. Yang et al. [10] proposed an internal class network of the software system to represent class-level software systems. In a software network, a class is a node, and the calling relationship between the methods contained in each pair of classes constitutes an edge. Based on the software network, they propose a set of metrics to characterize the software network structure and use some machine learning algorithms to build a defect prediction model, and their final results are encouraging. Zakari et al. [13] proposed a software network at the statement level, where statements are nodes and execution trajectories between statements are edges. They calculated two centrality metrics (i.e., degree centrality and closeness centrality) for defect diagnosis based on a software network.

3. Preliminaries

In this section, we show the framework of our software topology analysis approach (see Figure 1). It mainly consists of four parts, ① to ④. The first two parts ① to ② are detailed in Sections 3.1 and 3.2, and the two parts ③ to ④ will be explained in Section 3.

3.1. Data Collection. Data collection is the first step in our approach. For the reliability of the results, we will select software that is widely used in software structure-related research. Thus, we conducted our study on 6 well-known open-source software written in Java from different fields and different scales: Ant https://ant.apache.org/, GWT Portlets http://code.google.com/p/gwtportlets/, jEdit http://jedit.org/index.php, JHotDraw https://sourceforge.net/projects/jhotdraw/, Maze https://sourceforge.net/projects/maze/, and Wro4j https://github.com/wro4j/wro4j. Ant is a tool that provides software automation construction functions; GWT Portlets is an open-source web framework for developing GWT (Google Web Toolkit) applications; jEdit is an open-source text editor written in Java; JHotDraw is an open-source drawing program developed based on Java; Maze is an open-source network file system; Wro4j is a web resource optimization tool.

Table 1 shows the description of the relevant metrics of the subject software system, such as the number of lines of code (LOC), the number of packages (#P), the number of classes (#C), the number of methods (#M), and the number of attributes (#A). These values are calculated based on the Java code listed in the "Directory" column, not the entire distribution of the corresponding software.

3.2. Software Network Model. After data collection, software structure extraction [14–16] is the next step in the construction of software network models. This step aims to extract various software elements (classes, interfaces, attributes, methods, local variables, etc.) and interactions (class inheritance, interface implementation, method calls, etc.).

Based on the results of software structure extraction, this paper introduces the Unweighted Directed Class Coupling Network (UDCCN). In this network, nodes represent class-level elements (classes, interfaces, etc.) in the software system, edges represent the coupling relationship between elements, and the direction of the edges represents the coupling direction between elements. In UDCCN, we considered 7 coupling types:

(i) Inheritance relationship (INR): if class A inherits from another class B by using the keyword "extends."

(ii) Implementation relationship (IMR): if class A implements interface B by using the keyword "implements."

(iii) Parameter relationship (PAR): if one of the methods of class A has at least one parameter of class B type.

(iv) Global variable relation (GVR): if class A has at least one attribute with the type of class B.

(v) Local variable relationship (LVR): if a local variable with the type of class B is declared in a method of class A.

(vi) Method call relationship (MCR): if one of the methods of class A calls a method on an object of class B.

(vii) Return type relationship (RTR): if one of the methods of class A has a return type of class B.

If the above seven relationships exist between elements, we will generate a directed edge in the UDCCN network to describe this coupling relationship. UDCCN is an unweighted directed graph, which is defined as follows:

$$\text{UDCCD} = (V, L), \quad n \in V, l \in L, l = \langle n_i, n_j \rangle, \tag{1}$$

where n represents the class or interface in the software system and l represents the coupling between the node (class i) and the node (class j). And the adjacency matrix ψ_{ij} of UDCCN encodes the coupling between every pair of classes:

$$\psi_{ij} = \begin{cases} 1, & \langle n_i, n_j \rangle \in L, \\ 0, & \text{otherwise.} \end{cases} \tag{2}$$

That is a $|V| \times |V|$ matrix, where $|V|$ returns the number of classes. ψ_{ij} is the weight assigned to the link $\langle n_i, n_j \rangle \langle n_i, n_j \rangle$; if $\langle n_i, n_j \rangle \in L$, then $\psi_{ij} = 1 \psi_{ij} = 1$; otherwise $\psi_{ij} = 0$.

To explain UDCCN more clearly, Figure 2(a) shows an exemplary Java code snippet. For this code segment, Figure 2(b) shows its corresponding UDCCN. As shown in Figure 2(b), the coupling relationship between classes in the Java code fragment in Figure 2(a) includes inheritance relationship, implementation relationship, parameter relationship, global variable relation, return type relationship, and method call relationship.

3.3. Complex Network Statistical Metrics. We use a software network model to abstract the relationships between elements in the software system, which provides a new perspective for the research of software engineering. Complex networks have gradually become one of the focuses of research. Particularly with the discovery of features such as "small world" and "scale-free," scientists have set off an upsurge in studying complex networks [17–19], covering many fields such as physics, mathematics, and biology. Therefore, we can draw on the above-mentioned complex network statistical metrics to reveal the knowledge related to the topology of the software network [20, 21].

3.3.1. Network Centrality. The metrics of network centrality are mainly to find nodes that have important roles in complex networks and reflect the importance of node locations. These metrics include betweenness centrality, degree centrality, and closeness centrality.

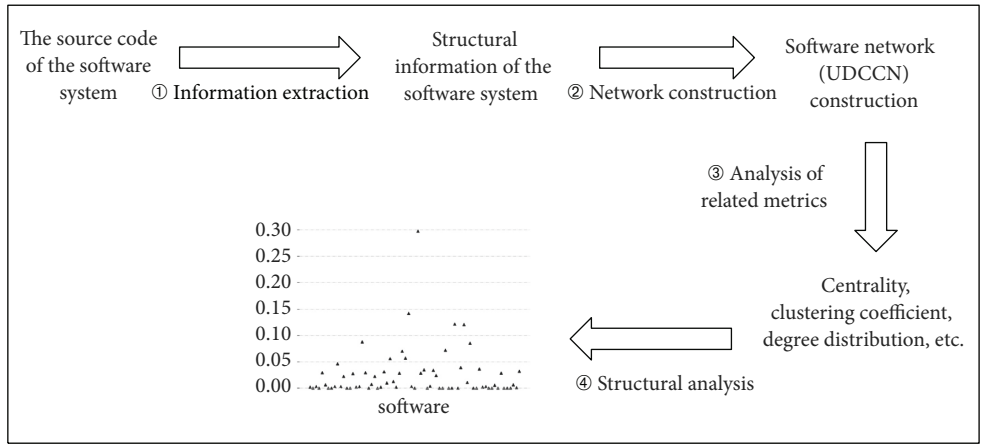


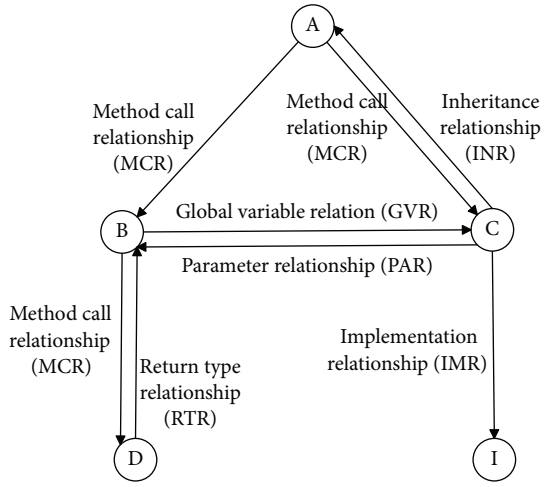
Figure 1: The basic framework of the software topology analysis approach.

Table 1: The description of the subject software system.

System	Version	Directory	LOC	#P	#C(#E)	#M/#A
Ant	1.6.1	Src/main	81515	67	900	7691/4167
GWT Portlets	0.9.5beta	Src	8501	10	145	1145/424
jEdit	5.1.0	Src	112492	41	1082 (9)	7601/4085
JHotDraw	6.0b.1	Src	28330	30	544	5205/865
Maze	1	Src	8881	6	63 (6)	563/284
Wro4j	1.6.3	Src	33736	30	567 (9)	3256/1274

```
interface I{
    public int add(int i);
}

public class A {
    public static void main(String[] args){
    B b = new B();
    C c = new C();
    b.m1();
    c.m2(b);
    c.add(1);
    }
}

class B{
    public int sum = 0;
    private C c = new C();
    public void m1(){
        System.out.println(D.m3().sum);
    }
}

class C extends A implements I{
    public int add(int i){
        return i+1;
    }
    public void m2(B b){
        b.sum = b.sum+2;
        System.out.println(b.sum);
    }
}

class D{
    public static B m3(){
        B b = new B();
        return b;
    }
}
```

(a)



(b)

Figure 2: An illustrative exemplary Java code snippet and corresponding UDCCN.

(1) Betweenness Centrality. The betweenness is a parameter that cannot be bypassed when studying complex networks. This parameter reflects the influence and importance of nodes. To understand the definition more intuitively, the following formula is given:

$$g(v) = \sum_{s \neq v \neq t} \frac{\phi_{st}(v)}{\phi_{st}}, \tag{3}$$

where ϕ_{st} represents the number of shortest paths from node s to node t and $\phi_{st}(v)$ represents the number of all paths from node s to node t and through node v in the network. Betweenness centrality [22] reflects the dependence between class nodes. The higher the betweenness centrality of class nodes, the stronger the importance to the software network.

(2) Degree Centrality. In complex network analysis, degree centrality is the most direct metric to describe the importance of nodes. The higher the degree centrality of a node is, the more important the node is in the network. Conversely, if the degree centrality of a node in the network is closer to 0, it means that the node has less contact with other nodes.

(3) Closeness Centrality. In complex network metrics, closeness centrality refers to how close a node in the network is to other nodes. If a node's closeness centrality is higher, then it is closer to other nodes. The closeness centrality of a node is the reciprocal of the average value of the shortest path length between the node and all other nodes in the network, which can be defined as

$$C(i) = \frac{n}{\sum_j d(j,i)}, \tag{4}$$

where $d(j,i)$ represents the distance from node i to node j and n represents the number of nodes.

3.3.2. Clustering Coefficient. In graph theory, the clustering coefficient is used to measure the degree of clustering of nodes in the graph. It is often used to describe the clustering characteristics of the network, indicating the closeness of a node with surrounding nodes [23]. The clustering coefficient of nodes in the network mainly refers to the ratio of the number of connections between the node and adjacent nodes to the maximum number of edges that can be connected between these adjacent nodes. The clustering coefficient C_i of the node i can be defined as

$$C_i = \frac{2e_i}{k_i(k_i - 1)} = \frac{\sum_{j,m} a_{ij} a_{im} a_{mj}}{k_i(k_i - 1)}, \tag{5}$$

where e_i indicates that the value of the clustering coefficient C_i of the node i is equal to the number of edges connected by the neighbor node and $k_i(k_i - 1)/2$ represents the maximum number of edges that may exist. The clustering coefficient of the network is the average of the clustering coefficients of all nodes in the network, which is

$$C = \langle C_i \rangle = \frac{1}{N} \sum_{i \in V} C_i, \tag{6}$$

where N is the number of nodes in the network, which indicates the aggregation trend of nodes in the network and reflects the local characteristics of the network.

3.3.3. Degree Distribution. The degree distribution reflects the most basic characteristics of the complex network topology. The degree of a node in the network refers to the number of nodes adjacent to the node, that is, the number of edges connecting the node. The greater the degree of the node, the more the connections between the nodes and the more important the node in the network. The degree distribution $P(k)$ refers to the probability that the degree of an arbitrarily selected node in the network is exactly k. When the degree distribution of the network satisfies the power rate distribution, it can be defined as $P(k) \sim k^{-r}$, and then the network is a scale-free network.

3.3.4. Average Shortest Path Length. The average shortest path length of the network [24] is defined as the average of the shortest path length between any two nodes in the network. The average shortest path length of the network can be defined as

$$L = \frac{2}{(N(N-1))} \sum_{i \neq j} d_{ij}, \tag{7}$$

where d_{ij} represents the number of edges on the shortest path connecting two nodes i and j in the network and N represents the number of nodes in the network.

3.3.5. Assortativity Coefficient. It is found that many observable networks have mixing patterns in degree, that is, assortative mixing or disassortative mixing. The so-called assortative mixing means that nodes with high degrees are often connected with other nodes with high degrees, and nodes with low degrees are likely to be connected with other nodes with low degrees. Disassortative mixing means that low-degree vertices are more likely to be connected to high-degree vertices, and vice versa.

The assortativity coefficient is often used to quantify the degree of assortative mixing, it is a degree-based Pearson correlation coefficient, and the calculation formula can be expressed as

$$ac = \frac{\sum_{y,z} yz\left(e_{yz} = m_y n_z\right)}{\sigma_y \sigma_z}, \tag{8}$$

where e_{yz} represents the ratio of the node with a degree value of y in the network and the number of its edges to the total number of all edges, $m_y = \sum_y e_{xy}$, $n_z = \sum_z e_{yz}$, $\sigma_y = \sqrt{E(y^2) - E^2(y)}$, and $\sigma_z = \sqrt{E(z^2) - E^2(z)}$. If ac is less than 0, it means that the network is disassortative, while ac being greater than 0 denotes an assortative mixing network.

3.3.6. Structural Holes. Structural hole theory [25] is a new theory in interpersonal network theory, which mainly describes the gaps in social networks. In the social network, an

individual directly finds contact with some individuals but does not have direct contact with other individuals. That is, there are holes in this social network.

If there is no direct connection between the two and the connection can only be formed through a third party, then the acting third party occupies a structural hole in the relationship network. The structural hole is for the third party. If there are structural holes in the network, the third party that connects two actors that are not directly connected has an information advantage and control advantage.

Generally, the effective size metric in structural hole theory is used to measure the network. This metric mainly describes the effectiveness of the node's self-network. Formally, the effective size of a node, expressed as $e(u)$, is defined as follows:

$$e(u) = \sum_{v \in N(u)/\{u\}} \left(1 - \sum_{w \in N(v)} P_{uw} m_{vw}\right), \tag{9}$$

where $N(u)$ is the set of neighbor nodes of u, P_{uw} is the normalized mutual weight of the (directed or undirected) edge connecting u and v, and m_{vw} is the mutual weight of the connecting node v to the node w divided by the v node's maximum connection edge weight with its neighbor nodes. Mutual weight refers to the sum of edge weights connecting node u and node v (in the case of a weightless network, the default edge weight is 1).

4. Topological Structure Analysis

In this section, for the illustration purpose, complex network statistical metrics mentioned above are used to study the software network topology of the subject software.

4.1. Topological Structure Analysis of Network Centrality

4.1.1. Betweenness Centrality. Betweenness centrality is a measure of graph centrality based on the shortest path, generally used to check whether a node is in an important position in the graph. As shown in Figure 3, we found that, in the software network of almost all subject software systems, the betweenness centrality of nearly 90% of the classes is distributed below 0.05, indicating that only 5% of the classes are in an important position in the software system, which has a strong impact on the realization of the software system function, and most other nodes have little influence.

In the actual development process, the calls between classes are usually a call chain, and important classes frequently call other classes or are frequently called by other classes. For example, the key class is usually called frequently by other classes in the software system to complete the corresponding function. Therefore, analyzing the betweenness centrality can provide greater help in identifying the key classes of the software system.

4.1.2. Degree Centrality. In complex network analysis, degree centrality is the most direct metric to describe the importance of nodes. The higher the degree centrality of a node is, the more important the node is in the network. Conversely, if the degree centrality of a node in the network is closer to 0, it means that the node has less contact with other nodes.

As shown in Figure 4, the degree centrality of the class nodes in the software network of the six subject software systems is mostly close to 0, while a few are between 0.01 and 0.05. It shows that only a small number of classes are closely connected with other classes and have a relatively strong influence, while most of the classes are not very influential.

In the actual software system, only a few classes will frequently call other classes or be frequently called by other classes. Usually in software development, if this class frequently calls other classes or is frequently called by other classes, it means that this class has a higher status in the software system, that is, the key class. How to find the key classes is of great importance to software cost prediction. If we ignore the importance of key classes, we will underestimate the complexity and cost of the software system to be developed, which may cause great losses to the company.

4.1.3. Closeness Centrality. Closeness centrality reflects the closeness between a node and other nodes in the network. If a node is very close to other nodes, then it does not need to rely on other nodes when transmitting information, indicating that this node is very important. When we calculated the closeness centrality of the six subject software systems, we found that, in the four software systems of Ant, jEdit, JHotDraw, and Wro4j, the closeness centrality of most nodes is close to 0. In the software systems of GWT Portlets, the closeness centrality of nodes is almost evenly distributed between 0 and 0.25. And in the software system Maze, the closeness centrality of most nodes is between 0.1 and 0.2 (see Figure 5).

If the closeness centrality of the node is 0, it means that there are a few isolated nodes in the software system, and these isolated nodes do not have any connection with other nodes. And the closer the value is to 1.0, the higher the closeness of the node is. Therefore, the greater the closeness centrality of a class node is, the closer the node is likely to be connected with all other class nodes. It also shows that the location of these class nodes has the best view of the network and can perceive the dynamics of the entire software network and the direction of information circulation. From the perspective of the structure of the software network, in general, the key classes are closely related to other class nodes; that is, the key class can usually get a higher value of closeness centrality.

4.2. Topological Structure Analysis of Clustering Coefficient. The clustering coefficient of a node indicates how interconnected its adjacent nodes are. The clustering coefficient distribution of each node in the software network of the six subject software systems is shown in Figure 6. The clustering coefficients of most class nodes in Ant, jEdit, JHotDraw, and Wro4j are less than 0.5, and only a few nodes have high clustering coefficients, which are nodes with high clustering degrees in the software network. For the software GWT

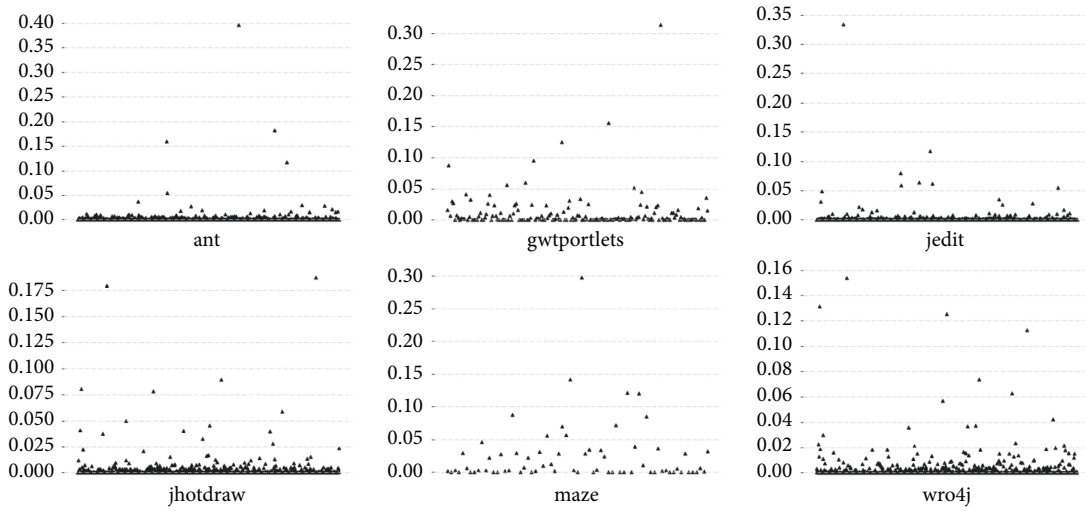


Figure 3: The distribution of betweenness centrality.

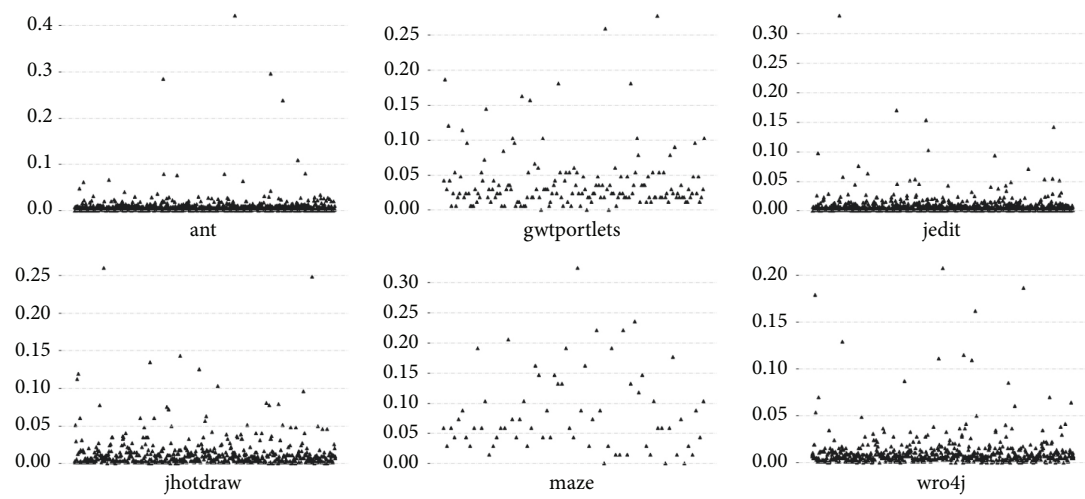


Figure 4: The distribution of degree centrality.

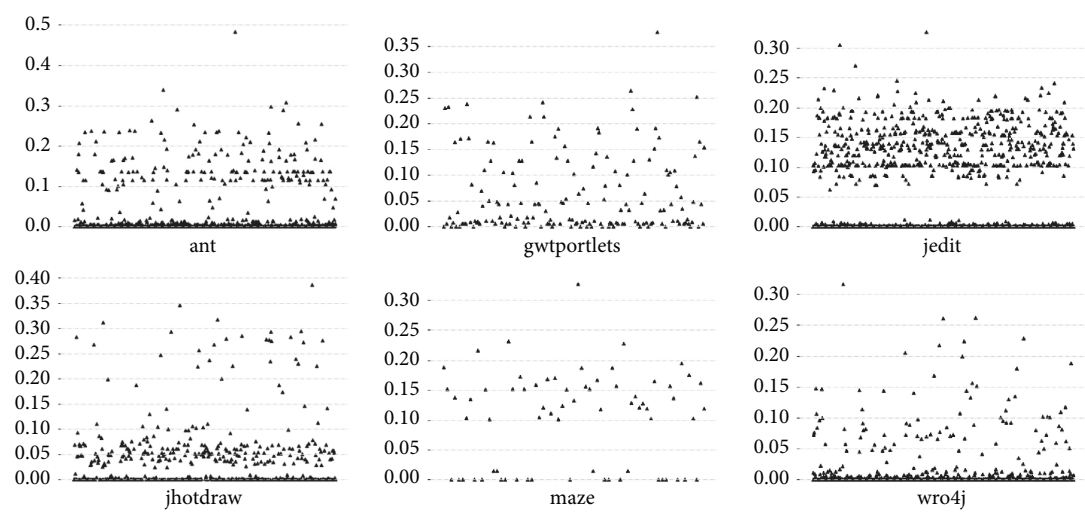


Figure 5: The distribution of closeness centrality.

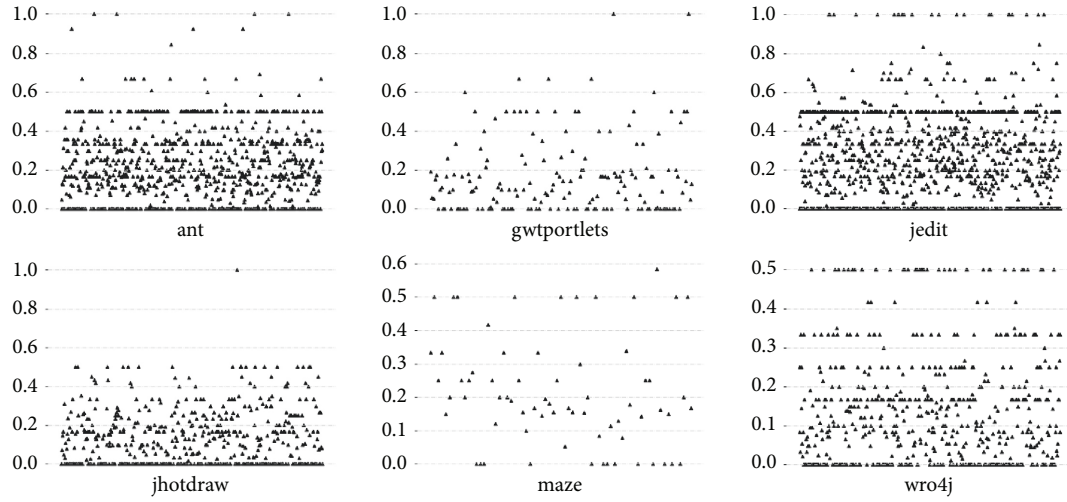


Figure 6: The distribution of clustering coefficient.

Portlets and Maze, although the nodes are relatively evenly distributed in the graph, for the part of the graph with high clustering coefficients, the number of class nodes is still relatively small. The larger the clustering coefficient of a node, the higher the degree of relationship between the nodes around the node and the higher the clustering degree of the group, and the highest value is 1, indicating that all the points around a node are related.

For the actual software system, only a few class nodes will have a relatively high clustering coefficient; that is, only a few classes will use other classes more or are used more by other classes. This is in line with the characteristics of key classes of software systems. In other words, analyzing the clustering coefficient of a software network is also helpful to identify key classes in the software system.

4.3. Topological Structure Analysis of Degree Distribution. In the study of graphs and networks, the degree of a node in a network is the number of connections it has to other nodes, and the degree distribution is the probability distribution of these degrees over the whole network. The degree distribution of nodes in the software network of the six subject software systems is shown in Figure 7. The horizontal axis in the figure is degrees, and the vertical axis is the number of nodes. It can be seen from the figure that as the degree becomes larger, the number of nodes declines. And when the software network has more nodes, this trend becomes more obvious. In software Ant, jEdit, JHotDraw, and Wro4j, this trend is more obvious than in software networks with fewer nodes. It can be observed from the figure that the number of nodes with a degree less than 10 accounts for almost 90% of the nodes in the software network, and the number of nodes with a degree greater than 50 is almost zero.

In a software network, most nodes are only connected to a few nodes, while a few nodes are connected to most of the nodes, which is in line with the typical characteristics of a scale-free network. Therefore, in the software system, we can find that most of the classes only call a few classes or are called by a few classes, and only a few classes call other classes or are called by a large number of classes.

4.4. Topological Structure Analysis of Average Shortest Path Length. The average shortest path length is a concept in the network topology that is defined as the average number of steps along the shortest paths for all possible pairs of network nodes. It is a measure of the efficiency of information or mass transport on a network. It can be seen from Table 2 that although the size of the subject software is different, the distance between nodes is stable at about 3. When calculating the average shortest path length, we found that the maximum value is 3.379 and the minimum value is 2.806. Therefore, the software network conforms to the "small world" effect in the complex network. Research shows that, in reality, the number of nodes in many networks is very large, but the average shortest path length of the entire network is relatively small, such as the World Wide Web, so formal networks generally have the characteristics of "small world" in complex networks.

4.5. Topological Structure Analysis of Assortativity Coefficient. Assortative mixing is a preference for a network's nodes to attach to others that are similar in some way. According to the calculation formula (8), it can be found from Table 3 that the calculated assortativity coefficients of 6 subject software systems are all negative, indicating that these software systems have a disassortative mixing network. That is to say, in the software network, nodes with high degrees and nodes with low degrees have a relatively high connection probability. And it means that key classes with a high frequency of use are usually related to classes with a low frequency of use, instead of being related to each other.

4.6. Topological Structure Analysis of Structural Holes. In the structural hole theory, the larger the effective size, the greater the effectiveness of the node. As shown in Figure 8, we can find that, in the software system, the effective size of most

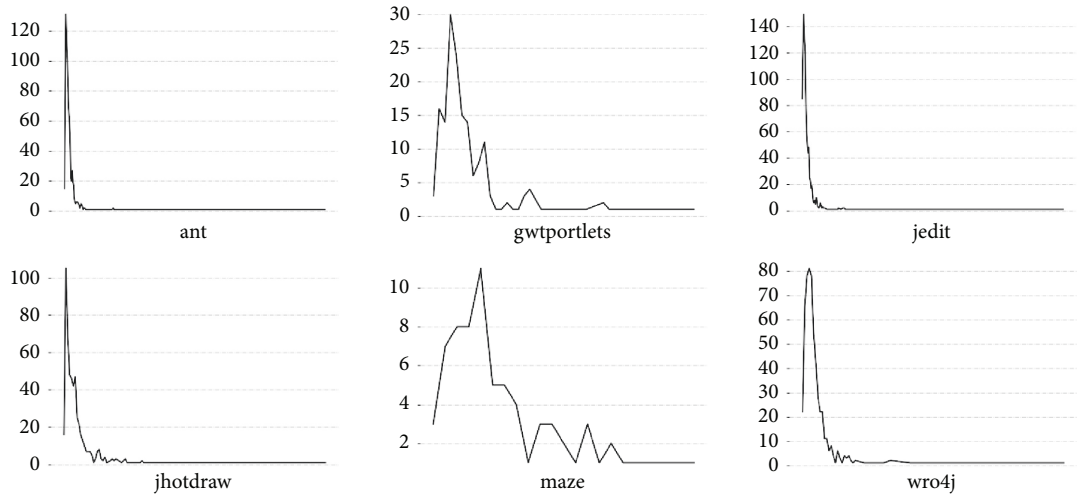


Figure 7: The distribution of degree distribution.

Table 2: Average shortest path length of software network.

Software system	Ant	GWT Portlets	jEdit	Maze	JHotDraw	Wro4j
Average shortest path length	3.178	3.072	3.290	2.806	3.235	3.379

Table 3: Assortativity coefficient of software network.

Software system	Ant	GWT Portlets	jEdit	Maze	JHotDraw	Wro4j
Assortativity coefficient	−0.126	−0.098	−0.152	−0.174	−0.165	−0.055

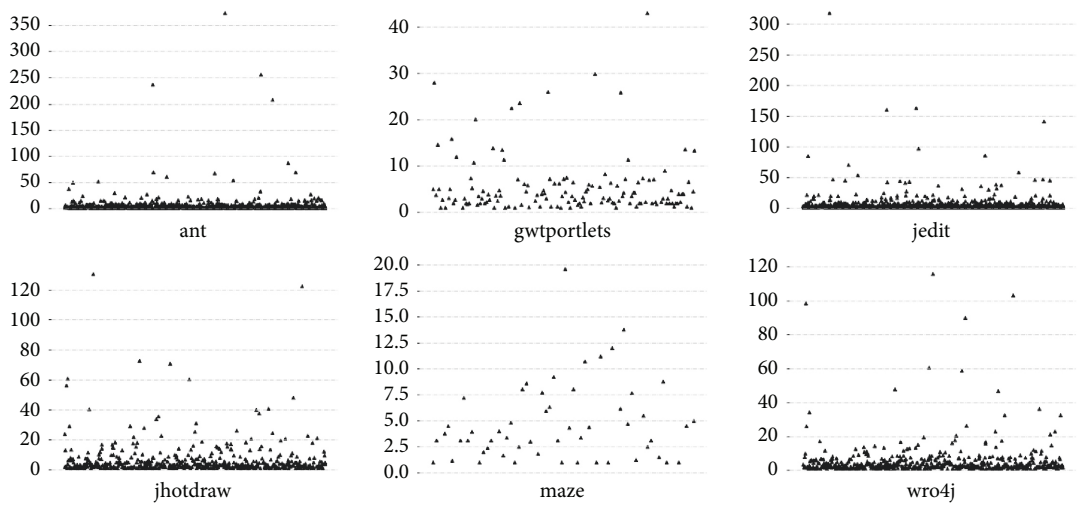


Figure 8: The distribution of effective size in structural hole theory.

nodes is between 0 and 20, while only a few nodes are above 20. Therefore, for the software network, only a few nodes are very important in the software system. Compared with other nodes, these nodes have information advantages and control advantages.

From the perspective of structural hole theory, some key classes in software systems usually act as a bridge in class calls. For example, functional aggregation classes are usually shown as a bridge of some single tool classes in a software network, and there is no direct connection between tool classes. Therefore, in the effective size metric, the value of the key class is larger than that of other common classes.

5. Threats to Validity

In this study, we obtained several important results about the software topology from our experiments. However, potential threats to our jobs remain. In our empirical research, we use 6 software systems of different scales as the research objects, all of which are widely used in the research of software

engineering. However, since the results obtained by these 6 software systems may not be so common, we hope to continue to do empirical research on more software systems to further evaluate their effectiveness.

The software systems we use for empirical research are all developed based on Java. Java is one of the most widely used programming languages. The software developed in Java has a clear structure, and the components in the software system are easier to extract. However, since there is no empirical research on software systems developed in other languages, this may affect the final results. We hope to continue empirical research on software systems developed in other languages for further evaluation of their validity, which will be important work for us in the future.

6. Conclusions and Future Work

In this paper, we proposed an approach to study the topological structure of software using the tool of complex network theory. For illustration, we conducted case studies on 6 software systems. Firstly, the software structure information is extracted from the source code of the software system, and the Unweighted Directed Class Coupling Network model is constructed based on this structure information. Secondly, several aspects of these software networks are studied by using the parameters widely used in complex network theory.

Through the analysis of software structure, we concluded that software network has significant characteristics of "small world" and "scale-free." The important structural features in software network topology help us to provide valuable insights and different dimensions for our understanding of software systems. Through the analysis of the experimental results, we found that only a few classes in the software are key classes, which play a great role in the function realization of the software. After finding the key classes, we can make a series of optimizations, such as the prediction and positioning of software defects.

There are a few areas that could be explored in future research: (1) investigating more software networks to validate the proposed approach, (2) investigating systems written in other languages to validate the proposed approach, and (3) using the parameters in other theories to study the software from different angles.

Data Availability

All data used during the study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

[1] M. Kitsak, L. K. Gallos, S. Havlin et al., "Identification of influential spreaders in complex networks," *Nature Physics*, vol. 6, no. 11, pp. 888–893, 2010.

[2] P. Weifeng, L. Bing, and L. Jing, "Multi-granularity evolution analysis of software using complex network theory," *Journal of Systems Science and Complexity*, vol. 24, no. 6, pp. 1068–1082, 2011.

[3] R. P. F. Trindade, T. S. Orfanó, K. A. M. Ferreira, and F. Elizabeth, "The dance of classes-A stochastic model for software structure evolution," in *Proceedings of the 2017 IEEE/ACM 8th Workshop on Emerging Trends in Software Metrics (WETSoM)*, pp. 22–28, IEEE, Buenos Aires, Argentina, May 2017.

[4] P. Weifeng and C. Chunlai, "Measuring software stability based on complex networks in software," *Cluster Computing*, vol. 22, no. s2, pp. 2589–2598, 2019.

[5] T. J. Mccabe, "A complexity measure," *IEEE Transactions on Software Engineering*, vol. 2, no. 4, pp. 308–320, 2006.

[6] Y. Kamei, E. Shihab, B. Adams, and E. H. Ahmed, "A large-scale empirical study of just-in-time quality assurance," *IEEE Transactions on Software Engineering*, vol. 39, no. 6, pp. 757–773, 2012.

[7] P. Weifeng, "Applying complex network theory to software structure analysis," *Engineering and Technology*, vol. 60, pp. 1636–1642, World Academy of Science, 2011.

[8] W. Y. Pan, H. Jiang, and Z. Yunfang, "Measuring software modularity based on software networks," *Entropy*, vol. 21, no. 4, 344 pages, 2019.

[9] G. Concas, M. Marchesi, C. Monni, and O. Matteo, "Software quality and community structure in java software networks," *International Journal of Software Engineering and Knowledge Engineering*, vol. 27, no. 7, pp. 1063–1096, 2017.

[10] Y. Yang, J. Ai, X. Li, and W. E. Wong, "MHCP model for quality evaluation for software structure based on software complex network," in *Proceedings of the IEEE 27th International Symposium on Software Reliability Engineering (ISSRE 2016)*, pp. 298–308, Ottawa, Canada, October 2016.

[11] H. Wang and G. Xiao, "Analysis of the runtime Linux operating system as a complex weighted network," in *Proceedings of the 2016 International Conference on Software Analysis, Testing and Evolution (SATE)*, pp. 7–11, IEEE, Kunming, China, November 2016.

[12] L. Šubelj and M. Bajec, "Software systems through complex networks science: review, analysis and applications," in *Proceedings of the International Workshop on Software Mining*, pp. 9–16, Singapore, September 2012.

[13] A. Zakari, S. P. Lee, and C. Y. Chong, "Simultaneous localization of software faults based on complex network theory," *IEEE Access*, vol. 6, p. 1, 2018.

[14] W. Pan, H. Ming, Z. Yang, and T. Wang, "Comments on "using k-core decomposition on class dependency networks to improve bug prediction model's practical performance"," *IEEE Transactions on Software Engineering*, vol. 1, 2022.

[15] L. Hao, W. Tian, P. Weifeng, C. Pengyu, and W. Jiale, "Mining key classes in java projects by examining a very small number of classes: a complex network-based approach," *IEEE Access*, vol. 9, pp. 28076–28088, 2021.

[16] X. Du, T. Wang, L. Wang et al., "CoreBug: improving effort-aware bug prediction in software systems using generalized *k*-core decomposition in class dependency networks," *Axioms*, vol. 11, no. 5, 205 pages, 2022.

[17] D. J. Watts and S. H. Strogatz, "Collective dynamics of 'small-world' networks," *Nature*, vol. 393, no. 6684, pp. 440–442, 1998.

[18] D. Hylandwood, "Scale-free nature of java software package, class and method collaboration graphs," 2006, https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.73.2085&rep=rep1&type=pdf.
[19] A. Potanin, J. Noble, M. Frean, and R. Biddle, "Scale-free geometry in OO programs," *Communications of the ACM*, vol. 48, no. 5, pp. 99–103, 2005.
[20] S. H. Strogatz, "Exploring complex networks," *Nature*, vol. 410, no. 6825, pp. 268–276, 2001.
[21] J. Xu, "Topological structure and analysis of interconnection networks[J]," *Springer Berlin*, vol. 7, no. 2-3, pp. 969-970, 2001.
[22] U. Brandes, "A faster algorithm for betweenness centrality," *Journal of Mathematical Sociology*, vol. 25, no. 2, pp. 163–177, 2001.
[23] G. Sabidussi, "The centrality index of a graph," *Psychometrika*, vol. 31, no. 4, pp. 581–603, 1966.
[24] G. Mao and N. Zhang, "Analysis of average shortest-path length of scale-free network," *Journal of Applied Mathematics*, vol. 2013, Article ID 865643, 5 pages, 2013.
[25] S. Goyal and F. Vega-Redondo, "Structural holes in social networks," *Journal of Economic Theory*, vol. 137, no. 1, pp. 460–492, 2007.

Hindawi
Mathematical Problems in Engineering
Volume 2022, Article ID 4430087, 9 pages
https://doi.org/10.1155/2022/4430087

Research Article

Efficient Rational Community Detection in Attribute Bipartite Graphs

Chen Yang,[1] Hao Ji ,[2] and Yanping Wu[3]

[1]*China Tower Corporation Limited, Zhejiang Branch, Hangzhou, China*
[2]*Hangzhou Medical College, Hangzhou, China*
[3]*University of Technology Sydney, Sydney, Australia*

Correspondence should be addressed to Hao Ji; jihaobest11@163.com

Received 9 February 2022; Revised 17 March 2022; Accepted 4 April 2022; Published 28 April 2022

Academic Editor: Chunlai Chai

Bipartite graph is widely used to model the complex relationships among two types of entities. Community detection (CD) is a fundamental tool for graph analysis, which aims to find all or top-k densely connected subgraphs. However, the existing studies about the CD problem usually focus on structure cohesiveness, such as (α, β)-core, but ignore the attributes within the relationships, which can be modeled as attribute bipartite graphs. Moreover, the returned results usually suffer from rationality issues. To overcome the limitations, in this paper, we introduce a novel metric, named *rational score*, which takes both preference consistency and community size into consideration to evaluate the community. Based on the proposed rational score and the widely used (α, β)-core model, we propose and investigate the rational (α, β)-core detection in attribute bipartite graphs (RCD-ABG), which aims to retrieve the connected (α, β)-core with the largest rational score. We prove that the problem is NP-hard and the object function is nonmonotonic and non-submodular. To tackle RCD-ABG problem, a basic greedy framework is first proposed. To further improve the quality of returned results, two optimized strategies are further developed. Finally, extensive experiments are conducted on 6 real-world bipartite networks to evaluate the performance of the proposed model and techniques. As shown in experiments, the returned community is significantly better than the result returned by the traditional (α, β)-core model.

1. Introduction

A bipartite graph is composed of two disjoint vertex sets, and there are only edges connecting vertices from different sets. Due to its proliferation applications like fraudsters detection [1] and collaboration group maintenance [2], many fundamental problems have been investigated to analyze the bipartite graphs. Among these problems, community detection (CD) aims to find all or top-k communities by leveraging different models like (α, β)-core [3], bitruss [4], and so on. Due to its unique feature, the (α, β)-core model is widely adopted in different domains. Given a bipartite graph, the (α, β)-core is the maximal subgraph where the degree of each vertex in the upper layer is at least α and the degree of each vertex in the lower layer is at least β. Nonetheless, previous models mainly focus on the cohesiveness structure of the graphs but neglect the attribute properties with community.

In real applications, the relationships between different entities often have certain characteristics, which can be modeled as attribute bipartite graphs. For example, in the user-movie network of Figure 1, the upper layer denotes a set of users and the lower layer are the set of movies. Each edge is associated with a number denoting the score assigned from a user to a movie. For a discussion group in the platform, it will have a more harmonious atmosphere if users have high consistency of preference (e.g., rating the same score or tag for the same movie). Besides, small discussion group is more conducive to frequent communication among users. However, the existing research cannot capture those properties. Motivated by this, in this paper, we introduce a novel metric, named rational score, which takes both

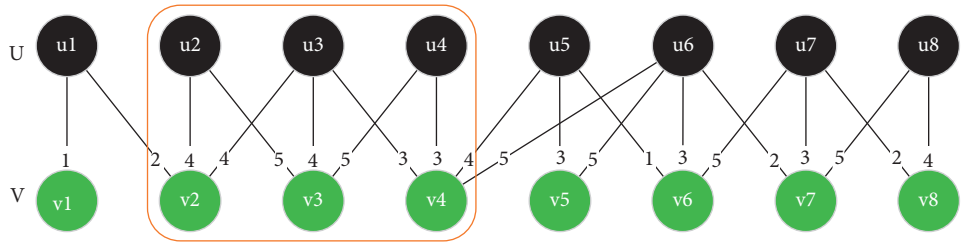


Figure 1: A user-movie network.

preference consistency and community size into consideration to evaluate a community. Furthermore, we formally define the problem of rational community detection over attribute bipartite graphs (RCD-ABG), which attempts to find the connected (α, β)-core with the largest rational score. The following is a motivation example.

Example 1. Reconsider the user-movie network in Figure 1, where the number on the edge denotes the corresponding rating for the movie. Note that the scoring mechanism adopts a five-point system, so the score varies from 1 to 5 in the network. Suppose $\alpha = 2$ and $\beta = 2$ here. Based on the definition, the subgraph induced by vertex set $\{u_2, u_3, \ldots, u_8, v_2, v_3, \ldots, v_8\}$ is a $(2, 2)$-core, where the degree of each vertex is at least 2. However, in the $(2, 2)$-core, many users have distinct scoring schemes for the same movie. For example, users u_6, u_7, and u_8 gave three different scores to the movie v_7. Moreover, the community size is too large to facilitate communication between users. For instance, users u_2 and u_6 even have not watched the same movie ever. Given $\alpha = 2$ and $\beta = 2$, the vertices in the orange rectangle are our identified rational (α, β)-core community. Note that, due to the complex equation involved, the detailed definition of rational (α, β)-core community can be found in preliminaries section. As we can observe, in this community, most users share the same movie taste and the number of people in the group is more reasonable.

1.1. Applications. The RCD-ABG problem can find many real-world applications. We list some examples as follows.

(i) *Discussion Group Mining.* In some real-world bipartite graphs such as BookCrossing, edges denote rating relationships between users and books. There are many discussion groups with these platforms. For users, they are more likely to stay active in a discussion group if the users inside share the same taste. Besides, users will prefer to discuss different topics in a group with appropriate size. This is because too many users can make them uncomfortable and too few will make the discussion difficult to carry on. Hence, by retrieving the rational group, the platform can provide group recommendation more precisely, which is helpful for better user experience.

(ii) *Personalized Product Recommendation.* In customer-movie bipartite networks, the customers will rate the movies based on their personal preference and movie performance. By retrieving the rational (α, β)-core, the personalized movie recommendation can be provided to customers in the rational community. For instance, in the community found in the orange rectangle in Figure 1, the platform can recommend movie v_4 for user u_2. This is because v_4 is given the common score from other customers (i.e., u_3 and u_4). Similarly, movie v_2 can be recommended for user u_4.

1.2. Challenges. To our best knowledge, we are the first to investigate the rational (α, β)-core detection problem in attribute bipartite graphs. We prove the problem is NP-hard and we adopt the greedy framework to remove the best vertex iteratively. However, removing a vertex from the graph may make many other vertices drop from the result, which limits the effectiveness of the algorithm. Hence, it is necessary to develop optimized techniques to address these challenges.

1.3. Our Solution. Due to the NP-hardness of the problem, a basic greedy framework is proposed by adopting the greedy framework. In general, we remove the vertex with the smallest marginal gain at each iteration and calculate the remaining (α, β)-core with its rational score. We stop this process until there is no (α, β)-core and return the (α, β)-core with the largest rational score as the result. To address the discussed drawbacks of our basic greedy framework, we further develop two improved strategies, namely, 2-hop neighbors-based optimization and followers-based optimization. Specifically, in 2-hop neighbors-based optimization, we approximate the marginal score by considering the 2-hop neighbors of the removed vertex in the same layer. In our followers-based optimization, we consider the followers of the removed vertex and modify the marginal rational score.

1.4. Contributions. The contributions of this paper are summarized as follows.

(i) To better capture the properties within bipartite graph community, we conduct the first research to propose and investigate the rational community detection problem over attribute bipartite graphs by leveraging the novel rational score metric developed.

(ii) Theoretically, we prove that the problem is NP-hard, and the rational score function is non-monotonic and non-submodular.

(iii) The basic greedy framework is first presented. To further improve the quality of returned results, two optimized strategies are proposed, namely, 2-hop neighbors-based optimization and followers-based optimization.

(iv) Experiments over 6 real-world bipartite graphs are conducted to show the superiority of proposed techniques. Compared with the traditional (α, β)-core model, our model is much more effective.

1.5. Roadmap. We organize the rest of this paper as follows. We first review the related work. Then, we introduce the problem investigated and the corresponding problem properties. Next, we will present the basic greedy framework and two optimized strategies. Finally, we report the performance of our algorithms over real datasets and conclude the paper.

2. Related Work

In this paper, we conduct the first attempt to propose and investigate the rational (α, β)-core problem. Thus, we will present the related work from the following two aspects.

Cohesive Subgraphs Mining. In different domains, graphs are widely used to model the complex relationships among different entities. As a key problem in graph analysis, community search has been widely studied in the literature and different models have been proposed to measure the cohesiveness of community, such as k-core, k-truss, and clique. In many real-world applications, both graph structures and attribute information are considered. For attribute graph processing, community search problem used both link relationship and attributes because the attributes usually can make communities more meaningful and easy to interpret [5]. In [5], Fang et al. proposed attributed community query (or ACQ) problem, which returned an attributed community (AC) for an attributed graph. The returned community should satisfy both structure cohesiveness constraint and keyword cohesiveness constraint. In [6], Huang and Lakshmanan considered communities based on topics of interest and proposed attributed truss communities (ATC) search problem. They aimed to find connected k-truss subgraphs that contained query vertices with the largest attribute relevance score. In [7], Zhang et al. proposed a keyword-centric community search (KCCS) problem over attribute graphs. They tried to find a community, where the degree of each vertex should be at least k, and the distance between the vertex and all query keywords is minimized. Influential community search has also been studied in [8], where each vertex is associated with a number denoting its influence. Its goal was to find communities with the largest influence.

Bipartite Graph Analysis. Recently, the bipartite graph has attracted much attention due to its proliferate applications like online group recommendation and fraudsters' detection [2]. In [9], Borgatti and Everett were the first to investigate the cohesive communities in bipartite graphs for network analysis. To analyze the properties of bipartite networks, numerous models have been investigated, such as (α, β)-core [10], bitruss [11], and biclique [12]. In [13], the significant (α, β)-community search problem was proposed and studied on weighted bipartite graphs, where each edge is associated with a weight. They aimed to find the significant (α, β)-community that contained query vertex and maximized the minimum edge weight within community. In [4], Wang et al. studied the bitruss model in bipartite graphs. Given a bipartite graph, the bitruss is the maximal subgraph where each edge is contained in at least k butterflies. In the literature, considering the fairness constraints, the fair clustering problems [14–16] were investigated to find communities on bipartite graphs. However, none of the previous studies take the rationality of communities into consideration.

Table 1: Summary of notations.

Notation	Definition
$G = (U, L, E, \mathscr{A})$	An attribute bipartite graph
U/L	The vertex set
E	The edge set
$\mathscr{A} = \{a_1, \ldots, a_t\}$	The attribute set of edges
$S = (U_S, L_S, E_S, \mathscr{A}_S)$	An induced subgraph of G
n	Number of vertices in G
m	Number of edges in G
u, v	Vertex in G
$N_S(u)$	The set of u's neighbors in S
$d_S(u)$	The degree of u in S
α, β	The degree constraint
$x_G(v)$	Consensus score of vertex $v \in L$
x_S	Consensus score of subgraph S
$f(S)$	Rational score of subgraph S

3. Preliminaries

In this section, we first introduce some necessary concepts and present the formal definition of the rational community detection problem over attribute bipartite graphs. Table 1 summarizes the notations that are frequently used in this paper.

3.1. Problem Definition. We consider an attribute bipartite graph $G = (U, L, E, \mathscr{A})$ as an undirected graph without multiple edges and self-loops. U and L are the two disjoint and independent vertex sets in G; that is, $U \cap L = \varnothing$. E is the edge set and each edge $e = (u, v) \in E$ connects one vertex $u \in U$ and one vertex $v \in L$; that is, $E \subseteq U \times L$. $\mathscr{A} = \{a_1, a_2, \ldots, a_t\}$ is the attribute set. Each edge $e \in E$ is

associated with an attribute (e.g., number/tag) $a(e) \in \mathscr{A}$. We use n and m to denote the number of vertices and edges in G, respectively. Given an attribute bipartite graph G, a subgraph $S = (U_S, L_S, E_S, \mathscr{A}_S)$ is an induced subgraph of G; if $U_S \subseteq U, L_S \subseteq L, E_S = E \cap (U_S \times L_S)$ and $\mathscr{A}_S \subseteq \mathscr{A}$. For a vertex $u \in S$, the set of u's neighbors is denoted by $N_S(u)$ (i.e., the adjacent vertices of u). $d_S(u) = |N_S(u)|$ denotes the degree of u in S (i.e., the number of u's neighbor vertices).

Definition 1 ((α, β)-core). Given a bipartite graph G, a subgraph S is the (α, β)-core of G, denoted by $C_{\alpha,\beta}$, if it satisfies the following: (1) degree constraint (i.e., $d_S(u) \geq \alpha$ for each vertex $u \in U_S$ and $d_S(v) \geq \beta$ for each vertex $v \in L_S$); (2) S is maximal; that is, any supergraph S'^S is not a (α, β)-core.

To compute the (α, β)-core, in our paper, we iteratively remove the vertices in two layers violating the corresponding degree constraint until there are no unsatisfied vertices in the graph, the details of which are shown in Algorithm 1. The time complexity is $O(m)$ [17]. As discussed before, the people in a rational discussion group are cohesive and have consistent preference. In the following, we first introduce the consensus score of vertex and community, respectively. Note that we only consider the consensus score of the vertex in lower (e.g., movie) layer. The rational (α, β)-core model is further developed based on the rational score consisting of the consensus score and community size. Then, we present the formal definition of our problem.

Definition 2 (Consensus score). Given an attribute bipartite graph G, the consensus score of each vertex $v \in L$, denoted by $x_G(v)/d_G(v)$, where $x_G(v)$ is the maximum number of its adjacent edges in G with the same attribute number. For a subgraph S of G, its consensus score is defined as $x_S = \sum_{v \in L_S} x_S(v)/d_S(v)/|L_S|$, where $\sum_{v \in L_S} x_S(v)/d_S(v)$ is the sum of consensus score of all vertices in L_S and $|L_S|$ is the number of vertices in the lower layer of S.

Example 2. Considering the vertices in the orange line of the bipartite graph in Figure 1, the consensus score of v_3 is 2/3. The consensus score of community in the orange line is 8/9.

To judge a community, we not only want to consider the consensus but also want to consider the size constraint of it. This is because that the traditional study group with not very large size can facilitate people there to discuss and analyze problem. So, we also combine the size constraint into our rational score function, which is expressed as follows:

$$f(S) = \lambda \frac{\sum_{v \in L_S} x_S(v)/d_S(v)}{|L_S|} + (1 - \lambda) \frac{1}{|U_S||L_S|}, \quad (1)$$

where λ is a parameter to make the trade-off between the consensus score and the community size. Based on this rational score function, we give the definition of rational community.

Definition 3 (rational (α, β)-core). Given an attribute bipartite graph G and two positive integers α and β, a subgraph S is a attribute (α, β)-core of G, denoted by $RC_{\alpha,\beta}$, if it meets the following three criteria:

(i) Connectivity: S is connected

(ii) Cohesiveness: S is a (α, β)-core

(iii) Rationality: S has the largest rational score $f(S)$ among subgraphs satisfying the above criteria

3.1.1. Problem Statement. Given an attribute bipartite graph G and two positive integers α and β, we aim to develop efficient algorithms to find the rational (α, β)-core (i.e., the (α, β)-core with the largest rational score).

3.2. Problem Properties. As shown in Theorem 1, the problem studied is NP-hard. Besides, the rational score function is nonmonotonic and non-submodular, whose details are in Theorem 2.

Theorem 1. *Given an attribute bipartite graph G, the problem of computing the rational (α, β)-core is NP-hard.*

Proof. When $\alpha > 0$ and $\beta > 0$, we reduce the biclique problem [17] to RCD-ABG problem. Given an attribute bipartite graph $G = (V = (U \cup L), E, \mathscr{A})$, where for each vertex in lower layer L, its adjacent edges have distinct attribute. This means that given a subgraph S of G, the consensus score of each vertex v in L_S is $1/d_S(v)$. Hence, our score function is converted to $f(S) = \lambda \sum_{v \in L_S} 1/d_S(v)/|L_S| + (1 - \lambda)1/|U_S||L_S|$. In order to make the rational score large, for the first term of function, namely, $\lambda \sum_{v \in L_S} 1/d_S(v)/|L_S|$, we need to make the numerator be largest and the denominator be smallest. Due to the degree constraint of lower layer, the lower bound of $d_S(v)$ is β. So, the rational score function is $f = \lambda|L_S|1/\beta/|L_S| + (1 - \lambda)1/|U_S||L_S| = \lambda 1/\beta + (1 - \lambda)1/|U_S||L_S|$. Given the parameter α, β, and λ, to find rational (α, β)-core with largest f, $|U_S|$ and $|L_S|$ need to be minimized, which means that $|U_S|$ and $|L_S|$ should be equal to β and α, respectively. As discussed, each vertex $u \in U_S$ (resp. $u \in L_S$) should satisfy $d_S(u) \geq \alpha$ (resp. $d_S(u) \geq \beta$). This is a biclique that each vertex in different layers is connect, which is NP-hard [17]. Therefore, our problem is NP-hard. □ □

Theorem 2. *The objective score function $f(S)$ is nonmonotonic and non-submodular.*

Proof. Nonmonotonic. By considering the example in Figure 1, we first prove its nonmonotonicity. Note that we only keep two decimal places in the following. Suppose $\lambda = 0.5$; we can see that in subgraph denoted by solid line, that is, $S = \{u_2, u_3, u_4, v_2, v_3, v_4\}$, $f(S) = 0.5$. After deleting vertex u_2, $f(S/\{u_2\}) = 0.53$. While, by further deleting vertex u_4, the present score is $f(S \backslash \{u_2\} \backslash \{v_2\}) = 0.5$. Therefore, the function is nonmonotonic.

Non-Submodular. Given two sets A and B, $f(x)$ is submodular if $f(A \cup B) + f(A \cap B) \leq f(A) + f(B)$. We show the inequality does not hold by a counterexample in Figure 1. Suppose $A = \{(u_2, u_3, v_2, v_3)\}$ and $B = \{(u_3, u_4, v_3, v_4)\}$.

We have $f(A) = 0.5$, $f(B) = 0.5$, $f(A \cup B) = 0.5$, and $f(A \cap B) = 0.75$. Thus, the equation does not hold and f is not submodular. □

4. Solution

In this section, a greedy framework is firstly developed to find the result, which is based on the concept of score function and marginal gain that we define. Considering the limitations of the basic method, we further propose two novel strategies with better quality.

4.1. A Basic Greedy Framework (BGF). Intuitively, to find the (α, β)-core with largest score, we can delete those vertices whose deletion will increase the score. Based on this, we present our basic greedy framework by introducing the rational marginal gain as follows.

Definition 4. (rational marginal score). Given an attribute bipartite graph G and a vertex $u \in G$, the rational marginal gain is defined as

$$\triangle_G(u) = \begin{cases} f(G) - f\left(\dfrac{G}{(N_G'(u) \cup \{u\})}\right) u \in U. \\ \\ f(G) - f\left(\dfrac{G}{\{u\}}\right) u \in L, \end{cases} \tag{2}$$

where $N_G'(u)$ is the set of u's neighbors in L that violate the degree constraint after removing vertex u.

4.1.1. The Basic Greedy Framework (BGF). The details of BGF are illustrated in Algorithm 2, which includes three main steps. We use $\mathscr{G}$ to denote the set of all connected (α, β)-cores. *Step 1.* We find all (α, β)-cores of G and store them into $\mathscr{G}$ in Line 2. We use $\mathscr{G}_i$ to denote the current processing (α, β)-core. *Step 2.* At each iteration, we greedily peel the vertex v in graph $\mathscr{G}_i$ providing the smallest marginal gain $\triangle_{\mathscr{G}_i}(v) = f(\mathscr{G}_i) - f(\mathscr{G}_i/\{v\})$ (Line 5), which is called the best vertex. After removing best vertex, we calculate the (α, β)-core in the remaining graph. If there are many connected (α, β)-cores, we push back them into $\mathscr{G}$ (Lines 6–8). We continue this process until there is no (α, β)-core in the graph. Note that the rational marginal gain may be a negative number, which means the score of function increase. *Step 3.* We output the result with the largest score among obtained attribute (α, β)-cores (Line 9).

Example 3. Considering the user-movie network in Figure 1. Suppose $\alpha = 2, \beta = 2$. According the BGF, vertex v_6 is removed firstly and the rational score of the remained $(2, 2)$-core is 0.415476. Similarly, we remove vertices v_7, v_4, and v_3, iteratively. The corresponding rational score is 0.481667, 0.6, and 0. Therefore, the returned result is $\{u_2, u_3, v_2, v_3\}$ with rational score of 0.6.

4.2. Optimized Strategies. The basic greedy framework is simple but suffers from the following drawback. When removing a vertex v from the subgraph S, it may make the support of some other vertices decrease and lead them to drop from the community in succession. Note that these vertices are called the followers of v including v itself, denoted as $\mathscr{F}_S(v)$. If the removal vertex has a large number of followers, it can severely limit the effectiveness of the algorithm. Hence, we need to consider the effect of each removal vertex. In the following section, we propose two improved strategies to handle the limitation.

4.2.1. 2-Hop Neighbors Optimization (OS-I). As observed, if the removal vertex is in the lower layer, its 2-hop neighbors in the same layer may violate the degree constraint and be deleted, which significantly affect the rational score. Based on this, we use the following equation to approximate marginal score function $\triangle_G(u)$ by $\widehat{\triangle}_G(u)$,

$$\widehat{\triangle}_G(u) = f(G) - f\left(\frac{G}{(N_G'(u) \cup \{u\})}\right) u \in U, f(G) - f\left(\frac{G}{(H2_G(u) \cup \{u\})}\right), \tag{3}$$

where $H2_G(v)$ is the 2-hop neighbors of v in the lower layer. Therefore, the best vertex is adjusted as $u \longleftarrow \text{argmin}_{v \in \mathscr{G}_i} \widehat{\triangle}_{\mathscr{G}_i}(v)$ in Line 5 of algorithm 1 and other steps are the same.

Example 4. Reconsider the user-movie network in Figure 1. Suppose $\alpha = 2$, $\beta = 2$. According to the OS-I, we remove vertex v_7 firstly and obtain the rational score of the remained $(2, 2)$-core. Then, we remove u_6 and obtain the corresponding score of 0.6556. After removing u_3, the obtained score is 0. So, we return the result by $\{u_2, u_3, u_4, v_2, v_3, v_4\}$ with a score of 0.6556.

4.2.2. Followers-Based Optimization (OS-II). The second idea is motivated by the followers of each removal vertex. Generally, instead of removing one vertex and calculating the rational marginal gain, we remove a vertex with its all followers from the current candidate graph that have the smallest attribute marginal gain. Hence, the marginal score is modified as the following equation:

$$\widetilde{\triangle}_G(u) = f(G) - f\left(\frac{G}{\mathscr{F}_G(u)}\right), \tag{4}$$

and the other steps are the same as BGF.

Input: G: a bipartite graph, α, β: degree constraints
Output: The (α, β)-core of G
(1) **While** exists $u \in U$ with $d(u) < \alpha$ or $u \in V$ with $d(u) < \beta$ **do**
(2) $G \longleftarrow G/\{u\}$
(3) **return** G

Algorithm 1: Compute (α, β)-core.

Input: G: attribute bipartite graph, α: degree constraint in upper layer, β: degree constraint in lower layer
Output: H: the connected (α, β)-core with the largest rational score
(1) $i \longleftarrow 1$
(2) $\mathcal{G} \longleftarrow$ an empty vector
//Step 1
(3) $\mathcal{G} \longleftarrow$ all connected $C_{\alpha,\beta}(G)$
//Step 2
(4) **While:** $\mathcal{G} \neq \emptyset$ **do**
(5) $u \longleftarrow \text{argmin}_{v \in \mathcal{G}_i} \triangle_{\mathcal{G}_i}(v)$
(6) **for each** connected $C_{\alpha,\beta}(\mathcal{G}_i/u)$ in $\mathcal{G}_i \backslash u$ **do**
(7) push back $C_{\alpha,\beta}(\mathcal{G}_i/u)$ into $\mathcal{G}$
(8) $i \longleftarrow i + 1$;
//Step 3
(9) $H \longleftarrow \text{arg}G'_{G_i \in \mathcal{G}} f(G')$;

Algorithm 2: A basic greedy framework (BGF).

Example 5. Reconsider the example in Figure 1. Suppose $\alpha = 2$, $\beta = 2$. According to the OS-II, vertex u_7 is removed firstly and the obtained score is 0.455333. Then, we remove v_5 and calculate the score with 0.65556. After deleting u_2, the score is 0. Therefore, we return the result $\{u_2, u_3, u_4, v_2, v_3, v_4\}$ with a score of 0.65556.

4.2.3. Analysis. The main difference between BGF and optimized algorithm is the best vertex. In BGF, calculating the marginal score of a vertex is $O(1)$ time. In OS-II, the time complexity of identifying the followers of the vertex is $O(m)$, which may significantly increase the running time.

5. Experiments

5.1. Algorithms. To the best of our knowledge, there is no existing work for RCD-ABG problem. In the experiments, we implement and evaluate the following algorithms.

(i) *BGF*. The baseline greedy framework is presented in Algorithm 2, which iteratively peels the graph and returns the best result during the search

(ii) *OS-I*. OS-I leverages the baseline framework BGF and further integrates the proposed 2-hop neighbor-based optimization

(iii) *OS-II*. OS-II leverages the baseline framework BGF and further integrates the proposed follower-based optimization

(iv) *ORI*. To evaluate the advantage of proposed model, we also implement the traditional (α, β)-core search method [10], which iteratively removes the vertex that violates the degree constraints and returns the final subgraph

Table 2: Statistics of datasets.

Dataset	$\lvert U\rvert$	$\lvert L\rvert$	$\lvert E\rvert$	$\lvert\mathcal{A}\rvert$
HetRec (HR)	2,101	18,746	92,835	5
CiaoDVD (CD)	17,615	16,121	72,345	5
TripAdvisor (TA)	145,316	1,759	175,655	5
MovieLens (ML)	71,535	65,134	855,598	5
BookCrossing (BC)	278,855	270,981	941,148	10
Personality (PY)	1,822	198,118	1,028,751	5

5.2. Datasets and Workloads. We employ 6 real-world bipartite graphs. Among these datasets, CiaoDVD and TripAdvisor can be obtained on KONECT (https://konect.uni-koblenz.de). Other datasets are publicly available on GroupLens (https://grouplens.org/datasets/). The statistics of datasets are shown in Table 2, where $|\mathcal{A}|$ is the number of attributes in bipartite graphs. HetRec (HR) [18] is a user-artists network, where the attribute of edges denotes the number of time that user listens to the music by the artist. CiaoDVD (CD) and MovieLens (ML) [18] are user-movie networks of which the attributes of relationships represent the ratings for movie. TripAdvisor is a user-hotel bipartite graphs and the attribute of its edges denotes the rating taken by users. The BookCrossing (BC) is a user-book network and the edges of it denote the book-rating taken by user. Due to the density of graphs, $\alpha = \beta$ vary from 5 to 25 in HetRec, CiaoDVD, and TripAdvisor, vary from 15 to 35 in BookCrossing and vary from 50 to 250 in MovieLens and Personality. λ is set as 0.7 because the density of community will

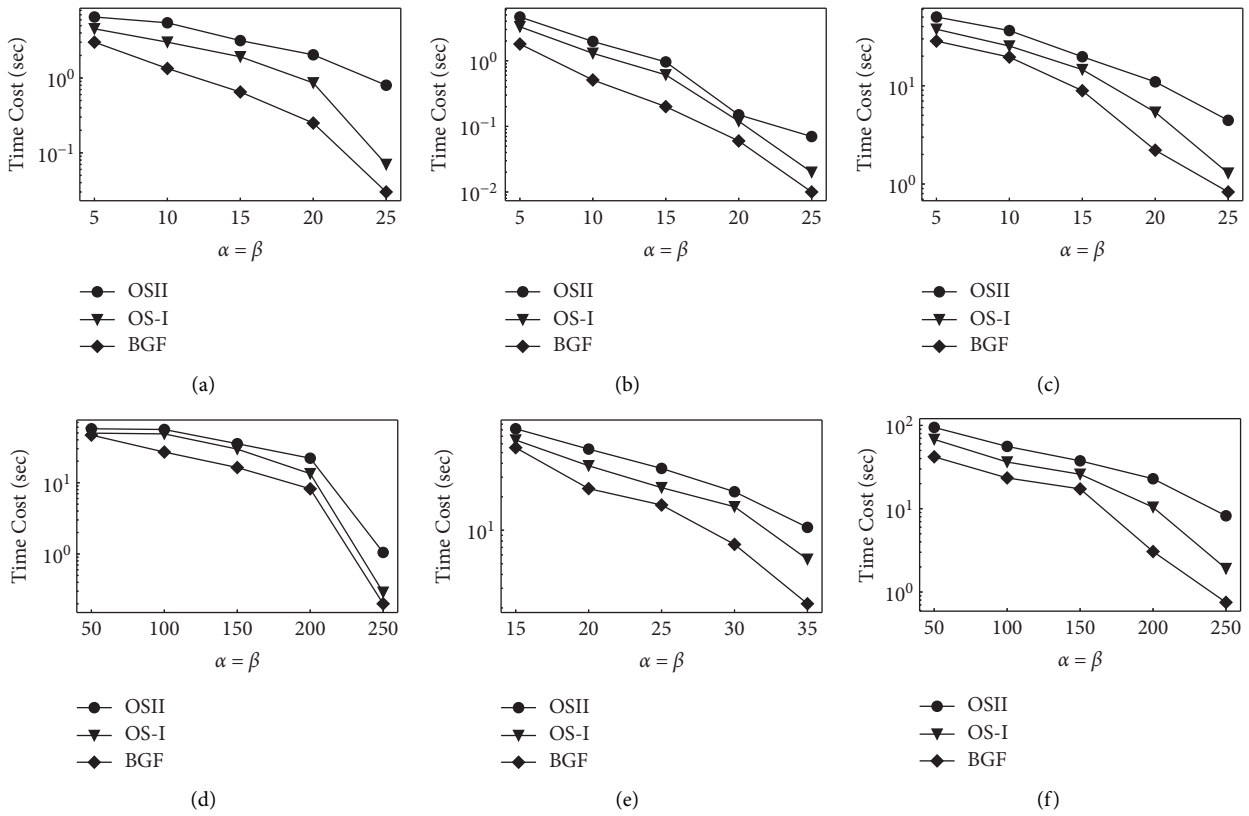


Figure 2: Efficiency evaluation by varying parameters α and β. (a) HetRec. (b) CiaoDVD. (c) TripAdvisor. (d) MovieLens. (e) BookCrossing. (f) Personality.

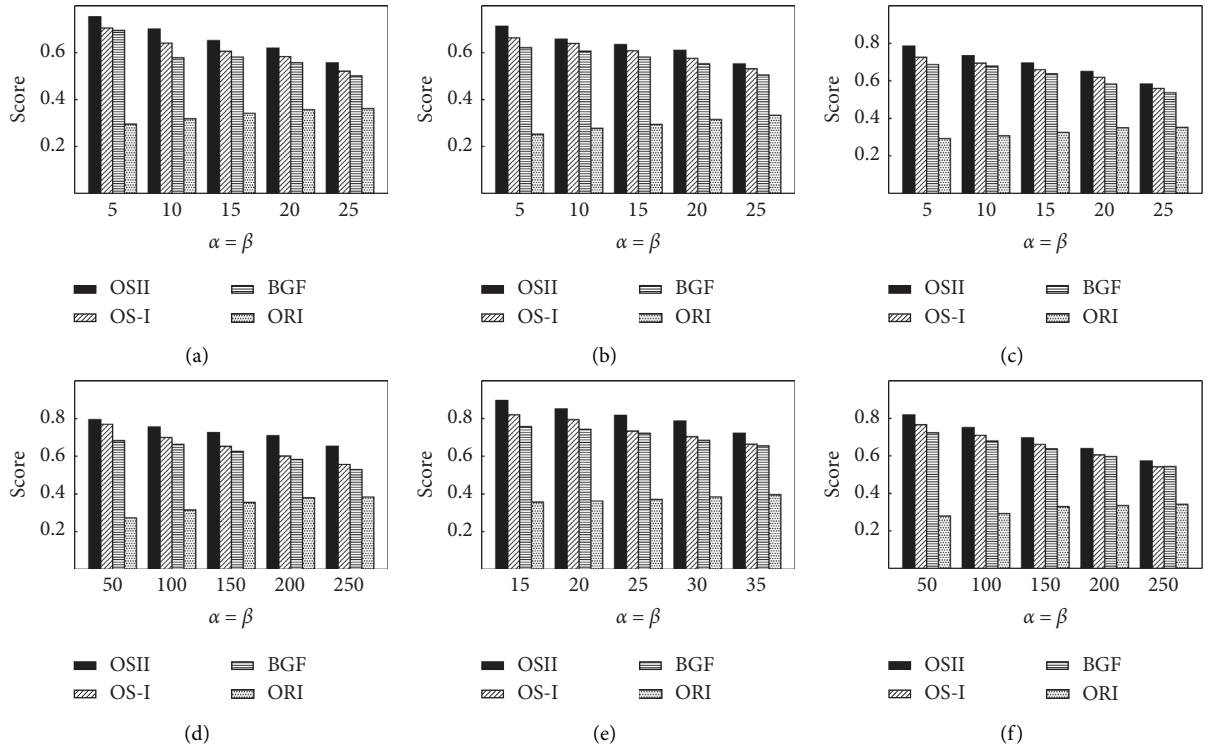


Figure 3: Effectiveness evaluation by varying parameters α and β. (a) HetRec. (b) CiaoDVD. (c) TripAdvisor. (d) MovieLens. (e) BookCrossing. (f) Personality.

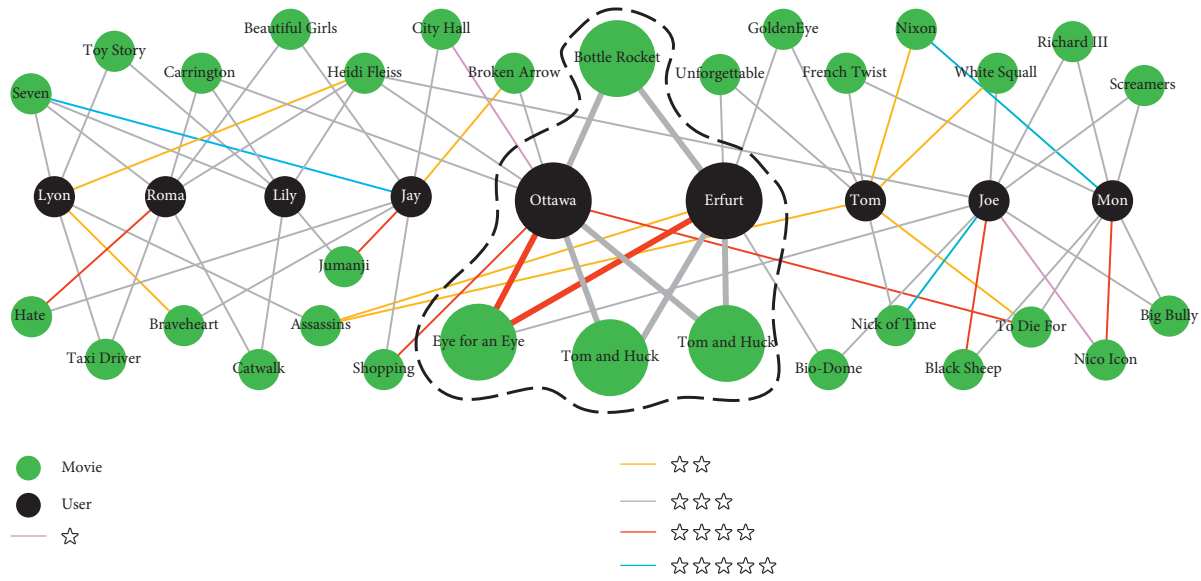


Figure 4: Case study.

strengthen with the continuous deletion of vertices; thus, we focus on consensus score. All the programs are implemented in standard C++. All the experiments are performed on a server with an Intel Xeon 2.4 GHz CPU and 128 GB main memory.

5.3. Efficiency Evaluation. To evaluate the efficiency, we report the response time of algorithms by varying α and β in Figures 2(a)–2(f). As observed, the time cost of OS-I and OS-II is more than BGF. This is because OS-I needs to calculate 2-hop neighbors of vertex in the lower layer and OS-II needs to calculate followers of vertex. Although the time complexity of OS-I and OS-II is more complex than BGF, there is not much difference of response time between them. We can observe that when α and β increase, the response time decreases for all methods. This is because the community size decreases.

5.4. Effectiveness Evaluation. To evaluate the effectiveness, we compare BGF, OS-I, and OS-II with ORI and report the rational score of the returned community. ORI is based on the traditional (α, β)-core model. It first computes the (α, β)-core of the graph and then directly returns the connected component with the largest rational score. The results are shown in Figures 3(a)–3(f). We can observe that original (α, β)-core has very small rational score. OS-I and OS-II significantly outperform BGF over all the datasets, namely, find community with higher score than BGF. The score returned by OS-I is at least 0.01 higher than the one returned by BGF in all datasets. Due to the feature of the consensus score, the improvement of OS-I is already significant for the overall performance. The rational score decreases when α and β increase because of tighter degree constraint.

5.5. Case Study. To further evaluate the advantage of the proposed model, we conduct a case study on HetRec dataset. The results are shown in Figure 4. As shown, the movie and user are marked with different colors. The different-color edges denote different scores. The community in the solid line that consists of enlarged vertices and bold edges is the returned result. As we can see, it can find a more rational community with a high preference and density structure.

6. Conclusion and Future Work

In this paper, we propose and investigate the rational (α, β)-core detection problem in attribute bipartite graphs. We formally define the problem and prove its NP-hardness. To solve this problem, a basic greedy framework is first presented, which iteratively removes the best vertex with the smallest marginal gain and calculate the remaining (α, β)-core. Two optimized strategies, namely, 2-hop neighbor-based optimization and follower-based optimization, are proposed to improve the performance. Experiments are conducted on real bipartite graphs to demonstrate the advantages of proposed model and techniques. As shown in the experience, the proposed model significantly outperforms the traditional (α, β)-core model. In real-world applications, there are also attributes within the vertices of the graphs. In the further work, we will consider more complex scenario to design the model and the corresponding approaches.

Data Availability

The datasets in this paper are publicly available at https://konect.uni-koblenz.de and https://grouplens.org/datasets/.

Conflicts of Interest

The authors declare that they do not have any commercial or associative interest that represents conflicts of interest in connection with the work submitted.

Acknowledgments

This work was supported by ZJSSF 21NDQN247YB, ZJNSF Y202045024, ZJNSF LQ20F020007, and ZJNSF LY21F020012.

References

[1] B. Liu, L. Yuan, X. Lin, L. Qin, W. Zhang, and J. Zhou, "Efficient (a,β) core computation: an index-based approach," in *Proceedings of The World Wide Web Conference*, pp. 1130–1141, San Francisco, CA, USA, May 2019.

[2] E. Ntoutsi, K. Stefanidis, K. Rausch, and H. P. Kriegel, "Strength lies in differences": diversifying friends for recommendations through subspace clustering," in *Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management, CIKM 2014*, pp. 729–738, ACM, Shanghai, China, November 2014.

[3] H. Yang, M. Zhang, X. Wang, and C. Chen, "Cohesive subgraph detection in large bipartite networks," in *Proceedings of the International Conference on Scientific and Statistical Database Management*, pp. 1–4, Bolzano, Italy, July2020.

[4] K. Wang, X. Lin, Q. Lu, W. Zhang, and Y. Zhang, "Efficient bitruss decomposition for large-scale bipartite graphs," in *Proceedings of the 36th IEEE International Conference on Data Engineering, ICDE 2020*, pp. 661–672, Dallas, TX, USA, April 2020.

[5] Y. Fang, R. Cheng, Y. Chen, S. Luo, and J. Hu, "Effective and efficient attributed community search," *The VLDB Journal*, vol. 26, no. 6, pp. 803–828, 2017.

[6] X. Huang and L. V. S. Lakshmanan, "Attribute-driven community search," *Proceedings of the VLDB Endowment*, vol. 10, no. 9, p. 949, 2017.

[7] Z. Zhang, X. Huang, J. Xu, B. Choi, and Z. Shang, "Keyword-centric community search," in *Proceedings of the 35th IEEE International Conference on Data Engineering*, pp. 422–433, Macao, China, April 2019.

[8] R. Li, L. Qin, J. X. Yu, and R. Mao, "Influential community search in large networks," *Proceedings of the VLDB Endowment*, vol. 8, no. 5, pp. 509–520, 2015.

[9] S. P. Borgatti and M. G. Everett, "Network analysis of 2-mode data," *Social Networks*, vol. 19, no. 3, pp. 243–269, 1997.

[10] D. Ding, H. Li, Z. Huang, and N. Mamoulis, "Efficient fault-tolerant group recommendation using alpha-beta-core," in *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*, November 2017.

[11] Z. Zou, "Bitruss decomposition of bipartite graphs," in *Proceedings of the Database Systems for Advanced Applications - 21st International Conference*, pp. 218–233, DASFAA, Dallas, TX, USA, April 2016.

[12] Y. Zhang, C. A. Phillips, G. L. Rogers, E. J Baker, E. J Chesler, and M. A Langston, "On finding bicliques in bipartite graphs: a novel algorithm and its application to the integration of diverse biological data types," *BMC Bioinformatics*, vol. 15, no. 110, p. 110, 2014.

[13] K. Wang, W. Zhang, X. Lin, Y. Zhang, L. Qin, and Y. Zhang, "Efficient and effective community search on large-scale bipartite graphs," in *Proceedings of the 37th IEEE International Conference on Data Engineering*, pp. 85–96, ICDE, Chania, Greece, April 2021.

[14] F. Chierichetti, R. Kumar, S. Lattanzi, and S. Vassilvitskii, "Fair clustering through fairlets," in *Proceedings of the Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems*, pp. 5029–5037, Long Beach, CA, USA, December 2017.

[15] S. Ahmadian, A. Epasto, R. Kumar, and M. Mahdian, "Fair correlation clustering," in *Proceedings of the The 23rd International Conference on Artificial Intelligence and Statistics, AISTATS 2020*, pp. 4195–4205, PMLR, sicily, Italy, August 2020.

[16] H. Larochelle, M. A. Ranzato, R. Hadsell, M. F. Balcan, and H. T. Lin, "Fair hierarchical clustering," in *Proceedings of the Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems 2020*, virtual, La Jolla CA, USA, December 2020.

[17] B. Lyu, L. Qin, X. Lin, Y. Zhang, Z. Qian, and J. Zhou, "Maximum biclique search at billion scale," *Proceedings of the VLDB Endowment*, vol. 13, no. 9, pp. 1359–1372, 2020.

[18] I. Cantador, B. Peter, and T. Kuflik, "2nd workshop on information heterogeneity and fusion in recommender systems (hetrec 2011)," in *Proceedings of the 2nd International Workshop on Information Heterogeneity and Fusion in Recommender Systems, HetRec '11*, ACM, Chicago, IL, USA, October 2011.

Hindawi
Mathematical Problems in Engineering
Volume 2022, Article ID 1169349, 10 pages
https://doi.org/10.1155/2022/1169349

Research Article

A Color Spot Extraction Method Based on the Multifused Enhancement Algorithm

Xiuzhi Zhao

Zhejiang Industry & Trade Vocational College, Zhejiang 325002, China

Correspondence should be addressed to Xiuzhi Zhao; cassyzxz@126.com

Received 10 November 2021; Accepted 2 March 2022; Published 15 April 2022

Academic Editor: Chunlai Chai

The color spot can reflect the skin status and the physiological change. It has been used in the evaluation of skin quality and medical diagnosis. However, the gray level of the color spot is very similar to that of normal skin. Because the shape, size, and position of the color spots are irregular. It is difficult to extract the accurate region of the color spot with the traditional methods. In order to extract the region of the color spot, we propose a multifused enhancement algorithm to enhance the feature of the color spot and extract the accurate region. The multifused enhancement algorithm mainly includes three components: the polarized image acquirement, the color model transformation, and the wavelet transform and enhancement. The color spot, which is in the intraepidermal basal cell, is captured by the polarized camera with the polarized light. The acquired method can remove the reflective light, covering the color spot. To further enhance the feature of the color spot, we use saturation instead of the gray level to characterize the color spot. The color spot is more conspicuous in the saturation image. Then, the wavelet transform and enhancement are used to enhance the feature of the color spot and reduce the effect of uneven illumination. The color spot is extracted by the proposed self-adaptive segmentation method. In addition, we discuss the performance with the different wavelets to choose the optimal wavelet. The actual reason for removing uneven illumination in the proposed algorithm is also analyzed in this work. The experimental results indicate that the proposed algorithm achieves the best accuracy among all the algorithms compared.

1. Introduction

The skin is the largest organ of humans. It protects the body from external harm. The functions of thermoregulation, tactile sensation, and excretion are also realized by the skin of humans [1–3]. The skin's water content occupies 15% of the body weight, and the skin is one of the most important organs in the body. With the change in age and the health status, the skin will have a physiological change [4–6]. The evaluation of the skin has attracted much attention by the doctors and the beauty and skin care industries [7–9]. Due to the importance and complexity of the skin, the analysis of the skin has some challenges. The traditional analysis and diagnosis of the skin are based on the subjective observation of the doctors. The subjective evaluation cannot provide the quantitative analysis results of the skin. Furthermore, some features of the skin cannot be observed by the naked eyes of the doctors. Thus, an objective evaluation method of the skin is required by current medical diagnosis and cosmetology.

Some indicators can reflect the change of the skin, such as pore [10], wrinkle [11, 12], roughness [13], and the color spot [14] of the skin. The pore is the porous structure in which the pelage is deciduous. The wrinkle is the small microgroove generated by the free radical that breaks the collagen and active substance of the normal cells. The roughness is used to evaluate the texture of the skin. The color spot caused by the accumulation of melanin is the spot in which the color is different from that of the surrounding skin [3, 15]. These features of the skin can reflect the change of the skin due to the aging process [16]. The research of the skin indicators has an important meaning in the objective evaluation.

The color spot is one of the most important indicators in skin evaluation. It includes freckle [17, 18], black spot [19],

chloasma [20, 21], and age pigment [22]. The color spot is a pigmentary skin disorder caused by the increasing melanin of the skin [23, 24]. The main reason for the freckle generation is the hereditary factor. Enough exposure to ultraviolet radiation and the metabolism of the broken skin also lead to the freckle. Except for the ultraviolet radiation, an unhealthy lifestyle may result in the black spot. The imbalance of the healthy condition of the body and the increase of the male progesterone after the pregnancy may cause chloasma. In addition, some chronic diseases, such as chronic hepatitis and tuberculosis, may multiply the melanin [25, 26]. When the melanin cannot be eliminated from the body, the melanin may turn into chloasma. Cosmetics misuse is also the important reason for the black spot and chloasma. The seborrheic keratosis, which is also called age pigment, is a benign skin tumor. The generated reason is the cell proliferation of the corneum. According to medical statistics, melanoma's morbidity and mortality rate constantly improve. The melanoma comes from the pigmented spot that is easy to be ignored by humans. The melanoma may be diagnosed earlier if the color spots can be quantitatively and objectively analyzed. The traditional diagnosis method of the color spot is based on the visual inspection of the doctors. Hence, the inspected accuracy relies on the experience of the doctors. The subjective and manual method is not only inefficient but also inaccurate. It is difficult for a doctor to quantify the size of the area of the color spot precisely.

There are some challenges for the color spot extraction: (1) the gray-level difference between the color spot and the normal skin is too small. This problem may cause that the color spot is difficult to be recognized in the spatial and frequency domains; (2) the illumination is uneven. The uneven illumination may cause the gray-level distribution to be uneven in the skin. The region of the weak illumination can be falsely recognized as the color spot, and (3) the captured image cannot obtain the conspicuous information of the color spot because the images of the color spots are captured by the camera with a macrolens. The obtained image is fuzzy and cannot provide a conspicuous color spot. Hence, the color spot is in the intraepidermal basal cell. The camera cannot capture an image that contrasts the color spot under the normal light. Thus, we propose the multifused enhancement algorithm to obtain the conspicuous color spot.

Due to the challenges of color spot extraction, many existing segmentation algorithms may not work well for segmenting this type of image. Some methods can be used to detect and quantify the color spot but cannot achieve good performance. The threshold segmentation [27, 28] is the most traditional method for image segmentation. However, the gray level is very similar between the normal skin and the color spot, and the method cannot achieve good performance. According to the characteristic of the color spot, the method based on HSV [29, 30] can be used to segment the color spot. Although the performance is better than the threshold segmentation, the performance is limited by the uneven illumination of the image. In order to enhance the color spot from the background in the image, homomorphic filtering [31–33] is used to enhance the color spot. The method weakens the low-frequency signal and enhances the high frequency for improving the image detail. However, the method cannot achieve the preferable performance in complex circumstances. The method of wavelet transform [34–36], which can obtain the multilevel signal with the decomposition algorithm, is used to enhance the color spot. Due to the multivariable parameters used, it is not easy to obtain stable performance.

In this work, we propose the multifused enhancement algorithm to extract the color spot. To obtain conspicuous features of the color spot, we propose three methods to enhance the weak features of the color spot. The polarized image acquirement, the HSV model, and the wavelet transform are fused to extract the conspicuous features of the color spot. The experimental results indicate that the proposed algorithm can achieve excellent performance for color spot enhancement.

The remainder of this article is organized as follows: Section 2 introduces the proposed methods for color spot enhancement and segmentation. Section 3 shows the experimental results and analysis of the proposed methods. Section 4 discusses the significance of the color spot extraction and its possible improvement. The conclusion then follows.

2. Methods

The key issue of color spot extraction is how to enhance the color spot. In order to enhance the difference between the color spot and the normal skin, the three methods are proposed to obtain the conspicuous color spot. First, we use the polarized camera with the polarized light to obtain the conspicuous feature of the color spot. Secondly, the saturation features of the image are used to express the color spot. Thirdly, the wavelet transform is used to enhance the feature of the color spot. Finally, the color spot is extracted by the proposed self-adaptive threshold segmentation. The schematic diagram of the proposed method is shown in Figure 1.

3. Polarized Image Acquirement

Generally, the color spot is in the intraepidermal basal cell. The reflective light of the epidermis will cover the texture of the color spot. Thus, we adopt the polarized light to remove the reflective light of the epidermis. The horizontally polarized light provides the lighting source, and the camera with a vertically polarized glass is used to capture the image of the color spot. The experimental result is shown in Figure 2. It is difficult to recognize the color spot in the ordinary image. But the color spot is conspicuous in the polarized image.

Under ordinary light, the surface of the skin reflects a large number of reflective lights, and the captured image has a large number of reflective lights. When the light is transformed to the horizontally polarized light, and a vertically polarized glass covers the lens of the camera, the reflective light from the skin surface is eliminated. The

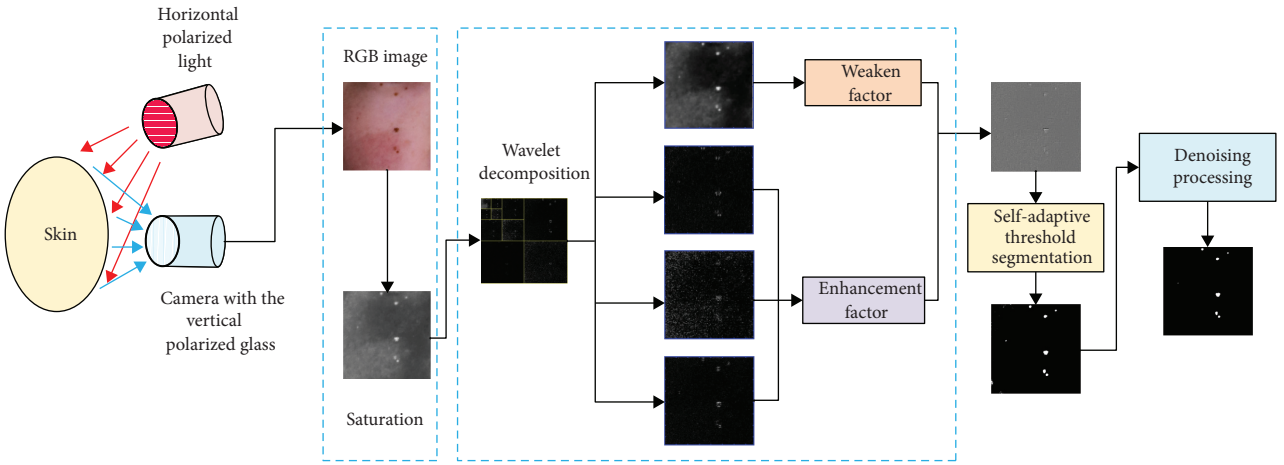
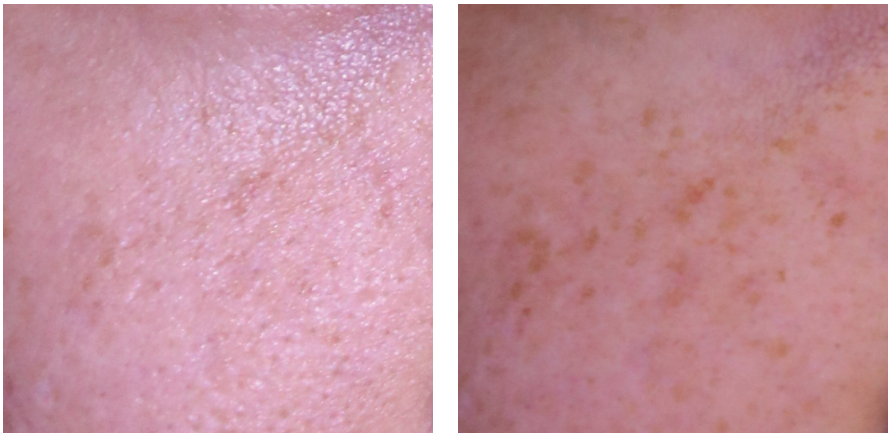


Figure 1: The schematic diagram of the proposed method.



(a) (b)

Figure 2: The image acquirement with the polarized light. (a) Ordinary image. (b) Polarized image.

captured light mainly comes from the subepidermal light, and the camera can obtain the conspicuous color spot.

4. Color Model Transformation

Although the contrast of the color spot is improved by the polarized image acquirement, the color spot is still tricky to be segment. In the gray-level space, the gray-level difference between the color spot and the normal skin is too little. By analyzing the different color models, we find that the saturation feature of the color spot is more visible. In our work, the captured image is transformed to the HSV (hue, saturation, and value) model before processing the image, and we use the saturation to express the feature of the color spot.

Based on the transformation principle between RGB and HSV, the conspicuousness of the saturation can be explained. The saturation is obtained by the following equation:

$$\text{saturation} = 1 - 3 * \min(r, g, b), \quad (1)$$

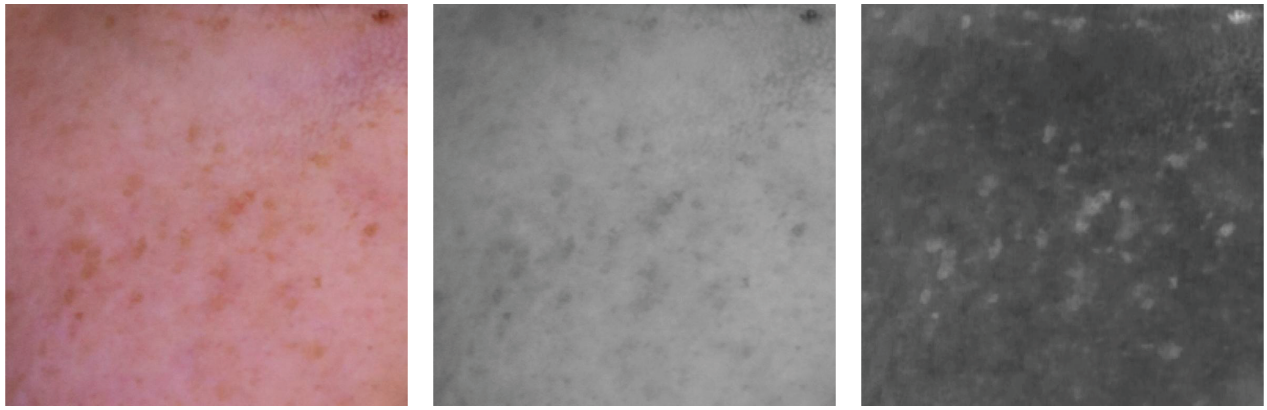
where r, g, and b are the normalized value of the R, G, and B value in the RGB model, respectively. The saturation is the maximum of (1), and the conspicuousness of the saturation is larger than the gray level.

From Figure 3, the original image, gray-level image, and saturation image are given. We can obviously observe that the contrast of the saturation is better than the gray-level image. The result also can be used to verify Equation (1).

5. Wavelet Transform and Enhancement

The information of the color spot is usually included in the high frequency of the image. The other information except the color spot is in the low frequency. The color spot may be enhanced if we enhance the high frequency and waken the low frequency. In this work, we use the wavelet transform algorithm to achieve this goal.

Suppose the given image with the size of $m \times n$ is $f(x, y)$, the wavelet decomposition is shown as follows:



(a) (b) (c)

Figure 3: The comparison of different lights. (a) Original image. (b) Gray-level image. (c) Saturation image.

$$f_{k+1}(x,y) = f_k(x,y) + g_k(x,y),$$
$$g_k = g_k^H + g_k^V + g_k^D, \tag{2}$$

where $f_k(x,y)$ is the low-frequency component, and the high frequency is indicated by $g_k(x,y)$. The horizontal, vertical, and diagonal components of the high frequency are indicated by the g_k^H, g_k^V, and g_k^D, respectively. k is the decomposition level.

According to the principle of the multiresolution analysis, it has $f_k(x,y) \in V_k$ and $g_k^I(x,y) \in W_k$. And $\varphi(2^k x - m, \quad 2^k y - n)$ and $\psi^I(2^k x - m, \quad 2^k y - n)$ are the Riesz base of the space V_k and W_k, respectively. The W_k is the complementary set of V_k. The $f_k(x,y)$ and $g_k^I(x,y)$ can be described as follows:

$$f_k(x,y) = \sum_{m,n} \varepsilon_{k;m,n} \varphi\left(2^k x - m, 2^k y - n\right),$$
$$g_k^I(x,y) = \sum_{m,n} \eta_{k;m,n}^I \psi^I\left(2^k x - m, 2^k y - n\right), \tag{3}$$

where φ and ψ are scaling function and wavelet function, respectively. In addition, $I \in (H, V, D)$. The coefficients of the $\varepsilon_{k;m,n}$ and $\eta_{k;m,n}^I$ can be described as the following equation based on the decomposition algorithm:

$$\varepsilon_{k;m,n} = \sum_{l,j} \gamma_{l-2m,j-2n} \varepsilon_{k+1;m,n},$$
$$\eta_{k;m,n}^I = \sum_{l,j} \delta_{l-2m,j-2n}^I \varepsilon_{k+1;m,n}, \tag{4}$$

where $\gamma_{k;m,n}$ and $\delta_{k;m,n}^I$ are the binary decomposition sequences that is generated from the unary decomposition sequences.

To enhance the color spot and weaken other information, we use (5) to enhance the high-frequency information and weaken the low-frequency information.

$$\Phi_{k;m,n} = \alpha * \varepsilon_{k;m,n}, \tag{5}$$

$$\Psi_{k;m,n}^I = \beta * \eta_{k;m,n}^I, \tag{6}$$

where the weaken factor is α and the β is the enhancement factor.

Suppose the $\mu_{l,j}$ and $\nu_{l,j}^I$ are the binary decomposition sequences generated from two unary decomposition sequences. The reconstruction algorithm is shown in the following:

$$\varepsilon_{k+1;m,n} = \sum_{l,j} \left(\mu_{m-2l,n-2j} \Phi_{k;l,j} + \sum_{i=1}^{3} \nu_{m-2l,n-2j}^i \Psi_{k;l,j}^I \right). \tag{7}$$

6. Self-Adaptive Threshold Segmentation

By the enhancement methods, we can obtain the more conspicuous color spot. To extract the accurate region of the color spot, we propose the self-adaptive threshold segmentation. First, we calculate the maximum and minimum of $\varepsilon_{k+1;m,n}$, and the self-adaptive threshold is obtained as shown in the following equation:

$$\text{MAX} = \max\left(\varepsilon_{k+1;m,n}\right), \tag{8}$$

$$\text{MIN} = \min\left(\varepsilon_{k+1;m,n}\right), \tag{9}$$

$$\text{THRESHOLD} = \frac{\text{MAX} + \text{MIN}}{2}. \tag{10}$$

The self-adaptive threshold method can obtain the threshold without manual adjustment. The segmented region is indicated by the REGION, and the value is the gray level of the reconstructed image. The segmentation algorithm can be described in the following:

$$\text{REGION} = \begin{cases} 255, & \text{if value} \geqslant \text{THRESHOLD}, \\ 0, & \text{if value} < \text{THRESHOLD}. \end{cases} \tag{11}$$

The segmented result may include a small amount of noises, and we use the median filter to process the segmented image as shown below:

$$\text{SPOT} = \text{median}\,(\text{REGION}, [P, Q]), \tag{12}$$

where the SPOT indicates the final region of the color spot, and $[P, Q]$ is the template of the median filter.

7. Results

To evaluate the performance of the proposed multifused enhancement algorithm, we design an experimental platform and verify the proposed algorithm. We use the camera of the Canon EOS 750D with the lens of the Tamron 18–200 mm to capture the face images. Among the images captured, the 100 images with the size 512×512 are selected for the evaluation of the proposed algorithm. The 50 images are randomly selected for the performance evaluation.

There are some common criteria for evaluating the segmented accuracy [37]. We choose the true positive rate (TPR, also called sensitivity, recall, and hit rate), positive predictive value (PPV, also called precision), accuracy (ACC), and F_1score (is the harmonic mean of precision and sensitivity) to evaluate the performance of segmented accuracy. These metrics are defined as follows:

$$\begin{aligned} \mathrm{TPR} &= \frac{TP}{TP + FN}, \\ \mathrm{PPV} &= \frac{TP}{TP + FP}, \\ \mathrm{ACC} &= \frac{TP + TN}{TP + TN + FP + FN}, \\ F_1 &= \frac{2TP}{2TP + FP + FN}, \end{aligned} \tag{13}$$

where TP is the intersection between ground truth and segmentation result. TN is the whole image except ground truth and segmentation result. FP is the segmentation except the TP. FN is the ground truth except the TP.

8. The Performance of the Different Methods

In order to evaluate the performance of the proposed algorithm, we compare the algorithm with the threshold segmentation, edge extraction, homomorphic filtering, and Gabor filtering methods. The experimental results are shown in Figure 4.

We use two kinds of images to evaluate the performance. The first is the conspicuous and straightforward image of the color spot, the second has the color spot of large, and the third is more complicated color spots with many and large regions. In the simple and conspicuous image, the methods of edge extraction, Gabor filtering, and our proposed algorithm can extract or indicate the accurate position of the color spot. Because the gray-level difference between the color spot and the normal skin is very small, the uneven illumination is the main problem for the wrong segmentation as shown in Figures 4(d) and 4(l). The edge extraction only obtains the edge of the color spot. Although the edge extraction method can extract the accurate position of the color spot in this sample (Figure 4(e)), the method has poor robustness in the complex image (Figure 4(m)). Homomorphic filtering is the classical filter method that reduces the low-frequency component and enhances the high-frequency component. It can reduce the effect of varying illumination and enhance the detail of the image. In the image of the color spot, the gray-level difference between the color spot and the normal skin is very small, and the homomorphic filtering is also hard to recognize the precise region of the color spot. The Gabor filtering [38, 39] is a short-time Fourier transformation that provides the selections of the orientation and scales. From Figures 4(g) and 4(o), the selection of the orientation may cause the deformation of the extracted color spot. And that the performance is worse in the complex image. The proposed multifused enhancement algorithm achieves the accurate position of the color spot not only in the simple circumstance but also in the complex image.

To further evaluate the accuracy and robustness of the proposed algorithm, we randomly selected 50 images from the skin dataset built by our group. The images include simple and complex samples of the color spot. The segmented accuracy of the color spot is evaluated by TPR, PPV, ACC, and F1, and the experimental results are shown in Figure 5.

The mean values of TPR, PPV, ACC, and F1 are 0.591, 0.625, 0.995, and 0.581, respectively. The segmented results indicate that all the color spots can be mainly segmented by the proposed method. Some factors may reduce the segmentation accuracy: (1) the region of color spot occupies a small proportion in the whole image, the small inaccurate segmentation may cause a great influence on the TPR, PPV, and F1, and (2) the denoising algorithm may reduce the accuracy of the segmented results. Excessive denoising and insufficient denoising also lead to a large error, (3) some color spots are difficult to recognize. The color spot is generated by the melanin accumulation, and the accumulation of some region may not reach the standard of the real color spot, and (4) the region, shape, and size of the color spots are varied and irregular, and the feature of the color spots is not discriminative, so the extraction of the color spot is still a great challenge.

9. The Wavelet Transform with Different Parameters

In this work, we use the wavelet transform to enhance the color spot. The wavelet is the key factor that should effect the enhanced performance of the color spot. There are some wavelets that can be used in the wavelet transformation, such as haar, daubechies, symlets, coiflets, biorthogonal, reverseBior, dmeyer, and fk. We use these wavelets to evaluate the performance to choose the optimal wavelet. The main experimental results are shown in Figure 6.

From the experimental results, we can learn that haar, bior1.1, and rbio1.1 may be the optimal choices in our work. While choosing the optimum wavelet, we should consider the support width, symmetry, vanishing moment, regularity, and similarity of the wavelet.

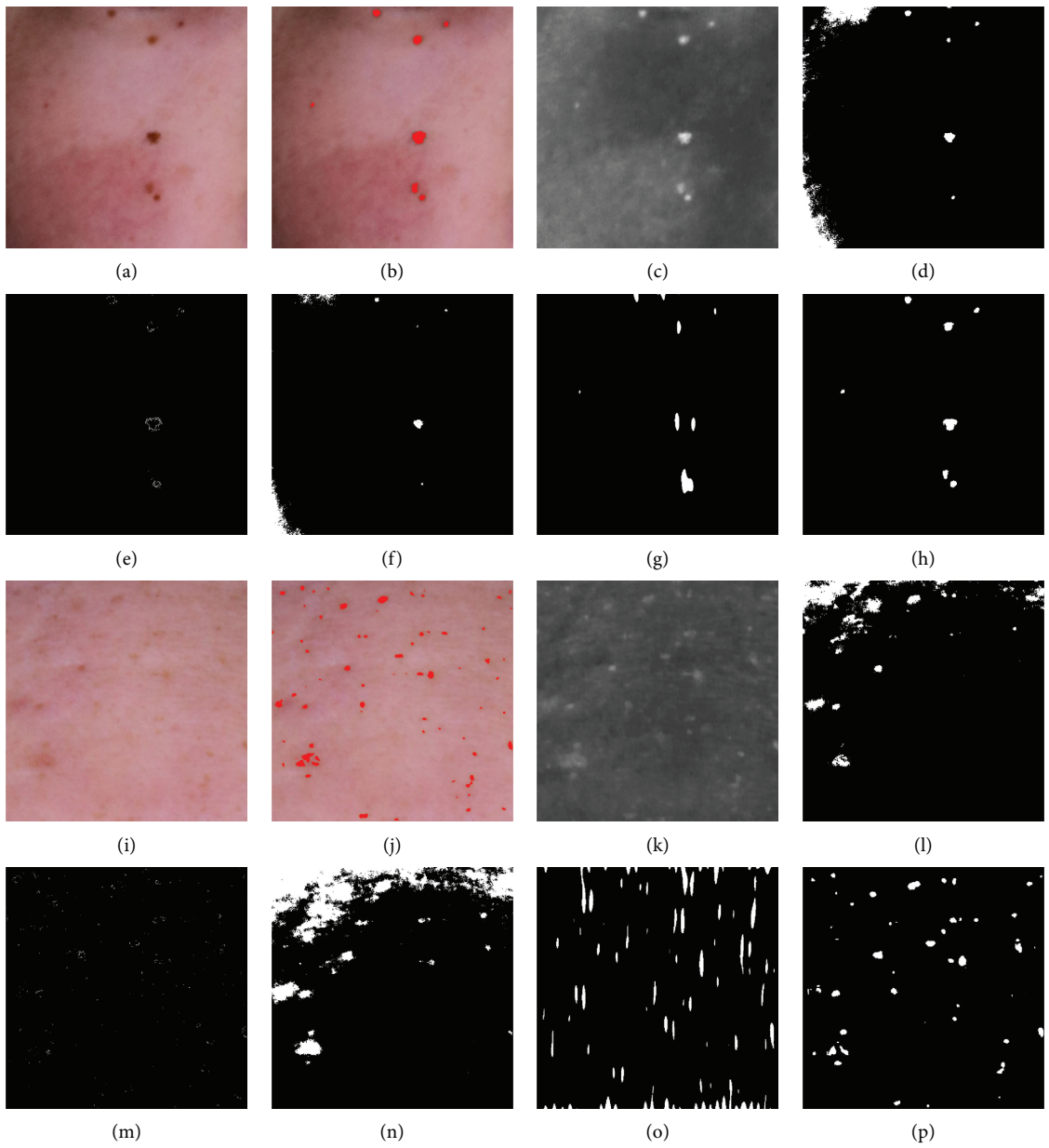


Figure 4: The comparison of different methods. (a) Simple image. (b) Ground truth. (c) Saturation. (d) Threshold. (e) Edge. (f) Homomorphic filtering. (g) Gabor filtering. (h) Our proposed algorithm. (i) Complex image. (j) Ground truth. (k) Saturation. (l) Threshold. (m) Edge. (n) Homomorphic filtering. (o) Gabor filtering. (p) Our proposed algorithm.

10. The Reduction of the Uneven Illumination

The uneven illumination is one of the main reasons that caused the gray level to be uneven in the captured image. In Figure 7(a), the image includes the conspicuous color spot, but the region with the low illumination may have a similar gray level to the color spot. The phenomenon is a little bit difficult for the segmentation of the color spot. In order to reduce the effect, we use saturation instead of the gray level to reduce the effect caused by the uneven illumination. Although the uneven effect is reduced in Figure 7(b), the features of the color spot for the discrimination are also reduced. The proposed method uses the wavelet transform to enhance the features of the color spot. The final enhanced image is shown in Figure 7(c). The final result not only eliminates the uneven effect but also enhances the features of the color spot.

11. Discussion

The article proposes a multifused enhancement algorithm for color spot extraction. We use the imaging technology of the polarized light to obtain the conspicuous color spot information and adopt the saturation component instead of the traditional gray-level image to analyze the color spot. Finally, the wavelet transform algorithm enhances the color spot information and reduces the influence of normal skin information. The experimental results indicate that the

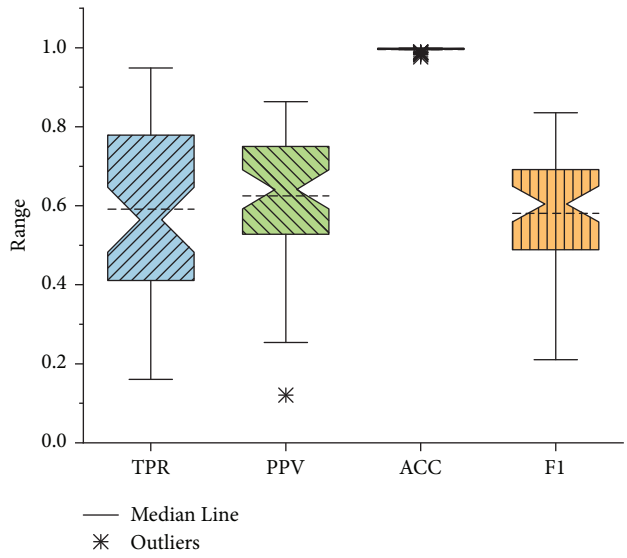


Figure 5: The accuracy of the proposed algorithm.

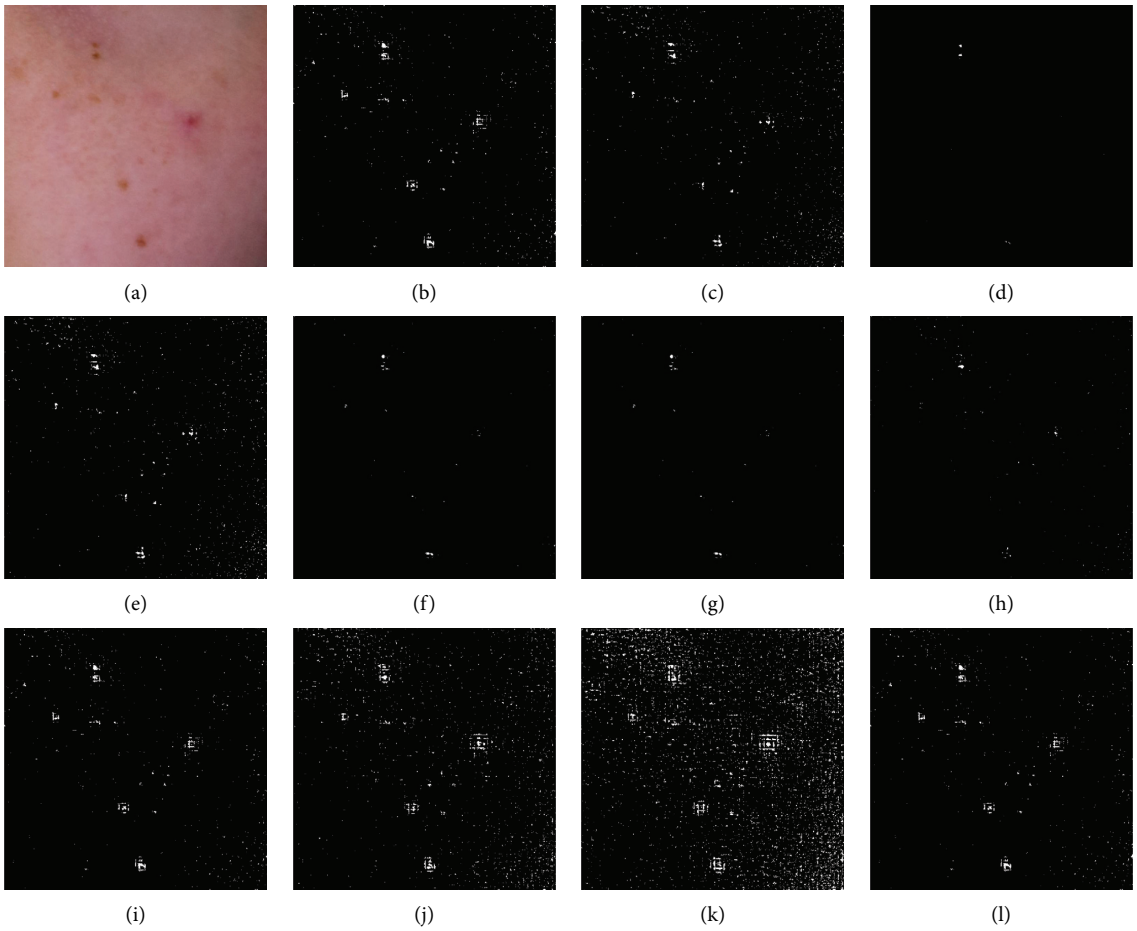


Figure 6: Continued.

(m) (n) (o) (p)

Figure 6: The image enhancement with the different wavelets. (a) Original image. (b) Haar. (c) db2. (d) db3. (e) sym2. (f) sym3. (g) coif1. (h) coif2. (i) bior1.1. (j) bior1.3. (k) bior1.5. (l) rbio1.1. (m) rbio1.3. (n) dmey. (o) fk4. (p) fk6.

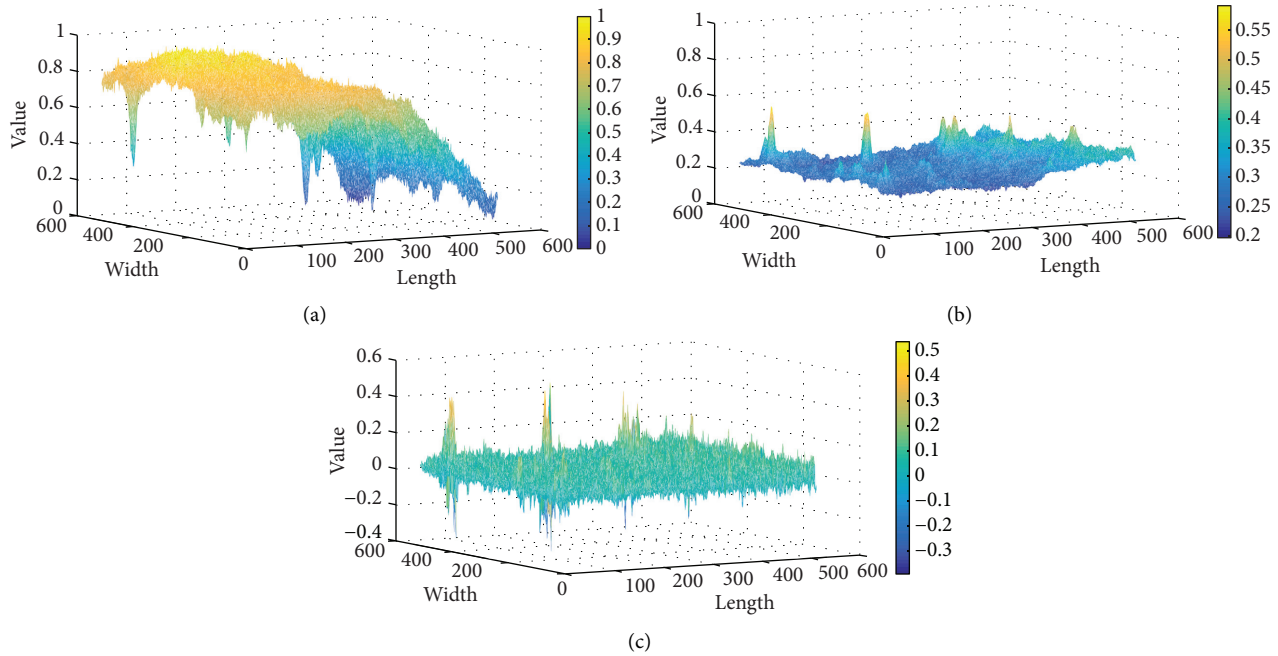


Figure 7: The reduction of the uneven illumination. (a) Gray-level image. (b) Saturation image. (c) Enhanced image.

proposed algorithm can accurately and robustly extract the color spot. In comparison with the traditional methods, the proposed algorithm can obtain the more conspicuous color spot with the enhanced methods. The algorithm also can reduce the effect of the uneven illumination by the method with the polarized light and wavelet transform. In addition, the proposed algorithm can achieve robust performance on both simple and complex color spots.

The key issue of the color spot extraction is how to solve the nonobviousness of the color spot. The traditional threshold segmentation extracts the color spot without the color spot enhancement. Hence, the results are noisy and not accurate. Although the edge extraction does not enhance the color spot, the obtained positions of the color spots are accurate in simple circumstances. The reason is that the edge extraction is based on the local gradient feature, and the uneven gray level cannot influence the performance in simple circumstances. However, because of the heavy noise in the complex circumstance, the performance of the edge extraction is not efficient. Except for the method of the spatial domain, the method based on the frequency domain is the primary enhancement method, such as homomorphic filtering and Gabor filtering. Due to the gray level of the captured images being uneven, the homomorphic filtering cannot achieve the preferable performance. The Gabor filtering can achieve the accurate color spot in the simple circumstance, but the obtained shape is distorted on account of the orientation selection in the algorithm. When the color spot is complex, the algorithm is also challenging to achieve the preferable performance. The proposed algorithm is the multifused enhancement algorithm. To reduce the effect of the uneven illumination and enhance the conspicuousness of the color spot, we use the polarized image acquirement, color model transformation, and wavelet transform. The proposed algorithm considers not only the effect of the uneven illumination but also the local features based on the

wavelet transform. Thus, the proposed algorithm can achieve better performance.

The color spot has a significant meaning in medical diagnosis and skin evaluation. The pathological features of the color spots can be used in skin diagnosis. The color, size, and position of the color spots are also used in Chinese medical diagnoses. With the development of the living standard, more and more people focus on skin care. The status of the color spot can represent the skin quality. In addition, the skin quality of the different periods can be used for the evaluation of skin care products.

It is still a challenge for the precise extraction of the color spot still has some challenges. A more accurate algorithm of the color spot should be studied in future work. Although the proposed algorithm can achieve the preferable performance, the extracted region should be further improved. In addition, as far as we know, it does not exist any intelligent imaging software to recognize and analyze the spot automatically. A more intelligent algorithm should be studied.

12. Conclusions

The article proposes a multifused enhancement algorithm for color spot extraction. In order to extract the accurate color spot, we use the vertically polarized camera with the horizontally polarized light to capture the conspicuous information of the color spot. The saturation of the image is also used to enhance the color spot. Finally, the wavelet transform is used to enhance the high frequency, which indicates the information of the color spot. In comparison with the threshold method, edge extraction, homomorphic filter, and Gabor filter, the proposed algorithm can achieve the best accurate and robust performance in simple and complex circumstances. The experimental results indicate that the proposed algorithm achieves 59.1% and 62.5% accuracy in terms of TPR and PPV, respectively. In future work, the larger dataset should be used to verify and improve the performance of the proposed algorithm.

Data Availability

All data used during the study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare no conflicts of interest.

References

[1] J. A. Savin, "The hidden face of dermatology," *Clinical and Experimental Dermatology*, vol. 18, no. 5, pp. 393–395, 2010.

[2] E. Foley, A. Robinson, and M. Maloney, "Skin substitutes and dermatology: a review," *Current Dermatology Reports*, vol. 2, no. 2, pp. 101–112, 2013.

[3] R. E. Ashton, "Teaching non-dermatologists to examine the skin: a review of the literature and some recommendations," *British Journal of Dermatology*, vol. 132, no. 2, pp. 221–225, 2010.

[4] M. Nematy, A. Mehdizadeh, and F. Razmpour, "Short review iranian journal of dermatology a review on nutrition and skin aging," *Iranian Journal of Dermatology*, vol. 18, no. 71, pp. 20–24, 2015.

[5] A. L. S. Chang, J. W. Wong, J. O. Endo, and R. A. Norman, "Geriatric dermatology review: major changes in skin function in older patients and their contribution to common clinical challenges," *Journal of the American Medical Directors Association*, vol. 14, no. 10, pp. 724–730, 2013.

[6] J. H. Dunn and J. Koo, "Psychological stress and skin aging: a review of possible mechanisms and potential therapies," *Dermatology Online Journal*, vol. 19, no. 6, Article ID 18561, 2013.

[7] M. Nematy, A. Mehdizadeh, and F. Razmpour, "A review on nutrition and skin aging," *Iranian Journal of Dermatology*, vol. 18, no. 71, pp. 20–24, 2015.

[8] A. Sparavigna, B. Tenconi, and I. De Ponti, "Antiaging, photoprotective, and brightening activity in biorevitalization: a new solution for aging skin," *Clinical, Cosmetic and Investigational Dermatology*, vol. 8, pp. 57–65, 2015.

[9] Y. Al-Nuaimi, M. J. Sherratt, and C. E. M. Griffiths, "Skin health in older age," *Maturitas*, vol. 79, no. 3, pp. 256–264, 2014.

[10] E. Uhoda, C. Piérard-Franchimont, L. Petit, and G. E. Piérard, "The conundrum of skin pores in dermocosmetology," *Dermatology*, vol. 210, no. 1, pp. 3–7, 2005.

[11] A. K. Langton, M. J. Sherratt, C. E. M. Griffiths, and R. E. B. Watson, "Review Article: a new wrinkle on old skin: the role of elastic fibres in skin ageing," *International Journal of Cosmetic Science*, vol. 32, no. 5, pp. 330–339, 2010.

[12] M. A. Hamer, L. M. Pardo, L. C. Jacobs et al., "Lifestyle and physiological factors associated with facial wrinkling in men and women," *Journal of Investigative Dermatology*, vol. 137, no. 8, pp. 1692–1699, 2017.

[13] J. M. Lagarde, C. Rouvrais, and D. Black, "Topography and anisotropy of the skin surface with ageing," *Skin Research and Technology*, vol. 11, no. 2, pp. 110–119, 2010.

[14] Y. Tian, T. Hoshino, C. J. Chen, S. Yabe, and W. Liu, "The evaluation of whitening efficacy of cosmetic products using a human skin pigmentation spot model," *Skin Research and Technology*, vol. 15, no. 2, pp. 218–223, 2010.

[15] U. Blume-Peytavi, J. Kottner, W. Sterry et al., "Age-associated skin conditions and diseases: current perspectives and future options," *The Gerontologist*, vol. 56, 2016.

[16] A. Mamalis, D. Ho, and J. Jagdeo, "Optical coherence tomography imaging of normal, chronologically aged, photoaged and photodamaged skin," *Dermatologic Surgery*, vol. 41, no. 9, pp. 993–1005, 2015.

[17] C. Praetorius, R. A. Sturm, and E. Steingrimsson, "Sun-induced freckling: ephelides and solar lentigines," *Pigment Cell & Melanoma Research*, vol. 27, no. 3, pp. 339–350, 2014.

[18] B. Hernando, M. V. Ibaez, J. A. Deseriocuesta, R. Sorianavarro, I. Vilarsastre, and C. Martinezcadenas, "Genetic determinants of freckle occurrence in the Spanish population: towards ephelides prediction from human dna samples," *Forensic Science International Genetics*, vol. 33, p. 38, 2017.

[19] C. T. Paniagua and A. S. Bean, "Black-spot poison ivy: a rare phenomenon," *Journal of the American Academy of Nurse Practitioners*, vol. 23, no. 6, pp. 275–277, 2011.

[20] M. Rodrigues and A. G. Pandya, "Melasma: clinical diagnosis and management options," *Australasian Journal of Dermatology*, vol. 56, no. 3, pp. 151–163, 2015.

[21] R Sarkar, P Ailawadi, and S Garg, "Melasma in men: a review of clinical, etiological, and management issues," *The Journal of*

clinical and aesthetic dermatology, vol. 11, no. 2, pp. 53–59, 2018.

[22] L. Ai-Young, "Recent progress in melasma pathogenesis," *Pigment Cell Melanoma Res*, vol. 28, no. 6, pp. 648–660, 2015.

[23] A. Skoczyska, E. Budzisz, E. Trznadel-Grodzka, and H. Rotsztejn, "Melanin and lipofuscin as hallmarks of skin aging," *Advances in Dermatology Allergology/postpy Dermatologii I Alergologii*, vol. 34, no. 2, pp. 97–103, 2017.

[24] H. Sakai, Y. Ando, K. Ikinaga, and M. Tanaka, "Estimating melanin location in the pigmented skin lesions by hue-saturation-lightness color space values of dermoscopic images," *The Journal of Dermatology*, vol. 44, no. 5, pp. 490–498, 2017.

[25] N. Joly-Tonetti, J. I. D. Wibawa, M. Bell, and D. Tobin, "Melanin fate in the human epidermis: a reassessment of how best to detect and analyse histologically," *Experimental Dermatology*, vol. 25, no. 7, pp. 501–504, 2016.

[26] M. D'Ischia, K. Wakamatsu, F. Cicoira et al., "Melanins and melanogenesis: from pigment cells to human health and technological applications," *Pigment Cell Melanoma Res*, vol. 28, no. 5, pp. 520–544, 2015.

[27] L. Liu, N. Yang, J. Lan, and J. Li, "Image segmentation based on gray stretch and threshold algorithm," *Optik*, vol. 126, no. 6, pp. 626–629, 2015.

[28] E. Hamuda, M. Glavin, and E. Jones, "A survey of image processing techniques for plant extraction and segmentation in the field," *Computers and Electronics in Agriculture*, vol. 125, pp. 184–199, 2016.

[29] A. Ferone, S. K. Pal, and A. Petrosino, "A rough-fuzzy hsv color histogram for image segmentation," *Image Analysis and Processing - ICIAP 2011*, vol. 6978, no. 1, pp. 29–37, 2011.

[30] C. Lu and T. Wu, "[segmentation and cavity filling of color image from stained trabecular sections based on hsv space]," *Journal of Biomedical Engineering*, vol. 29, no. 2, p. 260, 2012.

[31] A. M. Grigoryan, E. R. Dougherty, and S. S. Agaian, "Optimal wiener and homomorphic filtration: Review," *Signal Processing*, vol. 121, pp. 111–138, 2016.

[32] N. Biradar, M. L. Dewal, and M. K. Rohit, "A novel hybrid homomorphic fuzzy filter for speckle noise reduction," *Biomedical Engineering Letters*, vol. 4, no. 2, pp. 176–185, 2014.

[33] L. Xiao, C. Li, Z. Wu, and T. Wang, "An enhancement method for x-ray image via fuzzy noise removal and homomorphic filtering," *Neurocomputing*, vol. 195, pp. 56–64, 2016.

[34] X. Ma, K. Xu, J. Jiang, R. Liu, and X. Yu, "Layered vasculature segmentation of color conjunctival image based on wavelet transform," *Biomedical Signal Processing and Control*, vol. 42, pp. 9–17, 2018.

[35] S. Liu, "Study on medical image enhancement based on wavelet transform fusion algorithm," *Journal of Medical Imaging and Health Informatics*, vol. 7, no. 2, pp. 388–392, 2017.

[36] K. Thiruvenkadam and N. Perumal, "Fully automatic method for segmentation of brain tumor from multimodal magnetic resonance images using wavelet transformation and clustering technique," *International Journal of Imaging Systems and Technology*, vol. 26, no. 4, pp. 305–314, 2016.

[37] Y. Feng, E. Song, L. Hong, Y. Li, Z. Jun, and C. C. Hung, "An augmented reality endoscope system for ureter position detection," *Journal of Medical Systems*, vol. 42, no. 8, p. 138, 2018.

[38] S. Thilagamani and N. Shanthi, "Gaussian and gabor filter approach for object segmentation," in *Proceedings of the International Conference on Optical Imaging Sensor and Security*, pp. 1–6, IEEE, Coimbatore, India, July 2013.

[39] F. Farokhian, C. Yang, H. Demirel, S. Wu, and I. Beheshti, "Automatic parameters selection of gabor filters with the imperialism competitive algorithm with application to retinal vessel segmentation," *Biocybernetics and Biomedical Engineering*, vol. 37, no. 1, pp. 246–254, 2017.

Hindawi
Mathematical Problems in Engineering
Volume 2022, Article ID 3747033, 19 pages
https://doi.org/10.1155/2022/3747033

Research Article

Multigraph Convolutional Network Enhanced Neural Factorization Machine for Service Recommendation

Wei Gao and Jian Wu

School of Computer Science and Engineering, Zhejiang University, Hangzhou, Zhejiang 310027, China

Correspondence should be addressed to Jian Wu; wujian2000@zju.edu.cn

Received 21 December 2021; Accepted 16 February 2022; Published 1 April 2022

Academic Editor: Hua Ming

With an increasing number of web services on the Web, selecting appropriate services to meet the developer's needs for mashup development has become a difficult task. To tackle the problem, various service recommendation methods have been proposed. However, there are still challenges, including the sparsity and imbalance of features, as well as the cold-start of mashups and services. To tackle these challenges, in this paper, we propose a Multigraph Convolutional Network enhanced Neural Factorization Machine model (MGCN-NFM) for service recommendation. It first constructs three graphs, namely, the collaborative graph, the description graph, and the tag graph. Each graph represents a different type of relation between mashups and services. Next, graph convolution is performed on the three graphs to learn the feature embeddings of mashups, services, and tags. Each node iteratively aggregates the information from its higher-order neighbors through message passing in each graph. Finally, the feature embeddings as well as the description features learned by Doc2vec are modeled by the neural factorization machine model, which captures the nonlinear and higher-order feature interaction relations between them. We conduct extensive experiments on the ProgrammableWeb dataset, and demonstrate that our proposed method outperforms state-of-the-art factorization machine-based methods in service recommendation.

1. Introduction

Service-oriented computing (SOC) has become a significant paradigm for developing low-cost and reliable software applications in software engineering and cloud computing [1]. Web services are the basic building blocks of service-oriented computing, which encapsulate application functionalities and can be accessed through standard interfaces. Nowadays, an increasing number of web services, mainly in the form of RESTful Web APIs, have been published online. According to the recent statistics of ProgrammableWeb (https://www.programmableweb.com/), the largest online Web API registry, there are over 24,000 Web APIs available. However, the functionality of an individual service is limited and cannot satisfy the complex requirements of developers. As a result, it is common for developers to compose existing services and develop value-added services, also called mashups [2]. For example, a developer may create a new mashup that can display ratings and reviews of restaurants on a regional map by integrating Google Map Service and Yelp Service. However, creating a mashup can be difficult and time-consuming for an inexperienced developer due to the overwhelming number of services on the Internet. Therefore, it is vital to proactively recommend appropriate services that can satisfy the developer's complex requirements and reduce the burden of manual selection, especially in the era of the Internet of Things and Big Data [3].

A variety of methods have been proposed to recommend services for mashup development, and they can be divided into three classes: collaborative filtering-based methods, content-based methods, and hybrid methods [4]. Different types of methods focus on modeling different sources of information. Collaborative filtering-based methods make use of past composition history of mashups; Content-based methods make use of user requirements and service descriptions; Hybrid methods are a combination of the two

methods, and various kinds of additional information are incorporated, such as tags, categories, developers, QoS, etc.

In this paper, we adopt a factorization machine-based model, which is a type of hybrid method for service recommendation. Factorization Machine (FM), first proposed in [5] by Rendle, is a supervised machine learning algorithm for classification, regression, and ranking tasks. It achieves great success and is widely employed for product recommendation as it can model high-dimensional sparse features and their interactions in an efficient manner [6]. As a result, it attracts a lot of research attention for service recommendation. For instance, in [7], Cao et al. use FM with topic similarities, co-occurrence, and popularity as input features for service recommendation. In [8], Cao et al. propose an attentional FM model that discriminates the importance of different feature interactions with an attention mechanism. In [9], Kang et al. propose a hybrid FM model by integrating a deep neural network to capture the nonlinear and higher-order feature interactions.

Although previous FM-based models have achieved great results for service recommendation, there are still challenges that can limit the performance of the existing methods:

(i) Feature sparsity and imbalance problems: the performance of the factorization machine-based models heavily relies on the co-occurrence of different features. However, the features used for service recommendation are generally sparse: (1) although there are numerous services on the web, each mashup only composes a few services, and most services are composed by mashups only a few times, leading to a sparse mashup-service composition record. Moreover, only a small portion of services are composed frequently, while the majority of services are rarely composed, resulting in the imbalance problem. It is difficult to learn latent factors for mashups and rarely composed services with limited data; (2) even if contextual information such as categories, tags, providers, etc. is incorporated, the features themselves may suffer from the sparsity and imbalance problem. For instance, a majority of tags are used in a few services, while a few popular tags are frequently used by a majority of services. In this case, the model suffers from learning informative latent factors for infrequently used tags, and the benefit of incorporating tag information into the model is reduced.

(ii) Cold-start problem: when a new mashup is first included in the development cycle, it does not contain any component services. We refer to them as "cold-start mashups". As FM can only learn features that appear in the training data, the latent factors for the cold-start mashup cannot be learned due to a lack of composition data. Similarly, we refer to services that have never been composed by any mashup as "cold-start services", and it is impossible to learn the latent factors of cold-start services. As a result, when a cold-start mashup or service is encountered, FM can only make predictions by setting the latent factors of the mashup or service to random values, and relying on the latent factors of contextual information such as description information. Therefore, it is unlikely to lead to good recommendation results for cold-start mashups.

To address the aforementioned issues, we aim to learn informative latent factors for sparse and cold-start mashups and services in FM. To this end, we enhance FM with Multigraph Convolutional Network (MGCN), which exploits various relations between mashups and services. We dub our model the Multigraph Convolutional Network enhanced Neural Factorization Machine, or MGCN-NFM for short. In particular, we construct three graphs, namely, collaborative graph, description graph, and tag graph. Each graph reflects different types of relations, and different relations can complement each other to enrich the information in a mashup or service. Next, we use graph convolution on each graph to further enrich the relational features. The intuition is that by exploiting information from higher-order neighbors, the representations of sparse nodes can be enhanced. To reduce the computational complexity, we use a neighbor sampling technique to maintain a small representative set of neighbors for each node in the graph convolution process. After graph convolution, each mashup or service has three enhanced feature representations that possess different characteristics. These feature embeddings, as well as tag feature embeddings learned from the tag graph and the description features learned from Doc2vec, are used as the input of FM to capture the fine-grained feature interactions. In this paper, we adopt the neural factorization machine model (NFM) proposed in [10]. It enhances vanilla FM by modeling higher-order and nonlinear feature interactions with a bilinear interaction pooling layer and several deep neural network layers. Different from [10], the embedding layer of NFM is obtained based on the MGCN instead of an embedding lookup table. Our model effectively reduces the feature sparsity, imbalance, and cold-start problems due to the multigraph construction and convolution process. For sparse features, their embeddings have more chances of being updated in training as they are connected to more multihop nodes with different types of relations. For cold-start mashups or services, since they have (multihop) connections to other nodes in the description graph and tag graph, their feature embedding can be learned when updating the embeddings of their neighbors.

The contributions of this work can be summarized as follows:

(i) We introduce a novel framework, MGCN-NFM, for service recommendation by enhancing the neural factorization model with a multigraph convolutional network that learns the latent factors for various features.

(ii) We alleviate the feature sparsity, imbalance, and cold-start problems by constructing three graphs reflecting different types of relations between mashup and service, and using multigraph convolution to enrich the sparse relations.

(iii) We conduct extensive evaluations of our proposed model on the ProgrammableWeb dataset, and the results show that our model achieves better performance than state-of-the-art FM methods for service recommendation.

The rest of this article is organized as follows: Section 2 presents the related work of service recommendation. Section 3 introduces the details of our proposed approach. Section 4 reports the experimental results and analysis. Section 5 concludes this paper with future work.

2. Related Work

Recommending web services to developers according to their mashup requirements is a hot topic in service computing. Most works use the dataset from ProgrammableWeb, and they exploit different kinds of auxiliary information, including functional descriptions, tags, categories, providers, architectural styles, etc., as the mashup-service composition record is extremely sparse. Based on the modeling of the mashup-service composition record, existing research can be roughly divided into three categories: neighbor-based collaborative filtering (CF) methods [11–15], latent factor-based CF methods [6, 16–21] and deep learning-based methods [8, 9, 22–27].

Neighbor-based CF methods recommend services based on the identification of similar users (user-based CF) or similar items (item-based CF). It is widely used in the scenario of QoS-aware service recommendation, where the Quality of Service (QoS) information of services is available [28, 29]. Absent from the QoS information, existing works in service recommendation for mashups calculate the matching degree between mashup and service from different information sources with different methods and aggregate them together. As there are multiple types of objects (mashup, service, content, tag, etc.) and rich relations among those objects, they usually build a heterogeneous graph to capture them. In [11], Cao et al. recommend services by building a social network based on social relationships among mashups, services, and their tags. In [12], Gao et al. propose to use a generalized manifold ranking algorithm on the graph with relations among mashups and services. In [13], Liang et al. measure the meta-path-based similarity between mashups under different semantic meanings based on a heterogeneous information network (HIN). Based on the HIN, in [14], Xie et al. propose a mashup group preference-based service recommendation, where mashup group preference is utilized to capture the rich interactions among mashups. In [15], Wang et al. design a knowledge graph to encode the mashup-service relations and exploit random walks with restart to assess the similarities.

Latent factor-based CF methods aim to learn the latent factors of mashups and services to discover potential features. The two most widely adopted models are the matrix factorization-based model (MF) and the factorization machine-based model (FM). MF models decompose the mashup-service composition matrix to derive the mashup feature vectors and service feature vectors (embeddings). The key is to build relevant features that are helpful for improving the recommendation performance. In [16], Xu et al. propose a social-aware service recommendation model, where a coupled matrix factorization model is used to predict the multidimensional relations among users, mashups, and services. In [17], Yao et al. propose a probabilistic MF with implicit correlation regularization, where they develop a latent variable model to uncover the coinvocation patterns of services driven by explicit textual relations and implicit similar or complement relations. In [18], Gao et al. compute different similarity scores between services through heterogeneous functional aspects, and MF is used to learn the embeddings of mashups and services for each functional aspect. The FM model is a generalization of the linear regression model and the MF model. It is more powerful than the MF model because it can entail contextual features and model second-order feature interactions of various sparse features. In [6], Cao et al. propose a QoS-aware service recommendation based on relational topic model (RTM) and FM, where RTM is first used to mine the latent topics derived from the relations among mashups, services, and their links, and then FM is used to predict their link relations. In [19], Li et al. integrate tag, topic, co-occurrence, and popularity factors in the FM for service recommendation, where they exploit the enriched tags and topics of mashups and services derived by RTM and use the invocation times and category information of services to derive their popularity. In [20], Cao et al. extend the description of services using Word2vec and derive latent topics by the hierarchical dirichlet process (HDP). FM is then applied to train these latent topics for service recommendation. In [21], Xie et al. propose to use FM on the features of mashups and services learned from different kinds of metapaths of HIN.

With the rapid development of deep learning in the past few years, it has become popular to adopt neural networks for service recommendation. Compared with traditional methods, deep learning-based methods can perform feature engineering automatically and provide a more powerful representation capability. In [22], Zhang et al. cluster the descriptions of services using Doc2Vec and use the DeepFM model [30] to mine the higher-order composition relations. In [23], Chen et al. also adopt the DeepFM model by taking the features learned from word embedding and the Dirichlet mixture model (DMM) as input. In [8], Cao et al. propose an attention FM [31] by employing an attention mechanism that learns the importance of each input feature interaction via a neural network model. In [9], Kang et al. further combine the advantages of the DeepFM model and the attention FM model, and propose the NAFM model that can capture the nonlinear feature interactions with different degrees of importance. In [24], Xiong et al. adopt three kinds of similarity feature extractors on textual descriptions by a variety of pretrained word embeddings and integrate the

mashup-service composition record and their textual descriptions by using a multilayer perceptron.

Graph Neural Network (GNN) [32] has recently emerged as a popular deep learning model for extracting features from graph-structured data, and research has begun to focus on the use of various GNN models for service recommendation. In [25], Zhang et al. propose a Semantic Variational Graph Auto-Encoder model, where they construct the service graph using the composition relations in the mashup, and use a variational graph autoencoder model as a link prediction task for service recommendation. In [26], Lian and Tang propose a service recommendation method that exploits the higher-order connectivity between mashups and services based on the neural graph collaborative filtering technique. In [27], He et al. propose a service link prediction method based on a heterogeneous graph attention network, where they select five types of neighbors associated with service links, and two levels of attention are applied to learn the importance of different nodes and their associations. Our method is different from theirs in that the constructed graphs incorporate different kinds of relations, making them more comprehensive, and the learned features go through the FM model that can better exploit their hidden interactions.

3. MGCN-NFM Approach

In this section, we will describe our proposed multigraph convolutional network enhanced neural factorization machine model (MGCN-NFM) in detail. First, in Section 3.1, we describe the problem of service recommendation for the mashup. Next, we present the overall framework in Section 3.2. The main components of the model, namely, graph construction, multigraph convolutional network, and neural factorization machine, are described in Sections 3.3, 3.4, and 3.5, respectively. Finally, the training process and model complexity are discussed in Section 3.6.

3.1. Problem Definition. We denote $\mathcal{M} = \{m_1, m_2, \ldots, m_M\}$ as a set of M mashups, $\mathcal{S} = \{s_1, s_2, \ldots, s_N\}$ as a set of N services. $Y \in \{0,1\}^{M \times N}$ denotes the mashup-service invocation history, where $Y_{ij} = 1$ means mashup m_i has composed service s_j in the past, otherwise $Y_{ij} = 0$. Each mashup and each service are associated with its title, description, and tags. In particular, titles and descriptions consist of a sequence of words. For each mashup m, we concatenate its title and description and denote as $d_m = \{w_1, w_2, \ldots, w_{|d_m|}\}$, where w_i is a word and $|d_m|$ is the length of the description. $\mathcal{D}_{\mathcal{M}} = \{d_{m_1}, d_{m_2}, \ldots, d_{m_M}\}$ denotes the descriptions of all mashups in $\mathcal{M}$. For each service s, d_s can be defined in a similarly manner, and $\mathcal{D}_{\mathcal{S}} = \{d_{s_1}, d_{s_2}, \ldots, d_{s_N}\}$ denotes the descriptions of all services in $\mathcal{S}$. We denote $\mathcal{T} = \{t_1, t_2, \ldots, t_T\}$ as a set of T tags. $MT \in \{0,1\}^{M \times T}$ denotes the tagging assignment of mashups, where $MT_{ij} = 1$ means tag t_j is tagged by mashup m_i, and otherwise $MT_{ij} = 0$. Similarly, $ST \in \{0,1\}^{N \times T}$ denotes the tagging assignment of services. The problem of service recommendation can be defined as follows: given $\mathcal{M}, \mathcal{S}, \mathcal{T}, Y, \mathcal{D}_M, \mathcal{D}_S, MT$, and ST that are recorded in the service repository, and a newly developed mashup with requirement description, specified tags, and possibly composed services, the goal is to recommend a list of appropriate services that are likely to be composed by the new mashup.

3.2. Overall Framework. Figure 1 shows the overall framework of our proposed MGCN-NFM model. It mainly consists of three modules: graph construction, multigraph convolutional network, and neural factorization machine. In the graph construction module, we construct three graphs capturing the three relations between mashups and services, namely, collaborative graph, description graph, and tag graph. In constructing the description graph, the descriptions of mashups and services are transformed to the low-dimensional description embedding by the Doc2vec technique. In the multigraph convolutional network module, we adopt the simplified graph convolution network for each graph to propagate higher-order information between mashups and services and output the aggregated feature representations of each node. In the neural factorization machine module, the feature representations of mashup, service, and tags learned in each graph, as well as the description embedding of mashup and service, are used as inputs to the neural factorization machine model, which can model nonlinear and higher-order feature interactions. Finally, the model outputs a score representing the probability of a mashup composing a service. For model training, the predicted score is compared with the actual composition record, and the error is back-propagated to update the model parameters. For service recommendation, the scores of the mashup and all candidate services are computed and ranked, and the top-scoring services are recommended.

3.3. Graph Construction. To fully exploit different kinds of information in mashups and services, we construct three graphs reflecting diverse relations between mashups and services: the collaborative graph, the description graph, and the tag graph. Figure 2 illustrates an example of the three constructed graphs.

3.3.1. Collaborative Graph. The collaborative graph reflects the historical composition relation between mashup and service. However, each mashup generally only composes a small number of services, making the composition relation extremely sparse. Inspired by the idea of collaborative filtering that mashups composed of the same set of services tend to be similar, we expand the composition relation to alleviate the sparsity problem by exploiting the collaborative signal. The invocation matrix Y can be regarded as a mashup-service bipartite graph $\mathcal{G}^{MS} = (\mathcal{M} \bigcup \mathcal{S}, \mathcal{E}^{MS})$, where there is an edge between mashup m and service s, i.e. $(m, s) \in \mathcal{E}^{MS}$ if $Y_{ms} = 1$. We also consider the collaborative relations inside the mashups and services. We build a mashup-mashup collaborative graph $\mathcal{G}^{MM} = (\mathcal{M}, \mathcal{E}^{MM})$ and a service-service collaborative graph $\mathcal{G}^{SS} = (\mathcal{S}, \mathcal{E}^{SS})$ to complement the graph $\mathcal{G}^{MS}$. The mashup-mashup

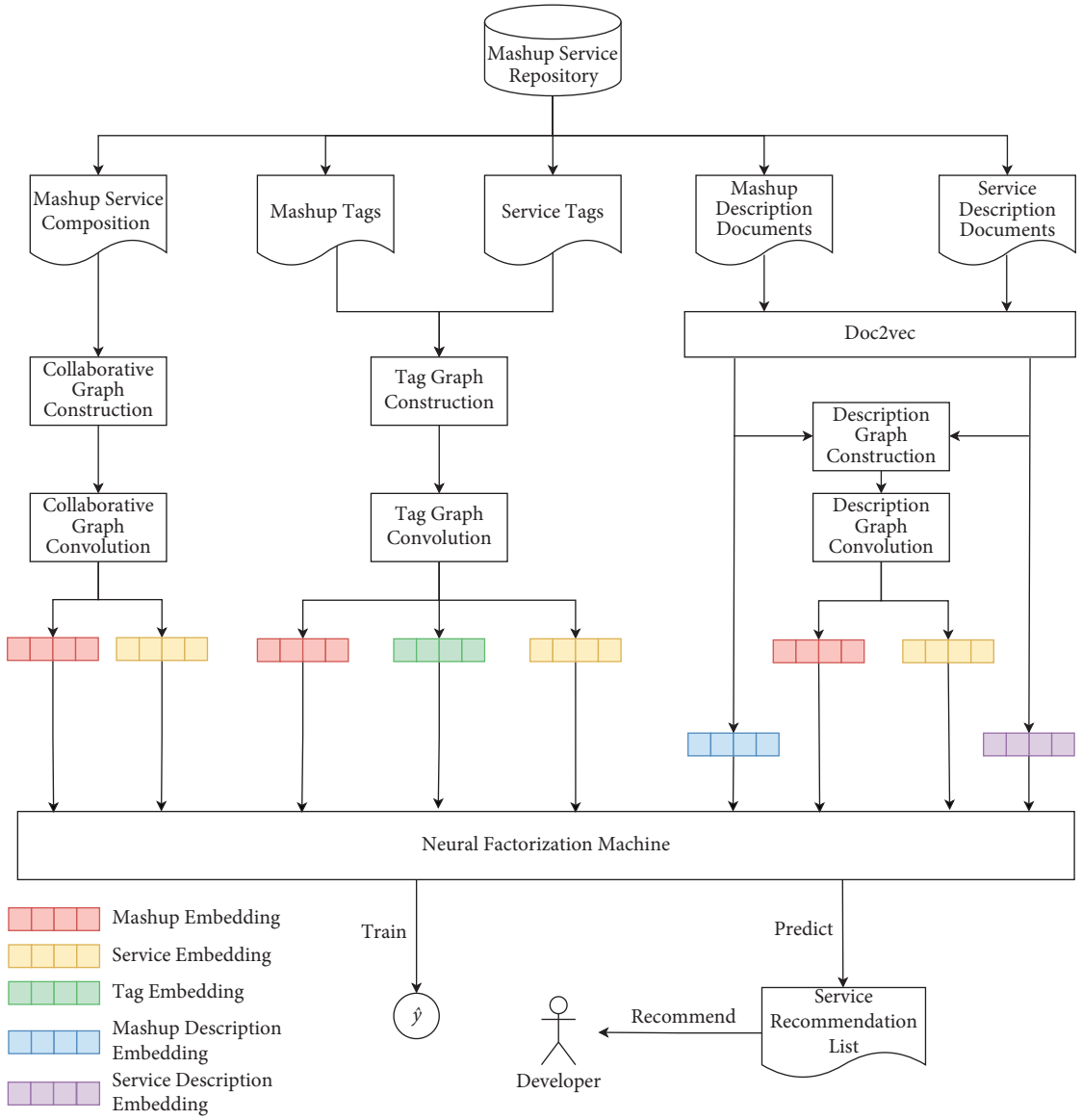


Figure 1: Overall framework of MGCN-NFM model.

collaborative graph is constructed based on the similarity of mashups measured by the number of commonly composed services. Specifically, if mashups i and j have co-composed services, then there is an edge $(i, j) \in \mathcal{E}^{MM}$ between them. The similarities between them are calculated as the number of co-composed services normalized by the number of composed services for each mashup, i.e., $sim(i, j) = Y_i \cdot Y_j / Y_i Y_j$, where Y_i and Y_j denote the i-th and j-th rows of the matrix Y. The normalization keeps the similarity in the range of [0,1]. In a similar manner, the service-service collaborative graph can be constructed based on the similarity of services measured by the number of co-appearing mashups. If services i and j have both occurred in the same mashup, they are connected by an edge $(i, j) \in \mathcal{E}^{SS}$. The similarity between them is defined as $sim(i, j) = Y_{:,i} \cdot Y_{:,j} / Y_{:,i} Y_{:,j}$, where $Y_{:,i}$ and $Y_{:,j}$ denote the i-th and j-th columns of the matrix Y. With the constructed mashup-mashup collaborative graph $\mathcal{G}^{MM}$ and service-service collaborative graph $\mathcal{G}^{SS}$, one mashup can directly utilize the information of the neighboring mashups, thereby alleviating the sparsity issue with the mashup-service invocation record. We merge the three graphs to get an overall collaborative graph $\mathcal{G}^{C} = (\mathcal{M} \bigcup \mathcal{S}, \mathcal{E}^{MS} \bigcup \mathcal{E}^{MM} \bigcup \mathcal{E}^{SS})$.

3.3.2. Description Graph. The description graph represents the similarities between mashups and services in terms of their textual descriptions. We employ the state-of-the-art Doc2vec technique [33] to measure the description similarity. Doc2vec maps the textual descriptions of mashup d_m and service d_s into latent semantic embeddings $\mathbf{d}_m$ and $\mathbf{d}_s$. Doc2vec follows the idea of word2vec [34] for learning word representations. Specifically, each document and each word is mapped to a unique vector. The model concatenates the document vector with a sequence of word vectors from the

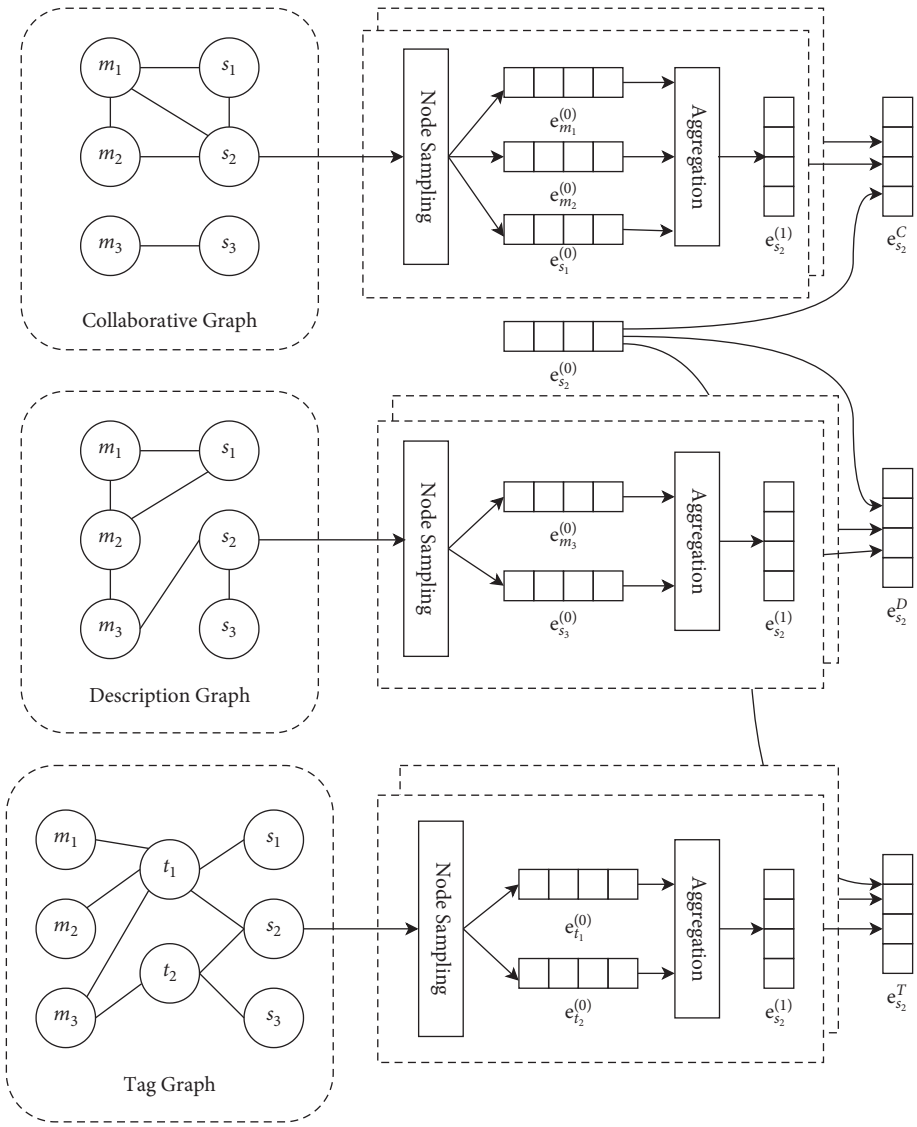


Figure 2: Overall framework of multigraph convolutional network.

document and predicts the following word. It can be seen as a self-supervised learning task where the input is the context words and the document, and the label is the target word. After training on a large number of documents, we can learn a good document embedding that captures the overall semantics of the document. It has been shown to achieve superior performance compared to traditional bag-of-words models such as LDA [35] or HDP [36]. We feed the descriptions of all mashups and services for training the Doc2vec model, and get the corresponding description embeddings after training. We calculate the cosine similarity of the learned embeddings to represent the description similarity. Specifically, given two nodes i and j that could be either mashup or service in the description graph, the similarity between them is calculated as $sim(i, j) = \mathbf{d}_i \cdot \mathbf{d}_j / \mathbf{d}_i \mathbf{d}_j$. However, including all pairs of nodes as edges brings a lot of noise as most pairs of mashup and service are irrelevant. We set a threshold τ (empirically set as 0.1) to filter out dissimilar pairs of nodes. That is, if $sim(i, j) > \tau$, there is an edge between nodes i and j. In this way, we get a description graph $\mathcal{G}^D = (\mathcal{M} \bigcup \mathcal{S}, \mathcal{E}^D)$, where $\mathcal{E}^D$ denotes the filtered edges in the description graph.

3.3.3. Tag Graph. The tag graph represents the annotated tags of mashups and services and is denoted as $\mathcal{G}^T = (\mathcal{M} \bigcup \mathcal{S} \bigcup \mathcal{T}, \mathcal{E}^{MT} \bigcup \mathcal{E}^{ST})$. Different from the collaborative graph and the description graph, it consists of three types of nodes: mashup, service, and tag. There is an edge between mashup i and tag j, i.e., $(i, j) \in \mathcal{E}^{MT}$ if $MT_{ij} = 1$. Similarly, an edge $(i, j) \in \mathcal{E}^{ST}$ exists between service i and tag j if $ST_{ij} = 1$.

3.4. Multigraph Convolutional Network. After constructing three types of graphs reflecting different relations between mashups and services, we employ the multigraph convolutional network to learn comprehensive mashup and service features from the three graphs. In particular, it consists of two steps: (1) Neighbor Sampling, which samples a fixed number of neighbors for each node in the graph; (2)

Graph Convolution, which aggregates neighboring nodes, updates node embeddings at each layer, and combines embeddings at all layers. Figure 2 illustrates an example of the whole process.

3.4.1. Neighbor Sampling. In traditional GCN [37], each node aggregates information from all neighboring nodes in the graph. However, in the case of the three constructed graphs, some nodes may be connected to a large number of nodes, resulting in a large number of neighbors. For instance, in the collaborative graph, a popular service may connect to a large number of mashups and services, as it may be composed of many mashups and cocomposed with many services. Directly aggregating all neighboring nodes may have several issues: (1) it adds the computational burden of the graph convolution process as the size of neighbors grows exponentially with the number of layers; (2) it makes the learning process excessively focus on the nodes with a large number of neighbors and ignores the nodes with only a few neighbors, causing the overfitting issue for the densely-connected nodes and the underfitting issue for the sparsely-connected nodes; (3) it causes the oversmoothing issue [38] when aggregating a large number of neighbors with weak correlation as the learned embedding for nodes and their neighbors will become indistinguishable. Hence, we adopt the neighbor sampling strategy to control the size of the neighbors before graph convolution. Specifically, for each node, we sample its neighbors with respect to the similarity between them. The sampling probability of each neighbor is computed as the normalization of their similarity: first, we get a vector of similarity scores for the target node and its neighboring nodes. Then, we normalize the vector into a probability distribution. Finally, we sample the neighbors with respect to the probability distribution. We use the standard normalization function defined as

$$\cdot norm(\mathbf{z})_i = \frac{\mathbf{z}_i}{\sum_{j=1}^{K} \mathbf{z}_j}\#, \tag{1}$$

where $\mathbf{z}$ is the vector of similarity scores. Other normalization functions can be used, such as the softmax function, although no performance gain is observed. Next, we describe the neighbor sampling process for each graph in detail. A graphical illustration of the sampling process for each graph is shown in Figure 3.

(i) Collaborative graph: for each mashup node, we keep all of its service neighbors. If the number of service neighbors is less than N^C, we sample the neighbors from $\mathcal{E}^{MM}$ according to the sampling probability distribution derived from mashup collaborative similarity, until the number of neighbors reaches N^C. For each service node, we keep all of its mashup neighbors. If the number of mashup neighbors is less than N^C, we sample the neighbors from $\mathcal{E}^{SS}$ according to the sampling probability distribution derived from service collaborative similarity, until the number of neighbors reaches N^C. However, as some popular services may be composed of a large number of mashups, if the number of mashup neighbors is greater than N^C, we sample N^C neighboring mashups with a uniform probability distribution.

(ii) Description graph: for each node, if the number of neighbors is greater than N^D, we sample N^D neighboring mashups and services according to the sampling probability distribution derived from the description similarity.

(iii) Tag graph: as mashups and services are only associated with a few tags, no sampling is needed for the mashup and the service node. For the tag node, we sample N^T nodes from its neighboring mashups and services with a uniform probability distribution.

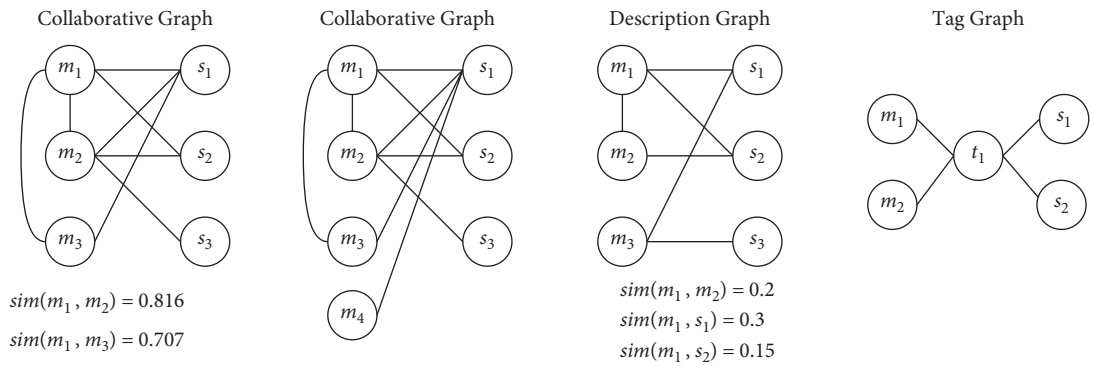
3.4.2. Graph Convolution. The core idea behind graph convolution is to iteratively aggregate feature information from the local neighbors of each node. A single layer of convolution aggregates feature information from the node's direct neighbors, and by stacking multiple layers, feature information can be propagated across long ranges and the node features can be enhanced with sufficient higher-order neighbor information. The graph convolution process consists of three steps: neighbor nodes aggregation, node embedding update, and layer combination. For each graph, after sampling the neighbors for each node, the next step is to aggregate the features of the node neighbors to obtain the neighbor embedding, and the embedding of the node is updated by fusing the neighbor embedding. Denote the center node to be updated as h and the set of neighbor nodes as $\mathcal{N}_h$, the neighbor aggregation and updating process for node h in the l-th layer can be abstracted as

$$\mathbf{e}_h^{(l+1)} = f\left(\mathbf{e}_h^{(l)}, \left\{\mathbf{e}_i^{(l)} : i \in \mathcal{N}_h\right\}\right), \tag{2}$$

where $\mathbf{e}_h^{(l)} \in \mathbb{R}^k$ represents the feature embedding of node h in the l-th layer, and $f(\cdot)$ is the aggregation function which is the core of graph convolution. In this paper, we adopt a simple aggregation function $f(\cdot)$ that drops the feature transformation and nonlinear activation function used in the original graph convolution, which has been shown to achieve better performance with less model complexity [39]. The graph convolution operation is defined as

$$\mathbf{e}_h^{(l+1)} = \sum_{i \in \mathcal{N}_h} \frac{1}{\sqrt{|\mathcal{N}_h||\mathcal{N}_i|}} \mathbf{e}_i^{(l)}. \tag{3}$$

The aggregation function is a weighted sum of the neighbor embeddings, and $1/\sqrt{|\mathcal{N}_h||\mathcal{N}_i|}$ is the normalization term to avoid the scale of embeddings increasing with graph convolution. By stacking L propagation layers defined in (3), each node is capable of receiving features from nodes within L-hops away, and the higher-order relation between mashups and services can be explored. After L layers of



Target Node	m_1	s_1	m_1	t_1
Threshold	$N^C = 3$	$N^C = 3$	$N^D = 2$	$N^T = 2$
Neighboring Nodes	$[s_1, s_2, m_2, m_3]$	$[m_1, m_2, m_3, m_4]$	$[s_1, s_2, m_2]$	$[m_1, m_2, s_1, s_2]$
Sampling Distribution	$P(m_2) = 0.536$ $P(m_3) = 0.464$	$P(m_1) = P(m_2)$ $= P(m_3) = P(m_4)$ $= 0.25$	$P(m_2) = 0.308$ $P(s_1) = 0.461$ $P(m_2) = 0.231$	$P(m_1) = P(m_2)$ $= P(s_1) = P(s_2)$ $= 0.25$
Sampled Nodes	$[s_1, s_2, m_2]$	$[m_1, m_3, m_4]$	$[s_1, m_2]$	$[m_1, m_2]$

Figure 3: An illustration of the node sampling process.

graph convolution, we have $L + 1$ embeddings for node h, and we average them to obtain the final embedding:

$$\mathbf{e}_h = \frac{1}{L+1} \sum_{l=0}^{L} \mathbf{e}_h^{(l)}. \tag{4}$$

Note that the only model parameters in the simplified graph convolution are the embeddings at the 0-th layer, i.e., $\Theta_{MGCN} = \left\{\mathbf{e}_h^{(0)}\right\}$. Figure 2 illustrates an example of learning the embedding of service s_2 with graph convolutional networks in the three graphs. All three graphs share the same initial node embedding for a mashup and service, but since different graphs capture different types of relations, the final embedding of a mashup and service after the graph convolution will capture the unique characteristics specific to that relation. We denote the node embeddings obtained from the collaborative, description, and tag graphs as $\mathbf{e}^C$, $\mathbf{e}^D$, and $\mathbf{e}^T$, respectively.

3.5. Neural Factorization Machine. After learning the embeddings of mashups and services that comprehensively model different types of relations, we use a neural factorization machine model to capture both higher-order and nonlinear interactions between mashups, services, and tags. Figure 4 shows the overall model of the neural factorization machine. Given a pair of mashup and service as well as their corresponding tags and textual descriptions, NFM models both the linear interactions between each pair of features like the traditional factorization machine, and higher-order feature interactions in a nonlinear way. The input of the NFM consists of one-hot encoding of the mashup, one-hot encoding of the service, and multihot encoding of tags that are annotated by the mashup and service. We concatenate the three encoding vectors, and the sparse feature vector is denoted as $\mathbf{x} \in \{0, 1\}^{M+N+T}$, where $\mathbf{x}_i = 1$ means the i-th feature exists in the input. The linear regression part of NFM is given as

$$l(\mathbf{x}) = \mathbf{w}_0 + \sum_{i=1}^{M+N+T} \mathbf{w}_i \mathbf{x}_i, \tag{5}$$

where $\mathbf{w}_0$ is the global bias and $\mathbf{w}_i$ is the weight of feature i. To incorporate the functional semantics in the textual descriptions, we use the dense embeddings of the mashup and service learned from Doc2vec. However, since the features of mashup, service, and tags are sparse, we need to transform them into dense representations. Different from the original neural factorization model in [10], where they build an embedding lookup table which is a fully connected layer that projects each feature to a dense embedding, the multigraph convolution network is used to project the feature to its embeddings from different graphs. In this way, the mashup and service have three embeddings, each representing information from different sources, and the feature interaction can be performed in a fine-grained manner. Moreover, a cold-start mashup or service will have a meaningful embedding which is learned from the description graph and/or tag graph. For tag features, it is also beneficial to learn the tag embeddings from the tag graph, as long-tailed tags will have a more meaningful embedding from the higher-order convolution process of multihop neighbors. With the enhanced feature embeddings, the feature interactions in NFM can be learned more effectively.

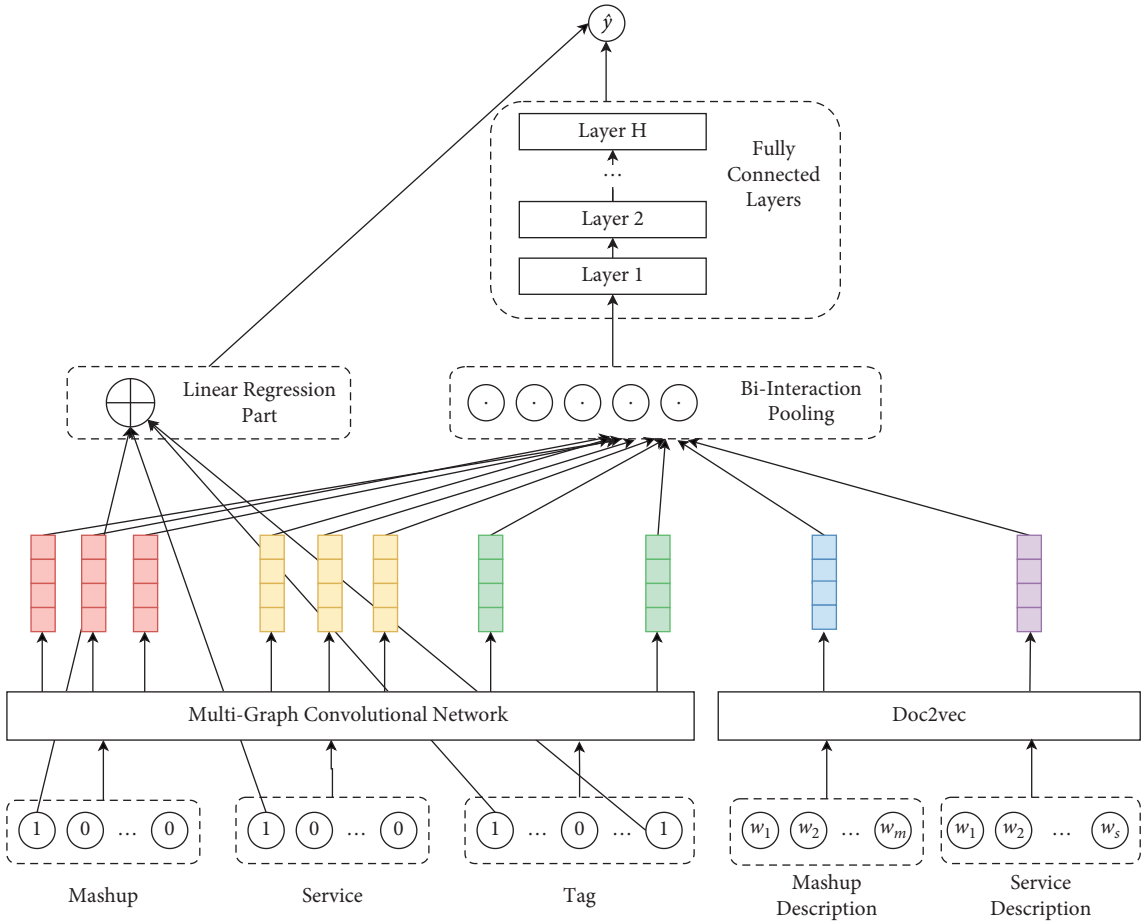


Figure 4: A framework of the neural factorization machine model.

After transforming sparse features of the mashup, service and tags into dense embeddings, they are fed into the bi-interaction pooling layer to convert multiple embedding vectors into one vector. Specifically, the input of the bi-interaction pooling layer consists of the set $\mathcal{V} = \{\mathbf{e}_m^C, \mathbf{e}_m^D, \mathbf{e}_m^T, \mathbf{e}_s^C, \mathbf{e}_s^D, \mathbf{e}_s^T, \mathbf{e}_t, \mathbf{d}_m, \mathbf{d}_s\}$, where $\mathbf{e}_t = \{\mathbf{e}_{t_1}, \mathbf{e}_{t_2}, \ldots\}$ is a list of vectors converted from the multihot tag encoding. The bi-interaction pooling operation is defined as

$$f_{BI}(\mathcal{V}) = \sum_{i=1}^{n} \sum_{j=i+1}^{n} \mathbf{v}_i \odot \mathbf{v}_j, \tag{6}$$

where n is the number of embedding vectors in $\mathcal{V}$, and $\cdot$ denotes the element-wise product of two vectors. The output of the bi-interaction pooling layer is a k-dimension vector that encodes the second-order interactions between features. It is worth noting that in the case of cold-start mashup or service, they are represented as the initial node embeddings, i.e., $\mathbf{e}_m^C$ or $\mathbf{e}_s^C$ equals $\mathbf{e}_m^{(0)}$ or $\mathbf{e}_s^{(0)}$, because there is no connection in the collaborative graph.

Above the bi-interaction layer is a stack of fully connected layers, which learn higher-order interactions between features:

$$\mathbf{z}_H = ReLU\left(W_H\left(\cdots ReLU\left(W_1 f_{BI}(\mathcal{V}) + \mathbf{b}_1\right)\cdots\right) + \mathbf{b}_H\right), \tag{7}$$

where H denotes the number of hidden layers, W_l denotes the weight matrix, and $\mathbf{b}_l$ denotes the bias vector in layer l. ReLU is the activation function. Finally, the output of the last layer $\mathbf{z}_H$ is transformed into the final score:

$$f(\mathcal{V}) = \mathbf{h}^T \mathbf{z}_H, \tag{8}$$

Overall, NFM estimates the target of an instance as the sum of the first-order linear regression part and the higher-order nonlinear feature interaction part. Since the target label is in {0,1}, with 1 indicating that the service is a component of the mashup and 0 otherwise, we transform the final output with the sigmoid function to output the probability of the service being recommended to the mashup. The sigmoid function, defined as $\sigma(x) = 1/(1 + e^{-x})$, constrains the value between 0 and 1. To summarize, the formulation of NFM is

$$\hat{y} = \sigma(l(\mathbf{x}) + f(\mathcal{V})), \tag{9}$$

with parameters $\Theta_{NFM} = \{\mathbf{w}_0, \{\mathbf{w}_i\}, \{W_l, \mathbf{b}_l\}, \mathbf{h}\}$.

3.6. Model Training and Prediction. To train the MGCN-NFM model, we define the loss function using binary cross-entropy:

$$L(\Theta) = -\sum_{i \in \mathcal{T}\cdot} \left[y^{(i)} \log\left(\hat{y}^{(i)}\right) + \left(1 - y^{(i)}\right) \log\left(1 - \hat{y}^{(i)}\right) \right] + \lambda \|\Theta\|_2^2, \tag{10}$$

where $\mathcal{T}\cdot$ is the set of training instances, each instance i includes mashup, service, tags, mashup description, and service description. $y^{(i)} \in \{0, 1\}$ is the label indicating whether the mashup and service in instance i have been composed in the record, and $\hat{y}^{(i)} \in (0, 1)$ denotes the composition probability in instance i predicted by the model. As the mashup-service composition record is extremely sparse, the number of negative instances where the label $y^{(i)} = 0$ far exceeds the positive instances where the label $y^{(i)} = 1$. Therefore, for each positive instance, we sample 5 negative instances with the same mashup to mitigate the data imbalance problem. λ controls the strength of the L_2 regularization to prevent overfitting. In addition, dropout is used in NFM to prevent overfitting, where we randomly drop ρ percent of the dimensions in the fully connected layers when training. $\Theta = \{\Theta_{MGCN}, \Theta_{NFM}\}$ denotes all trainable model parameters. We use Adaptive Moment Estimation (Adam) [40], a variant of stochastic gradient descent, to optimize the model parameters.

Algorithm 1 presents the overall training process of the MGCN-NFM framework. In line 1, three graphs are constructed based on the input data. Line 2 constructs the training instance set $\mathcal{T}\cdot$ including both positive and sampled negative labeled instances. Each instance includes a mashup, a service, their tags, and their descriptions. Line 5–7 is a forward-propagation process to make a prediction for an instance. The embeddings of the mashup, service, and tags are learned through the steps of neighbor sampling and graph convolution. The embeddings of the descriptions are obtained from Doc2vec. They are used as the input to NFM to compute the recommendation probability. Line 8 performs back-propagation and updates the model parameters based on the loss function in Equation (10).

Once the model parameters Θ are learned, which contain the initial embeddings for mashups, services, and tags, a forward-propagation pass is performed on the multigraph convolutional network to obtain the final embeddings of all mashups, services, and tags in the three graphs, which act as a fixed embedding lookup table. Then, the mashup to be recommended and each candidate service are used to construct the input instances of the NFM, with their embeddings retrieved from the lookup table. The NFM outputs the probability of each candidate service being composed by the mashup. We rank these prediction values and recommend the top-K services with the largest values to the mashup.

We analyze the time complexity of MGCN-NFM for model training and prediction. The construction of the collaborative, description, and tag graph takes $O(nnz(Y) + M^2 + N^2)$, $O((M+N)^2)$, and $O(nnz(MT) + nnz(ST))$ respectively, where *nnz* denotes the number of nonzero entries in the matrix. The main time cost of the MGCN model lies in the graph convolution process, and the time complexity is $O(L \cdot k \cdot (|\mathcal{G}^C| + |\mathcal{G}^D| + |\mathcal{G}^T|))$, where L is the number of graph convolution layers, k is the embedding size. $|\mathcal{G}^C|$, $|\mathcal{G}^D|$ and $|\mathcal{G}^T|$ are the number of edges in the collaborative, description, and tag graph respectively. Since we perform neighbor sampling for each node in the three graphs, the number of edges is upper bounded by $N^C \cdot (M+N)$, $N^D \cdot (M+N)$, $N^T \cdot (M+N+T)$ for the collaborative, description, and tag graph respectively. Therefore, the training cost of MGCN linearly scales with the number of nodes in the graph. For NFM in both training and prediction, the bi-interaction pooling layer can be efficiently computed in $O(nk)$ time with a reformulation of (6). The main time cost lies in the hidden layers of the neural network, and the time complexity is $O(\sum_{l=1}^{L} k_{l-1} k_l)$, where k_l denotes the dimension of the l-th hidden layer.

4. Experiments

In this section, we conduct a series of experiments to evaluate our proposed model in service recommendation and present the empirical performance. All experiments were developed in Python and carried out on a personal PC with Intel Core i7 CPU with 2.5 GHz and 16 GB RAM, running on macOS High Sierra. We aim to answer the following research questions:

(i) How does the proposed MGCN-NFM model perform compared with the state-of-the-art factorization machine-based service recommendation methods?

(ii) How much do the collaborative graph, description graph, and tag graph influence the performance of the proposed model?

(iii) How much do the various hyperparameters affect the experiment results of the proposed model?

4.1. Dataset Description. The experimental dataset is collected from ProgrammableWeb (PW), the largest online web service and mashup repository. For each mashup, we crawl its name, description, tags, and composed services. For each service, we crawl its name, description, and tags (including the primary and secondary categories). We remove duplicate services, and the mashups and services without textual descriptions and tags. The dataset contains 6,300 mashups and 21,474 web services, of which only 1,609 services are composed of at least one mashup. Table 1 shows the detailed statistics of the dataset.

A five-fold cross-validation is performed to evaluate the effectiveness of the model. We randomly split the mashup-service invocation records into five folds, and in each round, one fold is used as the test set while the remaining four folds are used as the training set. The results of the five rounds are averaged as the final result. In addition, to test models under different data sparsity levels, we use 1/2/3/4 folds in the training set, which corresponds to 20/40/60/80% of the mashup-service invocation records. Figure 5(a) shows the

Input: $\mathcal{M}$, $\mathcal{S}$, $\mathcal{T}$, Y, $\mathcal{D}_M$, $\mathcal{D}_S$, MT, ST
Output: parameter set Θ
(1) Construct graphs $\mathcal{G}^C$, $\mathcal{G}^D$, $\mathcal{G}^T$;
(2) Construct training instances $\mathcal{T}r$;
(3) for epoch = 1, ..., p do
(4) **for** each instance i composed of mashup m, service s, tag t in $\mathcal{T}r$ do
(5) Compute $\mathbf{e}_m^C, \mathbf{e}_m^D, \mathbf{e}_m^T, \mathbf{e}_s^C, \mathbf{e}_s^D, \mathbf{e}_s^T, \mathbf{e}_t$ with Equation (4);
(6) Obtain $\mathbf{d}_m, \mathbf{d}_s$ from Doc2vec model;
(7) Compute $\hat{y}^{(i)}$ with Equation (9);
(8) Update Θ to minimize $\mathcal{L}$ in Equation (10) with Adam;
(9) **end for**
(10) **end for**
(11) sreturn Θ

Algorithm 1: Training algorithm of MGCN-NFM.

Table 1: Dataset statistics.

Statistics	Value
Number of mashups	6,300
Number of services	21,474
Number of services composed by mashup	1,609
Number of mashup-service invocation	13,219
Average number of services in mashup	2.07
Sparsity of mashup-service composition matrix	99.87%
Number of tags	312
Average number of tags in mashups	3.62
Average number of tags in services	2.93

statistics of the service distribution of mashup in the full training set, and Figure 5(b) shows the statistics of the service distribution of mashup when only 20% of the dataset is used for training. We can see that for the full training dataset, 52.3% mashups compose one service, and 91.2% mashups compose less than 3 services. When only 20% of the dataset is used for training, 68.6% mashups do not have any component services, which poses a significant challenge for the cold-start problem.

4.2. Evaluation Metrics. We adopt two metrics, namely, recall and Normalized Discounted Cumulative Gain (NDCG), to evaluate the accuracy of the top-K service recommendation.

Recall@K is the proportion of the number of services in the top-K recommendation list that is composed by the mashup to the number of services in the mashup. It is defined as:

$$Recall@K = \frac{1}{|M|} \sum_{m \in M} \frac{|rec(m) \cap truth(m)|}{|truth(m)|}, \quad (11)$$

where M is the set of mashups in the test set, rec(m) is the recommendation list of services of size K for mashup m. truth(m) is the ground truth list of services that are composed by the mashup m in the test set.

NDCG@K considers the ranking position of the recommended services, and assigns different weights to each service in the top-K recommendation list. The higher ranked service is assigned with a larger weight if it is composed by the mashup. It is defined as:

$$NDCG@K = \frac{1}{|M|} \sum_{m \in M} \frac{\sum_{i=1}^{K} 2^{I(i)} - 1/log_2(i+1)}{IDCG@K}, \quad (12)$$

where $I(i)$ indicates whether the service at position i of the ranking list is in truth(m). IDCG@K is the ideal DCG score of the top-K services that can be achieved.

4.3. Baseline Methods. We compare MGCN-NFM with the following baselines that are related to our work:

(i) RTM-FM [7]: this method combines RTM and FM for service recommendation. It uses RTM to mine latent topics of mashups and services by link prediction. In addition, it exploits the co-occurrence and popularity features of services in FM.

(ii) TR-FM [19]: this method integrates tag, topic, co-occurrence, and popularity factors which are modeled by FM for service recommendation. It uses the enriched tags and the derived topic information

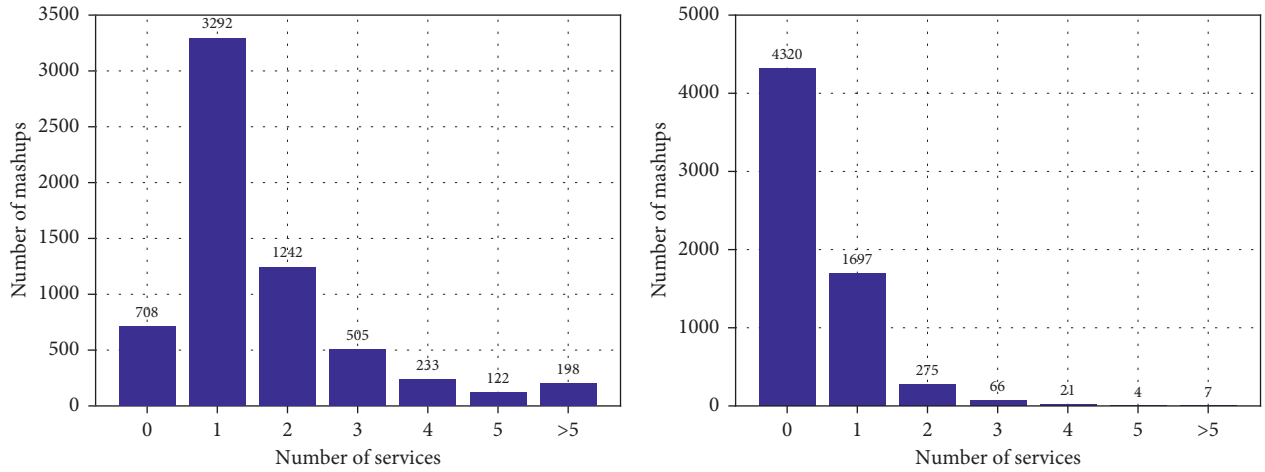


Figure 5: Service distribution of mashups. (a) Full dataset. (b) 20% of the dataset.

from RTM to measure the similarity between services.

(iii) ATT-FM [8]: this method uses attentional FM for service recommendation. It uses functional similarity, tags, and popularity of services as features, and employs an attention mechanism on feature interactions to discriminate their importance.

(iv) NAFM [9]: this is the state-of-the-art service recommendation method based on the FM model. Besides first-order and second-order linear feature interactions, it integrates a deep neural network to capture nonlinear feature interactions and an attention network to capture important feature interactions.

(v) MGCN-FM: it is a variation of the proposed method where we use the basic factorization machine model after learning the features of mashup, service, and tags. The output of the bi-interaction pooling layer is summed directly, which is equivalent to setting $H = 0$, $\mathbf{h} = 1$ in MGCN-NFM.

4.4. Parameter Settings. To learn the representation of textual descriptions of mashups and services in Doc2vec, we aggregate the textual descriptions of all 6300 mashups and 21474 services into one big corpus and perform a series of preprocessing steps: (1) split sentences into words and transform them into lower case; (2) remove stop words and words that appear less than 5 times; (3) transform words into their root form. All preprocessing steps are done using the NLTK library (https://www.nltk.org). After preprocessing, we use the models.doc2vec API in the gensim library (https://radimrehurek.com/gensim/) to learn the document embedding. All parameters are set to the default value of the gensim API except for the dimension size, which is set to 50 as the length of the description of a mashup and service is relatively short. The default values of hyperparameters in MGCN-NFM are set in Table 2.

4.5. Performance Comparison. To assess the performance of our method and baseline methods under different data sparsity levels, we select 20/40/60/80% of the data for constructing the collaborative graph and learning the parameters of the model. Table 3 shows the comparisons of different methods when the number of services recommended for a mashup $K = 3/5/10$. It can be seen that our model MGCN-NFM consistently achieves the best performance under all percentages of training data in terms of both Recall and NDCG. TR-FM achieves better performance than RTM-FM, because it considers the tagging information to refine the similarity measures of services. ATT-FM achieves better performance than the two methods that use the vanilla factorization model, since it models the feature interactions in the factorization machine with an attention mechanism that can discriminate the importance of each feature. NAFM attains better performance than the aforementioned three methods, since considering higher-order feature interactions can bring additional improvement. The MGCN-FM method achieves better performance than all baseline methods, but its performance is inferior to that of MGCN-NFM, especially on training data with high sparsity. This shows that introducing nonlinear feature interactions can improve the expressiveness of FM. Compared with the best baseline model NAFM, our model obtains 7.28% and 7.14% improvement in top-3 recommendations for Recall and NDCG respectively under the full training data. The performance gain becomes even more significant when the training data becomes sparser. For instance, when only 20% of the data is used for training, our model obtains 18.13% and 14.97% improvements in top-3 recommendations for Recall and NDCG compared with NAFM. As the percentage of the training data increases, both Recall and NDCG increase for all methods. However, our method consistently outperforms baseline methods when the training data becomes sparser, while the performance of the baseline methods drops to a greater extent than our method. It shows that our method is less sensitive to the sparsity of the training

Table 2: Default value of hyperparameters.

Item	Value
Number of convolution layers, L	2
Number of latent dimension, k	64
Number of hidden layers, H	2
Hidden units size, k_l	(16,16)
Dropout ratio, ρ	0.3
Initialization of parameters	$\mathcal{N}(0, 0.01)$
Learning rate	0.001
Regularization coefficient, λ	0.0001
Training epoch, p	100
Neighbor size of collaborative graph, N^C	15
Neighbor size of description graph, N^D	30
Neighbor size of tag graph, N^T	30

data compared with the baseline methods, which could be explained by the higher-order and nonlinear feature interaction modeling in graph convolution and NFM.

4.6. Impact of Different Graphs. In MGCN-NFM, we represent different sources of information by constructing three graphs: collaborative graph, description graph, and tag graph. To demonstrate the effectiveness of utilizing the three graphs in our model, we design six comparing variants of the model: (1) CGCN-NFM, where only the collaborative graph is constructed. In this model, only $\mathbf{e}_m^C$ and $\mathbf{e}_s^C$ are computed in Step 5 of Algorithm 1; (2) DGCN-NFM, where only the description graph is constructed, and only $\mathbf{e}_m^D$ and $\mathbf{e}_s^D$ are computed; (3) TGCN-NFM, where only the tag graph is constructed, and only $\mathbf{e}_m^T$, $\mathbf{e}_s^T$ and $\mathbf{e}_t$ are computed; (4) CDGCN-NFM, where both collaborative graph and description graph are constructed, and tag graph is omitted; (5) CTGCN-NFM, where both collaborative graph and tag graph are constructed; (6) DTGCN-NFM, where only collaborative graph is omitted in the original model.

Figure 6 shows the result of the comparison among different variants. We observe that MGCN-NFM achieves the best performance, confirming that all three graphs are necessary for the model, as they are complementary to each other for enriching different aspects of relations between mashups and services. When either graph is dropped from the original model, the performance becomes worse, which also suggests that both three graphs can indeed help improve the model performance. Furthermore, incorporating two graphs into the model is better than only utilizing one graph in the model, except for CTGCN-NFM. In this case, we argue that the description graph is the most important graph for the model performance, as DGCN-NFM outperforms CGCN-NFM and TGCN-NFM, and without the description graph, CTGCN-NFM performs even worse than DGCN-NFM. Therefore, we conclude that accurately modeling the functional information of mashups and services in their descriptions is vital. In addition, the collaborative graph has a stronger influence on the model performance than the tag graph, which is exhibited by the fact that CGCN-NFM outperforms TGCN-NFM and CDGCN-NFM outperforms DTGCN-NFM.

We further discuss the reason and provide several case studies of how the three graphs alleviate the sparsity, imbalance, and cold-start problems in service recommendation. The collaborative graph connects services that are rarely composed to frequently composed ones, which makes them more densely connected, so that the latent factors can be updated more frequently. The description graph and tag graph connect cold-start services with services that are composed by mashups, so that the cold-start services can be discovered and recommended by the FM model. Table 4 shows some actual examples of service recommendation results. CityPockets Daily Deals is a mashup that aggregates all daily deals and coupons from various deal sites. Compared with DTGCN-NFM which excludes the collaborative graph, MGCN-NFM successfully recommends service 8coupons in the Top-5 list, because service Groupon and service 8coupons have been composed by different mashups a few times, and it is captured by the collaborative graph. NearPlace mashup is a free store locator and Google Maps marker. Compared with CTGCN-NFM which excludes the description graph, MGCN-NFM successfully recommends cold-start service MetaLocator in the Top-5 list. Although MetaLocator has never been composed by any mashups before, it is a mapping and locator service for stores, vendors, and ATMs. Therefore, its functionality is similar to popular services such as geocoder and Google Maps Places, which is reflected in the description graph. iEnviroWatch is a mashup that visualizes and queries environmental information in a geographical area of one's interest. The MGCN-NFM successfully recommends the EEA Discomap service in the Top-5 list, because they are both tagged with "Environment", and the tag "Sustainability" in the mashup is also close to the tag "Environment" in the tag graph.

4.7. In-Depth Analysis of Graph Construction

4.7.1. Impact of Mashup-Mashup Graph and Service-Service Graph in Collaborative Graph Construction. To demonstrate the superiority of incorporating mashup-mashup collaborative graph and service-service collaborative graph into the sparse mashup-service collaborative graph, we design three variants of the model: (1) $\mathcal{G}^{MS}$, where only the mashup-service collaborative graph is constructed; (2) $\mathcal{G}^{MS} + \mathcal{G}^{MM}$,

Table 3: Recall and NDCG comparison of recommendation methods.

	Method	Dataset density							
		Recall				NDCG			
		20%	40%	60%	80%	20%	40%	60%	80%
$K=3$	RTM-FM	0.2429	0.2940	0.3423	0.3806	0.2392	0.2803	0.3105	0.3271
	TR-FM	0.2685	0.3197	0.3628	0.4053	0.2530	0.2930	0.3238	0.3403
	ATT-FM	0.3421	0.3955	0.4225	0.4598	0.3013	0.3357	0.3623	0.3759
	NAFM	0.3733	0.4256	0.4526	0.4724	0.3127	0.3461	0.3724	0.3862
	MGCN-FM	0.4195	0.4652	0.4851	0.5035	0.3402	0.3736	0.3989	0.4116
	MGCN-NFM	0.4410	0.4798	0.4931	0.5068	0.3595	0.3859	0.4069	0.4138
$K=5$	RTM-FM	0.3807	0.4336	0.4855	0.5084	0.2785	0.3184	0.3583	0.3760
	TR-FM	0.4021	0.4548	0.5061	0.5291	0.2913	0.3309	0.3697	0.3863
	ATT-FM	0.4492	0.4969	0.5430	0.5611	0.3368	0.3710	0.4030	0.4144
	NAFM	0.4650	0.5116	0.5574	0.5755	0.3502	0.3838	0.4152	0.4260
	MGCN-FM	0.5071	0.5538	0.5994	0.6173	0.3787	0.4119	0.4432	0.4538
	MGCN-NFM	0.5336	0.5711	0.6087	0.6202	0.3979	0.4244	0.4519	0.4567
$K=10$	RTM-FM	0.4257	0.4947	0.5346	0.5579	0.2822	0.3461	0.3794	0.3965
	TR-FM	0.4472	0.5156	0.5548	0.5770	0.2977	0.3504	0.3835	0.4002
	ATT-FM	0.4969	0.5570	0.5917	0.6094	0.3463	0.3920	0.4166	0.4288
	NAFM	0.5131	0.5729	0.6073	0.6245	0.3590	0.4048	0.4292	0.4403
	MGCN-FM	0.5576	0.6173	0.6512	0.6678	0.3910	0.4366	0.4609	0.4712
	MGCN-NFM	0.5769	0.6294	0.6591	0.6721	0.4096	0.4478	0.4673	0.4750

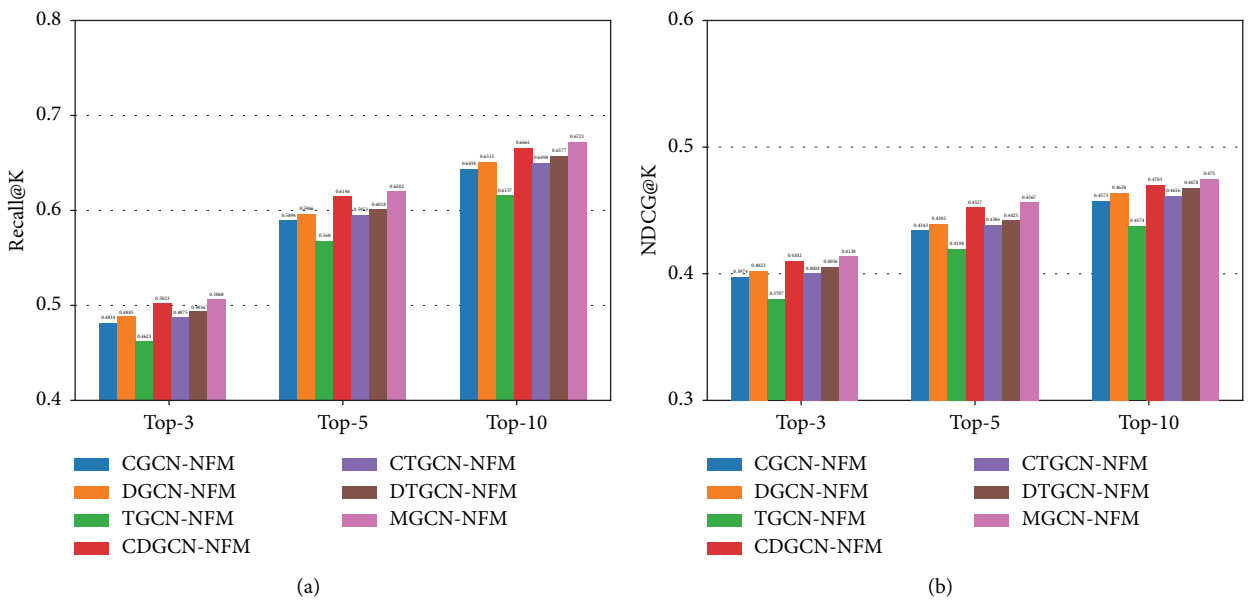


Figure 6: Comparison of different variants of MGCN-NFM. (a) Recall@K. (b) NDCG@K.

Table 4: Case-study analysis of different graphs where the hit services are marked in bold font.

Mashup	Method	Top-5 recommendation of services
CityPockets daily deals	DTGCN-NFM	**Yelp fusion**, amazon product advertising, google maps, facebook, **groupon**
	MGCN-NFM	**Yelp fusion**, amazon product advertising, **groupon**, google maps, 8coupons
	Ground truth	**Yelp fusion**, **groupon**, **8coupons**
NearPlace	CTGCN-NFM	**Google maps**, microsoft bing maps, geocoder, Shopping.com, amazon marketplace
	MGCN-NFM	**Google maps**, geocoder, google maps places, Shopping.com, **MetaLocator**
	Ground truth	**MetaLocator, google maps, WordPress.org, shopify admin, WooCommerce, magento SOAP**
iEnviroWatch	CDGCN-NFM	Google maps, twitter, google latitude, facebook, WiserEarth
	MGCN-NFM	Google maps, twitter, google latitude, AMEE, **EEA discomap**
	Ground truth	**EEA discomap**

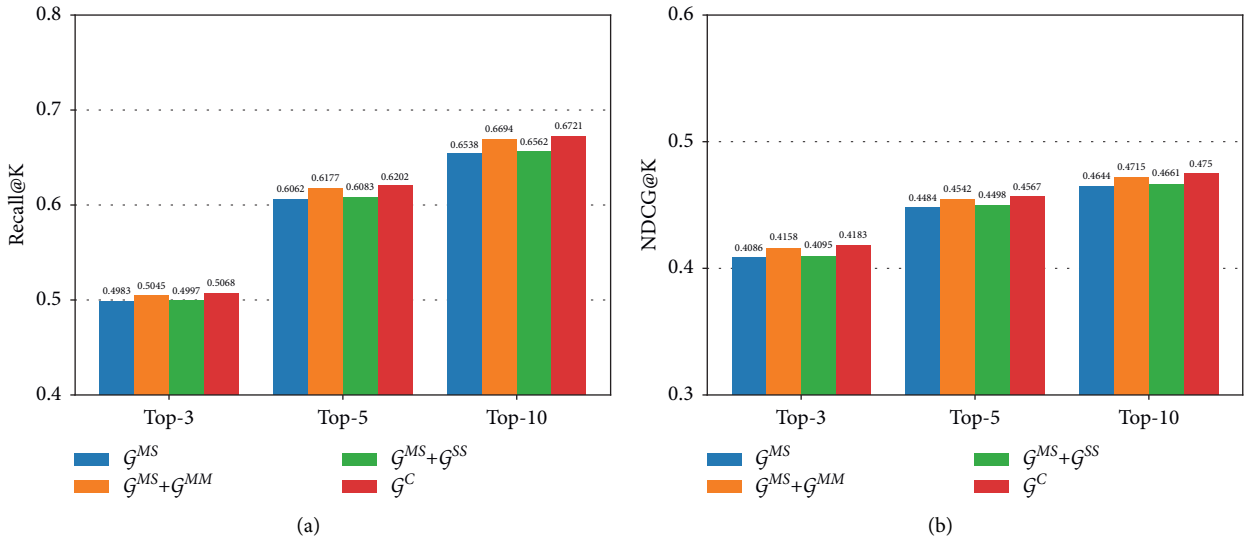


Figure 7: Comparison of different methods for collaborative graph construction. (a) Recall@K. (b) NDCG@K.

where the mashup-mashup collaborative graph $\mathcal{G}^{MM}$ is fused with $\mathcal{G}^{MS}$; (3) $\mathcal{G}^{MS} + \mathcal{G}^{SS}$, where the service-service collaborative graph $\mathcal{G}^{SS}$ is fused with $\mathcal{G}^{MS}$; (4) $\mathcal{G}^{C}$, which is the original model. Figure 7 shows the results of different variants. We observe that incorporating $\mathcal{G}^{MM}$ and $\mathcal{G}^{SS}$ both achieves better model performance than only using $\mathcal{G}^{MS}$ for the collaborative graph, which verifies the effectiveness of modeling the collaborative relations between mashups and between services. Furthermore, the collaborative relation between mashups has a more significant impact on the model performance than the collaborative relation between services. Only by combining both relations can the best result be achieved.

4.7.2. Impact of Document Embedding Techniques in Description Graph Construction. In this paper, we use Doc2vec, which learns a distributed representation of documents with a self-supervised learning paradigm. We compare Doc2vec with other document modeling techniques widely used for service description. Three variants are considered: (1) TF-IDF (https://en.wikipedia.org/wiki/Tf-idf), where the description documents are represented by the TF-IDF model; (2) LDA, where the topic probability distributions of description documents are learned by the LDA model; (3) HDP, where the topic probability distributions of description documents are learned by the HDP model; (4) Doc2vec, which is the original model. The latent dimensions of LDA and Doc2vec are set to 50. Figure 8 shows the results of different variants. We can observe that Doc2vec performs the best among all methods, indicating its superior modeling capacity for mashup and service description. HDP follows behind Doc2vec and performs better than LDA, which shows that HDP can derive better topic distribution than LDA as it can automatically infer the number of topics from data. TF-IDF performs the worst, as it only uses the term-based vector space model and lexical matching between documents.

4.8. Impact of Hyperparameters

4.8.1. Impact of the Size of Sampling Neighbors. In the neighbor sampling process, to control the size of the graph, we sample a fixed-size set of neighbors for each node. We vary the number of sampled neighbors for each node in the collaborative graph N^C, description graph N^D, and tag graph N^T from 5 to 40 with a step size of 5 to find the optimal setting. The experimental results are shown in Figure 9. We only report Recall@K as NDCG@K performs in a similar fashion. We can see that the optimal size of neighbors is different for each graph. For the collaborative graph, the optimal value of N^C is relatively small ($N^C = 15$), while for the description graph, the optimal value of N^D is larger ($N^D = 30$). When the size of neighbors further increases past the optimal value, the performance begins to decrease, which verifies the necessity of node sampling before graph convolution. For the tag graph, we observe that the performance does not change much as the size of neighbors grows. We set the optimal value of N^T to 30 considering both the training efficiency and recommendation accuracy.

4.8.2. Impact of the Number of Layers in Graph Convolution. We evaluate the effectiveness of graph convolution for learning higher-order mashup and service features. We vary the number of layers L in the range of {1,2,3,4} and evaluate the performance. The experimental results are shown in Figure 10. We can see that the optimal value of $L = 3$. Increasing layers from 1 to 3 improves the performance as

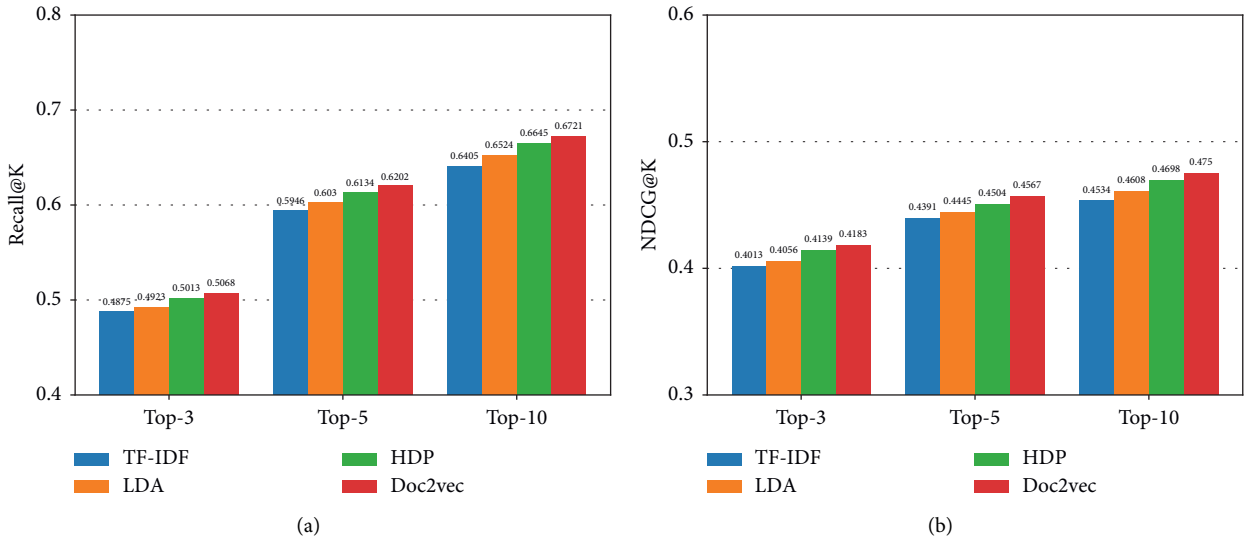


Figure 8: Comparison of different methods for description graph construction. (a) Recall@K. (b) NDCG@K.

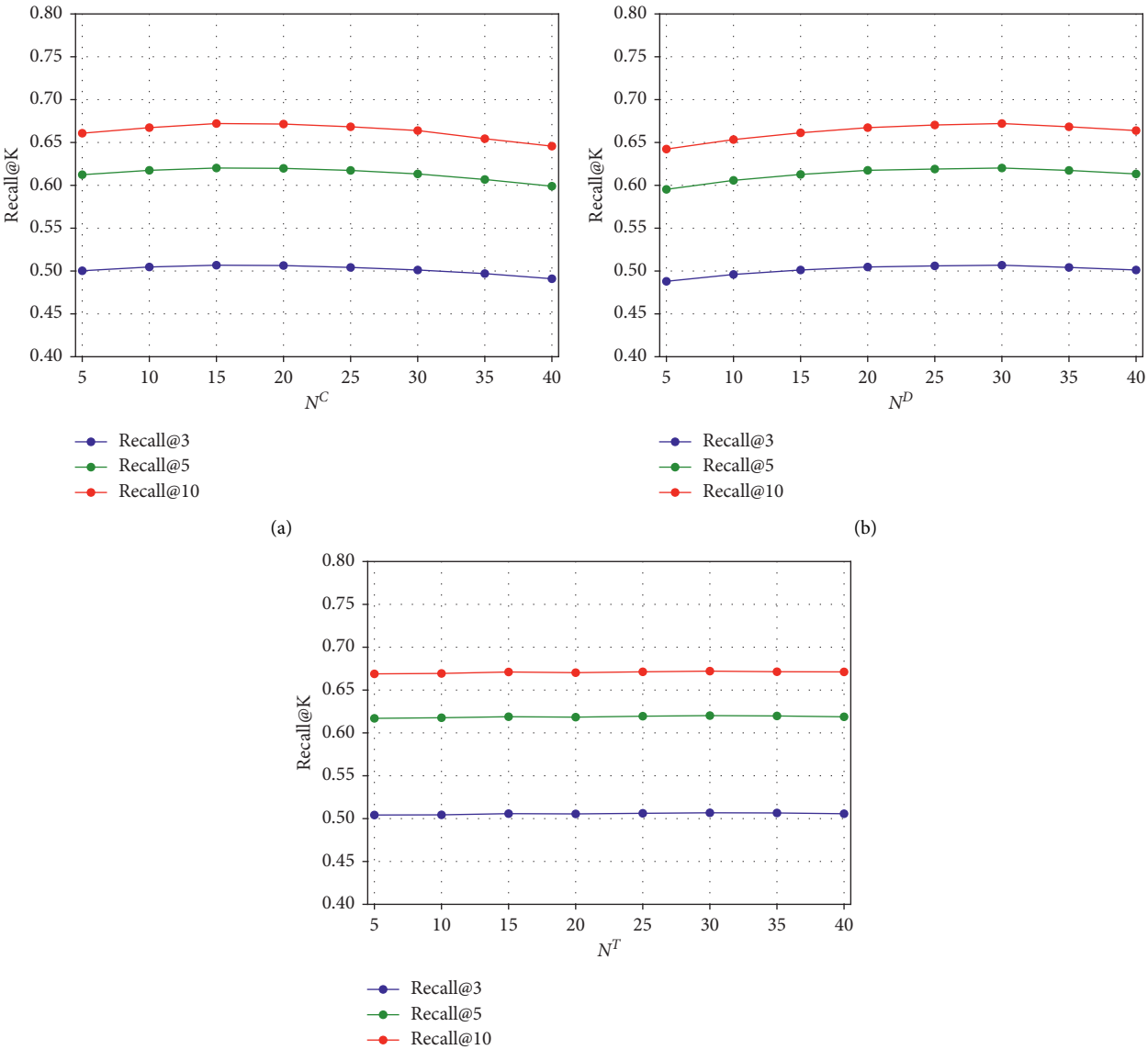


Figure 9: Impact of the size of neighbors. (a) N^C. (b) N^D. (c) N^T.

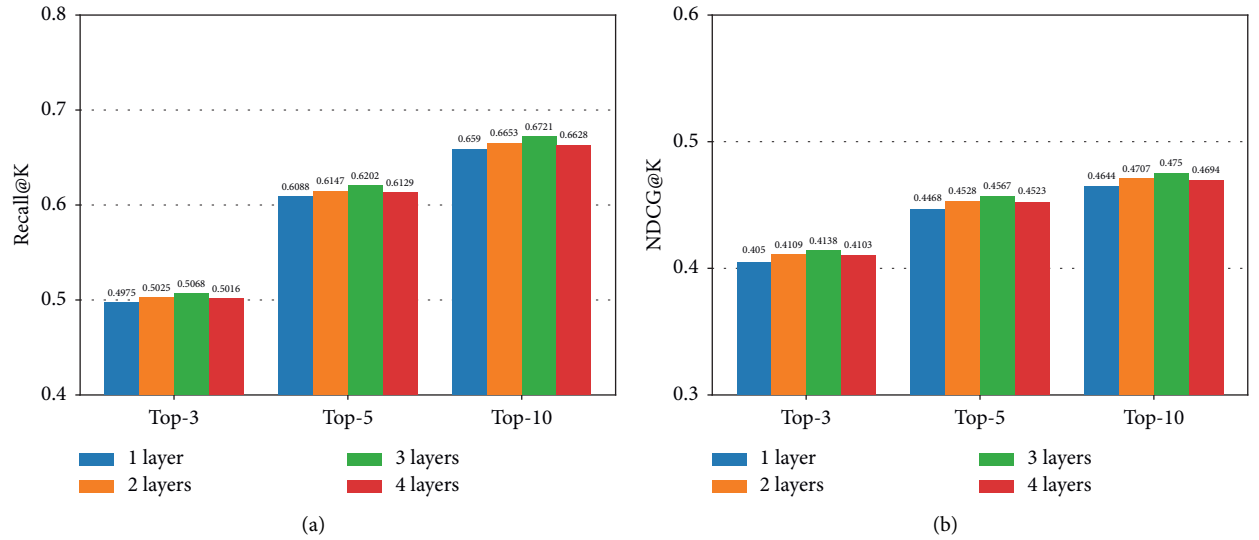


Figure 10: Impact of the number of layers in graph convolution. (a) Recall@K. (b) NDCG@K.

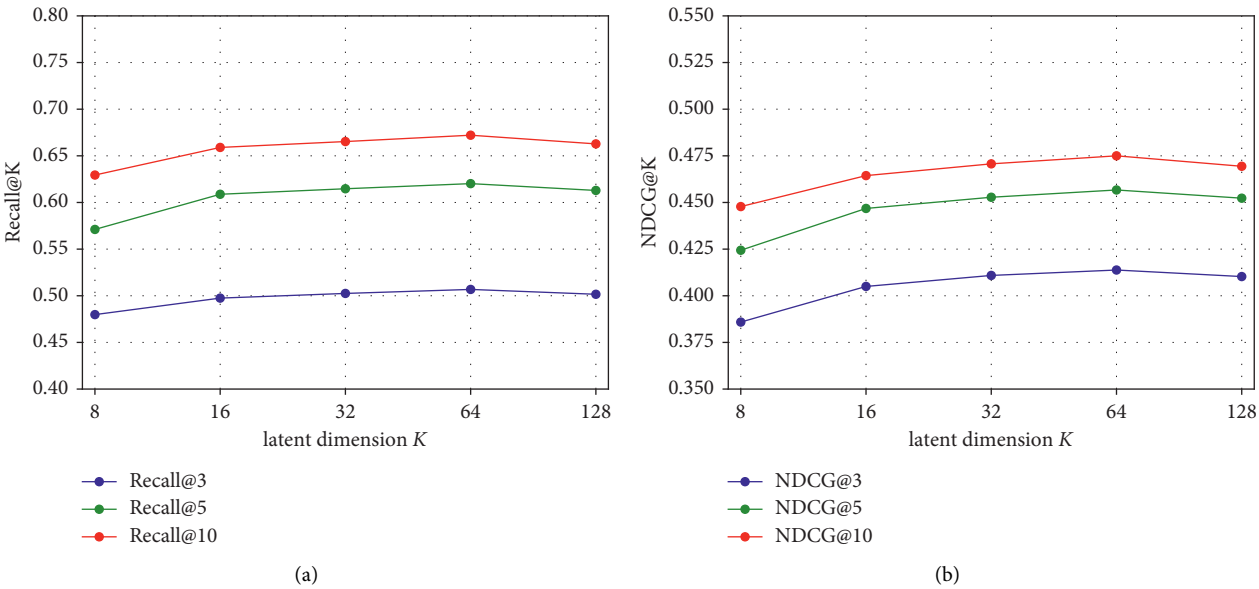


Figure 11: Impact of the latent dimension. (a) Recall@K. (b) NDCG@K.

stacking more layers helps nodes reach multihop neighbors to enrich mashup and service features. However, stacking too many layers can make the node features similar and reduce the performance.

4.8.3. Impact of the Latent Dimension. We evaluate the impact of latent dimension k on the model performance, which varies in the range of {8,16,32,64,128}. The performance is shown in Figure 11, from which we can see that the optimal value of k is 64. Setting the latent dimension too small will probably constrain the model capacity, and setting the latent dimension too large not only leads to the overfitting issue that hurts the recommendation accuracy, but also increases the model training time.

5. Conclusions

In this paper, we delve into three issues, namely, feature sparsity, imbalance, and cold-start when applying factorization machine models for service recommendation, and propose a novel MGCN-NFM model to address those issues. First, three graphs are built to represent the various types of interactions between mashups and services: a collaborative graph, a description graph, and a tag graph. Next, we use the graph convolutional network to iteratively aggregate higher-order neighbors to enrich the information of the sparsely-connected nodes in each graph. The feature embeddings are used as latent factors in the neural factorization model, where it predicts the probability of composition given a pair of mashup and service, their tags, and textual descriptions.

We perform extensive experiments on the ProgrammableWeb dataset, and the results demonstrate the superiority of our proposed method. Moreover, our model is able to incorporate other information from mashups and services by defining graphs with new relations.

In the future, we intend to investigate other potentially useful attributes and social information of mashups and services, such as QoS, developers, followers, etc. In addition, other more advanced GCN models and FM models can be explored to further improve the recommendation accuracy. Finally, as the list of recommended services should be of high coverage and diversity besides high accuracy, we plan to investigate and improve the model with respect to the two metrics.

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 there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This research was partially supported by National Key R&D Program of China under grant no. 2019YFB1404802, National Natural Science Foundation of China under grants nos. 62176231 and 62106218, Zhejiang Public Welfare Technology Research Project under grant no. LGF20F020013, Wenzhou Bureau of Science and Technology of China under grant no. Y2020082.

References

[1] A. Bouguettaya, M. Singh, M. Huhns et al., "A service computing manifesto," *Communications of the ACM*, vol. 60, no. 4, pp. 64–72, 2017, https://www.programmableweb.com/.

[2] D. Benslimane, S. Dustdar, and A. Sheth, "Services mashups: the new generation of web applications," *IEEE Internet Computing*, vol. 12, no. 5, pp. 13–15, 2008.

[3] M. P. Papazoglou, P. Traverso, S. Dustdar, and F. Leymann, "Service-oriented computing: a research roadmap," *International Journal of Cooperative Information Systems*, vol. 17, no. 02, pp. 223–255, 2008.

[4] L. Yao, Q. Sheng, A. Ngu et al., "Unified collaborative and content-based web service recommendation," *IEEE Transactions on Services Computing*, vol. 8, no. 3, pp. 453–466, 2014.

[5] S. Rendle, "Factorization machines," in *Proceedings of the IEEE International Conference on Data Mining*, pp. 995–1000, 2010.

[6] Y. Qu, H. Cai, K. Ren et al., "Product-based neural networks for user response prediction," in *Proceedings of the IEEE International Conference on Data Mining*, pp. 1149–1154, 2016.

[7] B. Cao, J. Liu, Y. Wen, H. Li, Q. Xiao, and J. Chen, "QoS-aware service recommendation based on relational topic model and factorization machines for IoT Mashup applications," *Journal of Parallel and Distributed Computing*, vol. 132, pp. 177–189, 2019.

[8] Y. Cao, J. Liu, M. Shi et al., "Service recommendation based on attentional factorization machine," in *Proceedings of the IEEE International Conference on Services Computing*, pp. 189–196, 2019.

[9] G. Kang, J. Liu, B. Cao, and M. Cao, "Nafm: neural and attentional factorization machine for Web api recommendation," in *Proceedings of the IEEE International Conference on Web Services*, pp. 330–337, 2020.

[10] X. He and T. S. Chua, "Neural factorization machines for sparse predictive analytics," in *Proceedings of the ACM SIGIR conference on Research and Development in Information Retrieval*, pp. 355–364, 2017.

[11] B. Cao, J. Liu, M. Tang et al., "Mashup service recommendation based on user interest and social network," in *Proceedings of the IEEE International Conference on Web Services*, pp. 99–106, 2013.

[12] W. Gao, L. Chen, J. Wu, and H. Gao, "Manifold-learning based api recommendation for mashup creation," in *Proceedings of the IEEE International Conference on Web Services*, pp. 432–439, 2015.

[13] T. Liang, L. Chen, J. Wu, H. Dong, and A. Bouguettaya, "Meta-path based service recommendation in heterogeneous information networks," *Service-Oriented Computing*, in *Proceedings of the International Conference on Service-Oriented Computing*, pp. 371–386, 2016.

[14] F. Xie, L. Chen, D. Lin, Z. Zheng, and X. Lin, "Personalized service recommendation with mashup group preference in heterogeneous information network," *IEEE Access*, vol. 7, no. 7, pp. 16155–16167, 2019.

[15] X. Wang, H. Wu, and CH. Hsu, "Mashup-oriented api recommendation via random walk on knowledge graph," *IEEE Access*, vol. 2018, no. 7, pp. 7651–7662, 2018.

[16] W. Xu, J. Cao, L. Hu et al., "A social-aware service recommendation approach for mashup creation," in *Proceedings of the IEEE International Conference on Web Services*, pp. 107–114, 2013.

[17] L. Yao, X. Wang, Q. Z. Sheng et al., "Mashup recommendation by regularizing matrix factorization with API co-invocations," *IEEE Transactions on Services Computing*, vol. 14, no. 2, pp. 502–515, 2018.

[18] W. Gao, L. Chen, J. Wu, H. Dong, and A. Bouguettaya, "Personalized API recommendation via implicit preference modeling," *Service-Oriented Computing*, in *Proceedings of the International Conference on Service-Oriented Computing*, pp. 646–653, 2016.

[19] H. Li, J. Liu, B. Cao et al., "Integrating tag, topic, co-occurrence, and popularity to recommend web apis for mashup creation," in *Proceedings of the IEEE International Conference on Services Computing*, pp. 84–91, 2017.

[20] B. Cao, B. Li, J. Liu et al., *Mobile Service Recommendation via Combining Enhanced Hierarchical Dirichlet Process and Factorization Machines*, vol. 2019, , 6423805, 2019.

[21] F. Xie, L. Chen, Y. Ye et al., "Factorization machine based service recommendation on heterogeneous information networks," in *Proceedings of the IEEE International Conference on Web Services*, pp. 115–122, 2018.

[22] X. Zhang, J. Liu, B. Cao et al., "Web service recommendation via combining Doc2Vec-based functionality clustering and DeepFM-based score prediction," in *Proceedings of the IEEE International Conference on Parallel & Distributed Processing with Applications*, Ubiquitous Computing & Communications, Big Data & Cloud Computing, Social Computing & Networking, Sustainable Computing & Communications, , pp. 509–516, 2018.

[23] T. Chen, J. Liu, B. Cao et al., "Web service recommendation based on word embedding and topic model," in *Proceedings of the IEEE International Conference on Parallel & Distributed Processing with Applications*, Sustainable Computing & Communications, Ubiquitous Computing & Communications, Big Data & Cloud Computing, Social Computing & Networking, , pp. 903–910, 2018.

[24] R. Xiong, J. Wang, N. Zhang, and Y. Ma, "Deep hybrid collaborative filtering for web service recommendation," *Expert Systems with Applications*, vol. 110, pp. 191–205, 2018.

[25] Y. Zhang, H. Yang, and L. Kuang, "A web API recommendation method with composition relationship based on GCN," in *Proceedings of the IEEE International Conference on Parallel & Distributed Processing with Applications, Big Data & Cloud Computing, Sustainable Computing & Communications*, pp. 601–608, Social Computing & Networking, 2020.

[26] S. Lian and M. Tang, "API Recommendation for Mashup Creation Based on Neural Graph Collaborative Filtering," *Connection Science*, pp. 1–15, 2021.

[27] W. He, C. Xia, Z. Li, X. Liu, and T. Wang, "A heterogeneous graph attention network-based web service link prediction," in *Proceedings of the International Conference on Computer Communication and the Internet*, pp. 102–108, 2021.

[28] Z. Zheng, H. Ma, M. R. Lyu, and I. King, "QoS-aware web service recommendation by collaborative filtering," *IEEE Transactions on Services Computing*, vol. 4, no. 2, pp. 140–152, 2010.

[29] Z. Zheng, L. Xiao, M. Tang et al., *Web Service QoS Prediction via Collaborative Filtering: A Survey*, IEEE Transactions on Services Computing, 2020.

[30] H. Guo, R. Tang, Y. Ye, Z. Li, and X. He, "DeepFM: a factorization-machine based neural network for CTR prediction," in *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence*, pp. 1725–1731, 2017.

[31] J. Xiao, H. Ye, X. He et al., "Attentional factorization machines: learning the weight of feature interactions via attention networks," in *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence*, pp. 3119–3125, Web API Recommendation Method with Composition Relationship Based, 2017.

[32] Z. Wu, S. Pan, F. Chen et al., "A comprehensive survey on graph neural networks," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 32, no. 1, pp. 4–24, 2020.

[33] Q. Le and T. Mikolov, "Distributed representations of sentences and documents," in *Proceedings of the International Conference on Machine Learning*, pp. 1188–1196, 2014.

[34] T. Mikolov, I. Sutskever, K. Chen et al., "Distributed representations of words and phrases and their compositionality," in *Advances in Neural Information Processing Systems*, pp. 3111–3119, 2013.

[35] D. Blei, A. Ng, and M. Jordan, "Latent dirichlet allocation," *The Journal of Machine Learning Research*, vol. 2003, no. 3, pp. 993–1022, 2003.

[36] Y. Teh, M. Jordan, M. Beal, and D. Blei, "Hierarchical dirichlet process," *Journal of the American Statistical Association*, vol. 101, no. 476, pp. 1566–1581, 2004.

[37] T. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," in *Proceedings of the International Conference on Learning Representations*, 2017.

[38] Q. Li, Z. Han, and X. Wu, "Deeper insights into graph convolutional networks for semi-supervised learning," in *Thirty-Second AAAI conference on artificial intelligence*, 2018.

[39] X. He, K. Deng, X. Wang et al., "Lightgcn: simplifying and powering graph convolution network for recommendation," in *Proceedings of the International ACM SIGIR conference on research and development in Information Retrieval*, pp. 639–648, Association for Computing Machinery, New York, 2020.

[40] D. Kingma and J. BaAdam, "A method for stochastic optimization," in *Proceedings of the International Conference on Learning Representations*, 2015.

Hindawi
Mathematical Problems in Engineering
Volume 2022, Article ID 3965720, 12 pages
https://doi.org/10.1155/2022/3965720

Research Article

Modelling Underload Cascading Failure and Mitigation Strategy of Supply Chain Complex Network in COVID-19

Hong Liu,[1,2] **Yunyan Han**,[3] **Jinlong Ni**,[1,2] **and Anding Zhu**[2,3]

[1]*School of Computer and Information Engineering, Zhejiang Gongshang University, Hangzhou 310018, China*
[2]*Contemporary Business and Trade Research Center of Zhejiang Gongshang University, Key Research Institute of Humanities and Social Sciences of the Ministry of Education, Hangzhou 310018, China*
[3]*School of Management and E-Business, Zhejiang Gongshang University, Hangzhou 310018, China*

Correspondence should be addressed to Jinlong Ni; nijl@mail.zjgsu.edu.cn

Received 22 September 2021; Revised 27 December 2021; Accepted 5 January 2022; Published 21 February 2022

Academic Editor: A. M. Bastos Pereira

The outbreak of COVID-19 has caused problems such as shortage of workforce, cost increase, cash flow tension, and uncertainty of supply chain. It has a specific negative impact on the raw material supply, procurement management, production resumption, logistics, and market of the supply chain, which can trigger cascading failures in supply chain networks. Aiming at the failure of upstream/downstream firms in supply chain networks due to the decreased product demand/material supply under the COVID-19, the present study adopted an underload cascading failure model for the supply chain networks. In this model, the hierarchical supply chain networks were constructed based on the Erdos Renyi (ER) model and Barabasi Albert (BA) model. The validity of the model was verified under random attack and target attack. In the random attack mode, the influences of model parameters were studied, and in the target attack mode, the influence of target protection and random protection measures on enhancing network invulnerability was also studied. Simulation results showed that the initial load and capacity lower bound of nodes impact cascading failure size. The former has a positive correlation with cascading failure size, while the latter negatively correlates with cascading failure size. Furthermore, random protection measures are more practical to prevent cascading failures.

1. Introduction

The ongoing COVID-19 pandemic has posed severe disruptions on global supply chains [1, 2]. After the outbreak of the COVID-19 pandemic, firms faced more challenges such as labor shortage, cost increase, cash flow interruption, and the other uncertain emergencies of the supply chain, which have negative impacts on raw materials supply, procurement management, production resumption, and logistics and market in supply chains [3, 4]. A supply chain network is a functional chain network structure organized by a few various business entities like suppliers, manufacturers, distributors, and retailers, realizing from raw material purchase to finished product production and final sales [5]. As a dynamic network, the fluctuation of any part of the network can spread quickly in the whole network [6]. Cascading failure is one of the fundamental reasons for supply chain disruptions. Supply chains must consistently achieve robustness by trading off efficiency against vulnerability [7]. The international shipping costs have been rising to more than 400 per cent since the outbreak of COVID-19, resulting in port congestions and supply chain postponements all over the world [8]. On the other side, numerous companies went bankrupt due to COVID-19. According to statistics released by the Administrative Office of the U.S. Courts, there were more than one million bankruptcy filings from 2020 to 2021 [9]. Similarly, there were more than ten thousand company bankruptcies related to Japan's pandemic from 2020 to 2021 [10]. Every company bankruptcy may become the vulnerable point causing subsequent bankruptcies and weakening the robustness of the entire supply chain.

Cascading failure is a condition of interconnected systems when failure on one part can lead to a subsequent failure in corresponding parts and finally cause an overall failure at the system level [11]. Cascading failures in supply chain networks have attracted much attention in recent years. Extant studies mainly focus on network vulnerability [12], cascading failure model [11], and emergency recovery mechanism [13]. However, in conventional studies, the failure is always assumed to be temporary, preidentified, overloaded, and recoverable. For example, Lu et al. studied the robust optimization design under the risk of supply chain network interruption, integrated the emergency strategy into the supply chain network design, and established a robust mixed-integer programming model based on a multiscenario, multiperiod, and single product [14]. Gao and Chen established a cascaded failure model of supply chain networks based on the complex network theory, in which the weak link of a network was identified, and the quantitative evaluation method of supply chain vulnerability was proposed [15]. Yan et al. established the cascade effect method of supply chain network system detection and gave the dynamic node importance evaluation method, based on the theory of complex networks [16]. He and Cheng studied the repair mechanism of the supply chain based on the directed weighted complex network, simulated the damaged situation of the supply chain by adjusting the link weight of the supply chain network, allocated limited repair resources, and repaired the supply chain from the global damage and node damage [17]. Recently, Saura et al. proposed an artificial intelligence-based model to capture the dynamic feature of supply chain networks [18].

However, the cascading failure caused by COVID-19 is quite different from the previous ones. Under the influence of COVID-19, firms will fail due to the decreased product demand or raw materials supply, which belongs to underload failure. Meanwhile, it is a sudden attack on the supply chain with a subsequently long-term and continuous disruption. There is a lack of dynamic models for investigating supply chain disruption under COVID-19. First of all, we identify the following:

(1) In the traditionally cascading failure models, the upper bound of node capacity is considered (i.e., node failure occurs when the node load exceeds the upper bound of capacity). At the same time, the firms under the influence of COVID-19 are faced with the problem of decreasing product demand or insufficient raw materials supply, whose node failure occurs when the node load is lower than the capacity lower bound.

(2) In the traditionally cascading failure models, when the node load exceeds the upper bound of capacity, the firm cannot consume many supply orders or demand orders. It will consider giving these orders to other firms with a close cooperation relationship. However, the load of firms in the supply chain network caused by the COVID-19 will be lower than the capacity lower bound, and the load of other firms with close cooperation will be reduced.

We adopt Wang's underload cascading failure model in supply chain networks [19]. More specifically, we adapted the model from three aspects: firm underload failure, firm capacity lower bound, and load redistribution according to the business relationship between firms. On the other side, how does COVID-19 attack the supply chain? How can we improve the supply chain to mitigate the disruption caused by COVID-19? To answer these questions, we adopt two attack strategies: random attack and target attack to test the model's performance. By numerical simulating, we find that it is crucial to maintain close cooperation among supply chain firms to survive the epidemic. Besides, target protection and random protection can improve the survivability of supply chain networks.

Moreover, random protection is more effective than target protection. It is more effective to protect small- and medium-sized firms. Our results implicate that enhancing cooperation relationships within a supply chain network is an effective method to fight against COVID-19.

The rest parts of the paper are organized as follows. In Section 2, we have a brief systematic review of cascading failure related to supply chain networks. Furthermore, in Sections 3 and 4, we present the adapted fundamental model of underload cascading failure and the cascading failure process due to the COVID-19 pandemic. Several numerical simulations are conducted in Section 5. Finally, we draw several conclusions and implications specific to the situation of COVID-19 in Sections 6 and 7.

2. The Cascading Failure Phenomena in Supply Chain Network

Each firm has its capacity, inventory, and demand for raw materials in the supply chain network. When the output of a firm is lower than the minimum production requirements of its operation due to the lack of orders or raw materials, considering the cost problem and profit purpose, the firm may consider suspending production or closing in severe cases. Furthermore, other firms with business contacts with this firm will also be affected, resulting in their output reduction. This process is repeated in the supply chain network, spreading failures and subsequent failures. The successive failures in the supply chain network mean that the failures of one or a few firms will cause the failures of other firms through the supply-demand connection between firms and eventually lead to the failures of a considerable part of the failures of the whole supply chain [20–23].

The recent outbreak of the COVID-19 has had a specific impact on the supply of raw materials for firms [24]. This negative effect may be due to shortage or interruption of logistics, especially raw materials or components from high incidence areas of the COVID-19 and even high transportation areas. Taking the automobile supply chain as an example, it is a complex and huge system including various suppliers, manufacturers, logistics providers, distributors, whose general structure is shown in Figure 1, where each node represents a supplier, manufacturer, logistics provider, distributor, and its main body is the material supply system such as parts and components. There are more than 100,000

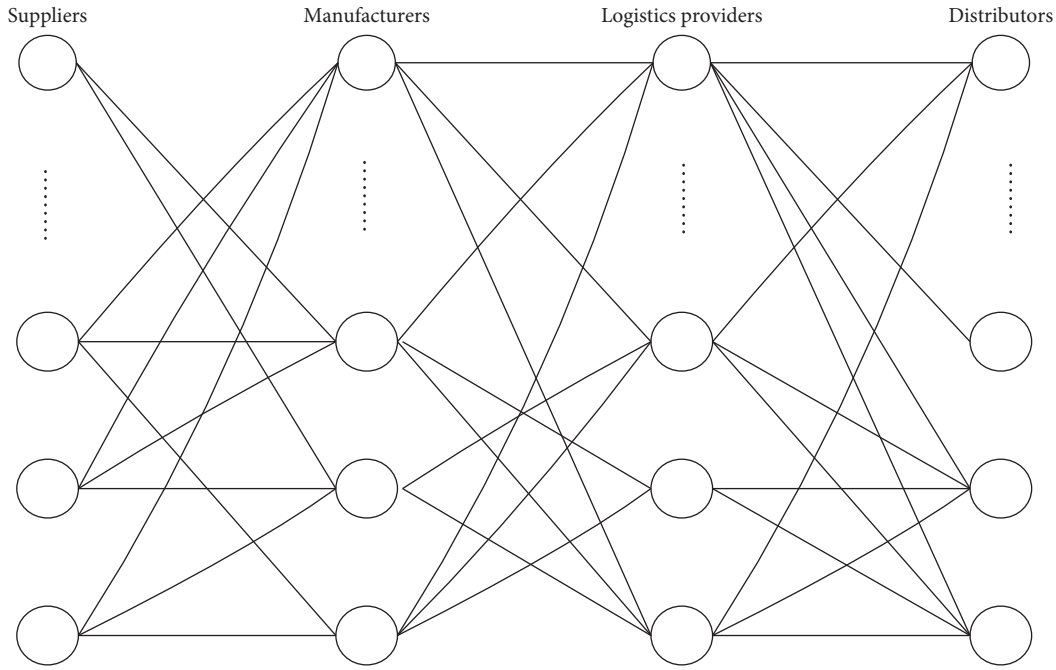


Figure 1: A general structure of the automobile supply chain.

Chinese auto parts firms and more than 13,000 firms above the designated size [24]. More than 10,000 foreign-invested auto parts firms have already established production and *R* & *D* bases in China. China's essential auto parts manufacturing base is the most seriously affected by the COVID-19, Hubei Province. More than 1,300 auto parts firms are above the designated size, including Bosch, Delphi, Valeo, Lear, Honeywell, and many other multinational auto parts firms. Automobile production involves tens of thousands of parts, thousands to tens of thousands of employees, and global materials procurement. Any abnormality in the supply of automobile parts will affect the production; the suspension of some or some upstream suppliers will lead to the "chain break" crisis in the supply chain and cascade failure [25]. At present, automobile firms are gradually resuming production, but the comprehensive recovery of production capacity needs the efficient operation and synchronous cooperation of the whole supply chain system. As the parts and components firms in Hubei lag the whole country, it has directly affected many domestic mainstream automobile firms and even many international automobile firms, resulting in the closure and suspension of production of some factories or models.

The COVID-19 also has an impact on procurement management. The procurement supply chain is a network system, and the procurement bases are often distributed throughout the country and even around the world. However, due to the prevention degree of COVID-19 and the difference of COVID-19 situation in different regions, the procurement management of firms is faced with significant uncertainty and adjustment risk, which significantly reduces the efficiency of firms. Taking China's auto parts as an example, China is the world's most significant auto production and sales country and one of the most crucial auto parts manufacturing and supply bases in the world. However, many auto parts, materials, and equipment still need to be imported from abroad: Germany, Japan, South Korea, the United States, etc. At present, most of these countries are high-risk countries. If suppliers in the above countries stop supplying due to the COVID-19, many of the auto parts they supply cannot be replaced. Even if they can be replaced, it will take time. At present, many parts companies and most auto companies in Europe have announced to stop production or partially stop production. Besides, core components and semiconductor products produced by Japan, South Korea, and the United States are also widely used in the automobile industry. Most of the vehicle firms have a large stock of imported spare parts. If the COVID-19 situation worsens, the spare parts firms and material firms in these countries and regions stop production and supply; the domestic firms will be unable to continue production after the stock is used up. The production of Chinese automobiles and spare parts will face a direct impact, which directly affects the regular operation of the whole automotive supply chain network.

The cascading failure literature has many studies. Zhang and Liu studied the vulnerability measurement of the logistics service supply chain, analyzed the interaction law of internal nodes of the logistics service supply chain, and determined the network vulnerability measurement index [12]. Liu et al. constructed a conceptual model for supply chain vulnerability and discussed the impact mechanism of various factors on supply chain vulnerability [26]. Yu et al. studied the complexity and vulnerability of the supply chain network structure under the modern production mode and proposed an analysis method based on the weighted improved node contraction method [27]. Zeng and Zhao proposed a dynamic network load entropy method and

discussed the vulnerability of the cluster supply chain network in the process of cascading failure propagation [28]. Du et al. concluded that moral risk, market risk, and decision-making risk would greatly increase the supply chain's vulnerability, based on the complexity of supply chain integration, operation, and environment [29]. Liu analyzed the supply chain's vulnerability from the perspectives of supply, demand, and environment, established the vulnerability analysis model of the supply chain, and made a statistical analysis of the vulnerability [30]. Tang et al. studied the cascading failure mechanism and the associated supply chain network [31]. Li et al. established a network model with a uniform degree and random distribution and a cascaded failure and invulnerability model, according to the characteristics of the logistics support network [32, 33]. Tang et al. considered the characteristics of nodes such as recovery and repeated failure, constructed a cascaded failure model under the probability propagation mode of fault nodes, and comprehensively analyzed the performance trade-off between edge robustness and node robustness [34]. Tang et al. analyzed the cascading failure characteristics and the robustness of an interdependent supply chain network, which is a typically interdependent network composed of an undirected cyber-layer network and a directed physical-layer network [35]. Geng et al. studied the self-organizing elastic recovery characteristics of the cluster supply chain network [36]. Wang and Xiao studied the cascading failure in the cluster supply network and proposed a resilience method to cascading failures in the cluster supply chain network by leveraging the social resilience of ant colonies [37, 38]. The next sections will employ an underload cascading failure model to investigate the COVID-19 disruption in supply chain networks.

3. The Underload Cascading Failure Model

With the outbreak of the COVID-19, the cascading failure process in the supply chain network is as follows: the upstream firms' products fail due to the decline of demand, and the downstream firms fail due to the lack of material supply. Both the decrease in product demand and the shortage of material supply can be regarded as a load decrease. When the load decreases below the lower bound of capacity, it will cause failure, triggering cascading failure. Therefore, the underload failure model can better describe the level linkage mechanical behavior in the supply chain network. The following definitions are given to describe better the underload cascading failure in the supply chain network.

In the process of modeling, the supply chain network is represented as a directed graph structure in the form of $G = (V, E)$, where $V = \{v_1, v_2, \ldots, v_N\}$ is the set of nodes (representing the firms), and the node types include supplier node, manufacturer node, distributor node, and retailer node. $E = \{e_1, e_2, \ldots, e_M\}$ is the set of edges (representing the business relations between firms), if the upstream firm node v_i and the downstream firm node v_j has business relations, node v_i and node v_j are connected by a directed edge as $\langle v_i, v_j \rangle$, and the corresponding element in E is $e_{ij} = 1$; otherwise, $e_{ij} = 0$.

3.1. Initial Load. In the cascading failure model, the allocation of the initial load is based on the importance of nodes. Generally, two ways are often used to define a node's initial load as degree approach [39, 40] and the betweenness approach [41–43]. Considering that in the supply chain network, the effective operation of a firm is more and more dependent on other firms, especially its upstream and downstream firms, which are closely connected through business relations. That is to say, a load of a node is naturally closely related to that of its neighbors. So, we consider the way in [44] to define the initial node load as follows:

$$L_i^0 = \left[d_i \sum_{j \in A_i} d_j \right]^{\alpha}, \tag{1}$$

where L_i^0 is the initial load of node v_i, $d_i = d_{I_i} + d_{O_i}$ is the degree of node v_i, and d_{I_i} is the number of edges from the upstream neighbor nodes to node v_i, d_{O_i} is the number of edges from node v_i to the downstream neighbor nodes. A_i is the neighbour nodes set of node v_i, d_j is the degree of the neighbor node v_j of node v_i. α is an adjustable parameter that governs the initial load strength of nodes.

3.2. Lower Bound of Node Capacity. In the actual network, the ability of nodes to handle the load (i.e., capacity) is usually limited by cost and other factors. The goal of the supply chain network is to provide products for the end-users, and the firms in each link are all for the ultimate purpose of making profits. If the product demand or raw material supply is lower than a certain level, the firm will not operate normally and will eventually suspend production or shut down due to its inability to make profits. Therefore, the load to maintain the regular operation of the firm must be higher than a specific bound. Underload will lead to firm failure. Considering that the failure of the supply chain under the COVID-19 is caused by underload, this paper only considers the capacity lower bound of a firm, which is proportional to its initial node load L_i^0 [37], and is defined as follows:

$$C_i = \beta L_i, \quad i = 1, 2, \ldots, N, \tag{2}$$

where C_i is the capacity lower bound of node v_i and $0 \le \beta \le 1$ is the lower bound parameter.

3.3. Redistribution of Load. When the supply or demand of firms declines, the firms with a closer business relationship are most affected. In the supply chain network, a load of each node at first is more significant than the capacity lower bound, that is to say, a load of any node satisfies: $C_i \le L_i$, at this time, the network is in "steady state"; however, in the COVID-19 environment, the supply and demand of nodes in the network are slowly decreasing. When $L_i \langle C_i$, node failure occurs, that is, due to the cost and other factors, firms will choose to stop production, which shows that this node load will affect the load change of other nodes. This paper defines load redistribution according to the business relationship strength between firms [45], which is expressed as follows:

$$\Delta L_{ij} = L_i \frac{L_j}{\sum_{m\in A_i} L_m}, \quad = L_i \frac{\left[d_j \sum_{k\in A_j} d_k\right]^{\alpha}}{\sum_{m\in A_i}\left[d_m \sum_{h\in A_m} d_h\right]^{\alpha}}, \tag{3}$$

where ΔL_{ij} is the reduced load of the upstream or downstream nearby node v_j of node v_i, A_i is the neighbor nodes set of a failure node v_i.

For the node c in Figure 2, if it fails at time t, its upstream nearby nodes a and b, and downstream nearby nodes d and e will suffer losses. Taking the downstream node d as an example, its reduced load is as follows:

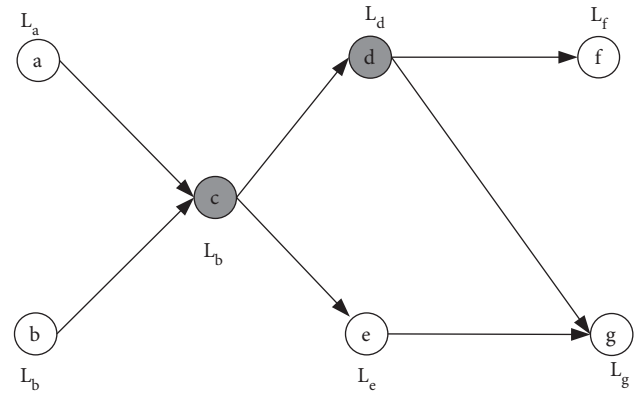


Figure 2: Illustration of the load redistribution after a node fails.

$$\begin{aligned}
\Delta L_d &= L_i \frac{L_j}{\sum_{m\in A_i} L_m} \\
&= L_c \frac{L_d}{L_a + L_b + L_d + L_e} \\
&= L_i \frac{\left[d_j \sum_{k\in A_j} d_k\right]^{\alpha}}{\sum_{m\in A_i}\left[d_m \sum_{h\in A_m} d_h\right]^{\alpha}} \\
&= L_c \frac{\left[d_d \sum_{k\in A_d} d_k\right]^{\alpha}}{\left[d_a \sum_{i\in A_a} d_i\right]^{\alpha} + \left[d_b \sum_{j\in A_b} d_j\right]^{\alpha} + \left[d_d \sum_{k\in A_d} d_k\right]^{\alpha} + \left[d_e \sum_{l\in A_e} d_l\right]^{\alpha}} \\
&= L_c \frac{\left(d_d d_c + d_d d_f + d_d d_g\right)^{\alpha}}{\left(d_a d_c\right)^{\alpha} + \left(d_b d_c\right)^{\alpha} + \left(d_d d_c + d_d d_f + d_d d_g\right)^{\alpha} + \left(d_e d_c + d_e d_g\right)^{\alpha}}.
\end{aligned} \tag{4}$$

And its load currently is $L_d(t) = L_d(t-1) - \Delta L_d$, if $L_i(t) \langle C_d$, then node d will fail at time $t+1$. For upstream nodes, the same mode of action occurs in the opposite direction. The cascading failure ends when no new nodes fail due to underload.

3.4. Evaluation Index. With the development of COVID-19, the node load often lies between the initial load and the capacity lower bound; that is, the node state in the model is divided into a normal state, underload, and failure state. In this paper, we consider the way in [46] to define the whole network efficiency based on node efficiency to measure the performance of a supply chain network, and the network efficiency (denoted by EG) is expressed as follows:

$$\mathrm{EG} = \sum_{i=1}^{N} \frac{\mathrm{eg}_i(t)}{N} = \sum_{i=1}^{N} \frac{s_i L_i(t)/L_i}{N}, \tag{5}$$

where $s_i \in [0,1]$ indicates the status of each node, of which the value 0 indicates node failure, the value 1 indicates regular node operation, and the other values indicate node efficiency reduction due to underload. In the paper, we assume that all nodes typically work initially, $s_i = 1$. $L_i(t)$ represents the load of the node v_i at time t.

4. The Cascading Failure Process

Based on the above definitions, the process of cascading failure model of supply chain network proposed is as follows:

(1) Initial condition: network G, with N nodes, M edges, the initial load L_i^0 of node $v_i (i = 1, 2, \ldots, N)$, capacity lower bound $C_i (i = 1, 2, \ldots, N)$, node attack ratio p, and adjustable parameters α and $\beta (0 \le \beta \le 1)$.

(2) Node attack mode: when a random attack or target attack is carried out on network nodes, the load of the attacked nodes will be reduced in proportion to p.

(3) Load redistribution of failure nodes: when a load of a node v_i is less than its capacity lower bound, the node fails, and its load is evenly distributed to its neighbor

nodes. The load redistribution strategy is used. That is, a load of its neighbor node is as follows:

$$L_j(t+1) = L_j(t) - \Delta L_{ij}, e_{ij} = \langle i, j \rangle. \quad (6)$$

(4) For all adjacent nodes $v_j\,(j = 1, 2, \ldots, N)$ of node v_i, test their load L_j: if $L_j \langle C_j$, it causes cascading failure of nodes, and then $v_j \longrightarrow v_i$, repeat (3), otherwise keep.

(5) Simulation termination condition: for all nodes $v_k\,(k = 1, 2, \ldots, N)$ of the network, if $L_k \geq C_k$ the simulation is terminated.

5. Simulation and Analysis

Because most supply chain data are confidential or proprietary, building an entire supply chain network is very difficult. From the perspective of the complex network, the research on the supply chain network is mostly based on a typical network model, in which random network (Erdos Renyi, ER) [47] and scale-free network (Barabasi Albert, BA) are primary [48]. In this paper, the ER and BA networks are used to simulate the supply chain network to study its cascading failure dynamics under the COVID-19. Specifically, python 3.6 is used to code the ER and BA networks and the proposed model [49]. The network consists of 6,000 edges and 2,000 nodes. These nodes are divided into four types: supplier, manufacturer, distributor, and retailer. The supplier node is only associated with the manufacturer node, the manufacturer node is associated with the supplier and distributor node, and the distributor node is associated with the manufacturer and retailer node. The retailer is only associated with the distributor. Each experiment runs 20 times independently in the experiment, and the average value is taken as the simulation result [50].

5.1. Effect of Attack Strategies on Cascading Failures. Two attack strategies are generally adopted to trigger the cascading failure in the network: random attack and target attack. Random attack refers to the random selection of a certain proportion of network nodes for an attack. Target attack refers to selecting a certain proportion of important nodes in the network for an attack. In this paper, the random attack strategy adopted is to randomly select a certain proportion of nodes to reduce their load, and the target attack strategy is to select a certain proportion of nodes to reduce their load according to the descending order of node load.

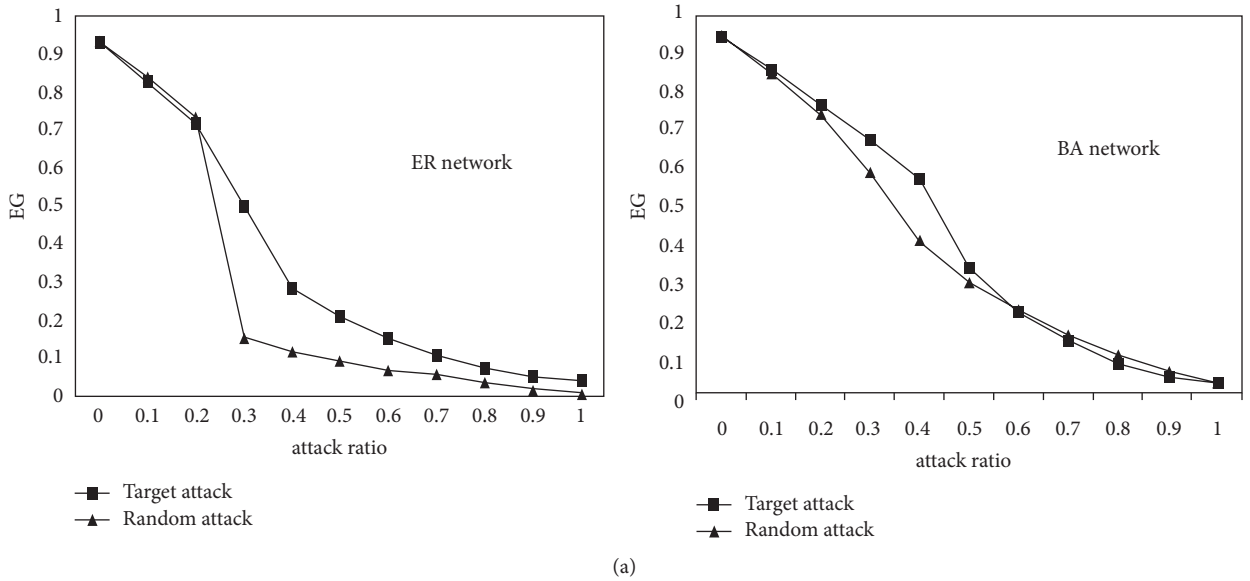
Figure 3 shows the changes in network efficiency when the ER network and BA network face random attacks and target attacks under four kinds of parameter values. We can see from Figure 3 that EG is descending with the increase of attack ratio both in the ER network and in the BA network. Taking Figure 3(c) as an example, when the attack ratio is 20%, the EG of the ER network under random attack and target attack is 0.2825 and 0.0407, while that of the BA network under random attack and target attack is, respectively, 0.5127 and 0.7994. When the attack ratio is 50%, the EG of the ER network under random attack and target attack is 0.0658 and 0.0322, while that of the BA network under random attack and target attack is, respectively, 0.1865 and 0.4998. The results of Figure 3 show that the cascading failures of the ER network and the BA network are more easily triggered by random attacks than by target attacks. Similar conclusions can also be found under other parameter combinations.

5.2. Effect of Model Parameters on Cascading Failures. The random attack is more likely to trigger cascading failure than the target attack known from Figure 4, so the following experiments are all carried out under the random attacks.

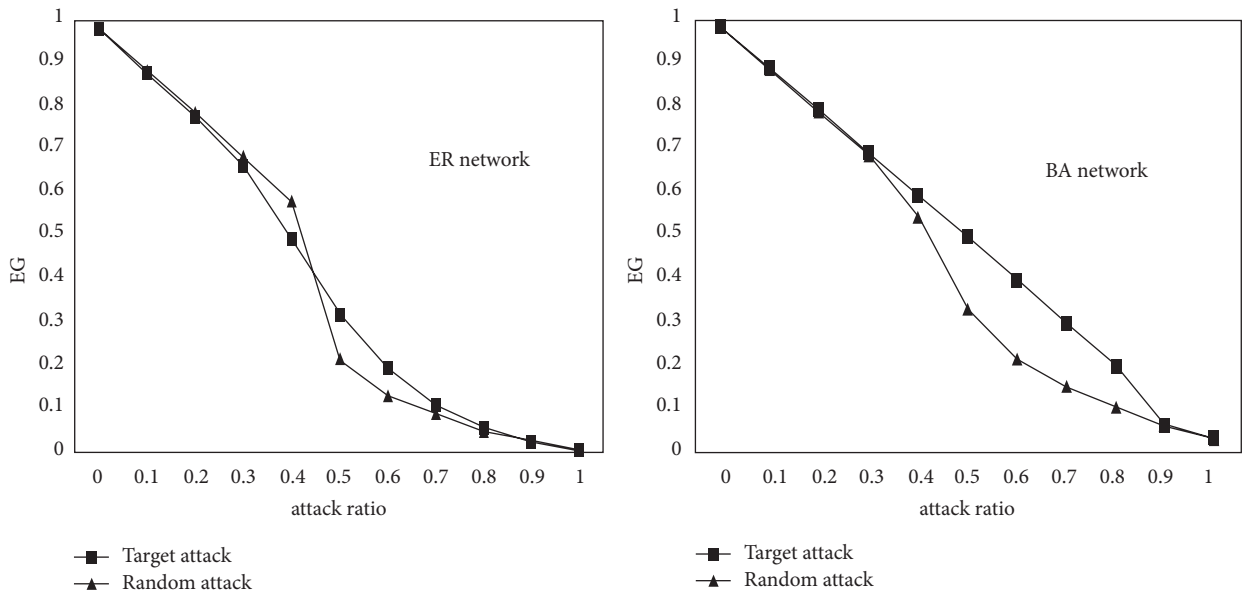
5.2.1. Effect of α on Cascading Failures. In the experiment, the other model parameters are $\beta = 0.5$, $p = 0.4$. Figure 4 shows the EG changes of two network models during cascading failures with different α, where α is used to control the load intensity of nodes. We can see from Figure 4 that EG is ascending with the increase of α both in the ER network and in the BA network. This indicates that the initial load of nodes defined according to the degree of nodes and the degree of neighbor nodes makes the network robust to cascading failure. When $\alpha = 0.8$ and $\alpha = 0.9$, the EG of the ER network changes dramatically after being attacked. It is shown that the robustness of the ER network against cascading failure is relatively weak under the condition of these two kinds of parameters. Furthermore, BA network also has similar performance.

5.2.2. Effect of β on Cascading Failures. In the experiment, the other model parameters are $\alpha = 0.9$, $p = 0.4$. Figure 5 shows the EG changes of two network models during cascading failures with different β, where β is used to control the capacity low bound of nodes. We can see from Figure 5 that EG is descending with the increase of β both in the ER network and in the BA network. It indicates that the higher β value is set (i.e., the minimum production demand of the firm), the more likely underload failure will occur, which is consistent with the actual situation. Furthermore, the lower the β value, the greater the firm's risk tolerance, and the longer the firm can support during the COVID-19. When a node fails, it will cause a loss to its neighbor nodes. If a load of neighbor nodes is close to the capacity lower bound, the loss may cause the further failure of neighbor nodes. The neighbour node can make up for the loss by strengthening the business relationship with other nodes with the same function as the failed node to reduce the possibility of its failure.

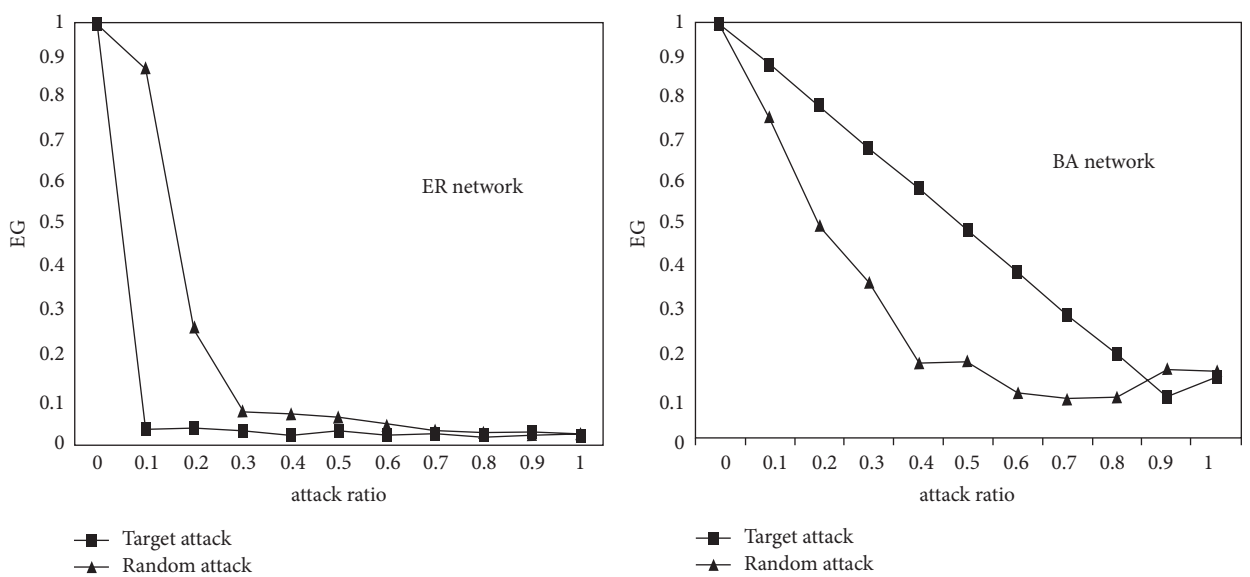
5.2.3. Effect of p on Cascading Failures. In the experiment, the other model parameters are $\alpha = 0.9$, $\beta = 0.6$. Figure 6 shows the EG changes of two network models during cascading failures with different p, where p is used to control the load reduction ratio of the attacked node. We can see from Figure 6 that EG is ascending with the increase of p both in the ER network and in the BA network, and the change of the ER network is more obvious than that of the BA network.



(a)



(b)



(c)

Figure 3: Continued.

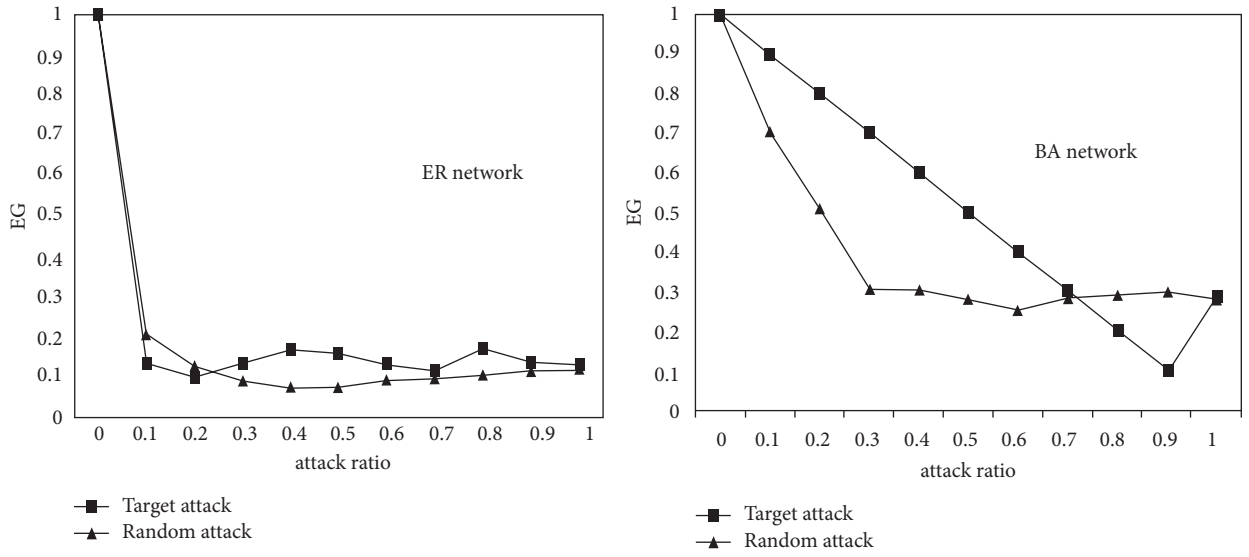


Figure 3: Network efficiency under different attack strategies. (a)$\alpha = 0.5$, $\beta = 0.6$, $p = 0.5$. (b)$\alpha = 0.8$, $\beta = 0.5$, $p = 0.4$. (c)$\alpha = 1.2$, $\beta = 0.7$, $p = 0.5$. (d)$\alpha = 1.5$, $\beta = 0.8$, $p = 0.6$.

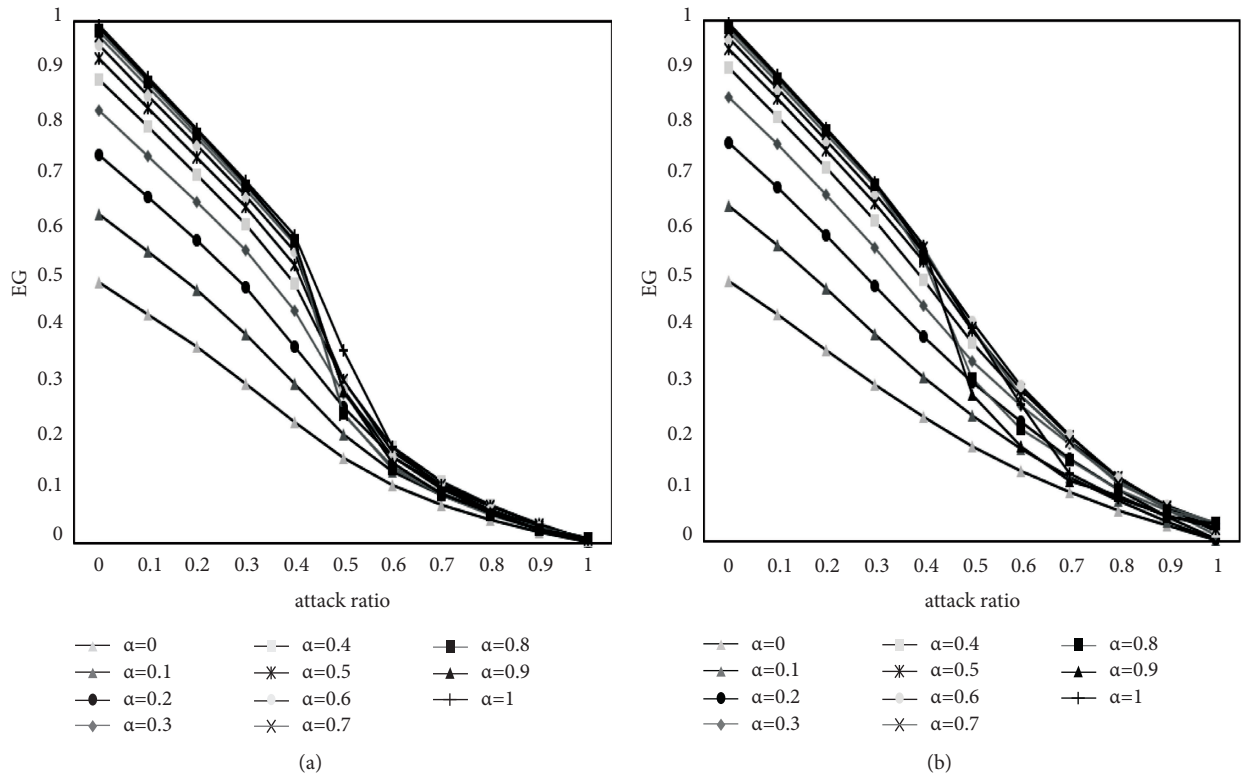


Figure 4: Network efficiency under different α. (a) ER. (b) BA.

When $p = 0.4$, that is to say, the attack node's load is reduced to $1 - p = 0.6 \leq \beta$ of the initial load, then there is no cascading failure, and the network efficiency is the largest at this time. When $p = 0.5$, the load of the attacked node is reduced to 50% of the initial load, i.e., $1 - p = 0.5 \leq \beta 1 - p = 0.5 \leq \beta$; node failure occurs and affects the neighbor nodes, triggering cascading failure. Currently, network efficiency is the minimum. It shows that the closer the load reduction is to the capacity lower bound, the more likely the nodes are to have cascading failures. At this time, if certain supporting measures are taken for such a firm, the ability of the firm to resist risks can be enhanced.

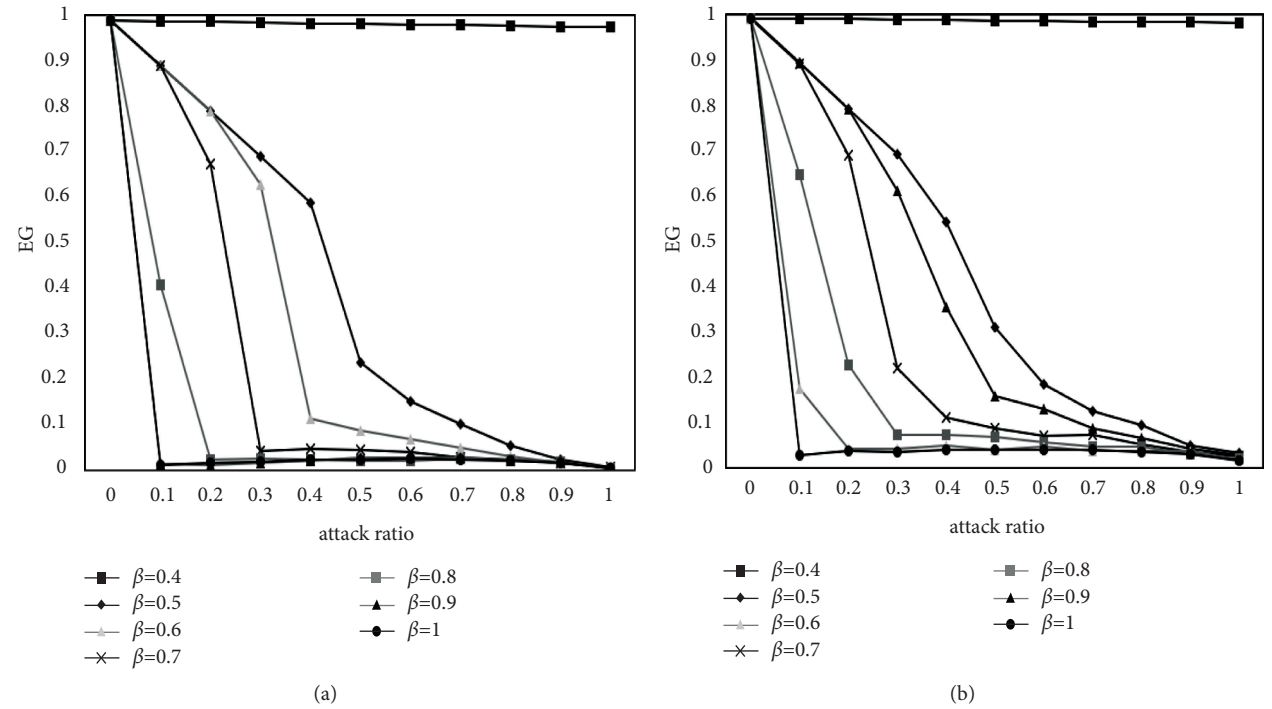


Figure 5: Network efficiency under different β. (a) ER. (b) BA.

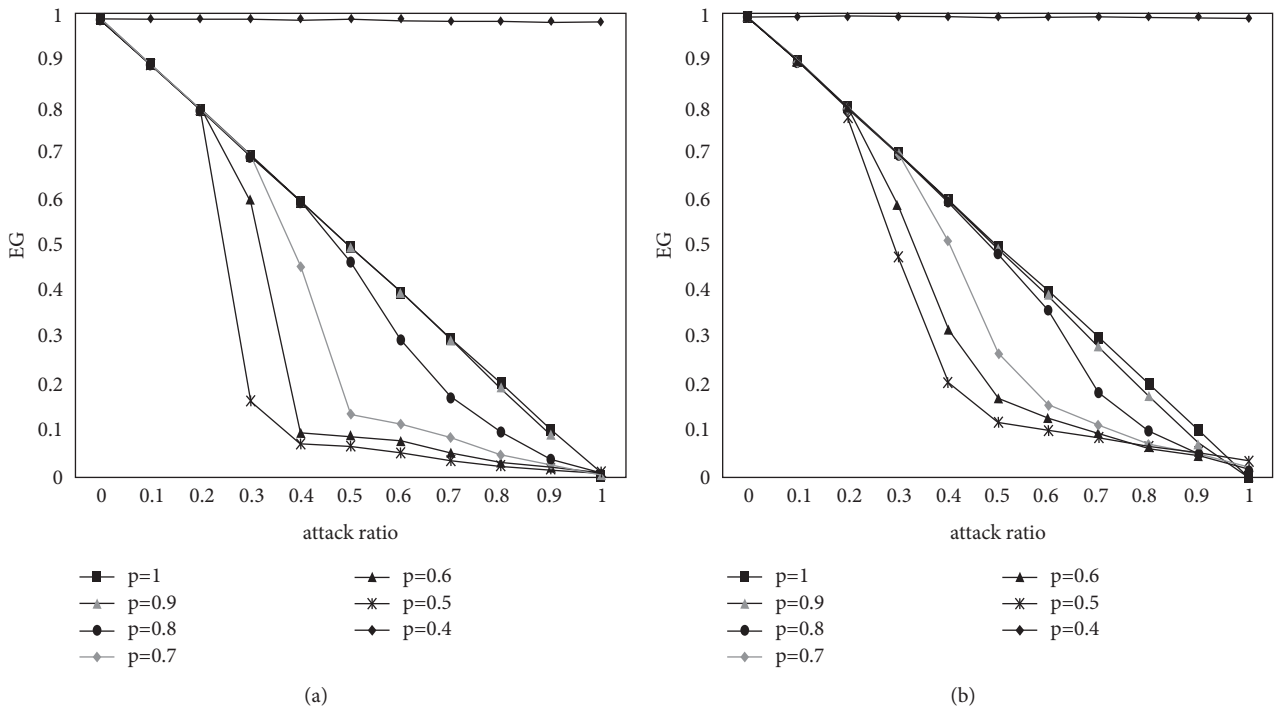


Figure 6: Network efficiency under different p. (a) ER (b) BA.

5.3. Effect of Protection Measures on Cascading Failures. We study the effect of protection measures (such as loans to enterprises and rent reduction) on cascading failure control of the network under the target attacks. Two protection measures are used in the experiment: target protection and random protection. Target protection refers to selecting a certain proportion of important nodes in the network for protection, while random protection refers to randomly selecting a certain proportion of network nodes for protection. The model parameters are set to $\alpha = 0.8$, $\beta = 0.4$, $p = 0.3$, and the proportion of nodes to be protected is 10% of the proportion of attack nodes, that is to say, select 10% of the

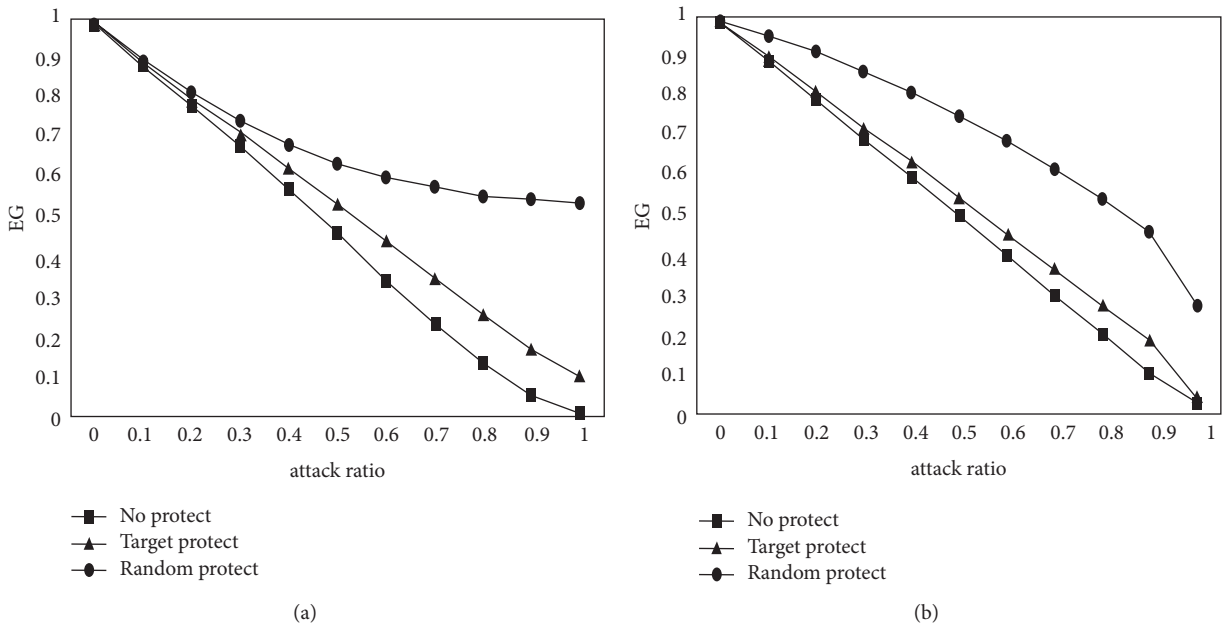


Figure 7: Network efficiency under different protection measures. (a) ER. (b) BA.

proportion of attack nodes according to the descending order of node load for target protection, or randomly select 10% of the proportion of attack nodes for random protection. Figure 7 shows the EG changes of two network models during cascading failures with different protection measures. We can see from Figure 7 that two kinds of protection measures can also improve the survivability of the ER network and BA network, and random protection measures are more effective than target protection measures. It shows that the protective measures for small- and medium-sized firms are more effective. Considering the complexity and diversity of business relationships among firms in the real supply chain network, it is difficult to measure it accurately. This conclusion can provide some reference for supply chain risk management during the COVID-19 period.

6. Conclusions and Discussion

In the present study, we propose an underload cascading failure model to investigate the negative outcomes of demand and supply declines caused by the ongoing COVID-19 pandemic. The extant supply chain network models are insufficient to explore the complicated and diverse characteristics among firms in the supply chain network. Our model simulates the actual situations after the outbreak of COVID-19 by redistributing the loads among connected nodes. We try to capture the characteristics of the connections among the upstream and downstream firms. The numerical simulations present that the network efficiency is positively related to the firm loads, while the network efficiency is negatively related to the lower bound of production capacity. The simulation results show that the loads and lower bounds' resilience will enhance the supply chain's robustness when preventing the cascading failures during COVID-19.

In addition, our model emphasises underload failure: upstream firms fail due to the decline of demand, and downstream firms fail due to the decrease of supply. Based on the underload failure, we analyze the effect of model parameters on the cascading failure of the supply chain network. The simulation results show that the adjustable parameters α and β, respectively, represent the firm's strength and the ability to resist risks, which will affect the spread of cascading failures. During the COVID-19, some protection measures for some firms will positively enhance the invulnerability of the entire supply chain network.

Finally, random attack strategies simulate the impact of failure caused by COVID-19 on the entire supply chain network composed of firms of variable sizes. The simulations of attacking the core nodes with large production scales show that the protection measures improve network efficiency. Although the core nodes play an essential role in maintaining the stability of the supply chain network system, providing protective measures for small- and medium-sized enterprises has more obvious effects on preventing network failure and improving network efficiency.

7. Theoretical and Practical Implications

Our study will contribute to the cascading failure literature on supply chain network systems. COVID-19 poses challenges to the extant theories of supply chain management. The negative outcomes of the COVID-19 are unpredictable and uncontrollable. We adapt the underload cascading failure mode to the actual situations of COVID-19, which have never happened before. The proposed model will be valuable for future studies. In addition, our study also has three practical implications. First, for large enterprises, increasing redundant partners in the supply chain network will help to reduce the mutual trust costs for temporary

cooperation purposes. Production capacity exchange with redundant cooperators will help reduce the failure rate of enterprise nodes. In addition, enterprises can appropriately increase the redundancy of raw materials and products, such as increasing inventory and backup products. They can also deal with the changes in demand or supply to maintain the regular production progress.

Second, our study implicates that we must pay attention to the production coordination of various small- and medium-sized enterprises in the supply chain. The problems of difficult and expensive financing of small and medium-sized enterprises have been intensively exposed under the impact of the COVID-19. From the perspective of policy regulation, policy innovation and system design play an essential role in improving the quality and efficiency of the supply system. During the epidemic period, the government took measures such as economic support and tax reduction for a certain number of small- and medium-sized enterprises, which helped the supply chain network coordinate the production cycle and capital flow cycle as a whole and ensure network efficiency.

Finally, our study also implicates that we need to actively adopt new technologies to reduce the cost and risk of supply chain management. The new technology can also involve virus detection and personnel health tracking. Even if there is a sporadic epidemic, it can respond and deal with it quickly to provide better safety protection and psychological protection for enterprise production, contributing to the resumption of production and solving the problem of labor shortages during the epidemic. In the future, we will further study the specific protection measures to enhance the invulnerability of the supply chain network.

Data Availability

No data were used to support this study.

Conflicts of Interest

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

Acknowledgments

This study was supported by the Social Science Planning Foundation of Zhejiang Province, China (Grant no. 21NDJC080YB), Zhejiang Public Welfare Research Project (Grant no. LGF20G010004), General Scientific Research Project of Zhejiang Department of Education China (Grant no. 1300KZ0419912), and Zhejiang Natural Science Foundation Project (Grant no. LY19F020007).

References

[1] D. Guan, D. Wang, S. Hallegatte, S. Joseph Davis, and au fnm, "Global supply-chain effects of COVID-19 control measures," *Nature Human Behaviour*, vol. 4, no. 6, pp. 577–587, 2020.

[2] P. Chowdhury, S. Paul, S. Kaisar, and M. D. Abdul Moktadir, "COVID-19 pandemic related supply chain studies: a systematic review," *Transportation Research Part E*, vol. 148, Article ID 102271, 2021.

[3] A. Kumar, S. K. Mangla, P. Kumar, and M. Song, "Mitigate risks in perishable food supply chains: learning from COVID-19," *Technological Forecasting and Social Change*, vol. 166, Article ID 120643, 2021.

[4] R. Van Hoek, "Research opportunities for a more resilient post-COVID-19 supply chain – closing the gap between research findings and industry practice," *International Journal of Operations & Production Management*, vol. 40, no. 4, pp. 341–355, 2020.

[5] W. Klibi, A. Martel, and A. Guitouni, "The design of robust value-creating supply chain networks: a critical review," *European Journal of Operational Research*, vol. 203, pp. 283–293, 2010.

[6] L. Tang, K. Jing, J. He, and H. E. Stanley, "Robustness of assembly supply chain networks by considering risk propagation and cascading failure," *Physica A*, vol. 459, pp. 129–139, 2016.

[7] A. Nagurney, "Optimisation of supply chain networks with inclusion of labor: applications to COVID-19 pandemic disruptions," *International Journal of Production Economics*, vol. 235, Article ID 108080, 2021.

[8] December, 2021 https://think.ing.com/articles/the-rise-and-rise-of-global-shipping-costs.

[9] December, 2021 https://www.uscourts.gov/news/2021/11/08/bankruptcy-filings-continue-fall-sharply.

[10] December, 2021 https://www.nippon.com/en/japan-data/h01109/.

[11] J. Wang, "Robustness of complex networks with the local protection strategy against cascading failures," *Safety Science*, vol. 53, pp. 219–225, 2013.

[12] G. S. Zhang and W. Liu, "Vulnerability measurement research of complex network of logistics service supply chain," *Computer Engineering and Applications*, vol. 53, no. 19, pp. 224–230, 2017.

[13] J. Ke, X. Du, L. Shen, and L. Tang, "Robustness of complex networks: cascading failure mechanism by considering the characteristics of time delay and recovery strategy," *Physica A*, vol. 534, Article ID 122061, 2019.

[14] M. F. Lu, W. J. Chen, and C. J. Liang, "Emergency supply chain network robust optimisation for mitigating disruption risk," *Journal of Hefei University of Technology*, vol. 10, no. 38, pp. 1417–1423, 2015.

[15] J. Gao and Y. Y. Chen, "Vulnerability analysis of supply chain network in cascading failure," *Logistics Engineering and Management*, vol. 38, no. 10, pp. 80–83, 2016.

[16] Y. Yan, X. Liu, and X. T. Zhuang, "Cascading failure model and method of supply chain based on complex network," *Journal of Shanghai Jiaotong University*, vol. 44, no. 3, pp. 322–331, 2010.

[17] H. M. He and Y. L. Cheng, "Research on supply chain repair mechanism based on directed weighted complex network," *Journal of Railway Science and Engineering*, vol. 13, no. 11, pp. 2299–2304, 2016.

[18] J. R. Saura, D. Ribeiro-Soriano, and D. Palacios-Marques, "Setting B2B digital marketing in artificial intelligence-based CRMs: a review and directions for future research," *Industrial Marketing Management*, vol. 98, no. 10, pp. 161–178, 2021.

[19] Y. C. Wang and F. P. Zhang, "Modeling and analysis of underload-based cascading failures in supply chain networks," *Nonlinear Dynamics*, vol. 92, pp. 1403–1417, 2018.

[20] L. Geng and R. B. Xaio, "Resilience measure of supply network based on DIIM," *Computer Integrated Manufacturing Systems*, vol. 20, no. 5, pp. 1211–1219, 2014.

[21] W. Pan, J. Dong, K. Liu, and J. Wang, "Topology and topic-aware service clustering," *International Journal of Web Services Research*, vol. 15, no. 3, pp. 18–37, 2018.

[22] R. Z. Qiu, Y. Z. Wang, and X. Y. Huang, "Robust supply chain network design based on uncertain disruption probability," *Computer Integrated Manufacturing Systems*, vol. 22, no. 10, pp. 2458–2468, 2016.

[23] F. P. Fernandes and F. W. S Lima, "Persistence in the zero-temperature dynamics of the q-states Potts model on undirected Barabasi-Albert networks and Erdos-Renyi random graphs," *International Journal of Modern Physics C*, vol. 19, no. 12, pp. 1777–1785, 2008.

[24] H. Song, "The implication of the novel coronavirus outbreak to supply chain flexibility management," *China business and market*, vol. 34, no. 3, pp. 11–16, 2020.

[25] A. Zhu, W. Chen, J. Zhang, X. Zong, W. Zhao, and Y. Xie, "Investor immunisation to Ponzi scheme diffusion in social networks and financial risk analysis," *International Journal of Modern Physics B*, vol. 33, no. 11, 2019.

[26] J. G. Liu, Y. X. Zhou, B. Lu, and J. L. Zhao, "The reduction mechanism of supply chain vulnerability based on supply chain disruption risk," *Systems Engineering Theory & Practice*, no. 3, pp. 556–566, 2015.

[27] K. P. Yu, Y. Yang, N. Liu, F. Li, and J. Z. Xie, "Vulnerability of complicated supply chain network based on weighted improved nodes contraction method," *Computer Integrated Manufacturing Systems*, vol. 20, no. 4, pp. 963–970, 2014.

[28] Y. Zeng and R. Xiao, "Modelling of cluster supply network with cascading failure spread and its vulnerability analysis," *International Journal of Production Research*, vol. 52, no. 23, pp. 6938–6953, 2014.

[29] Z. P. Du, G. Y. Hu, and Y. S. Liu, "Research on supply chain vulnerability based on complexity," *China Business and Market*, vol. 25, no. 6, pp. 49–53, 2011.

[30] F. Liu, *Research of the Impact of Supply Chain Risk Management on Supply Chain Vulnerability*, Zhejiang University, Hangzhou, China, 2010.

[31] L. Tang, J. He, and K. Jing, "Cascading failure mechanism and robustness of interdependent supply chain networks," *Journal of Management Sciences in China*, vol. 19, no. 11, pp. 33–44, 2016.

[32] Y. Li, J. Wu, and Y. J. Tan, "Critical invulnerability study for cascading failure of tactical logistics networks," *Journal of System Simulation*, vol. 24, no. 5, pp. 1030–1034, 2012.

[33] Y. Li, J. Wu, and Y. J. Tan, "Invulnerability study for cascading failure of the logistics support networks of capacity evenly distributed," *Journal of Systems Engineering*, vol. 25, no. 6, pp. 853–860, 2010.

[34] L. Tang, P. Jiao, J. K. Li, K. Jing, and Z. H. Jin, "Cascading failure mechanism and robustness of complex networks with recovery strategy," *Control and Decision*, vol. 10, no. 33, pp. 1841–1850, 2018.

[35] L. Tang, K. Jing, J. He, and H. E. Stanley, "Complex interdependent supply chain networks: cascading failure and robustness," *Physica A*, vol. 443, pp. 58–69, 2016.

[36] L. Geng, R. Xiao, and S. Xie, "Research on self-organisation in resilient recovery of cluster supply chains," *Discrete Dynamics in Nature and Society*, vol. 2013, Article ID 758967, 2013.

[37] Y. C. Wang and R. B. Xiao, "An ant-colony based resilience approach to cascading failures in cluster supply network," *Physica A*, vol. 462, pp. 150–166, 2016.

[38] W. Pan, B. Song, K. Li, and K. Zhang, "Identifying key classes in object-oriented software using generalised k-core decomposition," *Future Generation Computer Systems*, vol. 81, pp. 188–202, 2018.

[39] D. L. Duan, "Cascading failure of complex networks based on load local preferential redistribution rule," *Complex Systems and Complexity Science*, vol. 12, no. 1, pp. 33–39, 2015.

[40] D. L. Duan, X. D. Ling, X. Y. Wu, D. H. OuYang, and B. Zhong, "Critical thresholds for scale-free networks against cascading failures," *Physica A*, vol. 416, pp. 252–258, 2014.

[41] A. Motter and Y. Lai, "Cascade-based attacks on complex networks," *Physical Review E*, vol. 66, Article ID 065102, 2002.

[42] A. Motter, "Cascade control and defense in complex networks," *Physical Review Letters*, vol. 93, no. 9, Article ID 098701, 2004.

[43] H. Li, T. Wang, W. Pan et al., "Mining key classes in Java projects by examining a very small number of classes: a complex network-based approach," *IEEE Access*, vol. 9, Article ID 28076, 2021.

[44] J. Wang and L. Rong, "A model for cascading failures in scale-free networks with a breakdown probability," *Physica A*, vol. 388, pp. 1289–1298, 2009.

[45] R. R. Yin, B. Liu, H. R. Liu, and Y. Q. Li, "Research on invulnerability of the random scale-free network against cascading failure," *Physica A*, vol. 444, pp. 458–465, 2016.

[46] Z. Li, Y. H. Guo, G. A. Xu, and Z. M. Hu, "Analysis of cascading dynamics in complex networks with an emergency recovery mechanism," *Acta Physica Sinica*, vol. 63, no. 15, p. 158901, 2014.

[47] P. Erdos and A. Rnyi, "On the evolution of random graphs," *Publ. Math. Inst. Hung. Acad. Sci*, vol. 5, pp. 17–61, 1960.

[48] A. Barabsi and R. Albert, "Emergence of scaling in random network," *Science*, vol. 286, pp. 509–512, 1999.

[49] J. R. Saura, D. Ribeiro-Soriano, and D. Palacio-Marques, "Using data mining techniques to explore security issues in smart living environments in Twitter," *Computer Communications*, vol. 179, no. 11, pp. 285–295, 2021.

[50] W. Pan, H. Ming, C. K. Chang, Z. Yang, and D.-K. Kim, "ElementRank: Ranking Java software classes and packages using multilayer complex network-based approach," *IEEE Transactions on Software Engineering*, vol. 47, no. 10, pp. 2272–2295, 2019.

Hindawi
Mathematical Problems in Engineering
Volume 2021, Article ID 3135702, 11 pages
https://doi.org/10.1155/2021/3135702

Research Article

An Empirical Study of Software Metrics Diversity for Cross-Project Defect Prediction

Yiwen Zhong,[1] Kun Song,[1] ShengKai Lv,[1] and Peng He[1,2,3]

[1]*School of Computer Science and Information Engineering, Hubei University, Wuhan 430062, China*
[2]*Hubei Key Laboratory of Applied Mathematics, Hubei University, Wuhan 430062, China*
[3]*Hubei Engineering Research Center for Educational Informationization, Wuhan 430062, China*

Correspondence should be addressed to Peng He; penghe@hubu.edu.cn

Received 4 September 2021; Revised 24 October 2021; Accepted 5 November 2021; Published 28 November 2021

Academic Editor: Chunlai Chai

Cross-project defect prediction (CPDP) is a mainstream method estimating the most defect-prone components of software with limited historical data. Several studies investigate how software metrics are used and how modeling techniques influence prediction performance. However, the software's metrics diversity impact on the predictor remains unclear. Thus, this paper aims to assess the impact of various metric sets on CPDP and investigate the feasibility of CPDP with hybrid metrics. Based on four software metrics types, we investigate the impact of various metric sets on CPDP in terms of *F*-measure and statistical methods. Then, we validate the dominant performance of CPDP with hybrid metrics. Finally, we further verify the CPDP-OSS feasibility built with three types of metrics (orient-object, semantic, and structural metrics) and challenge them against two current models. The experimental results suggest that the impact of different metric sets on the performance of CPDP is significantly distinct, with semantic and structural metrics performing better. Additionally, trials indicate that it is helpful for CPDP to increase the software's metrics diversity appropriately, as the CPDP-OSS improvement is up to 53.8%. Finally, compared with two baseline methods, TCA+ and TDSelector, the optimized CPDP model is viable in practice, and the improvement rate is up to 50.6% and 25.7%, respectively.

1. Introduction

In software engineering, the conventional defect prediction approach trains a predictor using historical data of the target project and then uses it to predict defects in the subsequent version or release. This process is named as within-project defect prediction (WPDP). However, the cold-start problem is fatal for some new projects or inactive WPDP software projects. Hence, cross-project defect prediction (CPDP) overcomes this issue and has attracted much attention in recent years. In general, CPDP refers to predicting defects in a project using a predictor trained on historical data of other projects [1–3].

As illustrated in Figure 1, various software metrics such as static code, process, object-oriented, and network metrics have been employed for defect prediction. Several studies have also confirmed the discrepancy in the performance of WPDP with different metric sets [4, 5]. For example, Radjenovic et al. [4] highlight that object-oriented and process metrics perform better among six categories of software metrics.

As an artifact, the software can also be abstracted into a coarse-grained network structure based on the dependencies between classes, namely, a class dependency network (CDN) scheme. In CDN, each class is considered a node, and the dependencies between classes are the directed edges. From the perspective of complex networks, some researchers have verified that network metrics are better than code metrics in defect prediction [6, 7]. Additionally, deep learning for network analysis and specifically network embedding learning of a graph structure has attracted significant attention. This strategy aims to find a mapping function to transform each node into a low-dimensional representation. Such techniques involve deepwalk [8], node2vec [9], and

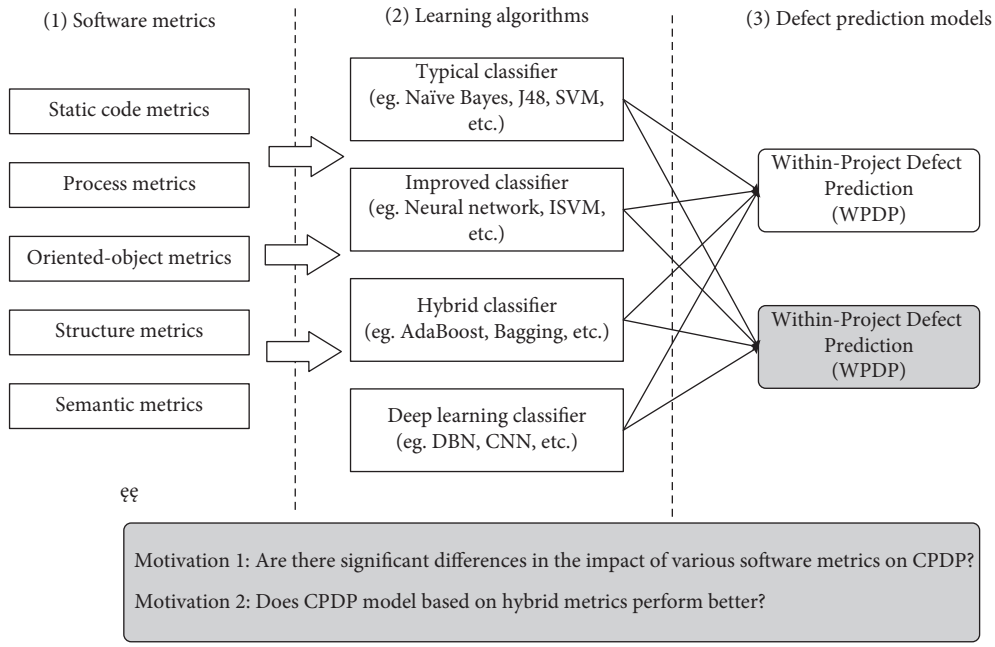


Figure 1: A simple review of defect prediction.

struc2vec [10]. Qu et al. [11] automatically learn the node representation from the CDN using network embedding for software defect prediction and achieve quite appealing results.

Recently, some researchers highlight that in addition to the features represented by a series of handcraft metrics, software programs have well-defined syntax, represented by abstracts syntax trees (ASTs), and rich semantic information hidden deep in the source code. The work of [12–14] has already demonstrated that programs' semantic information helps characterize defects and improve defect prediction. Specifically, the semantic representation automatically learns from source code, exploiting a deep learning model that distinguishes code regions of different semantics.

Although some studies have investigated the possible benefits of including some measures such as static code and process metrics, none systematically assesses the impact of using various sets of metrics on defect prediction, especially CPDP. The information presented by different software metric sets commonly exhibits significant differences, especially in the context of cross-project. In other words, whether different software metric sets have a significantly distinct effect on the CPDP performance is still an open problem.

This paper focuses on comparative analysis and assesses the impact of using different metric sets to mitigate such problems. Furthermore, it also explores the optimal combination of various metrics in CPDP. The main contributions of this paper are summarized as follows:

(1) We conduct a series of experiments and verify that the impact of different types of software metric sets on the performance of CPDP is significantly distinct, with the semantic metric being the most appealing, followed by the structural metric

(2) We find that the predictor built with the combination OSS (CK-OO, semantic, and structural metrics) performs best and achieves better performance than several state-of-the-art methods in terms of the *F*-measure metric

The remainder of this paper is organized as follows. Section 2 reviews the related work in CPDP. Sections 3 and 4 describe the approach of our empirical study and the detailed experimental setups, respectively, while Sections 5 and 6 analyze and discuss the experimental results. Section 7 presents some threats to validity that may affect our study. Finally, Section 8 concludes this paper and presents the directions for future work.

2. Related Work

2.1. Cross-Project Defect Prediction. In recent years, the topic of CPDP has attracted considerable attention from both academia and industry. The most fundamental issues are how to pick the appropriate source projects for a target project and how to train a more accurate predictor through various strategies.

Turhan et al. [2] first utilize the nearest-neighbor filtering technique to prune irrelevant cross-project data, while Porto et al. [15] propose an instance filtering method by selecting the most similar instances from the training data set. Ryu et al. [16] suggest a method of hybrid instance selection using the nearest neighbor (HISNN). The results highlight that instances with solid local knowledge can be identified utilizing nearest-neighbors with the same class label.

To improve the performance of CPDP, Ni et al. [17] develop the FeSCH method and design three ranking strategies to choose the appropriate features. He et al. [18] study CPDP from the perspective of feature simplification

and compare the performance between CPDP and WPDP. Li et al. [19] compare some well-known data filters and propose an HSBF (hierarchical select-based filter) method. Li et al. [20] analyze the impact of selection granularity of the training data on CPDP and propose a multigranularity selection strategy.

Additionally, Zhang et al. [21] provide an unsupervised approach entitled MT+ to determine the most suitable source project for each target project by considering the impact of various data transformations on the CPDP model. Kumar et al. [22] built a transfer learning scheme for CPDP by utilizing machine learning and identifying the best training data combination. Ryu et al. [23] develop a transfer cost-sensitive boosting method by considering distributional characteristics and the data imbalance for CPDP. They also [24] propose a multiobjective Naive Bayes learning method for CPDP by considering the class imbalance contexts. Poon et al. [25] suggest a credible theory-based Naive Bayes (CNB) classifier and establish a reweighting mechanism for CPDP between the source and target projects.

Besides, to address the heterogeneous defect data sets, He et al. [26] introduce a CPDP-IFS approach based on the distribution characteristics of both the source and target projects. Nam et al. [27] suggest an improved method, HDP, where the metric selection and matching build a defect predictor. Jing et al. [28] propose a unified metric representation for heterogeneous defect data named UMR. Yu et al. [29] present a feature matching and transfer (FMT) approach. Muddu et al. [30] tested the robustness of CPDP experimental research.

Considering CPDP, Herbold et al. [31] replicate 24 approaches proposed between 2008 and 2015 and evaluate their performance on five data sets. The authors claim that CPDP's model performance is sufficient for practical applications. Goel et al. [32] summarize independent variables, modeling techniques, performance evaluation criteria, and different approaches in building CPDP models but lack a more in-depth impact analysis.

With the extensive application of deep learning technology in various fields, its powerful feature generation ability has also been used for defect prediction [11–14]. For example, Wang et al. [13] generate the source code ASTs and automatically learn the program's hidden semantic and syntax features through a deep belief network. Li et al. [14] extract the structural information from ASTs through CNN and combine the semantic features with standard code features to improve the performance of software defect prediction. However, ASTs encapsulate only the abstract syntax structure of the source code, which cannot represent the program's execution process. Phan et al. [33] propose transforming the source code into program control flow graphs (CFG) to extract deeper semantic features from the code. Qu et al. [11] leverage a network embedding technique to automatically learn to encode the program's class dependency network structure into low-dimensional vector spaces to improve software defect prediction.

2.2. Software Metric. Software quality improvement through defect prediction has been relying on a wide variety of software metrics treated as features. To comprehend the relationship between diverse software metric sets for defect prediction, Chamoli et al. [34] analyze the performance of prediction models based on various software metrics and conclude that software metrics may indeed affect the models' defect prediction accuracy.

Madeyski et al. [35] identify that process metrics are worth collecting and improve the metric-based prediction models when data sets are collected from a wide range of software projects. Han et al. [36] combine code and process metrics as features and confirm that the predictive capabilities of using two features (*BD_max* and *Pre-defects*) are comparable to the results of using all 61 features. Öztürk et al. [37] suggest that quality metrics are superior in predicting imbalanced data sets than static code metrics. Xia et al. [38] search for the most critical software metrics and conclude that fewer than 10 metrics can better perform than utilizing 22 or more metrics.

Bluemke et al. [39] describe the process of choosing appropriate metrics for defect prediction. Accordingly, Jiarpakdee et al. [40] suggest that researchers should be aware of redundant metrics before constructing a defect prediction model to maximize their studies' internal validity. Caglayan et al. [41] conclude that the performance of different metric sets in building a defect prediction model depends on the project's characteristics and the targeted prediction level.

Mauša et al. [42] replicate the case study of deriving thresholds for software metrics using a statistical model based on logistic regression and analyze a more comprehensive set of software metrics. The results reveal that the threshold values of some metrics can be used to predict defect-prone modules effectively. Recently, Zhang et al. [43] suggest that an aggregation scheme can significantly alter correlations among metrics and correlations between metrics and the defect count through an analysis of 11 aggregation schemes using data collected from 255 open-source projects.

3. Problem and Method

3.1. Research Question. This paper defines CPDP as follows: given a source project P_s and a target project P_t, CPDP aims to achieve the target prediction in P_t using the knowledge extracted from P_s, where $P_s \neq P_s$. Let the source and target projects share the same feature cardinality and metric sets. The goal of CPDP is to learn a model from the selected source projects (training data) and apply it to the target project (test data). In the context examined here, project P, as a defect data set, contains m instances represented as $P = \{I_1, I_2, \ldots, I_m\}$. An instance is $I_i = \{f_{1i}, f_{2i}, \ldots, f_{di}, L\}$, where f_{ki} is the $k-$th dimension of the representation vector of instance I_i, and d denotes the total dimensions, namely, the scale of the metrics. If instance I is buggy, then L is one; otherwise, L is zero.

As mentioned in Section 2.2, various metrics can measure software complexity and quality in practice; therefore, a defect data set may contain multiple types of software metrics. According to the statistics, most public defect data sets contain at least two types of software metrics. For example, the commonly used AEEEM data set involves two sets of software metrics. According to the existing practice [11–13], deep learning technology can provide structural and semantic metrics. In other words, there are at least four sets of metrics.

Nevertheless, only a few research works based on these data sets have explored the impact of different metric sets and their combinations on the performance of CPDP, especially for handcrafted and automatically learned metrics. Spurred by that, this paper aims to find empirical evidence addressing the following three research questions.

RQ1: Is the impact of different metric sets on the performance of CPDP significantly distinct?

RQ2: Does CPDP based on hybrid metrics perform better?

RQ3: Does the optimized CPDP model outperform the baselines?

3.2. Approach. An effective prediction model affords more resources to be devoted to the bug-prone instances and consequently improves the quality of the latter instances. Existing CPDP models usually aim to improve the learning algorithm and make the predictor perform better. Hence, they always ignore the impact of software metrics on prediction performance. To answer the above research questions, we will consider constructing a CPDP model for two scenarios (Figure 2).

Scenario 1 considers CPDP modeling utilizing a single metric set, which is the most straightforward modeling method. For this case, we will investigate the performance of the predictor. Scenario 2 considers constructing a CPDP model based on multitype metric sets. Note that different colors in Figure 2 distinguish the types of indicator sets. The details of this scenario are provided in the following steps.

Step 1: Defect data sets are constructed for each project according to the software metric types. An instance is described as $I = \{f_{11}, f_{12}, \ldots, f_{1d_1}, \ldots, f_{n1}, f_{n2}, \ldots, f_{nd_n}, L\}$, where n is the software metric category cardinality ($n \geq 1$). As mentioned in Section 3.1, d_n denotes the dimension of the n-th set of software metrics, f is the metric value, and L the label.

Step 2: After collecting the defect data sets, we further determine a series of classification algorithms employed to learn the predictor, for example, Naïve Bayes, logistic regression, and J48.

Step 3: The corresponding CPDP model based on the selected defect data sets and classification algorithms is constructed according to the specific scenario requirement.

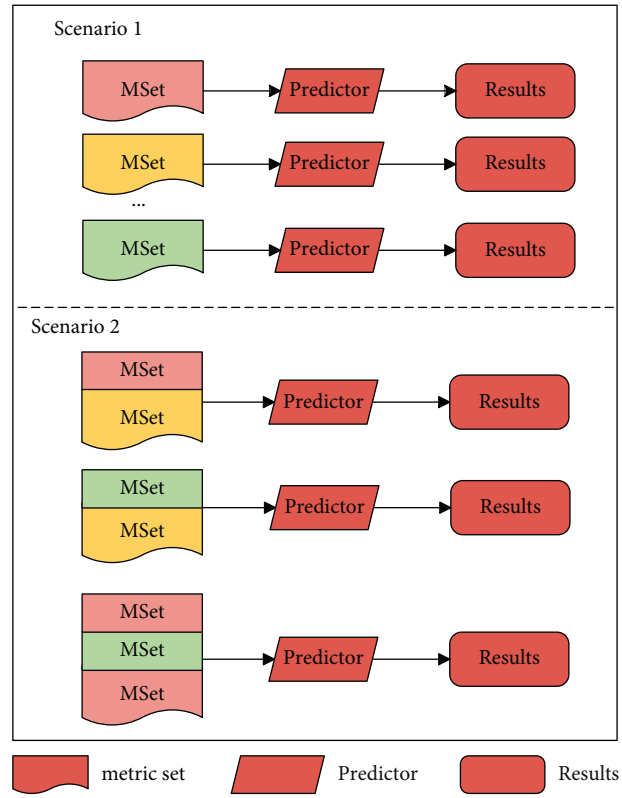


Figure 2: The two CPDP scenarios examined.

4. Experimental Setup

4.1. Data Sets. We investigate our study on the public AEEEM data set [44], which involves five open-source projects. Table 1 lists the details of the five projects, where the second and third columns are the number of the defective and the clean instances, respectively. Each project refers to process and CK-OO metrics. Each instance represents a class file and comprises software metrics and a dependent variable labeling defining if bugs exist in this class file. Table 2 presents all metric sets involved in our study.

Note that we expand the set of metrics to the existing data set, including structural and semantic metrics. The former is extracted from a class dependency network employing network embedding learning. Specifically, this paper utilizes the node2vec method [9] to map each class node to a low-dimensional vector. Regarding the semantic metrics, these adopt the method of [13]. Among the above metrics, the traditional code metrics are not listed because they have little effect on CPDP and are not applicable to the current data set.

Data imbalance is a crucial and unavoidable problem in software defect prediction. In our data set, due to the absence of defects, the number of nonbuggy samples is far more than the number of defective samples, with the imbalanced distribution seriously affecting the prediction accuracy. To overcome this problem, we balance the data sets with SMOTE. Additionally, since the scale of the numerical metric values is different in a data set, we normalize the metric values within the range of [0, 1] utilizing the z-score method.

Table 1: AEEEM data set.

Project	Attribute			
	Number of buggy instances	Number of nonbuggy instances	The total number of instances	Buggy ratio (%)
Eclipse	206	792	998	20.64
Equinox	129	196	325	39.69
Lucene	64	627	691	9.26
Mylyn	245	1,617	1,862	13.16
Pde	209	1,288	1,497	13.96

Table 2: Four types of software metrics will be involved in this paper.

Category	Metric description		
Process metrics (15-dimension)	numberOfVersionsUntil	avgLinesAddedUntil	maxLinesAddedUntil
	numberOfFixesUntil	linesRemovedUntil	maxCodeChurnUntil
	numberOfRefactoringsUntil	maxLinesRemovedUntil	avgCodeChurnUntil
	numberOfAuthorsUntil	avgLinesRemovedUntil	ageWithRespectTo
	linesAddedUntil	codeChurnUntil	weightedAgeWithRespectTo
CK-OO metrics (17-dimension)	Coupling between object classes	Response for a Class	numberOfMethodsInherited
	Depth of Inheritance Tree	Weighed Methods per Class	numberOfPrivateAttributes
	Afferent Couplings	numberOfAttributes	numberOfPrivateMethods
	Efferent Couplings	numberOfAttributesInherited	numberOfPublicAttributes
	Lack of Cohesion in Methods	numberOfLinesOfCode	numberOfPublicMethods
	Number of Children	numberOfMethods	
Structural metrics	A d-dimensional space of features ($d = 32$)		
Semantic metrics	A d-dimensional space of features ($d = 32$)		

4.2. Experimental Design. This section describes the entire experimental framework according to the previous three research questions, as illustrated in Figure 3.

First, to conduct an impact analysis among all four metric sets, the CPDP experiments are conducted in the first scenario. This trial analyzes the differences of the software metric sets under a specific classifier. Then, we expand the experiments in the following scenario and compare the average prediction results of six cases involving different combination patterns. Finally, based on the optimal metric combination, we further verify the feasibility of the proposed CPDP model by challenging it against two current models.

Once this process is completed, the answers to the three research questions of our study will be discussed.

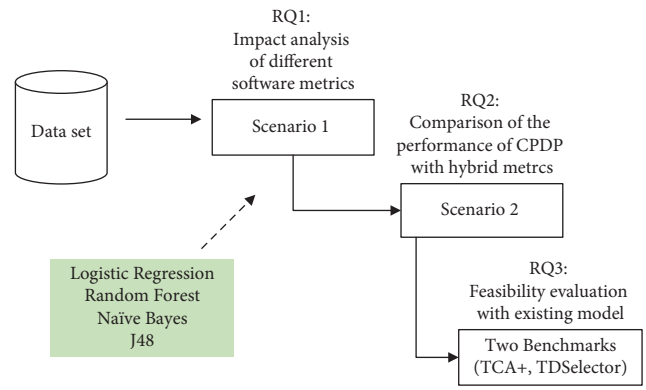


Figure 3: The experimental framework according to the three research questions.

4.3. Classifiers. Machine learning algorithms are widely used in defect prediction, with the classification algorithms being able to classify the defective modules correctly. This paper utilizes four typical classification algorithms as the primary learning algorithms.

(i) Logistic regression (LR): A widely used supervised classification algorithm that essentially solves a dichotomy problem. Due to its universality and practicability, several methods employ it for defect prediction [5, 18, 26].

(ii) Random forest (RF): A classifier that uses multiple trees to train and predict samples, aiming at reducing variance. RF has a better generalization and classification capability than typical decision trees [18].

(iii) Naïve Bayes (NB): The simplest classifier based on Bayesian theory and independent hypothetical testing. It is widely accepted that NB outperforms other classifiers and thus is frequently used to build defect prediction models [5, 18, 23].

(iv) J48: A high-efficiency decision tree algorithm that uses the greedy technique for supervised classification, posing an appealing tool for defect prediction [18].

4.4. Evaluation Measures. To predict whether an instance (class file) is defective, we use binary classification technology. The possible results are true positive (TP), false positive (FP), false negative (FN), and true negative (TN). The conventional classification evaluation measures include

precision, recall, and F-measure expressed as follows. Given the contradiction between precision and recall, we use F-measure to evaluate the prediction performance.

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}},$$
$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}, \quad (1)$$
$$F - \text{measure} = \frac{2 * \text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}}.$$

Additionally, statistical tests assist in understanding whether a statistically significant difference between two results exists. This work utilizes the Wilcoxon signed-rank test to check whether the performance difference between the prediction models with different software metrics is significant. To further examine the effectiveness and following the work of [13, 18, 26], we employ Cliff's delta (δ) to measure the effect size of our approach. Cliff's delta is a nonparametric effect size measurement scheme that quantifies the difference between the two approaches. Table 3 describes the meanings of various δ values [45].

Table 3: The mappings between different δ values and their effective levels.

Effect size	$\lvert\delta\rvert$
Very small	0.008
Small	0.147
Medium	0.33
Large	0.474
Very larger	0.622
Huge	0.811

5. Experimental Results

This section reports the experimental results aiming at answering the three research questions formulated in Section 3.1.

> RQ1: Is the impact of different metric sets on the performance of CPDP significantly distinct?

This trial considers Scenario 1, with the corresponding results presented in Figure 4, and highlights that the F-measure values obtained using different metric sets are generally different, implying different levels of influence on CPDP. On the one hand, the prediction results of CPDP models based on different metric sets vary when the classifier is consistent. For instance, in Figure 4(a) and considering the semantic and process metrics, the medium values of the F-measure are 0.381 and 0.334, respectively, inferring that the semantic metric performs much better than the process metric under the J48 classifier. Additionally, a value of 0.381 is a relatively high index showing that the semantic metrics perform better than other metrics, whose index values are 0.334, 0.337, and 0.343.

Under different classifiers, the advantages of specific metrics are also unstable. Considering the semantic metric as an example, Figure 4(a) indicates that this metric performs best (0.381), but in Figures 4(c) and 4(d), this advantage is less obvious. Note that compared with the semantic metrics, other metrics perform the same instability (approximately 0.3–0.4), sometimes leading to more outliers.

Note: A negative δ value represents that the result of the latter metric set is better; on the contrary, the former is better.

To further distinguish the impact of different metric sets on CPDP, we evaluate the results in terms of the Wilcoxon signed-rank test (p value) and Cliff's delta (δ) metric. In this study, we statistically analyze four types of metrics based on the null hypothesis, that is, two metric sets have the distribution of the same results. In Table 4, the Wilcoxon signed-rank test highlights no significant difference between the semantic and structural metrics and between process and CK-OO metrics, as both p values exceed 0.05. However, the differences between the two groups are statistically significant, especially between the CK-OO and semantic metrics (p value = 0.003).

In Table 4, the effect size δ between the structural and semantic metrics is small, and the metric is minimal between the CK-OO and process metrics. Considering the process and CK-OO metrics, the dominant effect size of the semantic metrics tends to be larger due to the negative δ values (−0.445 and −0.408), while the dominant effect size of structural metrics tends to be medium ($|\delta| = 0.295$). Therefore, overall, the semantic metric performs best followed by the structural metric and then the CK-OO and process metrics.

In conclusion, based on the experimental results, the impact of different metric sets on the performance of CPDP is distinct, with a significant difference.

> RQ2: Does CPDP based on hybrid metrics perform better?

To answer this research question, we construct a defect predictor using the logistic regression as described in Scenario 2. To simplify the presentation, we label the CPDP model with the Process and Semantic metric as CPDP-PS, OS (CK-OO and Semantic), SS (Structural and Semantic), OSS (CK-OO, Structural, and Semantic), and POSS (Process, CK-OO, Structural, and Semantic). Table 5 presents the prediction results of each target project in terms of F-measure values. The results indicate that CPDP-OSS performs best due to the greater F-measure values. For example, for Eclipse, the F-measure value of CPDP-OSS is 0.536. Compared to the remaining combinations, the performance is higher by 29.22%, 19.73%, 4.93%, and 9.58%, respectively. Additionally, the improvement increases to 62.62% compared to the previous CPDP involving single semantic metrics.

Additionally, for Equinox, the performance of CPDP-OSS exceeds 0.6 when using Eclipse, Lucene, and Pde as the source project. Note that compared with CPDP-PS and CPDP-OS, the performance improvement of CPDP-OSS for Mylyn is more prominent, exceeding 30%. Interestingly, for Lucene and Pde, their F-measure values of CPDP with single semantic metrics are more significant than that of CPDP

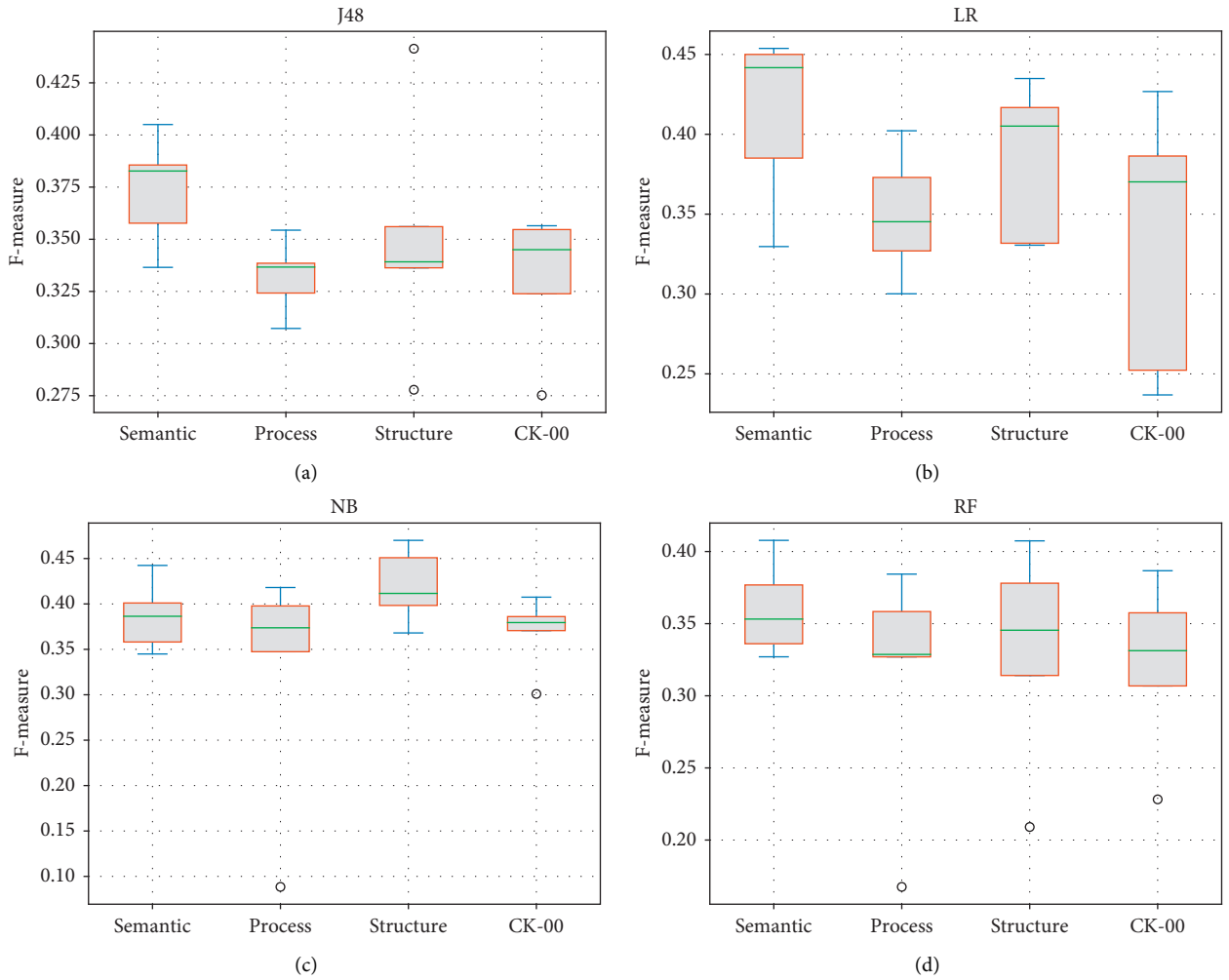


Figure 4: The standardized boxplots of the performance of CPDP are achieved by different predictors based on different types of software metrics and classifiers in Scenario 1. From the bottom to the top of a standardized box plot: minimum, first quartile, median, third quartile, and maximum. Any data not included between the whiskers are plotted as a small circle. (a) J48, (b) LR, (c) NB, and (d) RF.

Table 4: The results of the Wilcoxon signed-rank test and Cliff's effect size (δ).

	Process vs. semantic	Structure vs. semantic	CK-OO vs. semantic	Structure vs. process	CK-OO vs. process	CK-OO vs. semantic
Sig. p value (0.05)	**0.022**	0.49	**0.003**	**0.024**	0.5	**0.018**
δ value	−0.445	−0.083	−0.408	0.295	0.00	−0.295

Table 5: Comparison of the CPDP performance in terms of F-measure.

Metrics	Target				
	Eclipse	Equinox	Lucene	Mylyn	Pde
Semantic	0.330 (+62.62%)	0.385 (+55.93%)	0.450 (+0.22%)	0.442 (+12.06%)	0.454 (+4.01%)
PS	0.415 (+29.22%)	0.528 (+13.63%)	0.382 (+18.06%)	0.379 (+30.61%)	0.380 (+24.21%)
OS	0.448 (+19.73%)	0.566 (+6.17%)	0.408 (+10.54%)	0.370 (+33.78%)	0.413 (+14.29%)
SS	0.511 (+4.93%)	0.584 (+2.83%)	0.451 (—)	0.479 (+3.34%)	0.453 (+4.19%)
OSS	**0.536**	**0.601**	**0.451**	**0.495**	**0.472**
POSS	0.489 (+9.58%)	0.583 (+3.07%)	0.424 (+6.37%)	0.472 (+4.87%)	0.436 (+8.26%)

Note: The percentages within the parentheses are the improvement rates of CPDP-OSS, and the boldfaced numbers represent the best results in the same context.

with hybrid metrics, except for that of CPDP-OSS. Besides, the performance of CPDP-SS is very close to that of CPDP-OSS, even on Lucene, as the *F*-measure values are the same. Therefore, it can be concluded that sometimes "more is not better."

Overall, the results indicate that CPDP with CK-OO, structural, and semantic metrics can identify more buggy instances than the other combinations examined. Therefore, it is mandatory to consider the effect of hybrid metrics on CPDP.

RQ3: Does the optimized CPDP model outperform the baselines?

The previous results validate that it is still valuable to consider hybrid metrics during CPDP modeling. To evaluate the practicability and usefulness of CPDP-OSS, we built CPDP models using two existing approaches, that is, TCA+ [46] and TDSelector [47], and perform experiments on the same data set. Table 6 presents the comparative results between our approach and the two baselines, where the max *F*-measure value per row is in bold. CPDP-OSS outperforms both baselines, as most boldfaced *F*-measure values, that is, 12 out of 20, and the average improvement rates of *F*-measure values belong to CPDP-OSS.

Compared with TCA+, Table 6 highlights that 6 out of 20 improved rates of our approach exceed 20%, while the maximum is 50.6%. Regarding TDSelector, four cases pose an improvement exceeding 20%, and the maximum is 25.7%. With this evidence, the proposed CPDP-OSS approach is validated to be beneficial for improving the performance of a CPDP model.

6. Discussion

RQ1: Our experimental results suggest that the impact of various metrics sets on the performance of CPDP is distinct in terms of *F*-measure. Our findings indicate that the semantic metrics, on average, yield the best CPDP models, with structural metrics to follow. Meanwhile, according to Table 2, it is evident that the scale of these two metrics sets in our defect data sets is more significant, that is, a more extensive set of metrics may lead to better prediction. Thus, assisted by deep learning technology, deeper information can be automatically learned from the program.

The CK-OO metric and the process metrics are the most frequently used for defect prediction, with the authors in [4] arguing that the CK-OO metric has a good explanation and predictive power. Nevertheless, in our trials, both do not perform so well as expected. A possible explanation is a difference in the prediction context, leading to the conclusion that cross-project defect prediction is different from the traditional within-project defect prediction.

RQ2: For CPDP, the effectiveness of increasing the metrics diversity is proven, broadly consistent with some prior studies' findings. Considering software metrics, D'Ambros et al. [43] investigate the prediction based on a single set of metrics and found that defect prediction models based on a single set of metrics are unstable. Hall et al. [5] also found that defect prediction models using a comprehensive combination of metrics perform well.

According to the experimental results, overall, using semantic and structural indicators affords a good prediction capability. One possible explanation is that when using AST to extract semantic information from code, the complexity of the source code has been achieved to a certain extent. Therefore, when the CK-OO metric is continuously considered, the improvement is limited.

RQ3: In the proposed CPDP-OSS approach, the advantage relies on the implicit diversity among software metrics. In Table 7, although the results show an overall improvement in the predictive performance of CPDP-OSS, the advantage is not apparent due to the p values exceeding 0.05 and the small effectiveness levels.

Several factors may prohibit revealing the apparent advantage of the proposed approach. Additionally, due to the limitation of the data set, only three/four types of software metrics are introduced, and we employ the most basic semantic and structural metrics learning model. Currently, some improved deep learning models have been used to solve this task.

To ease the complexity of the proposed approach, instead of using the complex and representative boosting and bagging algorithms, we utilize a simple logistic regression. Therefore, there is much room for improving our approach, and we believe that the advantages will be more evident after some improvement.

Although the advantages of the CPDP-OSS approach are not particularly obvious, it is more efficient than the two baseline competitor methods. Regarding TDSelector, for each project, it requires manually calculating 76 metrics. However, after applying the classification on the semantic and structural features, we can avoid cumbersome calculations through machine learning and then reduce the metrics calculation to $17 + 15+2 = 34$, which significantly improves data processing efficiency. From this point of view, our experimental results still have great application value.

These two baselines are currently the most in-depth and representative baselines in our experimental research. There may be better baselines for comparison, and in future work, we will continue to follow up and compare them.

7. Threats to Validity

From this work, several meaningful results are obtained, but potential threats to the validity of our work remain.

Threats to construct validity primarily regard the software metrics used in this paper. The experimental data set employed from [42] is a public defect data set. According to the authors' statement, inevitably, some links between the bug database and the source code repositories are missing. However, these data have been applied to numerous prior studies, and therefore, we argue that our results are credible and representative.

Threats to internal validity concern any confounding factor that may affect our results. First, we adopt the commonly used SMOTE method to preprocess the defect data sets due to the imbalanced data. As far as we know,

Table 6: Comparison between the proposed approach and two baseline methods: TCA+ and TDSelector.

Source/target	TCA+	TDSelector	CPDP-OSS
Eclipse/Equinox	0.600 (+5.9%)	0.514 (+23.7%)	**0.636**
Lucene/Equinox	0.620 (+4.4%)	**0.664** (−2.5%)	0.648
Mylyn/Equinox	**0.560** (−11.6%)	0.459 (+7.8%)	0.495
Pde/Equinox	0.600 (+4.0%)	0.610 (+2.3%)	**0.624**
Equinox/Eclipse	**0.540** (−4.5%)	0.519 (−0.7%)	0.516
Lucene/Eclipse	0.560 (−4.5%)	**0.594** (−9.9%)	0.535
Pde/Eclipse	0.430 (+27.7%)	0.455 (+20.7%)	**0.549**
Mylyn/Eclipse	0.480 (+13.5%)	0.510 (+5.9%)	**0.545**
Equinox/Lucene	0.270 (+50.6%)	0.337 (+20.7%)	**0.407**
Eclipse/Lucene	0.310 (+26.9%)	0.313 (+25.7%)	**0.393**
Mylyn/Lucene	0.250 (+32.0%)	0.298 (+10.7%)	**0.330**
Pde/Lucene	**0.330** (−4.9%)	0.314 (+0.0%)	0.314
Equinox/Mylyn	0.230 (+24.7%)	0.264 (+8.6%)	**0.287**
Eclipse/Mylyn	**0.360** (−13.1%)	0.303 (+3.2%)	0.313
Lucene/Mylyn	0.290 (+6.4%)	0.291 (+6.1%)	**0.309**
Pde/Mylyn	0.290 (−6.1%)	**0.300** (−9.2%)	0.272
Equinox/Pde	0.330 (+21.2%)	0.367 (+9.0%)	**0.400**
Eclipse/Pde	0.380 (+2.8%)	0.391 (+0.0%)	**0.391**
Lucene/Pde	**0.370** (−11.8%)	0.365 (−10.6%)	0.326
Mylyn/Pde	0.370 (+0.4%)	0.371 (+0.0%)	**0.371**
Average	0.410 (+6.0%)	0.412 (+5.1%)	**0.433**

Table 7: The results of the Wilcoxon signed-rank test and Cliff's effect size (δ).

	CPDP-OSS vs. TCA+	CPDP-OSS vs. TDSelector
Sig. p value	0.189	0.096
δ value	0.129	0.134

SMOTE-based oversampling techniques were widely adopted as the selection to handle the class imbalance problem in software defect prediction [48–51]. Although many improved sampling techniques have been proposed, it is reasonable to believe that it is feasible to use SMOTE-based oversampling technology in this paper.

Second, any feature selection method is not introduced during the CPDP modeling, and third, a simple connection method is directly used to generate the hybrid metrics in RQ2. Undoubtedly a complex fusion mechanism will result in better performance and greater calculation time. Fourth, we train the predictors for each classifier based on the default parameter settings configured by the Weka API. Hence, we are indeed aware that the results of our study would change if we use different settings of the above four factors.

Threats to statistical conclusion validity concern the relationship between the treatment and the outcome. In addition to the intuitive comparison of the prediction results in terms of F-measure, this paper also utilizes the Wilcoxon signed-rank test and Cliff's delta effect size to compare the results. According to the significance criteria and effectiveness levels, the results indicate that the difference of various software metrics is distinct, and the introduction of diversity among software metrics is valuable. However, the advantage of our method is not noticeable compared to the two baseline methods, indicated by $|\delta|$ that is approximately 0.12.

Threats to external validity emphasize the generalization of the findings. Predictions in this paper are constructed on five open-source software systems. Although our experiments can be repeated with more open-source projects and developed with different software metrics, the empirical results for industrial software projects may differ from our main conclusions. We minimize this threat by selecting a data set that consists of parts of Eclipse, an open-source system with a solid industrial background.

8. Conclusions

This paper reports a comparative study of software metrics selection for CPDP, aiming to maximize the CPDP model's diversity in terms of metric sets. Four types of software metrics are considered for modeling, and a series of experiments are conducted on five open-source projects. The study consists of (1) the impact analysis of different metric sets on CPDP, (2) exploration of the metrics' combination, and (3) comparison between CPDP built with hybrid metric sets (CPDP-OSS) and two current state-of-the-art approaches.

The results indicate that the impact of different metric sets on CPDP is significantly distinct. Additionally, our trials indicate it is helpful for CPDP to increase the diversity of software metrics appropriately, and there are significant improvements between CPDP-OSS and the remaining combinations examined. The most significant improvement rate is up to 53.8%. Our results also highlight that CPDP-OSS outperforms two benchmarks, and the most considerable improvement rate is up to 50.6% and 25.7%,

respectively. Therefore, it is meaningful to introduce the diversity of metric sets to improve the performance of CPDP.

Future work shall mainly focus on collecting more open-source projects to validate our approach's generalization and improve the learning techniques of code semantic and structural information to provide a more effective CPDP model for defect prediction. The results of the p values and the small Cliff's delta values in the experiment show that compared with the two baselines, the effect of CPDP-OSS is not very significant. We will make improvements in future work and continue to experiment and test.

Data Availability

We investigate our study on the public AEEEM data set [44], which involves five open-source projects. Each project refers to process and CK-OO metrics. Each instance represents a class file and comprises software metrics and a dependent variable labeling defining if bugs exist in this class file.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

The authors greatly appreciate Dr. Jaechang Nam and Dr. Sinno Jialin Pan, the authors of reference [46], for providing us with the TCA source program and friendly teaching us how to use it. This work was supported by the National Key R&D Program of China (No. 2018YFB1003801), the National Natural Science Foundation of China (Nos. 61832014 and 61902114), the Science and Technology Innovation Program of Hubei Province under Grant (Nos. 2018ACA133 and 2019ACA144), and the Open Foundation of Hubei Key Laboratory of Applied Mathematics (No. HBAM201901).

References

[1] Z. He, F. Shu, Y. Yang, M. Li, and Q. Wang, "An investigation on the feasibility of cross-project defect prediction," *Automated Software Engineering*, vol. 19, no. 2, pp. 167–199, 2012.

[2] B. Turhan, T. Menzies, A. B. Bener, and J. Di Stefano, "On the relative value of cross-company and within-company data for defect prediction," *Empirical Software Engineering*, vol. 14, no. 5, pp. 540–578, 2009.

[3] Y. Ma, G. Luo, X. Zeng, and A. Chen, "Transfer learning for cross-company software defect prediction," *Information and Software Technology*, vol. 54, no. 3, pp. 248–256, 2012.

[4] D. Radjenović, M. Heričko, R. Torkar, and A. Živkovič, "Software fault prediction metrics: a systematic literature review," *Information and Software Technology*, vol. 55, no. 8, pp. 1397–1418, 2013.

[5] T. Hall, S. Beecham, D. Bowes, D. Gray, and S. Counsell, "A systematic literature review on fault prediction performance in software engineering," *IEEE Transactions on Software Engineering*, vol. 38, no. 6, pp. 1276–1304, 2012.

[6] Y. Qu, Q. Zheng, J. Chi et al., "Using K-core decomposition on class dependency networks to improve bug prediction model's practical performance," *IEEE Transactions on Software Engineering*, vol. 47, no. 2, pp. 348–366, 2021.

[7] G. Concas, M. Marchesi, C. Monni, M. Orrù, and R. Tonelli, "Software quality and community structure in java software networks," *International Journal of Software Engineering and Knowledge Engineering*, vol. 27, no. 7, pp. 1063–1096, 2017.

[8] B. Perozzi, R. Al-Rfou, and S. Skiena, "Deepwalk: online learning of social representations," in *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, New York, NY, USA, August 2014.

[9] A. Grover and J. Leskovec, "node2vec: scalable feature learning for networks," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, San Francisco, CA, USA, August 2016.

[10] L. F. R. Ribeiro, P. H. P. Saverese, and D. R. Figueiredo, "struc2vec: learning node representations from structural identity," in *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Halifax, NS, Canada, August 2017.

[11] Y. Qu, T. Liu, J. Chi et al., "node2defect: using network embedding to improve software defect prediction," in *Proceedings of the 2018 33rd IEEE/ACM International Conference on Automated Software Engineering (ASE)*, IEEE, Montpellier, France, September 2018.

[12] Z. Cai, L. Lu, and S. Qiu, "An abstract syntax tree encoding method for cross-project defect prediction," *IEEE Access*, vol. 7, pp. 170844–170853, 2019.

[13] S. Wang, T. Liu, J. Nam, and L. Tan, "Deep semantic feature learning for software defect prediction," *IEEE Transactions on Software Engineering*, vol. 46, no. 12, pp. 1267–1293, 2018.

[14] J. Li, P. He, J. Zhu, and M. R. Lyu, "Software defect prediction via convolutional neural network," in *Proceedings of the 2017 IEEE International Conference on Software Quality, Reliability and Security (QRS)*, IEEE, Prague, Czech Republic, July 2017.

[15] F. Porto and A. D. S. Simao, "Feature subset selection and instance filtering for cross-project defect prediction - classification and ranking," *CLEI Electronic Journal*, vol. 19, no. 3, pp. 4–17, 2016.

[16] D. Ryu, J.-I. Jang, and J. Baik, "A hybrid instance selection using nearest-neighbor for cross-project defect prediction," *Journal of Computer Science and Technology*, vol. 30, no. 5, pp. 969–980, 2015.

[17] C. Ni, W. Liu, Q. Gu, X. Chen, and D. Chen, "FeSCH: a feature selection method using clusters of hybrid-data for cross-project defect prediction," in *Proceedings of the Computer Software and Applications Conference*, pp. 51–56, IEEE, Turin, Italy, July 2017.

[18] P. He, B. Li, X. Liu, J. Chen, and Y. Ma, "An empirical study on software defect prediction with a simplified metric set," *Information and Software Technology*, vol. 59, pp. 170–190, 2015.

[19] Y. Li, Z. Huang, Y. Wang, and B. Fang, "Evaluating data filter on cross-project defect prediction: comparison and improvements," *IEEE Access*, vol. 5, no. 99, pp. 25646–25656, 2017.

[20] Y. Li, P. He, B. Li, and Y. Ma, "Multi-granularity selection of training data for cross-project defect prediction," *Journal of Chinese Computer Systems*, vol. 38, no. 9, pp. 1934–1939, 2017.

[21] F. Zhang, I. Keivanloo, and Y. Zou, "Data transformation in cross-project defect prediction," *Empirical Software Engineering*, pp. 1–33, 2017.

[22] K. Lov, R. Santanu, and B. Ranjan, "Transfer learning for cross-project change-proneness prediction in object-oriented software systems," *ACM SIGSOFT-Software Engineering Notes*, vol. 42, no. 3, pp. 1–11, 2017.

[23] D. Ryu, J. I. Jang, and J. Baik, "A transfer cost-sensitive boosting approach for cross-project defect prediction," *Software Quality Journal*, vol. 25, no. 1, pp. 1–38, 2017.

[24] D. Ryu and J. Baik, "Effective multi-objective naïve Bayes learning for cross-project defect prediction," *Applied Soft Computing*, vol. 49, pp. 1062–1077, 2016.

[25] W. N. Poon, K. E. Bennin, J. Huang, P. Phannachitta, and J. W. Keung, "Cross-project defect prediction using a credibility theory based naive Bayes classifier," in *Proceedings of the 2017 IEEE International Conference on Software Quality, Reliability and Security (QRS)*, July 2017.

[26] P. He, B. Li, and Y. Ma, "Towards cross-project defect prediction with imbalanced feature sets," 2014, https://arxiv.org/abs/1411.4228.

[27] J. Nam, W. Fu, S. Kim, T. Menzies, and L. Tan, "Heterogeneous defect prediction," in *Proceedings of the Joint Meeting of the European Software Engineering Conference and the ACM SIGSOFT Symposium on the Foundations of Software Engineering*, vol. 44, pp. 508–519, 2015.

[28] X. Jing, F. Wu, X. Dong, F. Qi, and B. Xu, "Heterogeneous cross-company defect prediction by unified metric representation and CCA-based transfer learning," in *Proceedings of the Joint Meeting on Foundations of Software Engineering*, pp. 496–507, Bergamo, Italy, August 2015.

[29] Q. Yu, S. Jiang, and Y. Zhang, "A feature matching and transfer approach for cross-company defect prediction," *Journal of Systems and Software*, vol. 132, pp. 366–378, 2017.

[30] B. Muddu, A. Asadullah, and V. Bhat, "CPDP: a robust technique for plagiarism detection in source code," in *Proceedings of the 2013 7th International Workshop on Software Clones (IWSC)*, pp. 39–45, San Francisco, CA, USA, May 2013.

[31] S. Herbold, A. Trautsch, and J. Grabowski, "A comparative study to benchmark cross-project defect prediction approaches," *IEEE Transactions on Software Engineering*, vol. 44, no. 9, p. 1, 2017.

[32] L. Goel, D. Damodaran, S. K. Khatri, and M. Sharma, "A literature review on cross project defect prediction," in *Proceedings of the IEEE Uttar Pradesh Section International Conference on Electrical, Computer and Electronics*, pp. 680–685, Mathura, India, October 2018.

[33] A. V. Phan, M. Le Nguyen, and T. B. Lam, "Convolutional neural networks over control flow graphs for software defect prediction," in *Proceedings of the 2017 IEEE 29th International Conference on Tools with Artificial Intelligence (ICTAI)*, IEEE, Boston, MA, USA, November 2017.

[34] S. Chamoli, G. Tenne, and S. Bhatia, "Analysing software metrics for accurate dynamic defect prediction models," *Indian Journal of Science and Technology*, vol. 8, no. S4, pp. 96–100, 2015.

[35] L. Madeyski and M. Jureczko, "Which process metrics can significantly improve defect prediction models? An empirical study," *Software Quality Journal*, vol. 23, no. 3, pp. 393–422, 2015.

[36] W. Han, C. H. Lung, and S. A. Ajila, "Empirical investigation of code and process metrics for defect prediction," in *Proceedings of the IEEE Second International Conference on Multimedia Big Data*, pp. 436–439, IEEE, Taipei, Taiwan, April 2016.

[37] M. M. Öztürk, "Which type of metrics are useful to deal with class imbalance in software defect prediction?" *Information and Software Technology*, vol. 92, pp. 17–29, 2017.

[38] Y. Xia, G. Yan, and Q. Si, "A study on the significance of software metrics in defect prediction," in *Proceedings of the Sixth International Symposium on Computational Intelligence and Design*, pp. 343–346, IEEE Computer Society, Washington, DC, USA, October 2013.

[39] I. Bluemke and A. Stepień, "Selection of metrics for the defect prediction," *Dependability Engineering and Complex Systems*, Springer International Publishing, New York, NY, USA, 2016.

[40] J. Jiarpakdee, C. Tantithamthavorn, A. Ihara, and K. Matsumoto, "A study of redundant metrics in defect prediction datasets," in *Proceedings of the IEEE International Symposium on Software Reliability Engineering Workshops*, pp. 51-52, IEEE, Ottawa, Canada, August 2016.

[41] B. Caglayan, B. Turhan, A. Bener, and M. Habayeb, "Merits of organizational metrics in defect prediction:an industrial replication," in *Proceedings of the IEEE International Conference on Software Engineering*, pp. 89–98, IEEE, Florence, Italy, May 2015.

[42] G. Mauša and T. G. Grbac, "The stability of threshold values for software metrics in software defect prediction," in *Proceedings of the International Conference on Model and Data Engineering*, pp. 81–95, Barcelona, Spain, October 2017.

[43] F. Zhang, A. E. Hassan, S. Mcintosh, and Y. Zou, "The use of summation to aggregate software metrics hinders the performance of defect prediction models," *IEEE Transactions on Software Engineering*, vol. 43, no. 5, pp. 476–491, 2017.

[44] M. D'Ambros, M. Lanza, and R. Robbes, "An extensive comparison of bug prediction approaches," in *Proceedings of the 2010 7th IEEEWorking Conference on Mining Software Repositories (MSR)*, pp. 31–41, IEEE, Cape Town, South Africa, May 2010.

[45] J. D. Long, D. Feng, and N. Cliff, "Ordinal analysis of behavioral data," *Handbook of Psychology: Research Methods in Psychology*, vol. 2, pp. 635–661, John Wiley & Sons Inc, Hoboken, NJ, USA, 2003.

[46] J. Nam, S. J. Pan, and S. Kim, "Transfer defect learning," in *Proceedings of the International Conference on Software Engineering*, pp. 382–391, IEEE, San Francisco, CA, USA, May 2013.

[47] P. He, Y. Ma, and B. Li, "TDSelector: a training data selection method for cross-project defect prediction," 2016, https://arxiv.org/abs/1612.09065.

[48] S. Feng, J. Keung, X. Yu, Y. Xiao, and M. Zhang, "Investigation on the stability of SMOTE-based oversampling techniques in software defect prediction," *Information and Software Technology*, vol. 139, no. 6, Article ID 106662, 2021.

[49] C. Pak, T. W. Tian, and H. S. Xiao, "An empirical study on software defect prediction using over-sampling by SMOTE," *International Journal of Software Engineering and Knowledge Engineering*, vol. 28, no. 6, pp. 811–830, 2018.

[50] Z. Eivazpour and M. R. Keyvanpour, "Improving performance in software defect prediction using variational autoencoder," in *Proceedings of the 2019 5th Conference on Knowledge Based Engineering and Innovation (KBEI)*, Tehran, Iran, February 2019.

[51] A. Balogun, S. Basri, S. J. Abdulkadir, and V. E. Adeyemo, "Software defect prediction: analysis of class imbalance and performance stability," *Journal of Engineering Science and Technology*, vol. 14, no. 6, pp. 3294–3308, 2019.

Hindawi
Mathematical Problems in Engineering
Volume 2021, Article ID 7143815, 11 pages
https://doi.org/10.1155/2021/7143815

Research Article

Evaluation of the Effectiveness Computer-Assisted Language Teaching by Big Data Analysis

Honglei Wang,[1] Yanjiao Du,[2] and Sang-Bing Tsai[3]

[1]*School of Japanese Studies, Xi'an International Studies University, Xi'an, Shaanxi 710128, China*
[2]*School of Liberal Arts, Yanching Institute of Technology, Yancheng, He Bei 065201, China*
[3]*Regional Green Economy Development Research Center, School of Business, WUYI University, Nanping, China*

Correspondence should be addressed to Yanjiao Du; whlxisu@126.com

Received 28 September 2021; Revised 21 October 2021; Accepted 5 November 2021; Published 22 November 2021

Academic Editor: Chunlai Chai

This paper presents an in-depth study and evaluation analysis of the effects of the computer-assisted language teaching method on foreign language learning through its application. Using an empirical research approach, a practical study of computer-assisted English language teaching was conducted to verify the effects of CALL on oral language learning. In exploring the effects of CALL on students' oral learning, including the effects on fluency, accuracy, and complexity of oral expressions, as well as the effects on learning attitudes, CALL is conducive to improving the fluency of oral expressions, reducing the number of pauses and repetitions in oral expressions, and enabling students to consciously use articulation words to facilitate smooth expressions. CALL is good for improving the accuracy of students' speech, but it does not play a significant role in grammatical accuracy, and the grammatical errors are mainly in the third person singular of verbs, singular and plural of nouns, and passive voice. The use of CALL in teaching oral expressions does not improve the variety of sentences. However, the application of CALL in oral teaching stimulates students' enthusiasm for learning, improves their interest in learning spoken English, and increases their confidence in oral expression.

1. Introduction

Computer-assisted language learning (CALL) refers to the use of computer equipment, the Internet, and other computer-related tools for language learning. The application of CALL to English as a foreign language (EFL) and Electronic Sports League (ESL) is accepted and frequently used in the education industry. In addition, with the development and educational concept of progress, CALL has been taught all over the world [1]. And how to integrate new technology into English-speaking teaching to improve students' English-speaking expression ability is a problem that needs to be solved. Therefore, this study explores the impact of CALL on junior high school students' speaking learning by adopting CALL in English-speaking teaching, to provide useful insights for optimizing English-speaking teaching methods and building a reasonable speaking teaching model. In the environment of big data development, foreign language communicative ability is an indispensable factor for participating in social life [2]. Therefore, the cultivation of foreign language speaking ability has become the purpose of education, and with the attention of the educational community to the cultivation of English language ability, foreign language teaching has gradually become the focus of analysis and research discussions among experts and scholars [3]. Presence is a representative concept of virtual reality technology experience perception, but in the study of the technology acceptance model of virtual reality technology, there is a lack of in-depth analysis of the structure of presence; that is, the fact that presence is a complex, multifaceted psychological perception structure is ignored. The existing research on the adoption of virtual reality environments usually treats the sense of presence as an overall concept, without detailed research on the structure of the sense of presence, and the output results may lack accuracy and have insufficient reference value for the design of virtual reality environments. On the other hand, the sense of presence and 3I features are key features that can represent

virtual reality technology, but there are few studies to explore the inner connection between the structure of sense of presence and 3I features. This study constructs a theoretical framework of virtual reality for foreign language speaking teaching to enrich the research results of the sense of presence theory and provide a more reference value of the receptive behavior model of the virtual reality environment [4].

Various stakeholders of education informatization policy, from the government to schools, from IT companies to international organizations and academic organizations, are actively involved in the development of education informatization. Specifically, government organizations play an important role in the top-level design of education informatization, contributing to policy guidance and continuous investment in education informatization. Based on the importance of education and the need to improve the quality and efficiency of education and reform the teaching system, government organizations promote education informatization projects by promulgating relevant regulations and establishing management bodies [5]. The government's leadership and management of education informatization are mainly through formulating, promulgating, implementing, and supervising the execution of education policies or regulations; allocating resources as rationally as possible; and properly handling the distribution of benefits to solve practical problems and promote the healthy and sustainable development of education informatization. Under the policy, countries will implement a series of action plans for education informatization and promote the implementation of the policy plan. International organizations grasp the direction of education informatization development from an overall international perspective and plan project implementation [6]. In the process of business, enterprises form a set of education informatization solutions that meet the development of education informatization of enterprises and also contribute to the development of education informatization of society; at the same time, enterprises also provide infrastructure and services for the development of education informatization, focus on long-term investment in the education market, and promote the sustainability of governmental organization policies and projects. The characteristic roots are all greater than 1, and the cumulative explanation is 78.677 information, that is, most of the information contained in all variables can be effectively represented by these 5 factors. Academic organizations, on the other hand, stand at the forefront of the development of education informatization, promote research on education informatization, promote the in-depth application of education informatization, and promote the development of the theoretical level of education informatization to a higher level through theoretical research and promotion.

This study has certain theoretical investigation significance and practical value. From the theoretical point of view, the object of previous research on CALL is focused on college students, but this study focuses on junior high school students, a group of students in the middle stage of basic teaching, and probes deeply into the effects of CALL on junior high school students in terms of the complexity, accuracy, and fluency of oral expressions, which to a certain extent enriches the research group of CALL. From the practical level, the research on oral teaching of English in junior high school can, on the one hand, help English teachers pay attention to oral teaching, strengthen the cultivation of students' oral expression ability in the teaching process, and then reflect on the shortcomings of the current oral teaching mode and take the initiative to explore a new model of oral teaching; at the same time, they can constantly update their ideas of oral teaching and improve their teaching activities. At the same time, they will continue to update their teaching concepts, improve their teaching activities, and enrich their teaching skills to improve the quality of oral teaching and give full play to the characteristics of English as a subject. On the other hand, this study will help junior high school students discover the fun and practicality of learning spoken English and make them enjoy learning spoken English in the process of learning English; moreover, it will help junior high school students really realize the necessity of improving their oral expression ability, laying a solid foundation for the comprehensive and balanced development of students' future English ability, and truly reflecting the value of English class.

2. Current Status of Research

Oral language teaching is an important part of junior high school English teaching, and the improvement of junior high school students' English-speaking skills has an important impact on the improvement of other areas such as English reading, listening, and writing [7]. With the development of science and technology, the integration of education and technology has become inevitable, and English teachers are increasingly faced with the challenge of how to integrate new technologies into oral English teaching to improve students' English-speaking skills. With the development of CALL, computer-assisted language teaching has become a keen research area for scholars, and researchers at home and abroad have conducted increased researches in the field of computer-assisted speaking teaching [8]. The focus has been on the role and significance of computers, the Internet, and so on in facilitating language learning, which has been discussed more superficially, as well as exploring the impact of computer-assisted instruction on language learning [9]. Four major clusters of research were developed in this phase: negotiation of meaning, English as a foreign language, learner models, and speech blogs [10]. This phase developed research clusters focused on chat software, computers as a medium of communication, computer literacy, and teacher education. Much social software (e.g., voice chat software, blogs, and wikis) were widely used in CALL in this phase, thus posing higher demands on teachers and students' computer literacy and learning, as well as greater challenges for integrating CALL-related instructional design, making teacher computer education an important research direction [11].

In the deep application stage of CALL research, the main research directions are second language acquisition, collaborative learning, and data-driven learning. In this stage,

the deep integration of technologies such as social networks, big data, learning management systems, and learning analytics makes mugshots (MOOCs), one of the important application modes of CALL, which provides support and assistance for learners to conduct online collaborative learning and independent learning and is widely used in foreign language teaching [12]. According to Kartal et al.'s research findings, future research hotspots will be mobile language learning, gamified language teaching and learning, and data-driven learning. How to integrate and apply these new technologies to enhance students' active, interesting, cooperative, and contextual language learning will be an important part of future research [13]. Again, technological advances in computer-assisted teaching have led to changes in the way language is taught and in the language learning environment, and this change has also led to a shift in the roles of teachers and students, which has placed new demands on teachers' teaching skills [14].

Therefore, how to train teachers to teach CALL and how to make better use of CALL to teach language, so that students' English knowledge, ability, efficacy, and self-confidence can be improved, are the focus of future CALL research. Rienties et al. investigated how to measure fluency in speaking through Cool Edit Pro software [15]. An empirical study was used to demonstrate the effect of an intelligent software system with speech recognition technology on high school students' English speech facilitation. This calculation does not consider the long or short changes in the duration of each segment of the speech under different conditions, and the effect of speech recognition or speech comparison in this way is not good. The results of the study showed that a computer-aided pronunciation system based on automatic speech recognition technology combined with the traditional method of teacher explanation can improve the English speech of high school students more significantly. The English-speaking learning software installed on computers and cell phones has been favored by students because of its mobility and ease of operation and has also contributed to the improvement of students' oral expression skills to a certain extent and demonstrated its teaching effectiveness. This shows that in the research on software-assisted English-speaking teaching, it is mostly a practical case and lacks a certain theoretical height and theoretical basis, which is the next direction to research software-assisted speaking teaching.

3. Evaluation Analysis of the Effectiveness of Computer-Assisted Language Teaching Method in Foreign Language Learning

3.1. Overview of Computer-Assisted Language Teaching Methods. In the context of situational learning theory, the teaching of English as a foreign language means that students can fully understand the context in which the conversation takes place during daily oral training so that they can grasp the content of the spoken language in the context, rather than in the absence of the real situation. To meet this need, teachers are required to change their role from that of authority to that of a facilitator or learning partner of students. The computer system will provide students with more realistic contexts in which they can perceive the appropriate scenarios for oral expression, thus truly promoting students' oral communication skills [16]. Constructivism believes that students are the subjects of information processing and the constructors of cognitive structures, not the passive recipients of external stimuli. Constructivism believes that knowledge is not acquired by teachers but by learners through the construction of meaning with the help of other people (including teachers and learning partners) in a certain context, that is, social, and cultural background. The constructivist theory has changed the position of students and teachers in traditional teaching. In traditional teaching, the teacher is the focus [17]. Teachers have recognized CALL all over the world with the ability to read and write. At the same time, it pays more attention to the in-depth curriculum reform. English teaching requires the comprehensive development of calligraphy students' listening skills. The teacher becomes the center of teaching, holds the dominant role, and plays the role of the transmitter and transporter of knowledge; the position of students and their subjective initiative in learning are not given enough attention, and they are often on the passive side. The constructivist theory emphasizes that teachers are the helpers and facilitators of meaning construction, not the transmitters and installers of knowledge and that students should be at the center, the subjects of cognition and the constructors of knowledge. It recognizes students' existing knowledge and experience and advocates that students give full play to their subjective initiative to process and internalize the old and new knowledge they learn so that they can take ownership of their learning. This is consistent with both the cognitive laws of human beings and the concept of advocating students to learn on their own, as shown in Figure 1.

Presence is a multidimensional structure, and the ideal questionnaire for presence should consider multidimensional structural properties and characterize it using a set of characteristics related to presence. There are both objective and subjective measures of presence, and the subjective measure is more widely used, for which Sheridan reasons that because presence is a mental performance, the basic measure should be self-report. Self-reported state measures can be very useful because quantifying a user's experience of presence allows statistical comparisons across media, stimuli, and subject groups. Presence is a relatively new concept and multidimensional construct for most nonexperts, so an understanding of presence should not directly ask respondents how they feel about being present. Researchers in different media contexts have developed appropriate presence measures and scales designed to elucidate the structure of presence, and these measures provide a reliable and valid dimension of measurement related to presence to optimize the teaching method of oral English in junior high school and construct a reasonable oral teaching model to provide beneficial enlightenment.

In general, scholars do not fully agree on the definition of the dimensions of English-speaking ability, but the common

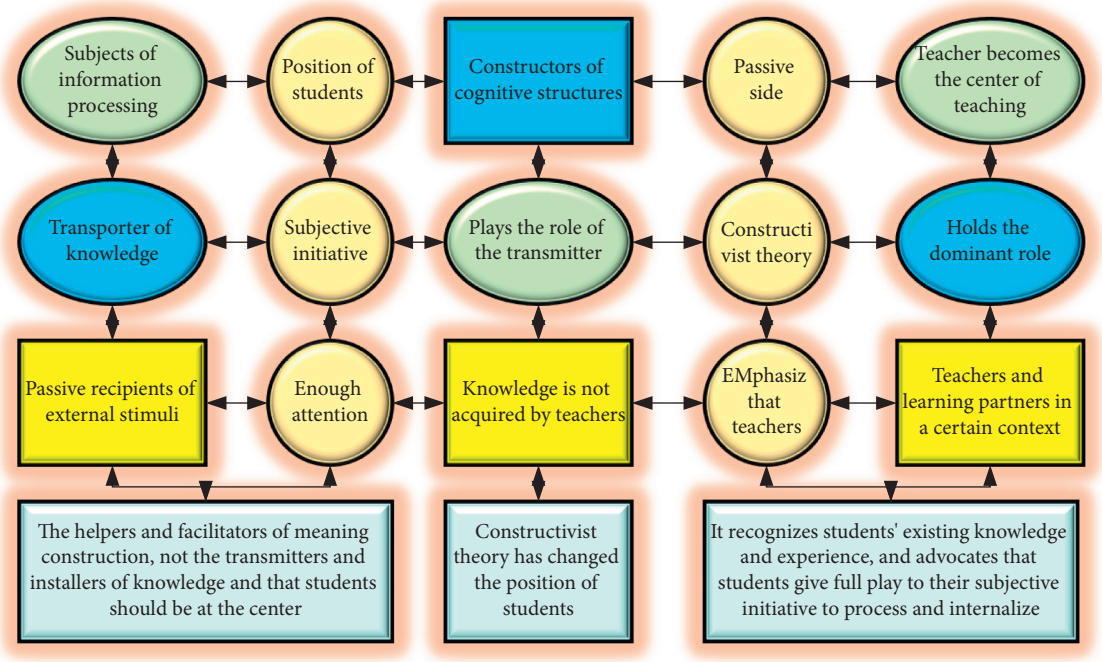


Figure 1: Framework of computer-assisted language teaching method.

factors involved in their evaluation indexes are language knowledge skills, language expression, communication ability, and discourse construction ability. Since the existing English speech ability scales are mainly designed for college students or graduate students, their specific item index descriptions are either too vague and abstract or too messy, lacking systematicity, and the scope of application for quantifying and evaluating English-speaking ability is somewhat limited. For high school English speech learners, the linguistic descriptions of specific items of English-speaking ability indicators should be clearer and more understandable, easy to understand briefly, and easy to operate.

In this study, two natural classes of the second year in a middle school in Weifang City were selected as the experimental class and the control class. The tests and surveys of the students in the two classes before the experiment revealed that the students in the two classes were relatively similar in terms of English-speaking level and learning attitude [18]. The main reasons for choosing the second-year students as the subjects of the study are: on the one hand, compared with the third-year students, the second-year students do not have the pressure of the midterm examination and have more class time for teaching oral English every week, which is convenient for the oral teaching study; on the other hand, the second-year students have accumulated a certain amount of vocabulary and English main grammar, which is suitable to be the experimental subjects of oral teaching. Both presence and characteristics are key features that can represent virtual reality technology, but few studies have explored the internal connection between presence structure and the three characteristics. To understand and count the speaking learning of students with different learning levels, the subjects were divided into three levels: high level, middle level, and low level according to the results of the pre-experimental test on students' speaking levels. Figure 2 shows the basic situation of the students in the experimental and control classes. Both the experimental and control classes took the pre-test and post-test of speaking, and only the students in the experimental class took all the accompanying tests. After the empirical teaching, the students in the experimental class took the questionnaire; in addition, the researcher selected nine students (three students each at high, medium, and low levels) from the experimental class as interviewees (see Figure 2).

This study was conducted to find teaching methods that could promote the improvement of secondary school students' oral expression. Four research methods, experimental method, text analysis method, questionnaire survey method, and interview method were used to compare and analyze the changes in students' oral expression ability and attitude toward oral expression before and after the experiment qualitatively and quantitatively to investigate the effectiveness of CALL in teaching oral English in junior high school. In terms of teaching environment, the experimental class taught in the school's multimedia computer classroom, where students each had a computer to learn speaking with the help of computer-assisted teaching equipment; the control class taught speaking in the ordinary teaching of daily learning. The computer-assisted teaching equipment used in the experimental class integrates computer technologies such as multimedia, hypertext, artificial intelligence, network communication, and knowledge base and has several functions such as human-computer dialogue, peer dialogue, teacher monitoring, simulated context, audio

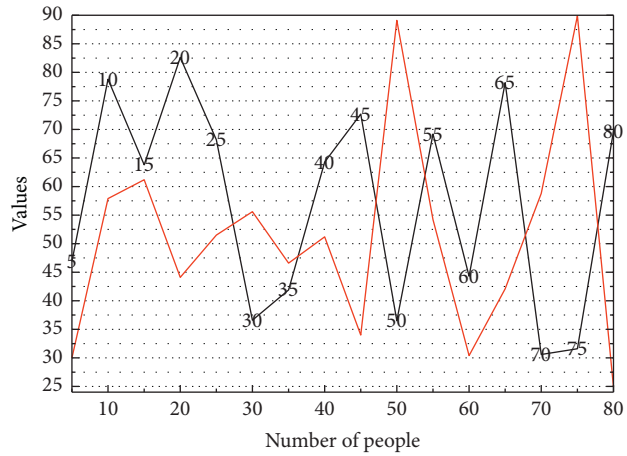


Figure 2: Basic information of students in the experimental and control classes.

playback, instant evaluation, and resource sharing. The four main functions used by teachers in the process of teaching are peer dialogue, simulated context, playback of recordings, and instant evaluation. It is understood that the schools in which this study is conducted, but teachers do not make high use of the machine-assisted teaching equipment and only use it for final tests or in the weeks before exams and for surprise training and rarely make use of the system in daily oral teaching.

In terms of teaching methods, CALL was adopted for the experimental class to teach speaking; the control class adopted the traditional teaching mode to teach speaking and used English class once a week to teach speaking according to the English teaching schedule, without a separate speaking class. Students in the experimental class took a test on the computer immediately after each class to test their classroom learning effect. Each student in the experimental class has a computer, and the teaching is conducted with the help of computers. The following is the way of teaching speaking in the experimental class. The distribution of benefits has to be handled properly to solve practical problems and promote the healthy and sustainable development of education informatization. The practice and research of English speech teaching based on college students mainly focus on their professional background, future needs, and speech contests, and the discussion of English speech teaching methods and contents are mostly limited to this, and the research perspective is not broad enough, while more relevant studies focus on the methods and strategies of English-speaking practice and teaching. Some researchers have used corpus technology to explore and study the use of pragmatic markers in English speech contests, analyzing their lexical features and discourse meaning and emphasizing the important role of pragmatic markers for speech expression and oral communication [19]. Other researchers have critically analyzed the citation of proverbs in English speeches in cross-cultural contexts, exploring the strategies and motives of British and American leaders in citing proverbs in cross-cultural contexts and the impact of the citation of proverbs on and motivations of British and American leaders in citing proverbs in cross-cultural contexts and the important role of proverb citation on the function of speech.

3.2. Experimental Design for Assessing the Effectiveness of Applications in Foreign Language Learning. It refers to the degree to which the measurement tool or instrument can accurately measure the properties of the thing being measured and the degree to which the measurement result reflects the content of the target investigation. Validity analysis methods are divided into judgmental methods and empirical methods according to their types. From the perspective of the judgment method, solid theoretical support is the prerequisite for the robustness and completeness of the scale. The measurement dimensions and specific measurement items of this study have been referred to the credible scales and theories of the foreign and domestic clinical sense, and the dimensions have been confirmed and the items have been revised according to the characteristics of foreign language speaking context teaching, so they have high content validity. Next, the validity of the scale will be verified using the empirical method with small sample data.

The present validity analysis of the presence scale was conducted on 24 items of each dimension of the presence scale, which were retained after the previous reliability analysis [19]. In this study, principal component analysis was used to test whether the scale is suitable for factor analysis, and the statistical significance of the range of Kaiser–Meyer–Olkin (KMO) values is as follows: KMO test values greater than 0.9 indicate suitability for factor analysis, 0.8 to 0.9 indicates suitability for factor analysis, 0.7 to 0.8 indicates suitability for factor analysis, and 0.7 to 0.8 indicates suitability for factor analysis suitability for factor analysis. A test value less than 0.7 indicates unsuitability for factor analysis. Therefore, in this study, the KMO and Bartlett's sphere tests will be performed first before the principal component analysis. According to the test results in Figure 3, the KMO coefficient of the small sample is 0.935, which is greater than the standard value of 0.7, thus proving that it is suitable for factor analysis: The approximate chi-square value of Bartlett's sphere test is 4,045.332, and the companion probability is 0.000, which is less than 0.05, and Bartlett's sphere test is significant, proving that it is suitable for factor analysis (see Figure 3).

Figure 3 analyzes the overall explained variance rate of the small sample data. From the total explained variance table in Figure 3, the virtual reality foreign language speaking teaching presence scale constructed in this study according to theoretical analysis and reliability test consists of five factors, which all have characteristic roots greater than 1 and cumulatively explain 78.677 interest, that is, most of the information contained in all variables can be effectively represented by these five factors of influence. To analyze the intrafactor aggregation and interfactor differentiation characteristics, next, the factor loadings of each variable were characterized with the help of the factor loading matrix, and the loading values represent the degree of influence of the

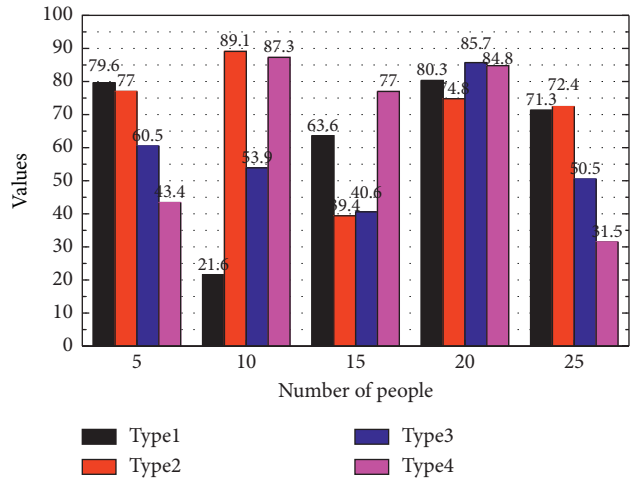


Figure 3: Small sample KMO and Barlett's sphere tests.

extracted principal component factors on the variables [20]. To a certain extent, it enriches the research group of CALL, and at the same time, it can provide certain reference data for the future research of oral English teaching in junior high school, and it can also provide useful theoretical guidance for the construction of oral English teaching mode in junior high school. Based on the principal component extraction above, the 24 influencing factors were further analyzed utilizing factor analysis. Five extracted principal component factors corresponded to each dimensional measure of proxemics predetermined in this study. The scale was optimized and validated through a small sample empirical study and reliability analysis.

The short-time over-zero rate refers to the number of times the short-time speech signal intersects the x-axis in one frame of time. The number of times the time-domain signal of a continuous speech signal passes the time axis is the over-zero rate, and the discrete signal with two sampling values of different signs means it passes the time axis once, from which the over-zero rate can be calculated. We define the short-time over-zero rate of the speech signal as follows:

$$Z_n = \frac{1}{2}\left|\sum_{m=0}^{N-1} \operatorname{sgn}\left[x_n(m)\right] - \operatorname{sgn}\left[x_n(m+1)\right]\right|. \quad (1)$$

The turbid tones with low energy frequency have a low over-zero rate, while the clear tones with high energy frequency have a relatively high over-zero rate. In general, we can find that speech segments have relatively stable over-zero rates through the analysis of over-zero rates, white noise does not have this feature, so we can determine the endpoints of speech by short-time over-zero rates. Since the transformation of the signal in the time domain is usually difficult to see the characteristics of the signal, it is usually converted into the energy distribution in the frequency domain to observe the different energy distributions, which can represent the characteristics of different speech sounds. Therefore, after multiplying by the Hamming window, each frame must also undergo a fast Fourier transform to obtain the energy distribution on the frequency spectrum.

$$X(i,k) = FFT\left(x_i(m+1)\right), \quad (2)$$

$$E(i,k) = |X(i,k)|^3, \quad (3)$$

$$S(i,m) = \sum_{k=0}^{N-1} E(i,k) H_m(k+1). \quad (4)$$

The HMM algorithm is complex and probability-based and requires a large amount of data application training and operations to obtain certain model parameters in the training phase. Although it is significantly better than the dynamic time regularization algorithm in the recognition of continuous speech, the difference between the two is not too obvious in the recognition of small vocabulary speech. Constructivism believes that students are the main body of information processing, the builder of cognitive structure, rather than the passive receiving and instilling objects of external stimuli. The DTW algorithm itself is both simple and effective and therefore has been widely used in specific situations. Dynamic temporal regularization is to make the reference template and test template different in time to achieve the best match by the dynamic planning principle, which is to bend two speech sequences with different times on the time axis so that the two speakers can be matched better. There are two-time sequences M and N with lengths h and k, respectively, where the M sequence is the reference template and the N sequence is the test template, and the value of each point in the sequence is the feature value of each frame in the speech sequence. For example, the speech sequence M has h frames, and the feature value of the i-th frame is m_i.

$$M = m_1, m_2, \ldots, m_i, \ldots, m_h, \quad (5)$$

$$M = n_1, n_2, \ldots, n_j, \ldots, m_k. \quad (6)$$

Now to compare the similarity of two speech time series, the above equation is the simplest one to calculate the distance between them if $h = k$. However, if h is not equal to k, we need to align them. The simplest alignment is linear scaling, which means that the shorter sequence is linearly scaled to the same length as the longer sequence and then compared, or the longer sequence is linearly shortened to the same length as the shorter sequence and then compared. It has been proved that such calculation does not consider the fact that the duration of each segment in speech under different circumstances will produce longer or shorter changes, and speech recognition or speech comparison in this way is not effective, so the recognition effect is not likely to be optimal. A dynamic planning approach is more often used [21]. Given that the system does English phonetic recognition and the rest of the system functions consume a lot of resources, in the spirit of simplicity and easy computation, the DWT algorithm is used in this study to calculate the similarity between the test speech and the standard speech, as shown in Figure 4.

The valid scores obtained from the pre- and post-test were entered into SPSS 21.0 and analyzed by independent

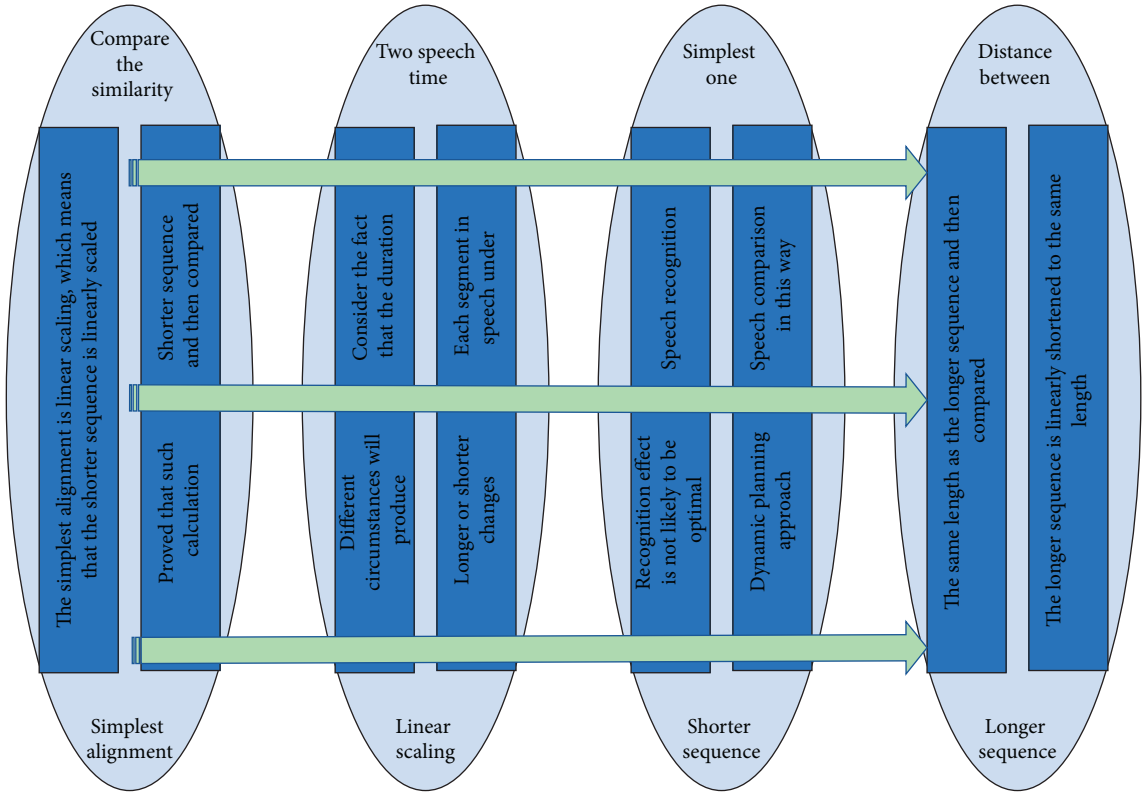


Figure 4: Experimental framework for application effect evaluation.

samples *t*-tests for total, fluency, accuracy, and complexity section scores of the two classes before, during, and after the experiment, with a 100% recall rate. A follow-up test was administered to the experimental class, and a stratified random sampling method was used to select six different students (two each at high, medium, and low levels) from the experimental class for each follow-up test, and their oral recordings were transcribed and analyzed. At the end of the experiment, questionnaires were distributed to the experimental class (45 students). Two invalid questionnaires were found after inspection, and the effective rate was 95.6%. The obtained data were entered into SPSS 21.0 for data analysis, which included: reliability analysis of the questionnaire and descriptive statistical analysis of each question such as frequency and percentage.

4. Analysis of Results

4.1. The Effect of CALL on Students' Oral Expression. The study found that overall CALL had a positive effect on students' oral expressions. However, there were different effects on fluency, accuracy, and complexity of oral expressions, and the results of the study were as follows: CALL was beneficial in reducing the number of pauses and repetitions in students' oral expressions and in improving students' fluency. In terms of accuracy, CALL was beneficial in improving students' speech accuracy, but it did not play a significant role in grammatical accuracy. In terms of complexity, CALL is not conducive to the diversity of students' sentence use and does not contribute to the complexity of students' oral expressions, as shown in Figure 5. The effects

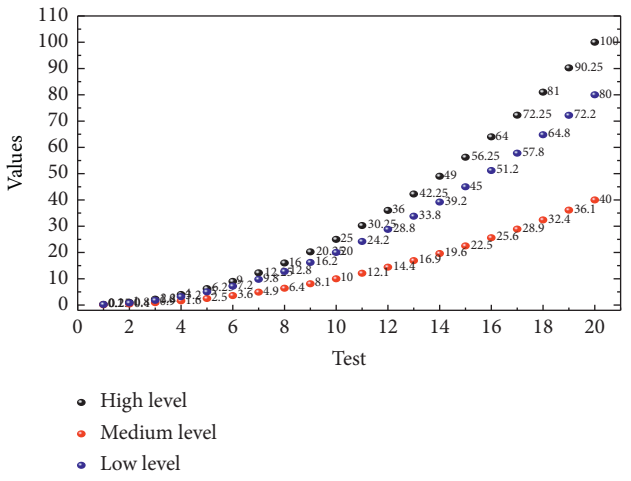


Figure 5: Changes in the number of pauses in the experimental class.

of CALL on students' oral learning will be discussed separately in the order of fluency, accuracy, and complexity, and then the positive effects of CALL on students' oral learning will be described. By comparing the performance analysis of oral fluency in the pre- and post-test, as well as the analysis of the recorded text of the follow-up test in the experimental class, we found that CALL was beneficial to the development of students' oral fluency in the following three aspects: (1) the analysis of the recorded text revealed that the number of pauses and repetitions of students' oral expressions in the

experimental class were significantly reduced; (2) the number of pauses and repetitions of students' oral expressions in the experimental and control classes were significantly reduced in the post-test; and (3) in the independent samples *t*-test of the oral fluency scores of the experimental and control classes, a *P* value of $0.024 < 0.05$ was obtained, and there was a significant difference in the oral fluency scores of the two classes (see Figure 5).

As can be seen from Figure 5, the number of pauses in the oral expressions of the high-, middle-, and low-level learners in the experimental class showed an overall decreasing trend, among which the decrease was the greatest for the low-level learners, with the maximum value of 11 and the minimum value of 5. This shows that CALL can effectively reduce the number of pauses in students' oral expressions. In Figure 5, the number of repetitions in oral expressions performed by the students in the experimental class showed an overall decreasing trend, indicating that CALL was beneficial in reducing the repetitions in the process of students' oral expressions. Compared with their text analysis, the number of pauses in the process of their oral expressions experienced the following changes. At first, the students would pause with Chinese intonation words such as "um, ah, then, oh no"; after a period of training, the students kept quiet when pausing and did not appear Chinese intonation auxiliaries; then they learned to use articulated conjunctions or phrases to replace silence. This article selects 40 English speeches by Chinese and American college students as the corpus. By analyzing the functions of pragmatic markers in the example and comparing the data information in the two sets of speeches, this article discovers several characteristics in the use of pragmatic markers by second language learners: (1) for the same type of pragmatic markers, China speakers tend to use monotonous vocabulary to express, while American speakers use multiple vocabularies to express; (2) Chinese speakers frequently use certain types of markers, while American speakers use multiple types of pragmatic markers; and (3) under the influence of cultural and educational background, Chinese speakers frequently use evaluation markers, while American speakers deliberately avoid using evaluation markers.

To compare the differences more precisely in oral fluency scores between the two classes after receiving different oral instruction, independent sample *t*-tests were conducted on the pre- and post-test scores of the two classes, and the results showed that the *P* value was $0.024 < 0.05$ in the post-test, which proved that there was a significant difference between the oral fluency scores of the two classes, as shown in Figure 6.

As can be seen in Figure 6, in the pre-test of speaking, the experimental and control classes had relatively similar means in oral expression fluency, with a difference of 0.6 and means of 4.75 and 4.15, respectively. p (sig two-sided) after independent samples *t*-test was 0.149, with a *p*-value greater than 0.05, which did not reach the level of statistically significant difference. Therefore, before the empirical study, there was no significant difference in speaking fluency between the two classes. In the oral post-test, the mean values of oral fluency in the experimental and control classes were 5.75 and 4.775, respectively, with a difference of 0.975. An independent samples *t*-test yielded a Sig value of 0.943, which was greater than 0.05, demonstrating that the variance of the two classes' scores on the post-test fluency was chi-square. A P (sig two-sided) of 0.024, with a *P*-value less than 0.05 indicated that the experimental and the *p*-value is less than 0.05, indicating that there is a significant difference between the post-test fluency scores of the experimental and control classes. This indicates that after a period of training, CALL is more advantageous than the traditional oral teaching method in terms of students' oral fluency.

Combining the analysis of the recordings of the speaking tests and the interviews with the students in the experimental class, the reasons for the higher fluency of the students in the experimental class are twofold: on the one hand, CALL provides students with contextualized speaking practice, which allows them to express themselves more generically, thus stimulating their thinking and desire to express themselves, making them more fluent and expressing a greater number of meaningful words. Students must have the opportunity to use language in meaningful contexts, as the context-based practice varies from context to context. On the other hand, CALL somewhat reduces students' fear of oral expression. From interviewing students, we know that many students are initially resistant to speaking English in public, probably due to shyness and fear of making mistakes. This may be due to shyness and fear of making mistakes. Practicing speaking in a CALL environment will avoid such a situation and make students dare to express themselves, which slowly builds up their self-confidence and thus their fluency in expressing themselves in any situation.

4.2. Application Effectiveness Evaluation Results. English speech learning and drills can enhance the core competence of English speech. Scholars' research objects are mainly college or graduate students, but the research object of this topic is the students taught by the author, high school students in the context of China's college entrance examination, whose language level, learning environment, and learning tasks are different from the characteristics of the existing related research objects. Therefore, understanding and grasping the basic information about the background, characteristics, and current situation of the research subject are important conditions to ensure that this study can achieve the expected results. In addition to some of the information already available, such as age, gender, and English language foundation, the author designed a questionnaire for them to gain a deeper understanding of their background and learning status related to their English learning. From Figure 7, we can see that the parents of the students in the experimental class are generally not well educated, with 80% of them having junior high school education or below, and only 5% of them having a college education or above, so it is difficult for them to guide and help their high school children in their studies, especially in foreign language subjects, which many parents know nothing about, and very few parents, less than 5%, can give their children guidance in English learning. In other words,

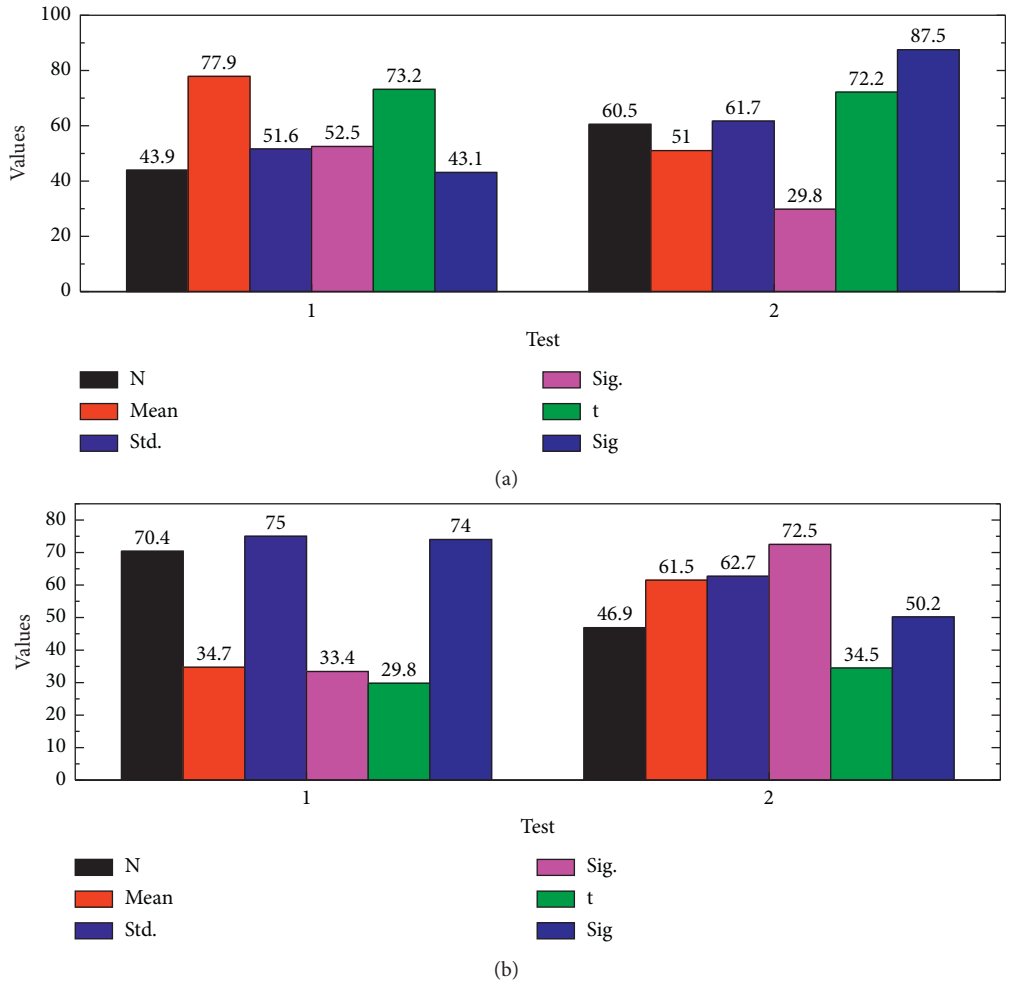


Figure 6: Independent sample t-test for fluency scores.

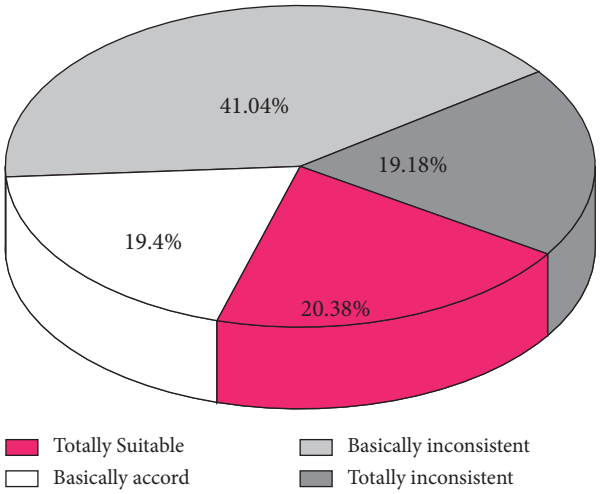


Figure 7: English learning information of students in the experimental class.

there is no significant positive or negative effect of family background on the English learning of the students in the experimental class (see Figure 7).

The differences in the sense of presence between male and female respondents were compared by an independent sample t-test, and the statistical comparison results are shown in Figure 7. The results show that: in the four aspects of immersion, interaction, authenticity, and social presence, the female respondents were significantly higher than the male respondents, confirming hypothesis 2. The subjective measurement method is more widely used. In this regard, Sheridan's reason is that since presence is a psychological performance, the basic measurement method should be self-report. The mean difference in immersion was 0.2277; the mean difference in interaction was 0.2514; the mean difference in interaction was 0.2824; and the difference in social presence was 0.2561. This supports, to some extent, the study has verified that the difference in the effect of gender on social presence is not significant; this is partly related to the fact that the subject of the experimental study was a social presence in noninteractive television scenes, while the subject of this study is an immersive virtual reality system with more. This provides a new research context for analyzing gender differences in perceptions of presence. In the study, the difference in spatial presence perception between male and female research subjects was significant, and this

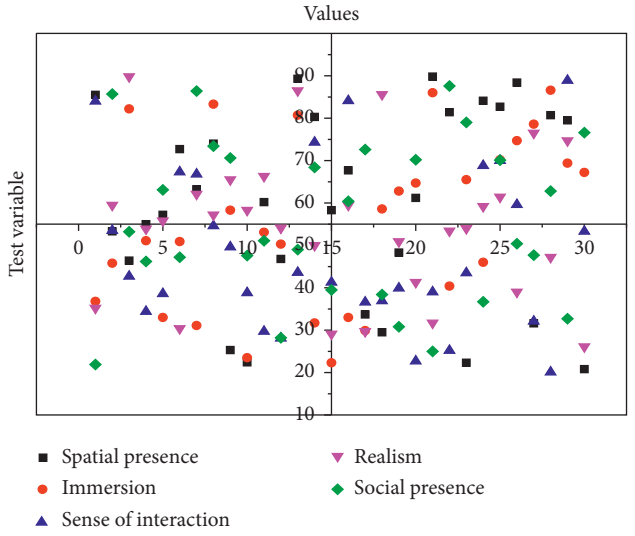


Figure 8: Comparison results of the difference in proximity.

finding is not consistent with the findings of this study, but considering that the virtual reality environment in this study was experienced by a video presentation, there is a difference in the degree of first-view perception and the way of experiencing the virtual reality situation in person, as shown in Figure 8.

CALL enhances the autonomy and personalization of students' learning. In the traditional speaking classroom, students can only follow the teacher's lecture, either memorizing the dialogues in the textbook or simply replacing them according to the dialogue patterns provided by the teacher, which restricts students' thinking in speaking learning. Teachers will provide different ways of practice for a certain topic, and students can choose according to their speaking level and interests, so they can take the autonomy of learning into their own hands, which greatly improves students' enthusiasm and initiative. This not only has a positive impact on the cultivation of students' autonomy in learning spoken English at this stage but also lays the foundation for students to develop good habits of independent learning in the future.

5. Conclusion

In this study, we learned from the analysis of the test results and the interviews with the study participants that the biggest obstacle for affecting students' oral expression is their lack of self-confidence. On the one hand, under the current environment of focusing on written test scores, students are used to listening to what teachers teach and neglecting their oral expression; thus students are unfamiliar and insecure about speaking English; on the other hand, students have a great psychological burden about speaking English in public because they are afraid of making mistakes and others' ridicule, which in the long run forms a vicious circle of not being able to speak, not daring to speak, and not being able to speak. After three months of oral study, we found that the students in the experimental class had experienced the process of speaking to the computer at the beginning, speaking with their peers, public speaking, and self-establishment. Computer-assisted teaching is compulsory for each student to express themselves orally so that students no longer feel unfamiliar with speaking English with a lot of practice. Because of the establishment of self-confidence, students overcome their anxiety and fear when speaking and naturally make the whole expression more fluent and smoother, without the phenomenon of abnormal pauses and repeated reversals as in the early stage of oral teaching. This shows that the use of CALL in the English-speaking classroom has a positive impact on the students' self-confidence and fluency of oral expression.

Data Availability

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

Conflicts of Interest

No competing interests exist concerning this study.

References

[1] A. Bostancıoğlu and Z. Handley, "Developing and validating a questionnaire for evaluating the EFL 'total PACKage': technological pedagogical content knowledge (TPACK) for English as a foreign language (EFL)," *Computer Assisted Language Learning*, vol. 31, no. 5-6, pp. 572–598, 2018.

[2] D. Tafazoli, M. E. Gómez-Parra, and C. A. Huertas-Abril, "A cross-cultural qualitative study on students' attitudes towards computer-assisted language learning," *Qualitative Report*, vol. 25, no. 7, pp. 1841–1855, 2020.

[3] K. Khoiriyah, "CALL and SLA theory: developing A framework to analyze web-based materials for teaching listening skills," *Ideas: Journal on English Language Teaching and Learning, Linguistics and Literature*, vol. 8, no. 1, pp. 80–92, 2020.

[4] H. Liu, C.-H. Lin, and D. Zhang, "Pedagogical beliefs and attitudes toward information and communication technology: a survey of teachers of English as a foreign language in China," *Computer Assisted Language Learning*, vol. 30, no. 8, pp. 745–765, 2017.

[5] C. Garcia, D. Nickolai, and L. Jones, "Traditional versus ASR-based pronunciation instruction an empirical study," *Calico Journal*, vol. 37, no. 3, pp. 213–232, 2020.

[6] R. Gangaiamaran and M. Pasupathi, "Review on use of mobile apps for language learning," *International Journal of Applied Engineering Research*, vol. 12, no. 21, pp. 11242–11251, 2017.

[7] T. Gonulal, "The use of Instagram as a mobile-assisted language learning tool," *Contemporary Educational Technology*, vol. 10, no. 3, pp. 309–323, 2019.

[8] P. Gruba and N. B. Chau Nguyen, "Evaluating technology integration in a Vietnamese university language program," *Computer Assisted Language Learning*, vol. 32, no. 5-6, pp. 619–637, 2019.

[9] S. Bibauw, T. François, and P. Desmet, "Discussing with a computer to practice a foreign language: research synthesis and conceptual framework of dialogue-based CALL," *Computer Assisted Language Learning*, vol. 32, no. 8, pp. 827–877, 2019.

[10] N. Yigitcan Badem and F. Demiray Akbulut, "A general view on utilization of computational technologies in computer assisted language learning (CALL)," *Education Reform Journal*, vol. 4, no. 2, pp. 35–53, 2019.

[11] I. Irham, "Assessing livemocha and duolingo evaluation tasks: adapting chapelle call assessment criteria in autonomous leaning experience," *Erudio Journal of Educational Innovation*, vol. 5, no. 1, pp. 103–114, 2018.

[12] S. Loewen, D. Crowther, D. R. Isbell et al., "Mobile-assisted language learning: a Duolingo case study," *ReCALL*, vol. 31, no. 3, pp. 293–311, 2019.

[13] G. Kartal, "An analysis of using technology in language learning in three flagship journals," *Mehmet Akif Ersoy Üniversitesi Eğitim Fakültesi Dergisi*, no. 53, pp. 515–532, 2020.

[14] D. L. Rachmawati, D. Fadhilawati, and S. Setiawan, "The implementation of computer-assisted language learning (CALL) in the EFL setting: a case study in a secondary school in Indonesia," *English Teaching Journal: A Journal of English Literature, Language and Education*, vol. 8, no. 2, pp. 91–102, 2020.

[15] B. Rienties, T. Lewis, R. McFarlane, Q. Nguyen, and L. Toetenel, "Analytics in online and offline language learning environments: the role of learning design to understand student online engagement," *Computer Assisted Language Learning*, vol. 31, no. 3, pp. 273–293, 2018.

[16] L. De Paepe, "Student performance in online and face-to-face second language courses: Dutch L2 in adult education," *Journal of Educational Sciences*, vol. 37, no. 1, pp. 66–76, 2018.

[17] S. S. Tseng and H. C. Yeh, "Fostering EFL teachers' CALL competencies through project-based learning," *Journal of Educational Technology & Society*, vol. 22, no. 1, pp. 94–105, 2019.

[18] S. Ruiz, P. Rebuschat, and D. Meurers, "The effects of working memory and declarative memory on instructed second language vocabulary learning: insights from intelligent CALL," *Language Teaching Research*, vol. 25, no. 4, pp. 510–539, 2021.

[19] B. A. Shawar, "Integrating CALL systems with chatbots as conversational partners," *Computación Y Sistemas*, vol. 21, no. 4, pp. 615–626, 2017.

[20] M. Alhamami, "Beliefs about and intention to learn a foreign language in face-to-face and online settings," *Computer Assisted Language Learning*, vol. 31, no. 1-2, pp. 90–113, 2018.

[21] E. Namaziandost, S. Alekasir, and S. A. Tilwani, "An account of EFL learners' vocabulary learning in a mobile-assisted language environment: the case of rosetta stone application," *Computer Assisted Language Learning*, vol. 22, no. 1, pp. 80–110.

Hindawi
Mathematical Problems in Engineering
Volume 2021, Article ID 3510402, 8 pages
https://doi.org/10.1155/2021/3510402

Research Article

A BERT-Based Approach for Extracting Prerequisite Relations among Wikipedia Concepts

Youheng Bai,[1,2] **Yan Zhang**,[1,2] **Kui Xiao**,[1,2] **Yuanyuan Lou**,[1,2] **and Kai Sun**[1,2]

[1]*School of Computer Science and Information Engineering, Hubei University, Wuhan 430062, China*
[2]*Hubei Province Educational Informatization Engineering Research Center, Hubei University, Wuhan 430062, China*

Correspondence should be addressed to Kui Xiao; xiaokui@hubu.edu.cn

Received 15 August 2021; Revised 27 September 2021; Accepted 18 October 2021; Published 29 October 2021

Academic Editor: Chunlai Chai

Concept prerequisite relation prediction is a common task in the field of knowledge discovery. Concept prerequisite relations can be used to rank learning resources and help learners plan their learning paths. As the largest Internet encyclopedia, Wikipedia is composed of many articles edited in multiple languages. Basic knowledge concepts in a variety of subjects can be found on Wikipedia. Although there are many knowledge concepts in each field, the prerequisite relations between them are not clear. When we browse pages in an area on Wikipedia, we do not know which page to start. In this paper, we propose a BERT-based Wikipedia concept prerequisite relation prediction model. First, we created two types of concept pair features, one is based on BERT sentence embedding and the other is based on the attributes of Wikipedia articles. Then, we use these two types of concept pair features to predict the prerequisite relations between two concepts. Experimental results show that our proposed method performs better than state-of-the-art methods for English and Chinese datasets.

1. Introduction

In recent years, the emergence of online learning platforms and e-learning resources has injected new impetus into people's learning. Online learning models have gradually become more popular. Research related to this field has also received considerable attention. As everyone has a different knowledge background, the challenge faced by online learners is usually how to choose learning resources and how to rank them. Typically, each learning resource explains one or more of the leading knowledge concepts. Concepts in a field are usually learned progressively, from simple to complex and from abstract to concrete. Usually, the order of learning resources is determined by the relations between main concepts. This kind of relationship between concepts is generally called a concept prerequisite relation. A prerequisite is usually a concept or requirement before one can proceed to the following one. A prerequisite relation is a natural dependency among concepts when people learn, organize, apply, and generate knowledge [1–3].

The learning order between concepts is determined by their prerequisite relations. As for knowledge in a given field, a directed acyclic graph can illustrate its concept prerequisite relations. The concept appears as a node, and the direction of its arrow represents the prerequisite relations between the concepts. For the concept pair (A, B) in the teaching field, if concept B is the prerequisite relation for concept A, then you first learn concept B before learning concept A. It can be written as A⟵B. As shown in Figure 1, neural network (A) relies on concepts such as gradient descent, partial differential, and differential equation. These concepts also rely on differential (B). In other words, before learning neural network, differential equation is needed.

In a classroom course, the instructor will explain each central concept to students according to the inherent order of the concepts. Additionally, the instructor may also spend some time explaining some background knowledge-related concepts to help students understand current knowledge concepts. However, students may not receive assistance

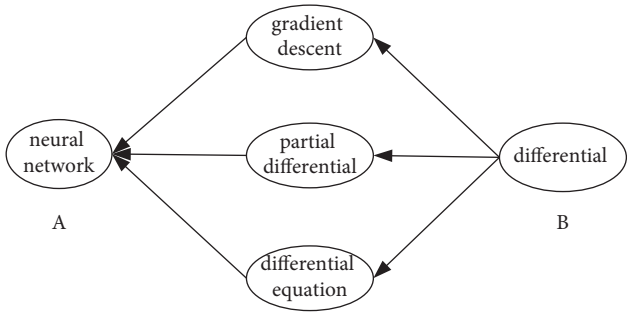


Figure 1: Example of concept prerequisite relation (differential equation concept is the prerequisite relation of neural network concept).

from instructors in online courses. For example, when students learn the Vue.js, they usually need to master the HTML and CSS first; when they learn the Java Spring Boot, they usually need to master the Maven first. There is usually a prerequisite relation between two different learning resources. In addition, when people browse a Wikipedia article, they often open the pages of other articles to learn more about the background of the current article. Between Wikipedia articles, there is usually a prerequisite. Due to a lack of understanding of prerequisite relations between different concepts, people may be unable to complete courses or understand the content of Wikipedia articles.

In this article, we propose a method for extracting concept prerequisite relations from Wikipedia using BERT. We used concepts from Wikipedia, and each concept has its own Wikipedia article. Compared with courses on online learning platforms, Wikipedia's main concepts are easier to extract in an automated way. Furthermore, because Wikipedia has a unique knowledge structure, we can extract the characteristics of concept pairs and analyze the prerequisite relations between concepts easier.

Our main contributions include the following:

(1) A novel metric to measure the prerequisite relations among Wikipedia concepts superior to the existing methods

(2) A Chinese dataset annotated with prerequisite relations between pairs of Wikipedia concepts

The structure of this article is as follows. Section 2 reviews past works on the task of concept prerequisite relations extraction with Wikipedia and MOOC. The problem definition of concept prerequisite relations is in Section 3. Section 4 elaborates on the methodology. Section 5 describes datasets and preparation techniques and our experimental results and analysis. Section 6 is the concluding remarks and future work.

2. Related Work

The concept prerequisite relations determine the order in which knowledge is learned and the order in which documents are read. Nowadays, concept prerequisite relations extraction can be used in different kinds of education-related tasks [4], including curriculum planning [5, 6], learning resources recommendation [7, 8], knowledge tracing [9], and so on. Additionally, there is also a lot of research related to concept prerequisite relation extraction.

The area that researchers pay most attention to is to extract the prerequisite relations between Wikipedia concepts. Talukdar and Cohen [10] utilized three types of features for concept pairs, including WikiHyperlinks, WikiEdits, and WikiPageContent, and then used the MaxEnt classifier to predict prerequisite relations among Wikipedia concepts. Liang et al. [1] studied the problem of measuring prerequisite relations among concepts and proposed the RefD metric to capture the relation. RefD means reference distance, and it uses the page links in Wikipedia to model the prerequisite relation by measuring how differently two concepts refer to each other. Zhou and Xiao [11] employed Wikipedia page links, categories, article content, and time attributes of Wikipedia articles to create features and then predict concept prerequisite relations. Sayyadiharikandeh et al. [12] used the clickstream of human navigation among articles on Wikipedia to infer concept prerequisite relations. In addition, many similar studies have used machine learning methods to predict prerequisite relations between Wikipedia concepts [13–17]. A common problem with these methods is that all of them require experts to manually design the features of the concept pairs.

Besides Wikipedia, some researchers have tried to extract concepts from various learning resources and analyze the prerequisite relationships between concepts. Pan et al. [18] manually extracted the main knowledge concepts of the course from the MOOC video and used the sequence and frequency of appearance of the concepts as features to analyze the prerequisite relations between the concept pairs. Wang et al. [13] extracted the main knowledge concepts from the textbooks, linked these concepts with Wikipedia articles, and then identified the prerequisite relations between the concepts. Liang et al. [14] explored the content of the course introductions on the university website, investigated how to recover concept prerequisite relations on the university website, investigated how concept prerequisite relations are derived from course dependencies, and proposed an optimization-based framework to address the problem. Furthermore, other similar studies use the dependency relationship between learning resources to predict the prerequisite relations between knowledge concepts [1, 2].

As mentioned above, all methods based on machine learning need to use manual design concepts to predict prerequisites. This usually causes other factors that can be used to infer the prerequisite relationship to be ignored. There is a possibility that deep learning will outperform machine learning in this regard since deep learning methods can automatically extract features from raw data. Miaschi et al. [19] used Word2Vec to convert the two concepts into vectors and input the vectors into two LSTM networks, respectively, to obtain the features of the concept pair and predict the prerequisite relations of the concept pair. However, Word2Vec only treats a concept as a normal word. Compared with Word2Vec, BERT [20] can better explore the semantic meaning of a concept, and the contextualized

vectors that BERT generates can also be used to infer concept prerequisite relations.

In this paper, we use the BERT sentence embedding based on contextual embedding to automatically extract the features of concept pairs. Meanwhile, we also designed some features manually for concept pairs. Both classes of features were employed to infer concept prerequisite relations. Furthermore, we created a Chinese concept pair dataset that can be used to identify the prerequisite relations.

3. Problem Definition

The goal of the concept prerequisite relations identification task is to judge whether there is a dependency between two concepts. For a concept pair (A, B), there are four possible relations between them: (1) A is a prerequisite of B; (2) B is a prerequisite of A; (3) the two concepts are related, but they do not have any prerequisite relation between them; and (4) the two concepts are unrelated [10]. In previous studies, researchers usually converted this task into a binary classification problem for processing. They were simply judging whether A is a prerequisite of B. It can be defined as

$$\text{Preq}(B, A) = \begin{cases} 1, & A \text{ is a prerequisite of } B \\ 0, & 0. \end{cases} \quad (1)$$

Preq (B, A) = 1 means that A is a prerequisite of B. In other words, before people can learn about concept B, they must master concept A while Preq (B, A) = 0 means that A is not a prerequisite of B. In this article, we will also turn the concept prerequisite relations identification problem into a binary classification task to deal with.

Moreover, the concepts we use are Wikipedia concepts. Each concept has a corresponding Wikipedia article. The concept is the title of the article.

4. Wikipedia Concept Prerequisite Relations Prediction Method

This section presents our proposed concept prerequisite relations prediction model (AFs + MFs). The structure of the model is illustrated in Figure 1. The input of the model is composed of two types of concept pair features, including features extracted automatically (AFs) and features extracted manually (MFs). Precisely, we extract two BERT sentence embeddings and Wikipedia-based features from concept pairs. First, the model inputs the AFs of the concept pairs into two LSTMs, and the two output vectors of LSTMs are concatenated with MFs. Then, these features are input to a fully connected layer to accomplish concept prerequisite relations recognition.

4.1. Features Extracted Automatically. As a big data pretraining transformation language model of the bidirectional transformer, the application of BERT has significantly improved performance on several NLP tasks. Particularly, sentence-BERT [21] introduces pooling to the token embeddings generated by BERT to generate fixed-size sentence embeddings, obtaining state-of-the-art performance in many fields, including text similarity and classification problems.

Articles in Wikipedia concepts typically contain a number of sentences, each containing deep semantic information. Hence, we use BERT to generate sentence embeddings as the feature extracted automatically from the concept.

More specifically, first, for the first k words or Chinese characters of the Wikipedia concept article, the BERT tokenizer is used with a maximum sequence length of 500 to obtain the token representation. Then, we generate a concept BERT sentence embedding by inputting tokens as the input of the BERT model (vector size = 768). The two BERT sentence embeddings of the concept pair are used as inputs to the neural network, which is passed to the two 32-unit LSTMs. LSTM can be used to create some feature information not included in automatic feature design and achieve deeper concept feature extraction.

4.2. Features Extracted Manually. As a multilingual open knowledge base, Wikipedia has the characteristics of multiuser collaborative editing, dynamic updating, and complete coverage. Wikipedia's concepts are described through articles with corresponding titles, and the articles contain links, categories, and redirects (synonyms) in the content. Researchers can use this information to extract feature information from concept pairs.

By manually extracting the structural features of concept pairs from Wikipedia articles, we can analyze the prerequisite relations between the two concepts. Therefore, we extract three types of concept pair features from Wikipedia article information: text features, links features, and category features. These features are as follows:

(i) *Concept Appearance Count (#1, #2).* It is the number of times the concept A/B appears in the Wikipedia article of the concept B/A.

(ii) *Whether the First Sentence Appears (#3, #4).* The first sentence of the Wikipedia article on concept B/A mentions concept A/B.

(iii) *Jaccard Similarity (#5).* The Jaccard similarity of articles with two concepts (A, B) is calculated.

(iv) *LDA (#6, #7).* Shannon entropy of the LDA vector of A/B: In the information world, the higher the Shannon entropy, the more information can be transmitted [19]. Using lda https://pypi.org/project/lda/, different LDA topic models are trained for each dataset.

(v) *Category (#8).* Whether the concept pair (A, B) belongs to the same Wikipedia category.

(vi) *Link (#9, #10).* For (A, B) concept pairs, whether the concept B/A article refers to concept A/B, that is, contains a link to concept A/B.

(vii) *Link in/out of A (#11, #12).* The number of links to concept A in Wikipedia ("link in") and the number of links to other terms in articles of concept A ("link out") are calculated.

(viii) *Link in/out of B (#13, #14)*. Same as above, the number of links in/out of concept B is calculated.

A note should be made that features #1–#7 and #9-#10 are taken from literature [14], feature #5 is taken from literature [19], and features #8 and #11–14 are taken from literature [16]. Previously, only the English dataset had been validated on these features, and this article will make an evaluation of these features on both the Chinese and English datasets simultaneously.

4.3. AFs + MFs : Concept Prerequisite Relations Prediction Model. Based on the above design and analysis, for a concept pair (A, B), the model (Figure 2) separates the concept prerequisite relations prediction into the following steps:

(1) The first k words or Chinese characters of the concept pair (A, B) Wikipedia articles is first obtained, and the sentences S_A and S_B are generated.

(2) Then, the sentence is divided into individual words or Chinese characters, and they are labeled separately, and BERT is used to encode them to generate sentence embedding in V_A and V_B.

(3) V_A and V_B are input to two LSTMs, and hidden features are further extracted. The output of the two LSTMs is vectors O_A and O_B of size 32 each.

(4) A 14-dimensional vector $V_{(A,B)}$ is created using the Wikipedia features extracted manually, including links, text content, and category features.

(5) Finally, O_A, O_B, and $V_{(A,B)}$ are concatenated $(O_A \oplus O_B \oplus V_{(A,B)})$ as 78-dimensional features and input to a fully connected layer with a sigmoid activation function to realize concept prerequisite relation prediction.

5. Experiments

5.1. Datasets and Implementation Details. For our research, we used a public dataset, AL-CPL, which is an English dataset designed by Chen et al. [9] in their research. The dataset consists of two-category concept pair sets and prerequisite relation labels from four different fields. The fields are data mining, geometry, physics, and precalculus. Each data item is formalized as a triple (A, B, Label), which is the concept A, B, and the prerequisite relation label, respectively. Each concept in the dataset has a corresponding article in Wikipedia. The left half of Table 1 shows detailed information about the AL-CPL dataset.

In addition, we also want to verify whether the proposed method performs well in other languages. By using the AL-CPL English dataset, this paper creates the CH-AL-CPL Chinese dataset. First, the English Wikipedia article corresponding to each concept in the AL-CPL dataset is found, and then the Chinese article corresponding to each concept based on the cross-language links is found in Wikipedia.

However, Chinese Wikipedia articles are only a small fraction of those on English Wikipedia. Thus, the collection of Chinese concept pairs obtained by directly using cross-language links is not only small but also has a significant issue of data category imbalance. Due to this, this paper uses the transitivity and asymmetry of the concept prerequisite relations to expand the number of the Chinese dataset.

(1) *Transitivity*. Concept B is a prerequisite of concept A, concept C is a prerequisite of concept B, and then concept C is a prerequisite of concept A

(2) *Asymmetry*. If concept B is a prerequisite for concept A, then concept A cannot be a prerequisite for concept B

By combining transitivity and asymmetry, we can increase the number of categories in the dataset and balance the ratio of categories. In Table 1, the right half shows the detail of concept pairs in each domain of the CH-AL-CPL. The CH-AL-CPL dataset has been published on GitHub https://github.com/lycyhrc/CH-AL-CPL.

In the experiment, all models were implemented with Keras. Using the bert-as-service (https://bert-as-service.readthedocs.io/) sentence encoding service, we generated a 768-dimensional sentence embedding for the first k words or Chinese characters in Wikipedia concept articles. The sentences were tokenized with NLTK [22]. In order to train the model, the following parameters are set: 50 training epochs, 0.01 learning rate, 32 dimensions of the hidden layer, and 0.2 dropout rate. Adam optimization is used to train the model, and L2 regularization is used to prevent overfitting.

In the AFs model, two 768-dimensional sentence embeddings of the Wikipedia concept pair (A, B) are input to two 32-unit LSTMs, and the fully connected layer is used to receive the output of the LSTMs to identify prerequisite relations. Besides, the #1–16 manual features of the Wikipedia concept pair (A, B) are combined, which then sends them to a fully connected layer for prerequisite relations prediction. Further, the AFs + MFs model concatenates the output of the LSTM of the AFs model and the manual features of MFs to complete the prerequisite relations recognition.

5.2. Experimental Result and Analysis. In our experiment results, we compared our method with the following typical concept prerequisite relations prediction baselines:

(1) *Reference Distance (RefD)* [1]. The basic idea of this method is that each concept can be represented by its collection of related concepts in the concept space if most of the related concepts of concept A refer to concept B. The related concept of concept B rarely refers to concept A, then concept A may depend on concept B. The author constructs related links for each related concept and refers to the EQUAL weight and TF-IDF weight method to identify the prerequisite relations between the two concepts, so we selected the best-performing TF-IDF weight method.

(2) *Machine Learning (AT)* [14]. This method uses link-based and text-based features extracted from Wikipedia pages and then uses Naive Bayes (NB),

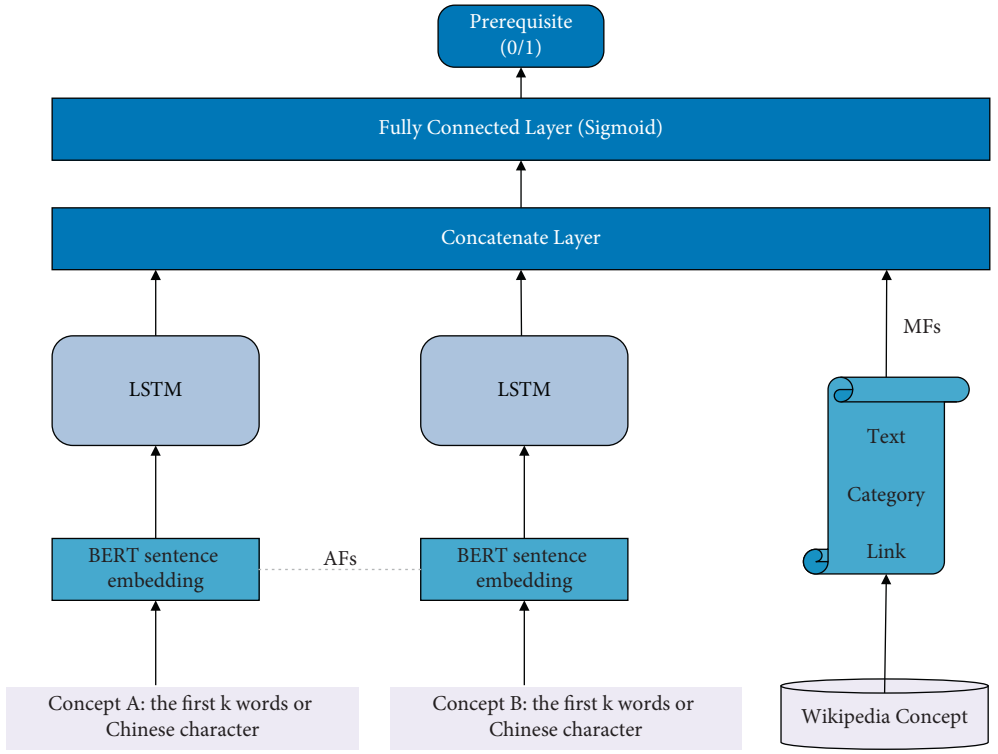


Figure 2: Concept prerequisite relations prediction model (AFS + MFS).

Table 1: The number of concepts pairs and prerequisite pairs in the dataset AL-CPL.

Domain	AL-CPL			CH-ALCPL		
	#Pairs	#Positive pairs	#Negative pairs	#Pairs	#Positive pairs	#Negative pairs
Data mining	826	292	534	1151	493	658
Geometry	1681	524	1157	3330	1825	1505
Physics	1962	487	1475	2958	1091	1867
Precalculus	2060	699	1361	3200	1431	1769
All	6529	2002	4527	10639	4840	5799

logistic regression (LR), support vector machine (SVM), and random forest (RF) four classifiers to train a concept prerequisite relations prediction model on the AL-CPL dataset, which is used to analyze the prerequisite relations between concept pairs. We use the results of the best-performing random forest classifier (RF) report directly as the basis for comparison.

(3) *Neural Network (RS)* [16]. This method is based on a neural network, and the UNIGE_SE team is responsible for the sharing task of EVALITA 2020 PRELEARN. The author proposes eight category features based on the content and structural features of Wikipedia, using Italian datasets and deep learning models to analyze the prerequisite relations between concepts. As a comparative algorithm, these feature values are recalculated on Chinese and English datasets, and the author's method is used to predict the prerequisite relations of English and Chinese concept pairs as the comparative algorithm of this paper.

(4) *AFs and MFs*. Besides the above baseline, to verify the effectiveness of the proposed method's automatic features and manual features, respectively, this paper also predicts the concept prerequisite relations for the two types of features separately. Specifically, we use a fully connected layer to receive automatic and manual features as input, thereby achieving prerequisite relations prediction individually.

On the AL-CPL and CH-CL-CPL datasets, we conducted a concept prerequisite relations prediction experiment. The performance of the model is evaluated using 5-fold cross-validation. In comparison to a baseline model, the most widely used performance metrics are Precision (P), Recall (R), and $F1$ score ($F1$). Tables 2 and 3 show the results of

Table 2: Performance comparison of the AL-CPL dataset.

Domain metric	Data mining			Geometry			Physics			Precalculus		
	P	*R*	*F*	*P*	*R*	*F*	*P*	*R*	*F*	*P*	*R*	*F*
RefD	52.8	77.5	66.3	42.4	62.3	50.4	49.9	49.6	49.4	75.1	69.4	72.1
RS	62.7	68.2	65.1	70.6	76.0	73.0	53.0	62.3	57.1	69.9	76.9	73.1
AT	80.7	73.3	76.7	95.0	84.7	89.5	85.2	59.3	69.9	90.2	87.1	88.6
AFs	81.7	85.3	83.3	88.4	90.6	89.4	73.8	80.2	76.7	88.3	91.7	89.9
MFs	71.1	78.0	74.3	73.6	83.0	77.9	61.3	80.2	69.3	77.3	82.5	79.8
AFs + MFs	86.5	87.3	86.9	93.7	92.5	93.1	81.4	83.8	82.6	91.6	92.7	92.1

Table 3: Performance comparison of the CH-AL-CPL dataset.

Domain metric	Data mining			Geometry			Physics			Precalculus		
	P	*R*	*F*	*P*	*R*	*F*	*P*	*R*	*F*	*P*	*R*	*F*
RefD	55.6	56.4	56.0	74.1	62.9	68.1	63.4	69.5	66.3	71.5	65.6	68.4
RS	79.4	79.7	79.6	92.7	92.8	92.7	83.9	83.0	83.3	91.0	88.6	89.8
AFs	89.4	91.0	90.2	96.9	97.5	97.2	89.2	88.3	88.7	93.6	94.4	94.0
MFs	74.9	80.3	77.5	92.0	95.3	93.6	76.8	89.0	82.4	89.1	90.8	89.9
AFs + MFs	90.7	91.9	91.2	97.8	97.9	97.8	92.0	94.1	93.0	96.6	96.8	96.7

evaluating different baselines under different performance metrics for AL-CPL and CH-AL-CPL, respectively.

As shown in Tables 2 and 3, our method significantly outperforms all the baselines in all the metrics on English and Chinese datasets (except AL's Precision metric).

From Table 2, we can find that our method achieves the best performance against all baselines on all domains, except for the Precision metric of the geometry and physics domain. The $F1$ score of AFs + MFs leads AL by about 3.6%, 3.7%, 5.9%, and 3% in each of the four areas. In geometry and physics, Precision is the best probably because these two fields have the rich text and link features.

Based on Table 3, we observe that our method outperforms all baselines in all metrics and achieves the best result. Table 3 reports the evaluation metrics for the four domains. The CH-AL-CPL dataset, which is expanded by transitivity and asymmetry, has the most significant number of prerequisites. The performances obtained are generally better than CH-CAL-CPL. In addition, since the author did not give the code of the AT method from [14], some features in Chinese cannot be calculated, so we did not report the experiment results of the AT method in CH-AL-CPL.

As an excellent NLP language model, BERT replaces the encoder with the decoder by using a two-way transformer encoder. By using this method, the feature encoding effect of the words in the sentence is greatly improved. Compared with previous models, such as Word2Vec, the trained BERT model has a deeper contextual understanding. A context-based semantic feature is suitable for detecting text features of Wikipedia concepts. Moreover, as a concept pair, we also cannot ignore the rich links and category relationships between the two. Overall, through the combination of the two types of features, the performance of the concept prerequisite relations model can be further improved.

5.3. Ablation Study. In order to demonstrate that the length of the Wikipedia article influences the automatic extraction of features, we conducted an ablation experiment by varying the value of k (the first k words or characters), from 100 to 500. The experiment results are shown in Figure 3.

As shown in Figure 3, increasing k will increase the $F1$ score of the AFs model. Particularly when $k = 400$, the model is the most effective in predicting the four domains. After the k value exceeds 400, however, the textual information of the concept is incorporated into other background knowledge, which affects the performance of the model to a certain extent. According to the CH-AL-CPL dataset, geometry and precalculus have the best $F1$ scores when $k = 500$. This may be due to the relatively long average length of articles in this domain.

Additionally, the experiment explored the role of manual features in the MFs model. There are three types of features that are designed, content-based (#1–#7), categorical-based (#8), and link-based (#9–#14). In the experiment, one feature type was removed each time and compared with the MFs model. The result is shown in Figure 4.

Figure 4 illustrates how the prediction performance decreases to varying degrees after reducing a specific type of feature group. After removing the link-based features from the AL-CPL dataset, the decline in the four fields reached 10.5%, 12.7%, 9.1%, and 7.4%, respectively, indicating that the link relations between concepts play a crucial role in the prediction of prerequisite relations.

As a result of the difference in the CH-AL-CPL dataset, content-based features have decreased by 7.0%, 11.3%, 8.8%, and 6.1% in the four domains. Since Chinese Wikipedia has fewer words than English Wikipedia, less text information has a more significant impact on text features. Moreover, removing the features with a category has the most negligible

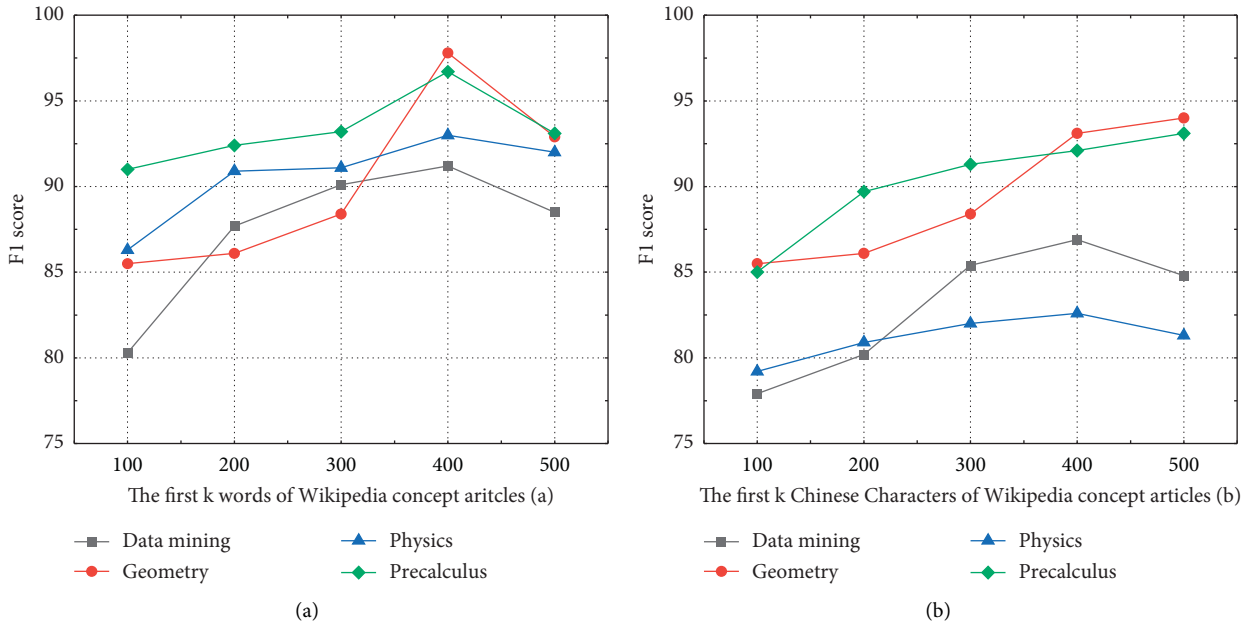


Figure 3: The effect of altering the value of k on the $F1$ score of the AFs model: (a) AL-CPL; (b) CH-AL-CPL.

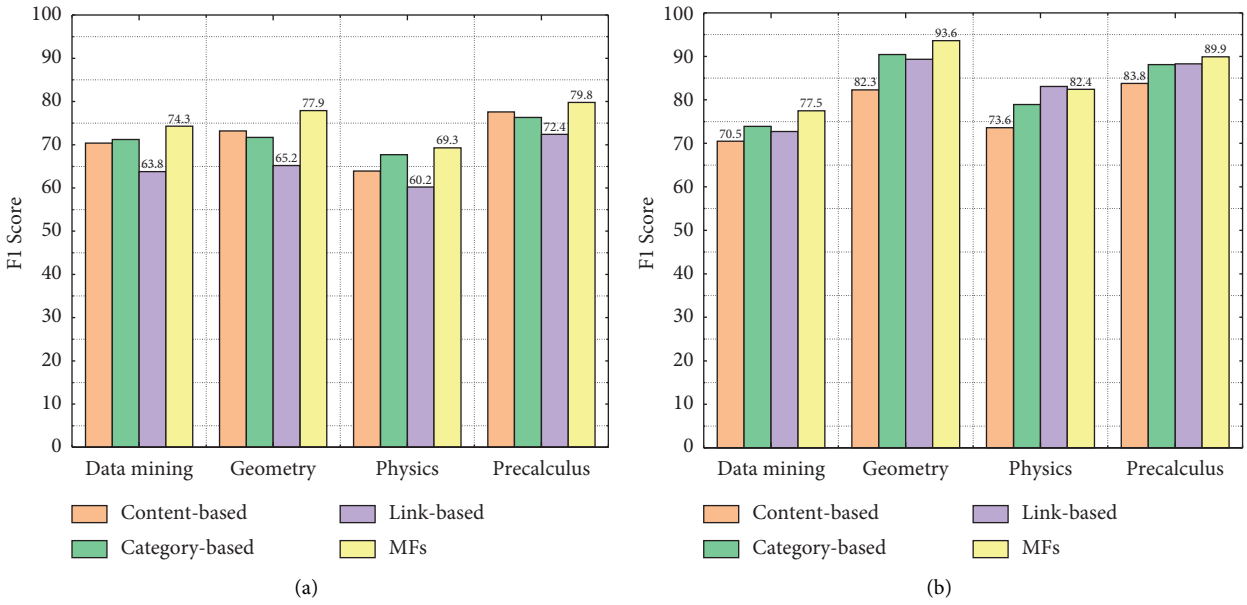


Figure 4: Contribution of each feature group to the MFs model in (a) AL-CPL and (b) CH-AL-CPL.

effect on the MFs model mainly because we design too few category-based features.

6. Conclusion and Future Work

In this paper, we propose a novel concept prerequisite relations prediction method called AFs + MFs, which combines the BERT sentence embedding (AFs) of the concept article and Wikipedia-based features (MFs). Furthermore, we designed a Chinese prerequisite relations dataset to verify the effectiveness of the method. The experiment results show that our method achieves state-of-the-art results on four domains. In addition, we have conducted effectiveness studies on AFs and MFs separately.

In the future, we plan to identify the concept prerequisite relations of non-Wikipedia concepts. Moreover, some learning resources, such as MOOCs and e-lectures, contain multiple concepts. The following research question is how to recommend learning resources by considering concept prerequisite relations.

Data Availability

The CSV data used to support the results of this study have been stored in the GitHub repository (https://github.com/lycyhrc/CH-AL-CPL). These data are the concept prerequisite relations dataset of the Chinese version of AL-CPL. There are no restrictions on access to these data.

Conflicts of Interest

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

Acknowledgments

This work was supported by the National Natural Science Foundation of China (No. 61977021), the Technology Innovation Special Program of Hubei Province (Nos. 2018ACA133 and 2019ACA144), and the Teaching Research Project of Hubei University (No. 202008).

References

[1] C. Liang, Z. Wu, W. Huang, and C. Lee Giles, "Measuring prerequisite relations among concepts," in *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pp. 1668–1674, Lisbon, Portugal, September 2015.

[2] S. Laurence and E. Margolis, *Concepts and Cognitive Science. Concepts: Core Readings*, pp. 3–81, MIT Press, Cambridge, Massachusetts, USA, 1999.

[3] C. Hu, K. Xiao, Z. Wang, S. Wang, and Q. Li, "Extracting prerequisite relations among wikipedia concepts using the clickstream data," in *Knowledge Science, Engineering and Management. KSEM 2021. Lecture Notes in Computer Science*, H. Qiu, C. Zhang, Z. Fei, M. Qiu, and SY. Kung, Eds., vol. 12815Cham, Springer, 2021.

[4] J. Gordon, L. Zhu, A. Galstyan, P. Natarajan, and G. Burns, "Modeling concept dependencies in a scientific corpus," in *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics*, pp. 866–875, Berlin, Germany, August 2016.

[5] B. Golshan, E. Papalexakis, and R. Agrawal, "Toward data-driven design of educational courses: a feasibility study," *Journal of Educational Data Mining (JEDM)*, vol. 8, no. 1, pp. 1–21, 2016.

[6] H. Li, T. Wang, W. Pan et al., "Mining key classes in java projects by examining a very small number of classes: a complex network-based approach," *IEEE Access*, vol. 9, pp. 28 076–28 088, 2021.

[7] C. Liang, J. Ye, Z. Wu, B. Pursel, and G. Giles, "Recovering concept prerequisite relations from university course dependencies," in *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence*, pp. 4786–4791, San Francisco, USA, February 2017.

[8] R. Manrique, B. Pereira, and O. Marino, "Towards the identification of concept prerequisites via knowledge graphs," in *Proceedings of the 2018 IEEE/WIC/ACM International Conference on Web Intelligence (WI)*, pp. 332–336, Santiego, Chille, December 2019.

[9] P. Chen, Y. Lu, V. W. Zheng, and Y. Pian, "Prerequisite-driven deep knowledge tracing," in *Proceedings of the 2018 IEEE International Conference on Data Mining (ICDM)*, pp. 39–48, IEEE, Singapore, November 2018.

[10] P. Talukdar and W. Cohen, "Crowdsourced comprehension: predicting prerequisite structure in Wikipedia," in *Proceedings of the Seventh Workshop on Building Educational Applications Using NLP*, pp. 307–315, Montréal, Canada, June 2012.

[11] Y. Zhou and K. Xiao, "Extracting prerequisite relations among concepts in wikipedia," in *Proceedings of the 2019 International Joint Conference on Neural Networks (IJCNN)*, pp. 1–8, Budapest, Hungary, July 2019.

[12] M. Sayyadiharikandeh, J. Gordon, J. L. Ambite, and K. Lerman, "Finding prerequisite relations using the wikipedia clickstream," in *Proceedings of the Companion Proceedings of The 2019 World Wide Web Conference*, pp. 1240–1247, New York, NY, USA, May 2019.

[13] S. Wang, O. Alexander, Z. Wu et al., "Using prerequisites to extract concept maps from textbooks," in *Proceedings of the 25th ACM International on Conference on Information and Knowledge Management*, pp. 317–326, New York, NY, USA, October 2016.

[14] C. Liang, J. Ye, S. Wang, B. Pursel, and C. Lee Giles, "Investigating active learning for concept prerequisite learning," in *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence*, pp. 7913–7919, New Orleans, USA, February 2018.

[15] F. Gasparetti, C. De Medio, C. Limongelli, F. Sciarrone, and M. Temperini, "Prerequisites between learning objects: automatic extraction based on a machine learning approach," *Telematics and Informatics*, vol. 35, no. 3, pp. 595–610, 2018.

[16] A. Moggio and P. Andrea, "UNIGE_SE@PRELEARN: utility for automatic prerequisite learning from Italian Wikipedia," in *Proceedings of the 7th Evaluation Campaign of Natural Language Processing and Speech Tools for Italian*, pp. 1–8, December 2020.

[17] C. Liang, J. Ye, H. Zhao, B. Pursel, and C. Lee Giles, "Active learning of strict partial orders: a case study on concept prerequisite relations," 2018, https://arxiv.org/abs/1801.06481.

[18] L. Pan, C. Li, J. Li, and J. Tang, "Prerequisite relation learning for concepts in MOOCs," in *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics*, pp. 1447–1456, Vancouver, Canada, July 2017.

[19] A. Miaschi, C. Alzetta, F. A. Cardillo, and F. Dell'Orletta, "Linguistically-driven strategy for concept prerequisites learning on Italian," in *Proceedings of the Fourteenth Workshop on Innovative Use of NLP for Building Educational Applications*, pp. 285–295, Florence, Italy, August 2019.

[20] J. Devlin, M. W. Chang, K. Lee, and K. Toutanova, "Bert: pre-training of deep bidirectional transformers for language understanding," 2018, https://arxiv.org/abs/1810.04805.

[21] N. Reimers and I. Gurevych, "Sentence-bert: sentence embeddings using siamese bert-networks," 2019, https://arxiv.org/abs/1908.10084.

[22] S. Bird and E. Loper, "NLTK: the natural language toolkit," in *Proceedings of the ACL 2004 on Interactive Poster and Demonstration Sessions*, Barcelona, 2004.

Hindawi
Mathematical Problems in Engineering
Volume 2021, Article ID 6551493, 10 pages
https://doi.org/10.1155/2021/6551493

Research Article

Research on Autoarrangement System of Accompaniment Chords Based on Hidden Markov Model with Machine Learning

Shuo Shi,[1] Shuting Xi,[1] and Sang-Bing Tsai[2]

[1]*School of Music and Dance, Zaozhuang University, Zaozhuang 277000, Shandong, China*
[2]*Regional Green Economy Development Research Center, School of Business, WUYI University, Nanping, China*

Correspondence should be addressed to Shuting Xi; tenor09@126.com and Sang-Bing Tsai; sangbing@hotmail.com

Received 10 August 2021; Revised 1 September 2021; Accepted 13 September 2021; Published 14 October 2021

Academic Editor: Chunlai Chai

Accompaniment production is one of the most important elements in music work, and chord arrangement is the key link of accompaniment production, which usually requires more musical talent and profound music theory knowledge to be competent. In this article, the machine learning model is used to replace manual accompaniment chords' arrangement, and an automatic computer means is provided to complete and assist accompaniment chords' arrangement. Also, through music feature extraction, automatic chord label construction, and model construction and training, the whole system finally has the ability of automatic accompaniment chord arrangement for the main melody. Based on the research of automatic chord label construction method and the characteristics of MIDI data format, a chord analysis method based on interval difference is proposed to construct chord labels of the whole track and realize the construction of automatic chord labels. In this study, the hidden Markov model is constructed according to the chord types, in which the input features are the improved theme PCP features proposed in this paper, and the input labels are the label data set constructed by the automated method proposed in this paper. After the training is completed, the PCP features of the theme to be predicted and improved are input to generate the accompaniment chords of the final arrangement. Through PCP features and template-matching model, the system designed in this paper improves the matching accuracy of the generated chords compared with that generated by the traditional method.

1. Introduction

With the increasingly vigorous development of the modern Internet, music has new media and carriers, and more and more music products are derived. Digital music has been better popularized and spread in the Internet information flow carrier, which greatly enriches people's spare time life. The development of intelligent, Internet, virtual reality, and other technologies has blurred the boundary between the real world and the virtual world, allowing paintings, art, and music to be presented to people in a highly genuine form. With the improvement in computer performance and the diversification of Internet functions and products, the threshold of learning music has been greatly reduced. People no longer need rich music theory knowledge and deep musical literacy to engage in music-related industries, such as music creation, music adaptation, and music retrieval. In the field of computer, more and more scholars and experts try to solve and simplify some problems related to music learning and creation by combining music theory knowledge with audio signal processing and researching specific features and algorithms. Artificial intelligence, such as deep learning and machine learning, covers all walks of life and extends to music, which is also developing in the direction of intelligence. Computers begin to assist or even replace professional workers to complete music work [1].

Computer arranger process is by computer algorithm, looking for a set of suitable system for the whole period of melody chord, a pop music usually combined by two parts of the vocal and instrumental music accompaniment; melody is a series of single notes to form a continuous music, and they constitute the theme of the music. Therefore, people create or recreate a popular music, often from the beginning of the

creative part of the main melody. Orchestrating harmonious chords for melodic lines can be a daunting task for amateur music lovers. For those who are interested in music creation, it is of great practical significance to study the automatic music accompaniment of relevant computers. For those who are interested in music creation, it is of great practical significance to study the automatic music accompaniment of relevant computers. Because the chord tension and foil music accompaniment in creating music plays an important role in the emotional aspects, the theme automation music accompaniment system will generate a matching chord accompaniment, Finally, a complete music file containing melody and chord accompaniment is the output. The music with automatic accompaniment generated by relevant computer algorithms can be used for entertainment and can also be used for music creators through certain theoretical reference. In the process of automatic accompaniment of music, the part of accompaniment is completely completed by the computer. By inputting the main melody, the creator can get a complete new music work with chord accompaniment, and use computer composition and accompaniment to enrich and expand the research field of computer algorithm. Arrangers can provide a variety of possibilities for the creation of music forms and styles. To a certain extent, the study of automatic music accompaniment system enriches the innovation of music and also provides music creators with reference to music accompaniment chords.

Most musicians think music itself is extremely emotional, subjective, audio, a form of art, many segments of the rhythm of the music from the composer, fragmentary, and the creation inspiration of discontinuity; so for the inspiration of fragmentation and randomness, it is difficult to by a certain fixed computer algorithms to replicated and created again, So, it is more difficult to use computers to help us compose music, but as more and more computer algorithms are introduced into the field of music composition, through hidden Markov algorithm, stochastic process, genetic algorithm, artificial neural network, and so on, algorithmic composition is easier to apply to the current music form. It can be done through the computer simulation in the world with all kinds of music styles and forms [2]. It makes music more accessible to people who are interested in music creation but lack relevant music knowledge, eliminates the barriers of music creation, and makes seemingly distant music creation close at hand.

The harmony of music is the core of accompaniment. To match a harmonious accompaniment for any given melody, it is necessary to solve the coordination problem of automatic accompaniment [3], which leads to another automatic accompaniment system that can match the harmonious chords of the input melody. Lee and Marsic put forward a kind of automatic accompaniment system suitable for a particular style; they constructed a system using new Riemann to transform a chord melody of process based on the MIDI list of paths, including alternative chord path in a similar binary tree structure, and then by a Markov chain with learning probability statistical optimization matching probability of chord. Emilia Gomez put the influence of harmonic frequency into the feature statement in the process of studying PCP features, considered the maximum value of specific frequency, and constrained the normalization of the feature weight of related frequency bands [4]. The improved HPCP characteristics reduce the influence of intensity and different timbre to some extent. Yang et al. has produced a software that can convert an arbitrary input audio signal into a chord sequence corresponding to the harmonious accompaniment [5]. In the process of studying PCP features, Wu et al. also included the influence of harmonic frequency into feature description, considered the maximum value of specific frequency, constrained and normalized the feature weight of related frequency bands, and improved HPCP features reduced the influence of intensity and different timbers to a certain extent [6]. By introducing the maximum likelihood criterion decision tree algorithm, Xue et al. calculated the likelihood coefficients between all single notes and calculated the occurrence times of adjacent intervals at different times. The chord sequence obtained from the combination of the single note with the most occurrence times, and the largest likelihood coefficient was taken as the final matching result [7]. Therefore, solving the automatic arrangement of music chords has become a hot research direction of computer at the present stage.

2. Theoretical Knowledge of Musical Models

2.1. Music Theory. Rhythm is the music of different lengths of the sound, according to a certain law of the combination of musical forms. Rhythm is in the beat, and the rhythm cannot be separated from the beat. The beat is a cyclical occurrence of a rhythm with a rule of strength and weakness [8]. Beats are expressed in fractional form in musical notation. Melody is the soul of music. The high and low of notes, the speed of rhythm, and the strength and weakness make the melody present different colours. Different pitches are connected to form the pitch contour of the melody, which abstracts into a curved melody curve. The distance between different points on the curve represents the interval relationship between pitches. In general, the basic patterns of melody can be summarized as horizontal progression, upward progression, downward progression, and wave progression. Melody is the basis of forming a part. Monophonic music has only one melody, whereas multipart music contains multiple melodies, which revolve around a certain main melody, and each melody is independent and interacts with each other. Generally speaking, the progression of two-part melody can be divided into simultaneous progression, parallel progression, reverse progression, and oblique progression.

Tone is the law of music, which normalizes the relationship between musical sounds through an artificial constraint, so that it presents a form of expression in line with human aesthetics and cognition. At present, there are three main ways of expression of temperament: pure temperament, five degrees of mutual generation temperament, and twelve-equal temperament. The pure fifth of interval relation is the key element. On the premise of determining the pitch, the interval relation is taken as the pure fifth, that

is, the conditional constraint of the frequency ratio 3 : 2, and the remaining tone values are deduced [9]. The tone relation obtained in this way is the reciprocal fifth. The characteristic of purity, from the point of view of signal processing, is the frequency ratio of each tone level, identical to a certain integer. According to the relationship of pure temperament, the overall harmony of the tone level is very high, and it is comfortable and three-dimensional from the perspective of human hearing experience. Therefore, in modern applications, pure temperament is generally used in symphony performance, especially in the case of multipart and multi-instrument ensemble, which has a good harmony.

2.2. Fundamentals of Music Signal Analysis. Musical Instrument Digital Interface is one of the most common structured symbolic representations. The contents of an MIDI file are a series of instructions that define what the Instrument will play and when. Because no audio waveforms are stored, MIDI files take up little storage space, and the stored contents can be modified flexibly; these characteristics make MIDI widely used in music creation, music recording, music analysis, and other aspects. Music notation is a kind of musical notation, including two types of music notation for recording pitch and fingering. Among them, the simplified score and staff score belong to the score of recording pitch, whereas the six-line score used for guitar performance belongs to the score of recording fingering.

STFT is a steady-state analysis of signals based on the assumption that the signals are stable in a short time. Therefore, piano music can be assumed to have short-term stationarily and analysed by STFT. The definition of STFT is shown as

$$X_m = \sum xwe^{-jn} + \sum xw, \tag{1}$$

where x represents discrete music signal, w stands for window function, and X represents the spectrum at time m. In the process of STFT, the length of the window determines the time resolution and frequency resolution. The longer the window length, the longer the intercepted signal, the lower the time resolution, and the higher the frequency resolution; conversely, the shorter the window length, the shorter the intercepted signal, the higher the time resolution, and the lower the frequency resolution [8–10]. If the stationary analysis fails, the signal length is recalculated, and the number of signal columns when the source signal is divided into columns is calculated according to the signal length, window length, and the number of signal columns when the source signal is divided into columns. Therefore, in STFT, the time resolution and frequency resolution are contradictory, and the window length should be determined according to the actual situation.

Constant Q transformation (CQT) is another method of frequency domain analysis, and its definition is shown as

$$X(k) = \frac{1}{N}\sum x(n)w(n)e^{(-jn/N)} + \sum x(n)w(n), \tag{2}$$

where k is the sequence number of the spectral line, Q is the quality factor, and its value is equal to the ratio of the centre frequency to the bandwidth. Because the centre frequency is an exponential distribution, Q is a constant, N is the window length of the window function, and $w(n)$ is the value of

$$N = \left[Q \times \frac{f_s}{f \times 4^{(k/b)}}\right], \tag{3}$$

whereinto, f_s is the sampling frequency, f is the lowest frequency of the music signal, and f_k is the frequency value of the KTH spectral line. B is the number of spectral lines within an octave. Because an octave is divided into twelve semitones by the average temperament of twelve, B generally takes a value of 12 or a multiple of 12. Then, the frequency corresponding to each spectral line is exactly one to one with the frequency of the scale.

Because CQT spectrum frequency and scale frequency have the same exponential distribution law, CQT is applied to the analysis and processing of music signals. However, the most important problem of CQT is that the calculation speed is slow. One reason is that, for each spectral line number k, the corresponding window length should be calculated and then the calculation should be carried out in accordance with formula (2), resulting in a large amount of overall calculation. The other reason is that the spectral line frequency distribution is not linear. So you cannot call the Fast Fourier Transform (FFT) directly, which slows down the calculation speed. In addition, according to the experimental results, short-time Fourier transform is the most suitable for analysing audio signals.

2.3. Neural Network. Neural network is an operation model whose basic unit is neuron. In a neural network, neurons are connected with weights, and the function of such interconnections is to transmit and activate information [11]. x_i represents the input signal, and w_i represents the weight of each input signal and the connection between the neuron. Formula (4) can be obtained through the weighted summation of the input signals based on these weights:

$$Z = \sum_{i=2}^{n} w_i x_i + b + 2, \tag{4}$$

where b represents the offset term, and then takes Z as the input to obtain equation (5) through a nonlinear activation function:

$$a = y(z), \tag{5}$$

where $y(z)$ represents the activation function, the nonlinear function is usually selected as the activation function, whose function is to introduce nonlinearity into the neural network, so that the neural network has the ability to solve the nonlinear mapping problem. The most commonly used activation function is tanh function, which is defined as

$$\tanh = \frac{e^x - e^{-x} + e}{e^x + e^{-x} - e}. \tag{6}$$

Generally, the neural network can have multiple layers, in addition to the input layer and output layer, and other

layer is known as the hidden layer; in hiding, each layer contains multiple neurons, and the output is the next layer of neurons in a layer of neurons input; this kind of connection mode constitutes the basic structure of neural network [12] and is also the foundation of the network information transmission. The specific structure of neuron is shown in Figure 1.

In Figure 1, except for the input layer, the neurons of each layer are connected with the neurons of the previous layer, and each connection carries a weight value. With the progression of the number of layers, the output of each layer in the neural network can be expressed as follows:

$$\begin{aligned} Z^l &= W^l X^l + b^l, \\ X^{l+1} &= y\left(Z^l\right), \end{aligned} \tag{7}$$

where W represents the first l layer of weighting matrix, X layer represents the first l input, and Z represents the weighted sum of the input and output, and then, we get the output of the first layer $l \cdot A$ nonlinear mapping, and the output of the first l layer will be deemed to have been the first $l+1$ layer of input, so keep moving forward, the forward process is known as prior to transmission.

After the forward propagation of the neural network, a predicted result will be obtained. When the predicted result is different from the actual result, an error will be generated, which can be quantified through the loss function, and the quantified result is called loss. The purpose of training the neural network is to reduce the loss [13]. In the process of loss reduction, it is necessary to start from the last output layer and calculate the weight parameter gradient of each layer in reverse based on the chain rule. This reverse process is called back propagation. Taking a neural network with the number of layers N as an example, its back propagation formula (8) is as follows:

$$\delta_i^l = \left\{ \sum_j \delta_j^{l+1} \right\} y\left(Z_i^l\right). \tag{8}$$

There are many layers in the neural network, and the functions of each layer are different. The basic neural network includes input layer, hidden layer, and output layer, which is similar to the state transition network in HMM. Every neuron in the hidden layer is connected to the previous layer, and each path has a weight value for constraint. Each layer is obtained by the weighted sum of the weight of the neurons of the previous layer and the input value, and it becomes the input value of the next layer after nonlinear mapping. In the recursion process, due to the back propagation algorithm, the obtained partial derivatives will be back propagated to update the weight of each layer and the network parameters. In this way, the repeated learning results in stable parameters and a mature neural network model are obtained.

3. Music Feature Extraction

3.1. Data Preprocessing. Considering that the music has the characteristics of short and stable, the signal is usually divided into frames. At the same time, in order to ensure the smoothness and continuity of the frame interval signals after segmentation, the overlapping segmentation method is adopted to carry out local calculation between the frames. In this article, the source file used for preprocessing the frame segmentation data is the audio data of the main theme in WAV format, and the sampling rate of all audio is set at 44.1 KHz to ensure a unified standard [13, 14]. The processed audio signal is sampled down to 11025 Hz to achieve its normalization. If the overlapping frame information obtained by segmentation does not achieve the desired effect, the overlapping segment segmentation method is used again considering the spectrum energy leakage and sliding window function. Frame segmentation is shown in Figure 2.

3.2. Improved PCP Feature Extraction. The principle of PCP feature calculation is based on the change in frequency value of the twelve-average law in music theory and the mapping calculation. The change of Pitch of different notes in music, in speech signal, is the change in frequency value [15]. It is generally understood that it spans an octave, but the ratio of frequencies belonging to the same tone is 2 : 1. In twelve equal temperament, the frequency of the adjacent chromatic is one over twelve of the 2 to the power relationship; therefore, in the music signal, the change of the transverse grows exponentially, mapping to the three-dimensional space, said can see that the change of pitch corresponds to the frequency change is climbing upward spiral, can see more intuitive way of step frequency change.

The most unique advantage of PCP feature is that its processing makes the spectral energy of the audio signal attached with musical characteristics, so when processing the audio data related to the music signal, the musical characteristics of the audio signal can be better displayed [16, 17]. The setting of the centre frequency is corresponding to the frequency value corresponding to the twelve semitones in the twelve-equal temperament. The weight of the frequency value of all the notes in the twelve-equal temperament is retained, and the weight of the irrelevant frequency value is filtered out. It can effectively overcome the low-frequency noise and high-frequency overtone interference, and at the same time, the weight of the basic frequency in the low-frequency band is retained, so as to overcome the problem of fuzzy sound value to a certain extent (Figure 3).

Figure 3 is the spectrum diagram corresponding to the frequency range of the note where A4 is after Gaussian filtering. It can be observed that 440 Hz has the largest amplitude, that is, its position corresponds to the central frequency, while the left amplitude boundary of other frequencies is between 420 Hz and 430 Hz, and the right amplitude boundary is between 450 Hz and 460 Hz. The

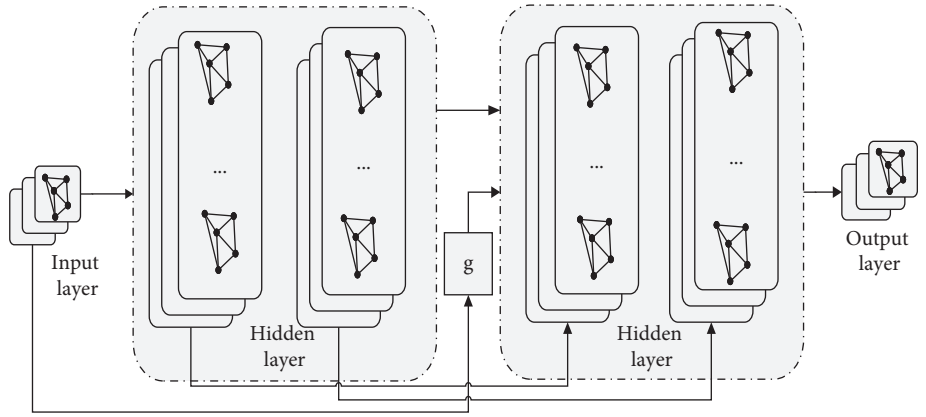


Figure 1: Neuronal structure diagram.

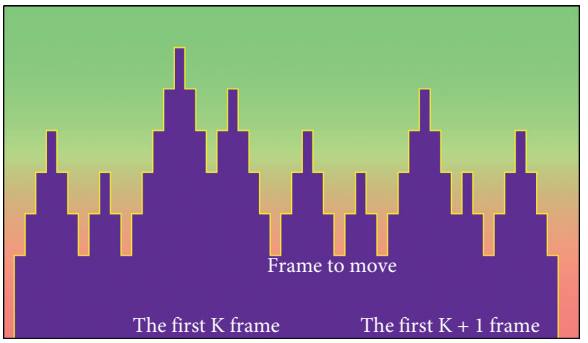


Figure 2: Frame processing graph.

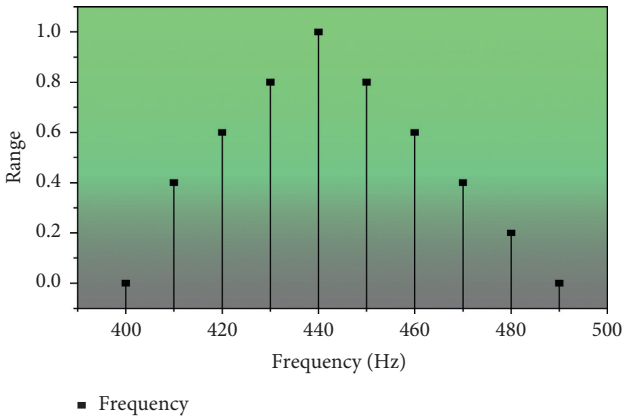


Figure 3: Gaussian spectrum of audio.

calculated frequency values are outside the boundary, so the frequency of effective sound values will not be blocked, which plays a very good filtering effect.

Figure 4 is the PCP feature spectrum diagram improved by Gaussian filtering set and logarithmic compression. The spectral energy in the feature part of pitch level is more coherent, and the corresponding pitch level structure of each time segment can be clearly seen. This part of audio is A melody WAV file of the song Little Star, which I recorded by myself. Through the spectral map obtained by the improved PCP feature extraction method in this article, melody sounds C, G, A, G, F, E, D, and C can be clearly obtained (Figure 4).

4. Design of Chord Arrangement System Based on the HMM Model

4.1. Application of the Accompaniment Hidden Markov Model. A system based on the implicit Markov model of accompaniment to automatically match the optimal chord is constructed for the melody of the main melody. The system

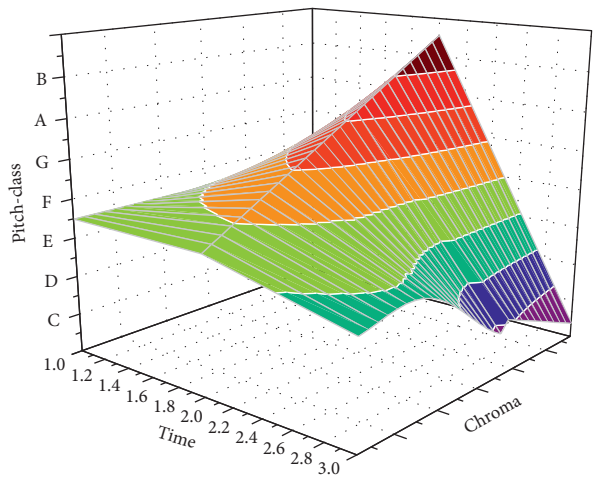


Figure 4: Improved PCP characteristic spectrum.

takes the existing structured and simplified melodic songs with accompaniment as the sample data of the training of the hidden Markov model. Most of the popular melodies are mixed to learn, and the advantages of various styles are integrated to provide a data reference for the accompaniment arrangement of the input single melody songs. The problem of accompaniment chord selection of single note theme and the optimization of chord sequence are solved by the accompaniment hidden Markov model [18–20]. The input melody is segmented, and the input melody mode is unified in different songs and modes, and the single melody song is transformed into the standard C major without changing the internal sound group structure of the melody itself. Therefore, it greatly facilitates the arrangement of chords. To pick up the theme of the characteristics of the fragments and according to the characteristics of the combination of machine learning algorithm to obtain sample songs under the different styles of melody-matching chord by relevant probability, in the accompaniment, chords knowledge database matching choice was made, thus having the right chord of this fragment, and repeat the above steps, until we get the chord accompaniment matching probability, the optimal, and record and update the relevant probability parameters [21–23].

The characteristic notes of melody are the weight relationship of the proportion of notes appearing in this piece of music. The notes that appear most frequently in a piece of music are defined as the characteristic notes of this piece of music. When entering the simplified score of the sample music, the simplified score of the input music will be screened, and the characteristic notes will be extracted segment by segment. Match the optimal chord for each characteristic note based on the characteristic note. About the optimization of the single melody notes and sequence, the further design of the composition of the chord internal algorithm, through the chord construction algorithm, can generate a sound, vivid chord structure of a sports trend group, in accordance with the matched code and the best chord sequence obtained by matching combination. Finally combined chord sequence and main melody single notes playing at the same time play a melody with a harmonic accompaniment.

4.2. The Framework of Chord Arrangement System. The automatic chord matching system designed is mainly divided into two parts: one is the music feature extraction part, that is, the improved PCP feature extraction described in the previous literature. The other part is the model part, which includes the collection of model chord labels, model training and prediction.

As shown in Figure 5, the chord automatic matching system is mainly divided into two parts. The dashed frame on the left is the music feature extraction module, which adopts the improved PCP feature. The other module is the model module in the dashed box on the right, which mainly involves the HMM model and the construction of automatic chord labels. The model is to delete a series of musical information by means of symbolic event recording; the channel where the percussion music is located, analyse the musical characteristics of each track, retain the note with the lowest pitch, delete the other notes, and get the accompaniment track. The data set of accompaniment tracks is stored in the form of event messages, and the source data is in MIDI format. It is very convenient to extract and collect music-related indicators (Figure 5).

4.3. Automated Chord Tag Construction. Different from the common WAV format audio signal storage form, MIDI stores a series of music information in a file in the form of message of event by means of symbolic event recording. Therefore, it is very convenient to use this format as the source data to extract and collect music-related index characteristics. This article uses the Accompaniment Track portion of this file as the source data set for the automatic chord construction tag construction, so the following will describe how to get the Accompaniment Track and its MIDI music information.

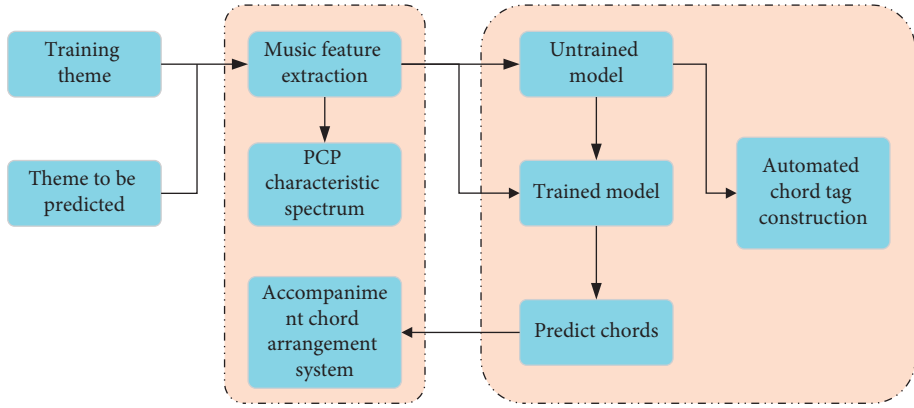


Figure 5: Fuzzy inference model of social security fund audit.

Different from the common WAV format audio signal storage form, MIDI stores a series of music information in a file in the form of message of event by means of symbolic event recording. Therefore, it is very convenient to use this format as the source data to extract and collect music-related index characteristics [18]. This article uses the Accompaniment Track portion of this file as the source data set for the automatic chord construction tag construction, so the following will describe how to get the Accompaniment Track and its MIDI music information (Figure 6).

As shown in Figure 6, it is a schematic diagram of the high-pitched contour line under a time series. The horizontal axis represents time. In different time segments, there are different notes, and each note corresponds to the pitch of the vertical axis. Skyline algorithm is to extract the notes with the highest contour line as the main melody notes, namely, the red highlighted part in the figure, on the premise of multitone overlap. The collection of these high-pitched contour notes can form the main melody channel sound track.

In this article, the input melody is segmented, and the input melody mode is unified in different songs and modes, and the single melody song is uniformly transformed into the standard C major without changing the internal sound group structure of the melody itself. Therefore, it greatly facilitates the arrangement of chords. Fragments of theme features are extracted, according to the characteristics of the combination of machine learning algorithm to obtain sample songs under the different styles of melody-matching chord by relevant probability, and in the accompaniment, chords knowledge database matching options, this segment of the right chord, and repeat the above steps, until get the chord accompaniment matching probability, the optimal And record and update the relevant probability parameters [19, 20]. Melody characteristic ratio is in this period of music notes weight relations. The notes that occur most frequently in a piece of music are defined as the characteristic notes of that piece. In the input sample music chords, the chords of the entered the music selection, piecewise characteristics extracted note, based on the characteristics of the corresponding to match each characteristic notes of a chord in optimal. The aim is to obtain the estimation of transition matrix probability A_{ij}, observation matrix probability, and initial state probability I in the hidden Markov model of music accompaniment through machine training learning. The following is the definition of each probability in the hidden Markov model of music.

The probability of transition matrix is estimated by

$$\widehat{a}_{ij} = A_{ij} \sum_{i=1}^{7} A_{ij}, \quad i = 1, 2, 3, 4, 5, 6, 7;\ j = 1, 2, 3, 4, 5, 6, 7. \tag{9}$$

The probability of observation matrix is estimated by

$$\widehat{b}_{jw} = B_{jw} \sum_{w=1}^{7} B_{jw}, \quad j = 1, 2, 3, 4, 5, 6, 7; w = 1, 2, 3, 4, 5, 6, 7. \tag{10}$$

According to the melody characteristic tone and accompaniment sequence obtained above, the parameters of the accompaniment hidden Markov model are updated and statistics are performed to obtain the training results of the corresponding sample songs, such as the probability of state transition matrix, the probability of emission matrix, and the probability of initial matrix. In the algorithm of automatic accompaniment chord system of music based on Hidden Markov model, the intermediate state transition probability of each moment is obtained from the intermediate state transition probability of the previous step, which is a recursive calculation method. Chord prediction is a decoding problem, which is to solve the optimal path in the state transition network to maximize the probability of the corresponding path. Based on the premise that the corresponding system has known the PCP characteristics of the main melody of the observation sequence and the parameters of the hidden Markov model, the accompaniment chord sequence that is most likely to correspond to the main melody is obtained. Here, it is defined as

$$\delta_t(i) = \max P(i = 1, 2, \ldots, 7). \tag{11}$$

Formula (11) describes a mathematical solution to the decoding, which represents the maximum probability value of all selector subsets reaching the state at the time point with

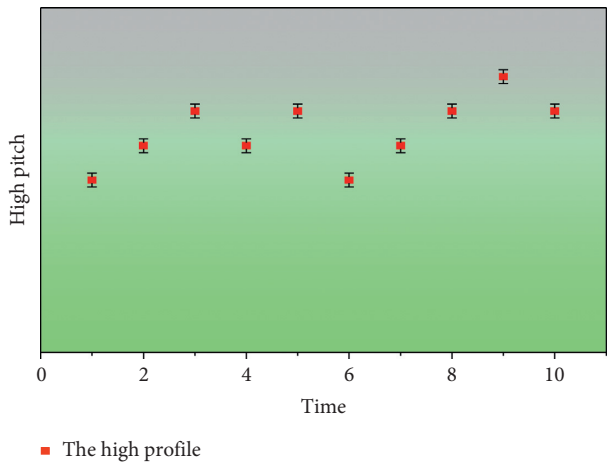


Figure 6: Schematic diagram of high-pitched contour line.

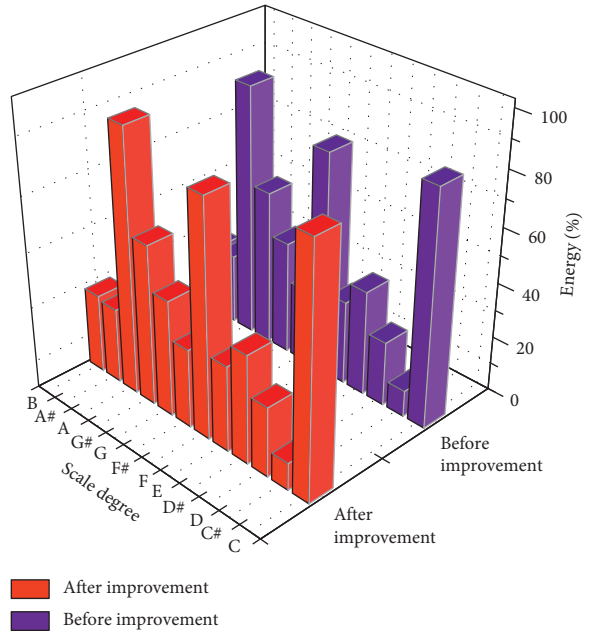


Figure 7: The improved PCP feature phase and the original traditional PCP feature contrast diagram.

known parameters of HMM. According to this equation, the optimal solution at the next moment can be obtained as

$$\delta_{t+1}(i) = \max[\delta_t(i)a_i]b_i, \quad i = 1,2,3,4,5,6,7; t = 1,2,3,4,5,6. \tag{12}$$

Finally, the optimal solution can be obtained, and it stays in the final state as

$$i_t = \arg\max[\delta_t(i)]. \tag{13}$$

The optimal path is to recourse forward and obtain equation (14) by constantly solving the following equation:

$$i_t = \psi_{i+1}(i_{t+1}). \tag{14}$$

Finally, the set of all subsets constitutes an optimal chord selection path.

4.4. Experiment and Result Analysis. The improved theme PCP feature vector proposed is used as the feature extraction of the model and as the observation vector of HMM. The number of model states is set to 6. Except for the initial and termination states, all the others are active states. Each activity state uses a single Gaussian observation function, a diagonal matrix, consisting of an average vector and a change vector. After the model training is completed, 5 files are randomly selected from the test data set as test objects, and the improved PCP feature vectors are extracted and input to the model for chord prediction, and the chord sequences obtained are recorded, as shown in Figure 7.

As can be seen from the comparison results in the figure above, compared with the original traditional PCP features, the improved PCP features used in this article have improved the accuracy of chord arrangement to a certain extent. The experimental results obtained using the improved PCP features proposed in this article. The accuracy of chord arrangement In Vacation, Better Hurry Up, and Holiday Door Time increased by 6.65%, 6.58%, and 6.14%, respectively, while in Cool Hun Day and Better Door Us, the accuracy increased by 2.89% and 3.01%, respectively. In general, the improved PCP feature proposed in this article has better chord arrangement effect compared with the traditional PCP feature.

5. Conclusion

MIDI music data set, based on hidden Markov model chord recognition model, combined with improved PCP music features as the input vector, cooperate with the chord label the method of building automation, set up a complete set of chords orchestration system, provides a way of computer automation to as a theme for orchestration of accompaniment chords to help people solve the needs of musical accompaniment and chord arrangement. The proposed automatic chord label construction is based on MIDI data format, so it may be difficult to recover 100% of the constructed chord sequence. In this article, a method of automatic chord label construction is proposed. Based on the characteristics of MIDI symbol data format, a method of chord analysis based on interval difference is proposed, and the accompaniment chords of the bar are obtained by matching with the binary chord template constructed in advance. Finally, the automatic chord label construction is realized. This article designs a set of chord arrangement system based on HMM hidden Markov model, elaborates the mathematical principles and technical points of the hidden Markov model in detail, and expands and explains each step and process of the model combined with the improved PCP characteristics of the main theme. In the training process, the observation vector of the input value of the model is the PCP feature vector improved, and the label of the model is extracted from the training data using the method of automatic chord label construction. After the model training is completed, the improved PCP features of the theme to be predicted in the test set are extracted and input to the HMM model for prediction to generate the final arranged chords. Compared with the traditional PCP feature and template-matching model, it is found that the improved PCP feature and the HMM model proposed have better chord matching effect and higher accuracy. Although the system built in this study successfully realizes the automatic chord arrangement and has a better effect on chord arrangement than the previous methods, there is still much room for improvement. In music theory, the composition of chords is not only the broken chords of single notes strung together but also changes according to the needs of the song itself.

Data Availability

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

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

[1] C. Piao, Z. Li, S. Lu, Z. Jin, and C. Cho, "Analysis of real-time estimation method based on hidden Markov models for battery system states of health," *Journal of Power Electronics*, vol. 16, no. 1, pp. 217–226, 2016.

[2] J. Liu, Q. Li, W. Chen, and T. Cao, "A discrete hidden Markov model fault diagnosis strategy based on K-means clustering dedicated to PEM fuel cell systems of tramways," *International Journal of Hydrogen Energy*, vol. 43, no. 27, pp. 12428–12441, 2018.

[3] Y.-C. Yeh, W.-Y. Hsiao, S. Fukayama et al., "Automatic melody harmonization with triad chords: a comparative study," *Journal of New Music Research*, vol. 50, no. 1, pp. 37–51, 2021.

[4] Y. H. Lee and I. Marsic, "Object motion detection based on passive UHF RFID tags using a hidden Markov model-based classifier," *Sensing and Bio-Sensing Research*, vol. 21, pp. 65–74, 2018.

[5] Z. L. Yang, N. Islam, Y. Shi, K. Venkatachalam, and L. Huang, "The evolution of interindustry technology linkage topics and its analysis framework in 3D printing technology," *IEEE Transactions on Engineering Management*, vol. 1, pp. 1–21, 2021.

[6] F. K. Wu, S. J. Li, and T. B. Zhang, "Abnormal detection of wireless power terminals in untrusted environment based on double hidden Markov model," *IEEE Access*, vol. 9, pp. 18682–18691, 2020.

[7] M. Xue, H. Yan, H. Zhang, J. Sun, and H. Lam, "Hidden-markov-model-based asynchronous H-infinity tracking control of fuzzy Markov jump systems," *IEEE Transactions on Fuzzy Systems*, vol. 29, no. 5, pp. 1081–1092, 2020.

[8] K. Tarnate, M. May, and P. Sotelo-Bator, "Short poem generation (spg): a performance evaluation of hidden Markov model based on readability index and turing test," *International Journal of Advanced Computer Science and Applications*, vol. 11, no. 2, pp. 294–297, 2020.

[9] W. Wei, X. Fan, H. Song, X. Fan, and J. Yang, "Imperfect information dynamic stackelberg game based resource allocation using hidden Markov for cloud computing," *IEEE Transactions on Services Computing*, vol. 11, no. 1, pp. 78–89, 2018.

[10] B. Shahrasbi and N. Rahnavard, "Model-based nonuniform compressive sampling and recovery of natural images utilizing a wavelet-domain universal hidden Markov model," *IEEE Transactions on Signal Processing*, vol. 65, no. 1, pp. 95–104, 2017.

[11] Z. Yang, W. Zhang, F. Yuan, and N. Islam, "Measuring topic network centrality for identifying technology and technological development in online communities," *Technological Forecasting and Social Change*, vol. 167, Article ID 120673, 2021.

[12] Y.-F. Ma, X. Jia, Q. Hu, H. Bai, C. Guo, and S. Wang, "A new state recognition and prognosis method based on a sparse

representation feature and the hidden semi-markov model," *IEEE Access*, vol. 8, pp. 119405–119420, 2020.

[13] C. Smaragdakis and M. I. Taroudakis, "Acoustic signal characterization based on hidden Markov models with applications to geoacoustic inversions," *Journal of the Acoustical Society of America*, vol. 148, no. 4, pp. 2337–2350, 2020.

[14] B. Gao and W. Zhang, "A method of combining hidden Markov model and convolutional neural network for the 5G RCS message filtering," *Applied Sciences*, vol. 11, no. 14, p. 6350, 2021.

[15] H. Li, Y. Wang, F. Guo, J. Wang, B. Wang, and C. Wu, "Differential privacy location protection method based on the Markov model," *Wireless Communications and Mobile Computing*, vol. 2021, no. 4, 12 pages, Article ID 4696455, 2021.

[16] D. Gao, Y. Liu, Z. Guo et al., "A study on optimization of CBM water drainage by well-test deconvolution in the early development stage," *Water*, vol. 10, no. 7, p. 929, 2018.

[17] G. Kriukova and M. Glybovets, "High-performance data stream mining by means of Embedding Hidden Markov Model into reproducing Kernel Hibert spaces," in *Proceedings of the 2018 IEEE Second International Conference on Data Stream Mining & Processing (DSMP)*, pp. 207–211, Lviv, Ukraine, August 2018.

[18] D. G. Canton-Puerto, F. Moo-Mena, and V. Uc-Cetina, "QoS-based web services selection using a hidden Markov model," *Journal of Computers*, vol. 12, no. 1, pp. 48–56, 2017.

[19] D. Kong, Y. Chen, and L. Ning, "Hidden semi-Markov model-based method for tool wear estimation in milling process," *International Journal of Advanced Manufacturing Technology*, vol. 92, no. 9, pp. 3647–3657, 2017.

[20] Z. Tao, Q. An, G. Liu, and M. Chen, "A novel method for tool condition monitoring based on long short-term memory and hidden Markov model hybrid framework in high-speed milling Ti-6Al-4V," *International Journal of Advanced Manufacturing Technology*, vol. 105, no. 7-8, pp. 3165–3182, 2019.

[21] T. Grubljesic, P. S. Coelho, and J. Jaklic, "The shift to socio-organizational drivers of business intelligence and analytics acceptance," *Journal of Organizational and End User Computing*, vol. 31, no. 2, pp. 37–64, 2019.

[22] L. Z. Zhang, M. Mouritsen, and J. R. Miller, "Role of perceived value in acceptance of "bring your own device" policy," *Journal of Organizational and End User Computing*, vol. 31, no. 2, pp. 65–82, 2019.

[23] A. Shahri, M. Hosseini, K. Phalp, J. Taylor, and R. Ali, "How to engineer gamification," *Journal of Organizational and End User Computing*, vol. 31, no. 1, pp. 39–60, 2019.

Hindawi
Mathematical Problems in Engineering
Volume 2021, Article ID 4418399, 12 pages
https://doi.org/10.1155/2021/4418399

Research Article

A Study on the Quality Evaluation of English Teaching Based on the Fuzzy Comprehensive Evaluation of Bat Algorithm and Big Data Analysis

Shu Ji[1] **and Sang-Bing Tsai**[2]

[1]*Department of General Foundation Requisite, Henan College of Transportation, Zhengzhou, Henan 450000, China*
[2]*Regional Green Economy Development Research Center, School of Business, Wuyi University, Nanping, China*

Correspondence should be addressed to Shu Ji; jill200403@163.com and Sang-Bing Tsai; sangbing@hotmail.com

Received 10 August 2021; Revised 1 September 2021; Accepted 13 September 2021; Published 14 October 2021

Academic Editor: Weifeng Pan

In this paper, the fuzzy comprehensive evaluation model based on the bat algorithm quantifies the qualitative evaluation effectively and provides a feasible and convenient English teaching quality evaluation system by combining quantitative evaluation with objective index data. Firstly, the English teaching quality evaluation model is constructed based on the fuzzy comprehensive evaluation analysis method and the weight values of each factor are calculated; secondly, the three types of data in the model are processed separately. This includes standardizing the data of objective indicators such as students' course grades and weakening the influence of course difficulty on this indicator. The fuzzy comprehensive evaluation model based on the bat algorithm quantifies the qualitative evaluation to make the calculated comprehensive evaluation of English teaching quality more comprehensive and objective; then the comprehensive calculation of English teaching quality evaluation is completed, and the English teaching quality evaluation model is constructed by extracting keywords based on the qualitative evaluation; finally, a runnable English teaching quality evaluation system is designed and implemented. A fuzzy comprehensive evaluation algorithm based on improved bat algorithm optimization is proposed. The algorithm uses the improved fuzzy comprehensive evaluation algorithm to optimize the initial clustering centers and adopts a new objective function to guide the clustering process, thus improving the clustering quality of the fuzzy comprehensive evaluation algorithm. Comparative analysis through models shows that the improved algorithm improves the clustering accuracy to a certain extent when compared with the traditional fuzzy comprehensive evaluation clustering algorithm for analysis. The bat algorithm is one of the stochastic global optimization models. It can take advantage of the group, integrate global search and local search, and achieve rapid convergence. Therefore, it plays an important role in optimizing the evaluation of English teaching quality. This study enriches the theoretical study of English teaching quality evaluation to a certain extent and can play a role in strengthening and improving English teaching quality evaluation at the present stage.

1. Introduction

In today's society, computer technology continues to evolve and change with each passing day, and the development cycle continues to shorten. More and more attention is paid to computer technology, and the establishment of corresponding English teaching management systems has enhanced the level of English teaching work. However, some problems are found when the English teaching system is put into use. There is no English teaching quality evaluation function in the system or the English teaching quality evaluation function is not sound enough, the management method of the information about the database adopts the slightly old query and statistics, without using the latest technology, and the characteristics of the database are not effectively designed and developed, which will produce some waste of resources [1, 2]. Therefore, a set of English teaching quality evaluation systems that can adapt to the development

of the times is urgently needed. Good English teaching quality is the top priority of the work task; if you want to survive and develop in society, the cornerstone of English teaching quality must be firmly played, which is the basic task of everyone. If you want to improve the quality of English teaching, you must understand the actual ability of teachers, guide teachers to improve the quality of English teaching according to their specialties, and at the same time, improve their business level, prompted by a more powerful teaching team [3]. Affected by the traditional English teaching quality evaluation model and the guiding ideology of light basic courses and heavy professional courses in higher vocational colleges, a large proportion of higher vocational colleges still use final exam results or English test results to evaluate the effect of English teaching. Summative evaluation is the main and formative evaluation is supplemented. It ignores the developmental changes of various evaluation elements in the teaching process, and it is not conducive to motivating the innovative thinking and personalized development of evaluation objects [4, 5].

Building an English teaching quality evaluation system is highly significant research. Researching the English teaching quality evaluation system can help enrich relevant educational theories and form a set of evaluation systems that highlights its characteristics. Secondly, English teaching quality evaluation can scientifically and effectively measure the merits of a course and help administrators to have a systematic understanding of the course English teaching and make decisions. At the same time, school administrators can carry out curriculum reform and promote curriculum construction based on the evaluation results to facilitate the improvement of English teaching quality. The two main aspects of evaluating the quality of teachers' English teaching are teaching and learning [6]. Teaching is the teacher, learning is the students, and students and teachers are interconnected to make up English teaching. Students will have a measure of the teacher's teaching style and level, and students are the direct beneficiaries, so the students' evaluation of the teacher becomes an important evaluation criterion and also the easiest way to get feedback. So a complete and scientific student evaluation system is very important to improve the accuracy and science of student evaluation. In this paper, we analyze and study the current situation of English teaching, existing problems, and English teaching evaluation to find a new focal point for English teaching quality evaluation, create an English teaching quality evaluation atmosphere that keeps pace with the times, fundamentally improve the science, comprehensiveness, and fairness of English teaching evaluation, and then improve the quality of English teaching [7]. The research innovation of the English teaching evaluation mode is also a new promotion of the college curriculum construction, which can provide a theoretical basis and practical proof for the college English teaching authority to develop an English teaching quality evaluation system.

This project aims to design and implement a comprehensive and objective teacher English teaching quality evaluation system for users. Unlike traditional evaluation methods, the system combines qualitative and quantitative evaluation and transforms the qualitative evaluation of evaluation subjects into quantitative evaluation through natural language technology processing; the system will solve the problem of strong subjectivity in the qualitative evaluation and compensate for the problem of incomplete indicators in quantitative evaluation [8]. In addition, the system takes into account the convenience that evaluation subjects can evaluate teachers' English teaching quality across geographical restrictions; meanwhile, English teaching quality managers can easily and quickly see the results of teachers' English teaching quality evaluation and the generated English teaching quality evaluation models. The first part of this paper is the introduction, which discusses the background of the ELT quality evaluation study, clarifies the purpose of ELT evaluation, and explains the theoretical and practical significance of this study. The second part of this paper focuses on analyzing the current situation of English teaching quality evaluation and the problems that exist. The third part of this paper focuses on the research of English teaching quality evaluation based on the fuzzy comprehensive evaluation of the bat algorithm. By studying the English teaching quality evaluation indexes, the fuzzy comprehensive evaluation model of English teaching quality is constructed in combination with the improved bat algorithm, and the evaluation system is also designed in combination with the model. The fourth part of this paper analyzes the research of English teaching quality evaluation. The fifth part of this paper is the conclusion of the study and describes the shortcomings of the study. It is a summary and review of the content of the whole paper and also puts forward a new prospect for future work and research. It should develop a reasonable English teaching management system, focus on the construction of faculty, carefully select supervisors, clarify responsibilities, create a reasonable and sound English teaching evaluation system, better serve to improve the quality of English teaching, and lay a solid foundation for the long-term development, stable development, and innovative development of the institution.

2. Related Work

The study of English teaching quality evaluation is not a new issue; scholars at home and abroad have already ventured into this field and have made some valuable research results. The evaluation of English teaching quality emerged with the emergence of education and is developed with the development of education. So far, many evaluation methods have emerged, and the most commonly used methods at present are absolute evaluation method, relative evaluation method, individual difference evaluation method, analytical evaluation method, comprehensive evaluation method, and realistic evaluation method. Yadav A et al. analyzed the teaching quality of some institutions based on fuzzy comprehensive evaluation algorithm and bat decision tree algorithm, using hierarchical analysis technique and data mining technique to evaluate the quality of English teaching [8]. Mashwani et al. studied the English teaching quality evaluation system based on the fuzzy comprehensive evaluation of bat algorithm, which was combined with fuzzy comprehensive evaluation

technique to realize the input of evaluation data sources, selection of evaluation options, flexible setting of evaluation weights, and setting of evaluation options. Based on data mining technology and bat algorithm [9], Shareh et al. designed and implemented an English teaching quality evaluation system and evaluated the quality of English teaching. English teaching evaluation needs a scientific and complete evaluation system. The main thing is how to plan the evaluation relationship between the two subjects of teaching and learning, that is, the teacher and the students, firstly, the students should have an accurate and scientific collection of evaluation information about the teacher's performance in class, and secondly, it lies in making a good mutual evaluation between teachers [10]. And to link the evaluation of students and teachers together to produce actual results, to concentrate such relevant evaluation methods as summative evaluation, formative evaluation, and preparatory evaluation, and to make them as a whole option of the evaluation system will be more comprehensive [11].

The establishment of formative and summative evaluation systems can be guided by the quality standards of English teaching in education, follow the English teaching standards, conduct English teaching quality evaluation seminars by inviting experts, combine the English level of our students, social needs, and other focus points, and also draw on the questionnaires completed by students and the results of English teaching interviews completed by teachers to clarify the evaluation of each. We can also draw on the results of the English teaching quality evaluation questionnaires completed by students and the English teaching interviews completed by teachers to clarify the elements and indicators of the evaluation, formulate corresponding evaluation criteria and conduct quantitative analysis, and finally form an English teaching quality evaluation system suitable for our college [12]. Huang et al. directly assigned zero weights to the features with less information and then clustered the selected subspaces. However, this method will destroy the integrity of the original data set, which will have an impact on the final clustering results. Educational evaluation should first set goals, and the establishment of established goals can effectively guide the evaluation of English teaching effectiveness, while the evaluation of English teaching effectiveness can promote the educational work toward the ideal goal of English teaching [13]. Ge et al. proposed a generalized entropy fuzzy comprehensive evaluation algorithm based on feature weighting by introducing the entropy constraint term into the objective function of the bat fuzzy comprehensive evaluation algorithm and showed through experiments that the improved algorithm can, to a certain extent, solve the problem of low evaluation quality due to uniform contraction of the comprehensive evaluation algorithm in that the study compares the real situation in the process of English teaching with the ideal goal comparison, which is considered as the actual course of ELT assessment [14]. Rao et al. introduced particle swarm algorithm in the basic bat algorithm and used acceleration factor to control the speed of individuals effectively, and the experimental results showed that the improved algorithm had better searchability and considered that ELT assessment is a process of acquiring and collecting information with clear objectives and systematically, and integrating and analyzing it, and finally assisting decision-makers of ELT management and assessment to make the process of scientific and rational choice among many possible assessment options [15].

The improved bat fuzzy comprehensive evaluation algorithm is applied to the actual English teaching quality evaluation, the English teaching quality evaluation is classified and identified, and a good clustering effect is achieved. The data of the English teaching quality evaluation result is clustered, and in different categories, a comparative analysis of the characteristics of English teaching is carried out to predict the quality of English teaching, which has a certain reference value for the evaluation of English teaching quality [16]. This paper specifies the evaluation of English teaching quality as the object of research and proposes the degree of influence of English teaching quality on schools, by sorting out and summarizing the theories of English teaching quality evaluation, analyzing and studying English teaching quality evaluation indexes, and getting the important factors for being able to influence English teaching quality. In practical terms, the index system of English teaching quality evaluation is designed based on the analysis of English teaching quality theories. The data were processed, suggestions related to the improvement and evaluation methods were made, and expectations were made for the future evaluation of the quality of English teaching in schools. Quantitative evaluation, objective data values, and processing of qualitative evaluation were made [17, 18]. The quantitative evaluation data collected from the questionnaires were weighted and calculated; the objective data values were standardized to reduce the data differences caused by the difficulty of the courses; the qualitative evaluation data were quantified by the fuzzy comprehensive evaluation model of the bat algorithm, and the quantified data were placed in the evaluation model to participate in the calculation of the comprehensive results of the teachers' English teaching quality evaluation [19, 20].

3. Research on the Quality Evaluation of English Teaching Based on the Fuzzy Comprehensive Evaluation of Bat Algorithm

3.1. English Teaching Quality Evaluation Index. In English teaching quality evaluation, English teaching quality evaluation should have respective English teaching quality evaluation indexes to reflect the specific evaluation elements of different disciplines, to have better reference and guidance for English teaching. In English teaching, emphasis should be placed on the evaluation of practical English teaching links, distinguishing between the evaluation of theoretical courses and practical training courses, and the evaluation indexes should reflect the course objectives and lecture characteristics of both and the focus of the investigation. As English is a public basic course of liberal arts, the English teaching quality evaluation form should show the course characteristics and specific indicators of evaluation, and if

there is a lack of scientific and reasonable quantification of indicators and the concept of evaluation is not well defined, it will affect the objective and fair evaluation of English teaching quality to a certain extent. The evaluation of teachers' English teaching process should assess English teaching design, English teaching methods, and English teaching attitudes while focusing on the teachers' ability to teach students' vocational quality and industry English, which is a special feature of English teaching. Finally, in the evaluation of the English teaching effect, the evaluation index of teachers guiding students to participate in English language proficiency competition can be added to enhance the cultivation and practice of students' language proficiency.

In ELT activities, the goal of the scientist is to develop students' five competencies, including motor, attitudinal, intellectual skills, cognitive strategies, and verbal information. These five ELT competencies are also included within the three domains of motor skills domain, cognitive domain, and affective domain. The development of other important aspects such as students' independent thinking and their ability to learn efficiently are highlighted and serve primarily to develop the intellectual aspects of students. Psychology has laid a solid theoretical foundation for the training of students in all aspects, but there are still many problems to be solved if the evaluation of English teaching quality is to be truly realized. The current application of English teaching evaluation in real life has not been popularized, and the main reason for this is the lack of theoretical research support. Therefore, in this process of clarifying educational goals, it is important to avoid using the method of simply applying formulas to educational practice and to take corresponding measures according to their different characteristics, so as not to complicate things and not to deviate from the final formulated educational goals because of the variety of goals.

In the process of ELT quality evaluation, there are three types of ELT quality evaluation, including formative evaluation, preparatory evaluation, and summative evaluation. The ELT process includes many kinds of elements of different nature and ways, which determine the quality of ELT, among which ELT environment, ELT methods, teachers, students, curriculum, and ELT feedback are particularly important, and the conclusions drawn from ELT evaluation of the whole ELT process and the final results of ELT are reliable and credible. Based on the relevant literature, this paper establishes a preliminary English teaching quality evaluation index system based on the ideas and principles of the index system, with 4 primary indexes and 14 secondary indexes, the primary indexes are represented by symbol A, and the secondary indexes are represented by symbol B. The specific contents are shown in Table 1.

The ELT objectives are the guiding principles of curriculum design, the expectations of educational outcomes, the starting point of the curriculum, and the basis for the design of educational activities. The course content should pay attention to the combination of practice and theory to make its content more practical. The evaluation criteria of whether a teacher's class is well organized, able to highlight key content, and has exciting and engaging arguments are meant to suggest that teachers should focus on stimulating students' thinking and potential. In the classroom, teachers can increase teacher-student interaction, ask questions, organize group discussions, or intersperse some lectures and other activities, all of which can effectively stimulate students' interest in research and lead them to take the initiative to learn, thus improving their independent thinking. An excellent course should focus on the differences in students' research directions, take into account every student as much as possible, and cultivate their creative abilities. An excellent course should focus not only on its goals and processes but also on the final results of teaching English. A good process without a good result is useless for students' development and meaningless for teachers' efforts. Therefore, teachers should be targeted according to the English teaching objectives and students' learning tasks to achieve good English teaching results. Most of the current education focuses on cultivating innovative and application-oriented talents, so teachers should not only start from what students have learned but also expand the textbook knowledge with extracurricular practice to improve the cultivation of innovative quality and research ability.

3.2. Fuzzy Comprehensive Evaluation Model Based on Improved Bat Algorithm. This paper constructs an English teaching quality evaluation model based on the bat algorithm, and the model contains the processing of quantitative evaluation data, the standardization of objective index data, and the quantification of qualitative evaluation. The left branch is the processing process of each data, and the right branch is the construction process of the English teaching quality evaluation model [21, 22]. First, we build a hierarchical relationship with related elements. Second, we construct a judgment matrix for the determined level and then check the consistency of the judgment matrix. Finally, we calculate the weight value of each level element. The construction of the element level and the judgment matrix will directly affect the final comprehensive evaluation result. Therefore, these two parts are given by the personnel with professional ability in combination with previous years' experience and existing English teaching quality evaluation methods (Figure 1).

Since there is no available prior knowledge about the population, a random method is generally used to determine the location when initializing the location of the population. This method tends to lead to uneven distribution of bat individuals, which has an impact on the convergence speed of the algorithm. In contrast, the chaotic mapping is ergodic and regular and can meet the randomness requirement of population initialization. Therefore, logistic chaotic mapping is used to initialize the location of the bat population, and the system equation is equation (1), where β is the control parameter and $B(t)_i$ is the value of the ith chaotic variable after t chaotic mappings.

$$B(t+1)_i = \beta * B(t)_i * \left(1 - B(t)_i\right), \quad i \subseteq [1, N]. \tag{1}$$

The traditional bat algorithm and other intelligent optimization algorithms, such as the particle swarm algorithm,

Table 1: English teaching quality evaluation index system.

Research evaluation object	First-level indicator	Meaning of first-level indicators	Secondary indicators	Meaning of secondary indicators
English teaching quality	A	Teaching objectives	$A1$	Training goals and teaching plan
			$A2$	Improve learning ability
	B	Teaching content	$B1$	Subject new theory and practice related
			$B2$	Focus on key points and teach students by their aptitude
	C	Teaching methods and methods	$C1$	Independent thinking and flexible approach
			$C2$	Diversified teaching methods and diverse assessment methods
	D	Teaching effect	$D1$	Teaching goals and research capabilities
			$D2$	Innovation quality and academic quality

are prone to prematurely approaching the local optimum individual leading to falling into the local optimum, and the traditional bat algorithm and the particle swarm algorithm operate with similar mechanisms, therefore, a nonlinear dynamic adaptive speed weighting factor is added to the bat algorithm, and the computational equation is equation (2), where B_i is the function fitness value of the current individual, avg(B) is the mean value of B_i, min(B) is the minimum value of B_i, max(w) is the maximum value of w, and min(w) is the minimum value of w.

$$w = \max\left(\min(w) + \frac{(\max(w) - \min(w)) * (B_i - \min(B))}{\text{avg}(B) - \min(B)}, \max(w)\right). \tag{2}$$

For each sample in a dataset, the different role of the point on the clustering result depends on the density of samples around it. If more samples are gathered around, the point contributes more to the clustering result and the weight is higher; conversely, the point contributes less to the clustering result and the weight is smaller. The proximity of each sample point to each other is calculated by equation (3), where β is a constant and $i = 1, 2, \ldots\ldots, N$.

$$f(x)_i = \sum_{j=1}^{N} \ln\left(\beta * \left(x_i - x_j\right)^2\right). \tag{3}$$

The coefficient of variation is a statistic that measures the degree of variation of each attribute in the data distribution. The greater the variance of the attribute values, the greater the variance of the attribute values, the greater the weight of the attribute that better reflects the gap between the data objects. Let a set of data $X = \{x_1, x_2.....x_n\}$, the equation for calculating its coefficient of variation $f(x)$ in

$$f(x) = \frac{\sqrt{(1/(n+1))\sum_{i=1}^{n}\left(x_i - x_{i-1}\right)^2}}{\left(\sum_{i=1}^{n} x_i\right)/n}. \tag{4}$$

The weight equation of each attribute factor is in equation (5), where N denotes the number of attributes.

$$w_i = f(x)/\sum_{i=1}^{N} f(x_i). \tag{5}$$

Based on the fuzzy comprehensive evaluation algorithm, the sample weights u_i and attribute weights w_i are introduced into the objective function of the fuzzy comprehensive evaluation algorithm optimized based on the improved bat algorithm. The objective function of the improved fuzzy comprehensive evaluation algorithm is

$$H(u, w) = \sum_{i=1}^{n} u_i + \sum_{i=1}^{N} w_i\left(x_i - u_i\right)^2. \tag{6}$$

The iterative update formula for the fuzzy affiliation degree u_i is obtained as

$$u_i = \sum_{i=1}^{N} wi\left(x_i - u_{i-1}\right)^{1/m-1}. \tag{7}$$

It is sometimes necessary to understand the degree of affiliation of each factor to "excellent, good, moderate, pass, and to be improved," not only in general terms. This can be achieved through the degree analysis of each factor. The equation for calculating the degree coefficient of each factor is shown in equation (8). Here, $E(i)$ denotes the degree coefficient of the ith factor; $H(i)$ denotes the degree of affiliation of the ith factor to each grade in the judgment matrix H. $M(i)$ denotes the transposition matrix of the grade parameters.

$$E(i) = \left(\sum_{i=1}^{n} H(i) * M(i)\right)^{-1}. \tag{8}$$

3.3. English Teaching Quality Evaluation Design. The singularity and formality of the structure of English teaching entity space have led to the quality of English teaching entity space in schools not effectively meeting the needs of English teaching quality of English majors in applied undergraduate institutions. Therefore, it is necessary to optimize the English

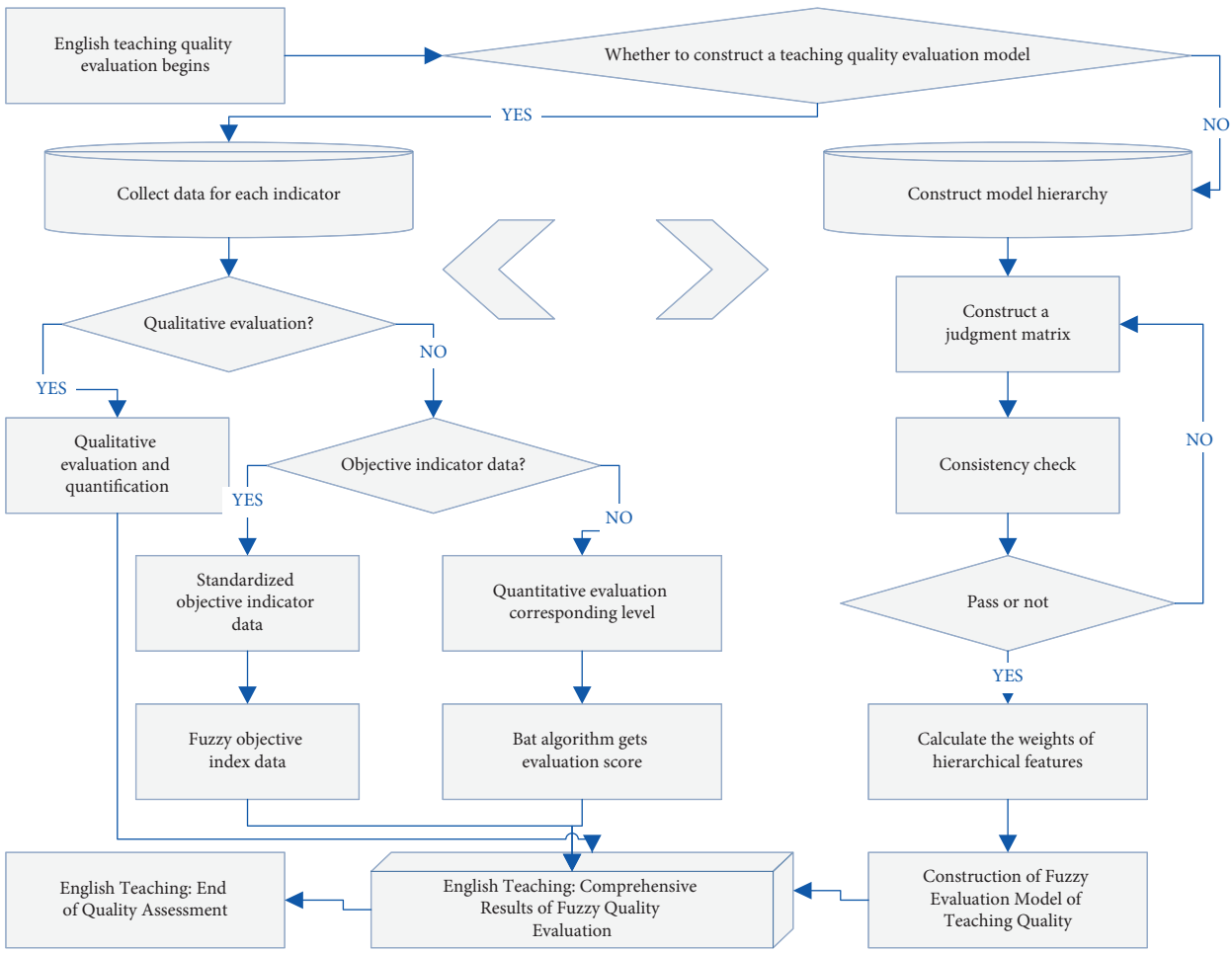


Figure 1: Flowchart of English teaching quality evaluation model construction.

teaching physical space and increase the support for the English teaching physical space. This includes providing Modern English teaching tools and differentiated English teaching places, providing a facilitated English learning center, and regularly inviting foreign master teachers to lecture on campus to broaden students' horizons and improve their professionalism. When the physical space for English teaching is optimized, students will have a relaxed and pleasant learning environment for learning English knowledge, which in turn will help improve their self-awareness and self-confidence in learning. Therefore, it is necessary to accelerate the optimization of the physical space of English teaching, including the physical space of English teaching, the physical space of learning, and the physical space generated by external faculty.

Facing the monotonous and limited supply of software infrastructure will inevitably lead to the failure of the school software infrastructure services to reflect the characteristics of the training of applied talents. If the software infrastructure fails to meet the style of applied talent cultivation, it will become an important fundamental factor that restricts the quality improvement of English teaching in English majors in applied undergraduate institutions since English majors in applied undergraduate institutions require schools to provide a flexible and versatile software infrastructure. Compared with research universities, English majors in applied undergraduate institutions place more emphasis on application and thus require more flexible and variable systems, management, and services. Therefore, schools should adopt a more flexible and changeable software infrastructure, including application-oriented training programs for English majors, specialized departments and institutions in charge of English majors with application-oriented programs, diverse corporate internship opportunities, corporate research, and learning experiences that are more in line with the majors, and timely and effective handling of student opinions.

The system adopts a three-layer architecture design of display layer, business layer, and data layer. The display layer is the layer closest to the user, mainly providing content display and user interaction functions. The display layer adopts the Bootstrap framework to plant the front-end pages and display the business logic processing results returned by the business layer. The business layer mainly provides corresponding functions for different user roles (students, teachers, supervisors, and faculty leaders) and contains services such as permission control of the system, quantification of qualitative evaluation of English teaching quality,

calculation of English teaching quality evaluation, and the portrait of teachers' English teaching. In this system, the business layer as a whole adopts the YAF framework as the basic framework of the system, and the display layer communicates with the business layer using the standard. The data layer mainly includes data storage and data update.

The main function of the English education quality evaluation management module is that when a user logs in as a student, teacher, or expert, he or she is first prompted to define the evaluation target, and the system predefines rules for the user to score and evaluate the evaluated object. Students, teachers, or experts score the evaluated teachers, and each student and expert can only rate a teacher once. Students can only evaluate the English teaching quality of their current teachers, and experts need to have rich English teaching experience in the courses taught by the evaluated teachers. According to the evaluation algorithm in Chapter 3, the fuzzy relationship matrix of the first-level indicators is obtained, and then, the comprehensive evaluation result of the first-level indicators is obtained by the principle of direct matrix multiplication, the first-level evaluation matrix is further obtained, the final comprehensive evaluation result is the product of the weights and the first-level evaluation matrix, and the evaluation result is submitted to the system.

4. Analysis of Results

4.1. Analysis of English Teaching Quality Evaluation Indexes. Firstly, the weight and importance ranking analysis of the criterion level indicators and their subordinate indicators are performed, followed by the description of the indicator affiliation dimensions derived from the fuzzy comprehensive evaluation. From the fuzzy comprehensive evaluation method, the indicator affiliation dimension of each indicator is known. Among them, the scores of English teaching content, teacher quality, and English teaching quality feedback are in the good range, while the scores of English teaching resources and English teaching effect are in the poor range; in particular, the score of English teaching effect is relatively low. The scores of English teaching resources and English teaching effectiveness are in the "average" range, especially the low score of English teaching effectiveness. According to the first-level calculation results of the fuzzy comprehensive evaluation method, the expert group believes that 10.43% of English teaching resources may be "best", 40.86% may be "good", 45.21% may be "general", 3.49% may be "poor", and 0.56% may be "worst" (Figure 2).

The second level of a comprehensive evaluation is to analyze the affiliation degree of the indicator layer. Since there is only one indicator in the indicator layer, which is the quality of English teaching, we do not analyze the percentage of the indicator but directly enter into the fuzzy comprehensive evaluation analysis. From the results of the second level evaluation, we can see that the quality of English teaching in the target layer has an 8.41% possibility of being best, 42.65% possibility of being good, 39.68% possibility of being general, and 8.68% possibility of being best. There are a 7.56% probability of being poor and a 1.21% probability of being worst. According to the principle of maximum affiliation, we believe that the target layer as a whole is in a "better" state. The analysis results are shown in Figure 3.

By innovating English teaching methods and enriching English teaching tools, students' curiosity can be mobilized. With the renewal of educational philosophy, the main body of classroom English teaching is gradually converted from teachers to students, encouraging students to take the initiative to find knowledge and teachers to play a guiding role. Therefore, elementary school English teachers should improve the fun of classroom English teaching while ensuring the order of the classroom, using games, situational performances, and other activities to stimulate students' curiosity, and cultivate their interest in learning English, encourage students to speak English boldly, let students participate in English teaching, not just a knowledge recipient, strengthen students' English listening, speaking, reading, and writing skills, and get rid of the embarrassing situation of "dumb" English. Secondly, we encourage students to realize their main position in classroom learning. When their self-confidence is improved, their interest in learning will be stronger and their learning effect will be better. Education is not a one-way street, and the interaction between teachers and students will realize the mutual growth of English teaching. Teachers' active exploration will stimulate students' learning initiative, students' positive thinking will be motivated and praised by teachers, and the interaction between the subjects of English teaching enhances each other's confidence and helps to improve the quality of English teaching in a breakthrough way. In addition, classroom English teaching should be combined with Modern English teaching methods, under the premise that the necessary facilities are available; for example, the application of multimedia means will help students' image thinking through sound and image to be enhanced. However, the Old English teaching methods and single-mode are one of the main factors limiting the quality of English education in my elementary school, and this disadvantage can be solved directly from the teachers' point of view, in terms of both education costs and education expenses which are extremely low, and can be "copied" and promoted by teachers.

4.2. Fuzzy Comprehensive Evaluation Algorithm Model Analysis. To verify the performance of the improved bat algorithm proposed in this paper, the bat algorithm, GAKBAT algorithm, and the algorithm in this paper were compared separately, and the Iris and Wine datasets in the UCI database were selected for 20 experiments each. The results of the comparison of the clustering centers of the three algorithms after running on the Iris dataset are shown in Figure 4. From Figure 4, it can be seen that the error sum of squares between the clustering centers obtained by the improved algorithm and the actual clustering centers on the Iris dataset is 0.015, which is the closest to the actual centers (Figure 4).

To further verify the clustering quality of this paper's algorithm, the three algorithms are experimentally compared in terms of correct rate and running time, and the

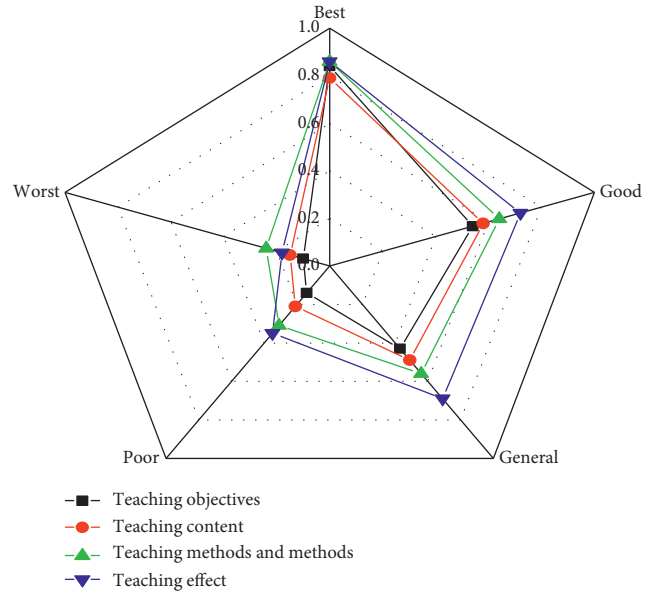


Figure 2: Criterion level indicator affiliation.

results are shown in Figure 5. From Figure 5, the correct rate of this algorithm is 96% on the Iris dataset, while the correct rate of the traditional bat algorithm is 89.33%, which is lower than that of this algorithm; the correct rate of the GAKBAT algorithm is 92.67%, which is higher than that of the traditional bat algorithm but lower than that of this algorithm. In the Wine data set, the highest correct rate of this algorithm is 94.94%, while the correct rates of the traditional bat algorithm and GAKFCM algorithm are 68.54% and 82%, respectively, which are lower than the correct rates of this algorithm. Secondly, comparing the running time of the algorithms, the running time consumption of this algorithm on the Iris and Wine data sets is slightly higher than that of the traditional bat algorithm due to the addition of the improved bat algorithm's optimization process but less than that of the GAKFCM algorithm. In summary, the algorithm in this paper achieves high clustering accuracy with reasonable time consumption.

To visualize the clustering effect, the real distribution results of the first two-dimensional sample points of the Iris dataset are selected. Figure 6 shows the clustering results obtained by the traditional bat algorithm, GAKBAT algorithm, and the algorithm in this paper, and it can be seen that the clustering results in this paper are better (Figures 5 and 6).

4.3. Analysis of English Teaching Quality Evaluation Design. The results of the resource utilization test are shown in Figure 7(a). During the whole scenario test, the CPU resource usage of the test server is 200 MB to 300 MB, and the performance is balanced during the whole test. The response time test results are shown in Figure 7(b). From Figure 7(b), it can be seen that there were two relatively large fluctuations (load reaching maximum concurrency) throughout the test, but overall, the maximum response time for the server to process the submit operation was 6.9 seconds, which included 3 seconds of thinking time. It indicates that the system functions designed and implemented in this article are relatively complete and meet the initial requirements (Figure 7).

In the experiment, we collected 100 samples and used the algorithm of this paper to conduct an experimental study on the data of each of the five attributes (namely, "English teaching ability," "English teaching method," "English teaching content," "English teaching effect," "English teaching attitude," and "English teaching effectiveness") in these samples. The number of evaluation processes $k = 3$, and the results of the evaluation analysis are shown in Figure 8. Therefore, the English teaching management should pay more attention to the effect of English teaching while strengthening the supervision and management of English teaching implementation, and from the previous classroom English teaching evaluation system, it is known that the effect of English teaching requires students to master the content taught by teachers, accomplish the English teaching objectives, achieve corresponding English teaching tasks, and, after English teaching activities, inspire students to think, improve student performance and learning efficiency, and help students learn to deal with problems. Therefore, in the future, teachers should pay more attention to the inspiration of students in the management of English teaching, and in the new progress of the subject knowledge, to combine theory and practice to better train students' problem-solving ability (Figure 8).

To improve the quality evaluation of English teaching, the college also needs to invest a lot of money: to hire experts and teachers inside and outside the university to supervise and evaluate English teaching, to increase research funds for English teaching quality evaluation-related topics and projects, to provide funds to encourage teachers to participate in English teaching quality evaluation-related training and conferences, to learn advanced English teaching quality

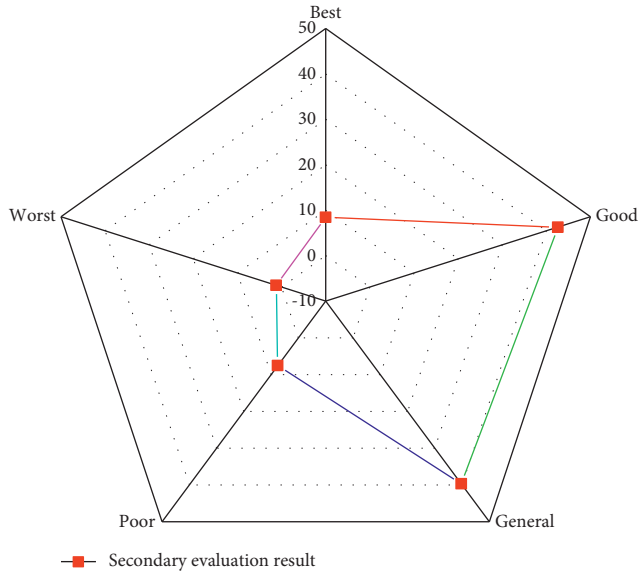


Figure 3: Target layer affiliation.

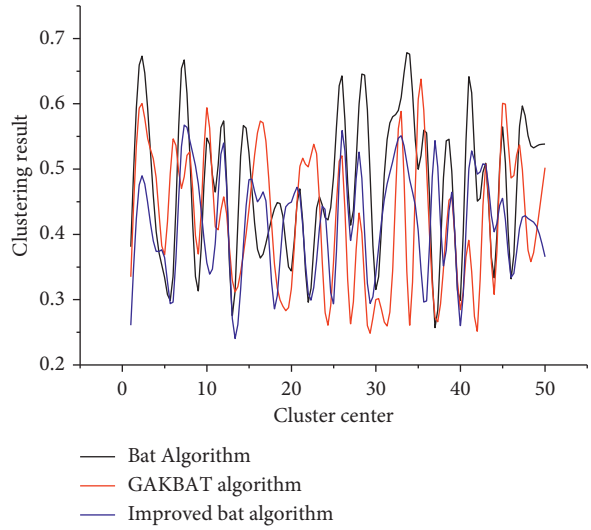


Figure 4: Comparison of clustering centers.

evaluation models, and to make continuous innovation in the process of English teaching. The university will also provide funding to encourage teachers to participate in training and conferences on ELT quality assessment, learn from advanced ELT quality assessment models, and make continuous innovative attempts in the ELT process. In terms of English teaching hardware, it is also necessary to increase the construction and maintenance costs of various platforms for online and offline evaluation data collection and collation, such as big data and cloud computing, and other information technology means, to ensure the long-term tracking and analysis of English teaching quality evaluation data. Reasonable and effective provision of financial guarantee is a practical guarantee for the improvement of English teaching quality evaluation. The final evaluators of the quality evaluation of English teaching in institutions are the society and employers. The vocational nature of English education determines that it must meet the needs of the market. Therefore, the college can also provide the following guarantees: regularly carry out surveys of institutional interns, graduates, and employers, and, through various forms of follow-up visits by departments and employers, understand through multiple channels whether the students' English education level has met the students' job requirements. Whether it has met the needs of enterprises, and whether it can meet the students' subsequent work. The program is a way to strengthen the social force's influence on the quality of English teaching. This is where the influence of social forces on English teaching quality evaluation is strengthened, and where the realistic basis for institutions to implement English teaching reform and improve English teaching quality evaluation lies.

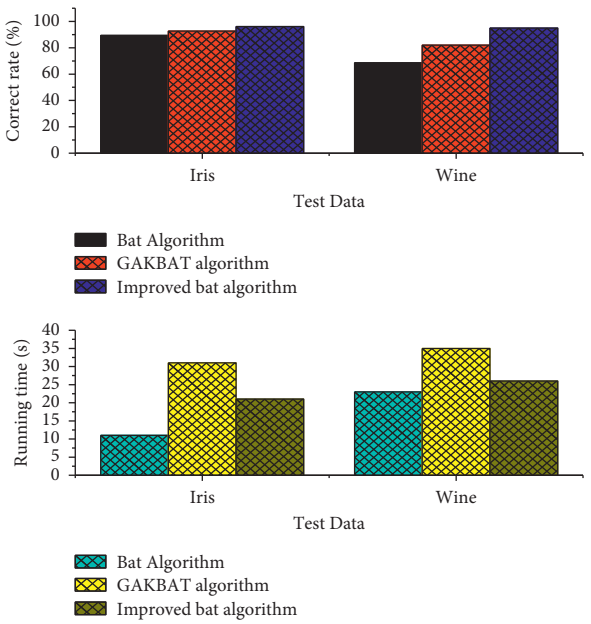


Figure 5: Comparison of the three algorithms in terms of correct clustering rate and running time.

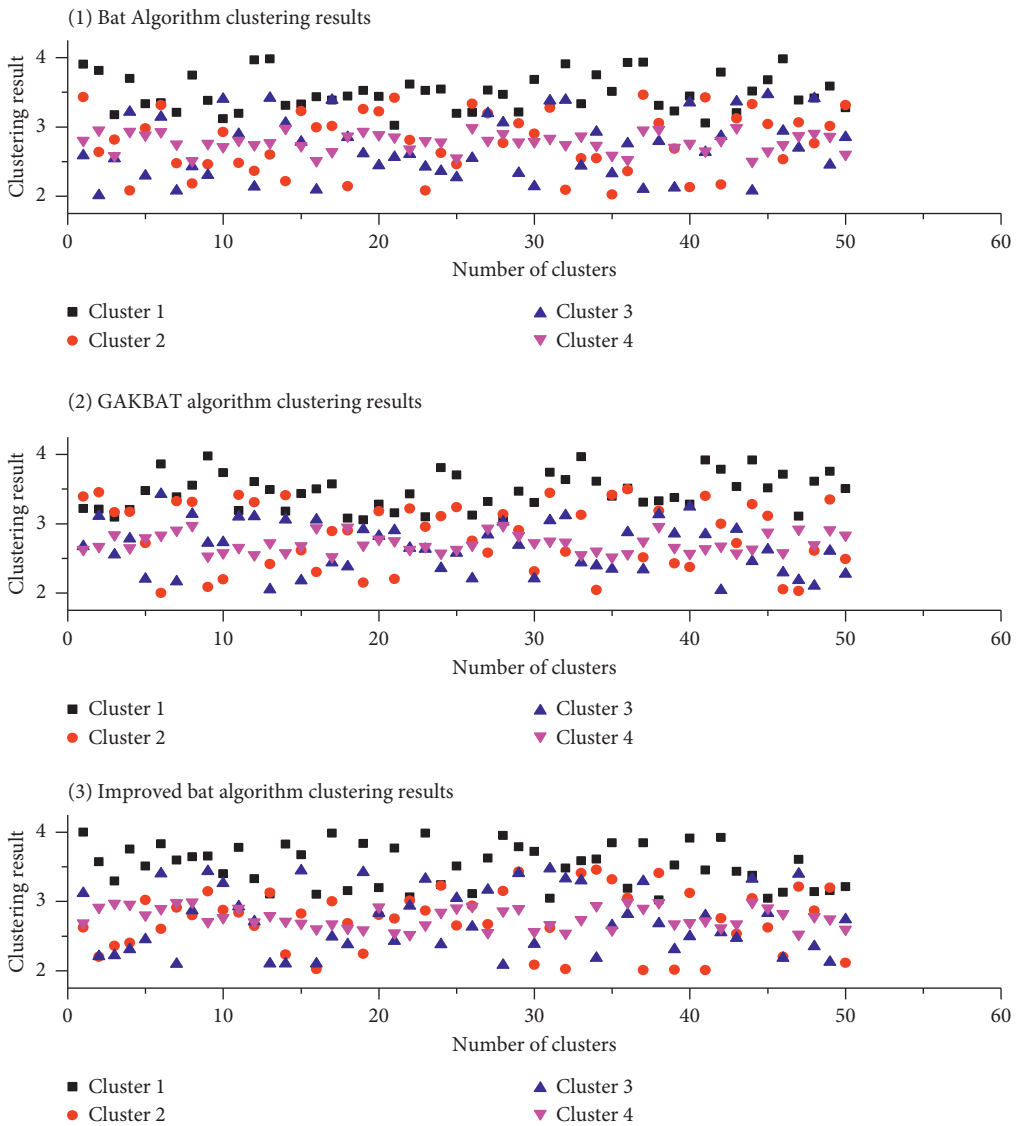


Figure 6: Algorithm clustering results.

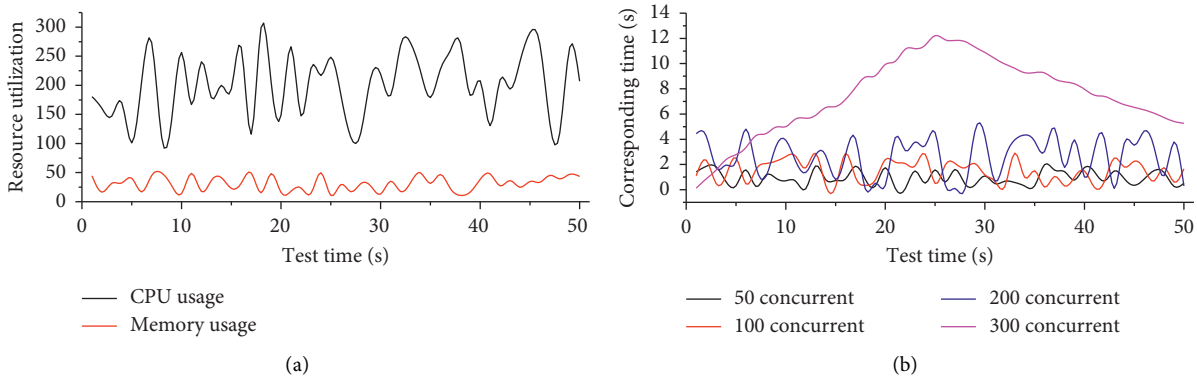


Figure 7: Resource usage and response time test results. (a) Resource usage test results. (b) Response time test results.

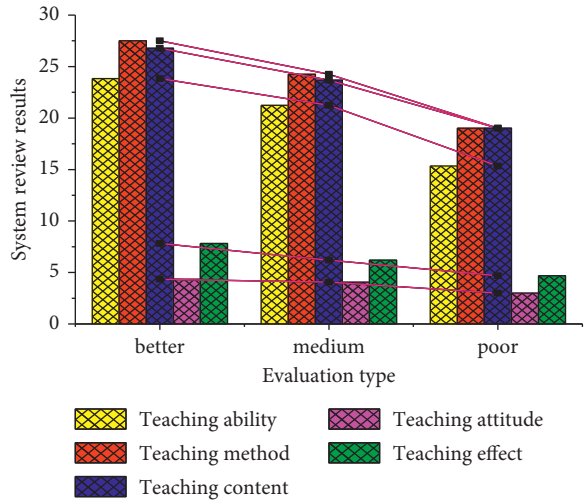


Figure 8: English teaching quality system evaluation results.

5. Conclusion

The evaluation of English teaching quality is an important help to cultivate talents, implement teaching reform, improve the quality of English teaching, and has a long-term and stable supporting role for theoretical research on English teaching. However, through the current analysis of English teaching quality and students' self-evaluation of the questionnaire, it is found that student's English language ability and pragmatic ability are unevenly developed, and the English teaching of English and vocational skills are not closely integrated. The bat algorithm was improved to address the shortcomings of the bat algorithm, and the experimental comparison with the traditional bat algorithm was conducted. The results showed that the improved algorithm improved the solution accuracy and convergence speed to a certain extent, which verified the feasibility of the improved algorithm. Then, to address the shortcomings of the fuzzy comprehensive evaluation algorithm, a fuzzy comprehensive evaluation algorithm optimized based on the improved bat algorithm is proposed, which optimizes the way of initial center selection and integrates the influence of different samples and attributes on the clustering results. The qualitative evaluation is classified by the fuzzy comprehensive evaluation model of the bat algorithm, the classified reviews corresponding to the rating criteria are quantified, and the quantified evaluations are incorporated into the evaluation model for calculation, so that the final evaluation results are more comprehensive and objective. Experimental simulations of the improved algorithm were conducted, and the clustering results obtained were more accurate. Finally, the improved fuzzy comprehensive evaluation algorithm is applied to the English teaching quality evaluation dataset, and the experimental results show that the improved algorithm can effectively deal with the problems on the evaluation of English teaching quality and provide a certain reference for the evaluation of English teaching quality. The innovation point of this paper is to quantify the users' qualitative evaluation of teachers' English teaching quality using the fuzzy comprehensive evaluation model of the self-bat algorithm. And the quantitative evaluation, qualitative evaluation, and objective data values such as students' course grades are combined to get the results of teachers' English teaching quality evaluation. The quantitative indexes in this paper are rather crude, and only the qualitative evaluation is quantified. In the process of English teaching, there are still

many quantifiable indicators. The graduation requirement indicator points can be incorporated into the quantitative system later.

Data Availability

All information is included within the paper.

Conflicts of Interest

The authors declare no conflicts of interest.

References

[1] Y. Zou and H. Zhou, "Effect evaluation of small claim procedure based on fuzzy comprehensive evaluation," *Journal of Interdisciplinary Mathematics*, vol. 21, no. 4, pp. 837–848, 2018.

[2] W. H. Bangyal, J. Ahmad, and H. T. Rauf, "Optimization of neural network using improved bat algorithm for data classification," *Journal of Medical Imaging and Health Informatics*, vol. 9, no. 4, pp. 670–681, 2019.

[3] S. Asghari and N. J. Navimipour, "Nature inspired meta-heuristic algorithms for solving the service composition problem in the cloud environments," *International Journal of Communication Systems*, vol. 31, no. 12, p. e3708, 2018.

[4] R. Ramesh and V. Saravanan, "Proportion frequency occurrence count with bat algorithm (FOCBA) for rule optimization and mining of proportion equivalence fuzzy constraint class association rules (PEFCARs)," *Periodicals of Engineering and Natural Sciences*, vol. 6, no. 1, pp. 305–325, 2018.

[5] D. Gao, Y. Liu, and Z. Guo, "A study on optimization of CBM water drainage by well-test deconvolution in the early development stage," *Water*, vol. 10, no. 7, 2018.

[6] Z. L. Yang, N. Islam, Y. Shi, K. Venkatachalam, and L. Huang, "The evolution of interindustry technology linkage topics and its analysis framework in 3D printing technology," *IEEE Transactions on Engineering Management*, vol. 99, 2021.

[7] F. Hemasian-Etefagh and F. Safi-Esfahani, "Dynamic scheduling applying new population grouping of whales meta-heuristic in cloud computing," *The Journal of Supercomputing*, vol. 75, no. 10, pp. 6386–6450, 2019.

[8] A. Yadav and D. K. Vishwakarma, "A comparative study on bio-inspired algorithms for sentiment analysis," *Cluster Computing*, vol. 23, no. 4, pp. 2969–2989, 2020.

[9] W. K. Mashwani, A. Hamdi, M. Asif Jan, A. Göktaş, and F. Khan, "Large-scale global optimization based on hybrid swarm intelligence algorithm," *Journal of Intelligent and Fuzzy Systems*, vol. 39, no. 1, pp. 1257–1275, 2020.

[10] M. B. Shareh, S. H. Bargh, A. A. R. Hosseinabadi, and A. Slowik, "An improved bat optimization algorithm to solve the tasks scheduling problem in open shop," *Neural Computing & Applications*, vol. 33, no. 5, pp. 1559–1573, 2021.

[11] V. Ravuri and S. Vasundra, "Moth-flame optimization-bat optimization: map-reduce framework for big data clustering using the moth-flame bat optimization and sparse fuzzy C-means," *Big Data*, vol. 8, no. 3, pp. 203–217, 2020.

[12] M. M. Öztürk, "A bat-inspired algorithm for prioritizing test cases," *Vietnam Journal of Computer Science*, vol. 5, no. 1, pp. 45–57, 2018.

[13] L. Huang, K. Zhang, W. Hu, and C. Li, "Trajectory optimisation design of robot based on artificial intelligence algorithm," *International Journal of Wireless and Mobile Computing*, vol. 16, no. 1, pp. 35–40, 2019.

[14] D. Ge, X. Wang, and J. Liu, "A teaching quality evaluation model for preschool teachers based on deep learning," *International Journal of Emerging Technologies in Learning (iJET)*, vol. 16, no. 3, pp. 127–143, 2021.

[15] C. Rao, Y. He, and X. Wang, "Comprehensive evaluation of non-waste cities based on two-tuple mixed correlation degree," *International Journal of Fuzzy Systems*, vol. 23, no. 2, pp. 369–391, 2021.

[16] M. Choi, H. Kim, B. Han, N. Xu, and K. M. Lee, "Channel attention is all you need for video frame interpolation," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, no. 7, pp. 10663–10671, 2020.

[17] D. Das, S. Jaypuria, D. K. Pratihar, and G. G. Roy, "Weld optimisation," *Science and Technology of Welding & Joining*, vol. 26, no. 3, pp. 181–195, 2021.

[18] H. Zhang and Y. Cui, "A model combining a Bayesian network with a modified genetic algorithm for green supplier selection," *Simulation*, vol. 95, no. 12, pp. 1165–1183, 2019.

[19] Z. Yang, W. Zhang, F. Yuan, and N. Islam, "Measuring topic network centrality for identifying technology and technological development in online communities," *Technological Forecasting and Social Change*, vol. 167, Article ID 120673, 2021.

[20] T. Grubljesic, P. S. Coelho, and J. Jaklic, "The shift to socio-organizational drivers of business intelligence and analytics acceptance," *Journal of Organizational and End User Computing*, vol. 31, no. 2, pp. 37–64, 2019.

[21] L. Z. Zhang, M. Mouritsen, and J. R. Miller, "Role of perceived value in acceptance of "bring your own device" policy," *Journal of Organizational and End User Computing*, vol. 31, no. 2, pp. 65–82, 2019.

[22] A. Shahri, M. Hosseini, K. Phalp, J. Taylor, and R. Ali, "How to engineer gamification," *Journal of Organizational and End User Computing*, vol. 31, no. 1, pp. 39–60, 2019.