

An online reviews-driven large-scale group decision-making approach for evaluating user preference of fresh fruit on e-commerce platform

Xiaolei Wang

Northwest A&F University

Bin Yang

bingyang0906@nwsuaf.edu.cn

Northwest A&F University

Research Article

Keywords: Large-scale group decision-making, Consensus reaching process, Fuzzy covering-based rough set, Fuzzy β -neighborhood, User preference, Online review

Posted Date: September 8th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10791181/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.
[Read Full License](#)

Additional Declarations: No competing interests reported.

An online reviews-driven large-scale group decision-making approach for evaluating user preference of fresh fruit on e-commerce platform

Xiaolei Wang¹ and Bin Yang^{2*}

¹College of Economics & Management, Northwest A&F University, Yangling, 712100, PR China .

^{2*}College of Science, Northwest A&F University, Yangling, 712100, PR China .

*Corresponding author(s). E-mail(s): binyang0906@nwsuaf.edu.cn;

Abstract

As an increasing number of consumers rely on AI-powered shopping assistants for online purchase decisions, understanding consumer preferences has become essential for improving the shopping experience and supporting personalized purchase guidance. This task is particularly challenging for fresh produce because of its inherent uncertainty. Taking fresh fruits as an illustrative case, this study collects online reviews, extracts preference-related keywords, and constructs fuzzy user preference sets. The study employs a large-scale group decision-making approach to identify and reconcile preference heterogeneity among users, thereby generating a representative group consensus. An intelligent decision-support system focused specifically on decision-making problems was developed to provide knowledge-grounded question answering and algorithmic support for decision analysis. Firstly, to improve consensus efficiency, the double-restricted Louvain model is proposed to divide user preference networks into multiple subgroups based on social network analysis. Second, in the consensus reaching process, discordant subgroups and individuals are managed. A particle swarm optimization algorithm with community relation is designed to achieve maximum consensus at minimum cost. This process aims to complete large-scale group consensus. Notably, this paper places greater emphasis on individual adjustment preferences. Finally, the neighborhood-related fuzzy covering-based rough set models are employed to determine decision-maker weights, and fuzzy TOPSIS is used to rank the importance of UPs. The findings provide decision support for e-commerce platforms in developing intelligent shopping assistants.

Keywords: Large-scale group decision-making, Consensus reaching process, Fuzzy covering-based rough set, Fuzzy β -neighborhood, User preference, Online review

1 Introduction

Driven by advances in artificial intelligence and natural language processing, e-commerce platforms are increasingly introducing intelligent shopping assistants, and consumers are becoming more dependent on these assistants to obtain product information, receive personalized recommendations, and make purchase decisions through interactive question answering [1]. However, in the online sales of fresh fruits, consumers often face greater uncertainty than in other product categories. Due to taste uncertainty, perishability, quality fluctuations, and logistics timeliness, fresh fruits are more likely to cause consumer dissatisfaction and increase the difficulty of purchase decision-making [2]. Therefore, accurately identifying consumers' preferences for fresh fruits is essential for improving the effectiveness of intelligent shopping assistants. Online reviews contain rich information about consumers' real purchase experiences and preference perceptions, which can provide an important data basis for preference analysis. In this context, it is necessary to develop an online reviews-driven decision-making approach that can extract user preferences from massive review data and transform them into usable decision support for intelligent recommendation. By embedding such an approach into intelligent shopping assistants, e-commerce platforms can better match consumers' personalized needs with suitable fresh fruit products, support consumers' purchase decision-making, and increase the likelihood of purchase [3]. This may further improve online sales volume and sales revenue, thereby benefiting multiple stakeholders, including farmers, suppliers, merchants, and e-commerce platforms.

User satisfaction has been widely regarded as an important factor affecting the performance of online stores [4]. In the B2C e-commerce model, user satisfaction is also one of the priority concerns of online shops, as it directly influences consumer retention, purchase intention, and the sustainable development of the platform [5]. For intelligent shopping assistants, whether consumers' preferences can be accurately identified and satisfied is crucial for improving recommendation effectiveness and shopping experience. In the purchase process of fresh fruits, user preferences may involve freshness, taste, price, packaging, logistics timeliness, and after-sales service. If these preferences are not properly understood, the assistant may provide unsuitable recommendations, which may reduce user satisfaction and weaken consumers' trust in the platform. Therefore, accurately obtaining user preferences (UPs) is a fundamental task for intelligent shopping assistants to provide personalized purchase guidance. Online reviews provide a natural and data-driven way to capture UPs. Compared with traditional survey data, online reviews are generated from real consumption experiences and can reflect consumers' satisfaction, dissatisfaction, and preference perceptions after purchase [6, 7]. Consumers often express their evaluations of fresh fruit products in online reviews. These review texts can be transformed into structured preference information and further used by intelligent shopping assistants to support interactive question

answering and personalized recommendation. More importantly, a data-driven method based on online reviews can continuously update preference information in dynamic e-commerce environments and help intelligent shopping assistants obtain UPs of fresh fruits more efficiently and objectively. However, online reviews are usually large-scale, unstructured, subjective, and fuzzy. Different consumers may express different or even conflicting opinions toward the same product attribute. Therefore, how to effectively extract, organize, and evaluate UPs from massive online reviews remains a challenging problem. Large-scale group decision-making (LSGDM) provides a suitable framework for this problem. Online reviews are generated by a large number of consumers, and each consumer can be regarded as a decision maker expressing preferences toward fresh fruit products. Existing studies have widely demonstrated the feasibility and effectiveness of using online reviews to analyze user preferences. In particular, review information has been applied to UPs importance evaluation, consumer decision support, and product recommendation [8, 9].

However, according to our knowledge, few studies have developed a complete online reviews-driven LSGDM method for analyzing UPs of fresh fruits on e-commerce platforms. Moreover, existing studies have rarely transformed LSGDM-based preference analysis results into decision support for platform-based intelligent shopping assistants. To fill these gaps, this paper investigates an online reviews-driven LSGDM approach for analyzing UPs of fresh fruits. Moreover, the weighting problem is an important issue in LSGDM. Existing studies on weights mainly focus on two aspects: the weights of experts or attributes and the weights of subgroups after clustering [10, 11]. In general, LSGDM aims to cluster the evaluations of large-scale individuals, adjust group consensus, and obtain reliable decision-making results, and it has been widely applied in various fields [12, 13]. However, assigning the same weight to all objects or decision makers is often unreasonable in practical decision-making problems, especially when consumers differ in preference expression, review credibility, and subgroup influence. Therefore, how to objectively determine weights in LSGDM is a vital issue. Fuzzy covering-based rough set models have advantages in processing uncertain and imprecise information and have shown good performance in decision-making evaluation. By using the evaluation coefficient obtained from fuzzy covering-based rough set models as the weight of decision makers, the equal-weight problem can be alleviated, and the reliability of LSGDM results can be improved. Reliable preference weights and ranking results are also essential for intelligent shopping assistants to generate explainable recommendations and provide effective purchase guidance. Based on the previous analysis, the following two aspects serve as the primary motivation of this paper.

- Information representation process. In terms of LSGDM method in UPs of fresh fruit reviews on e-commerce platform, (1) no previous study has used a complete data-driven method to obtain preference scores of fresh fruit on e-commerce platform. (2) The LSGDM method evaluates UPs of fresh fruit on e-commerce platform is lacking.
- decision-making process. Fuzzy covering-based rough set models are generalized to LSGDM problem to evaluate the importance scores of objects. Consistent object weights may lead to unreasonable decision outcomes, so it is necessary to adopt a reasonable way of assigning weights, the fuzzy covering-based rough set models

and their extended fuzzy β -covering-based rough set models are assigned weights to individuals combined with their UP.

The present study aims to develop a general LSGDM model that can extract UPs from online reviews and determine their importance. The obtained preference information can be embedded into intelligent shopping assistants to improve recommendation accuracy, support interactive purchase guidance, enhance user satisfaction, and further promote the online sales of fresh fruits. The detailed study is developed from the following aspects:

- On the basis of UPs score information of fresh fruit, we develop the UP fuzzy set by online reviews-driven instead of expert scoring. The small number of decision makers studied is a limitation of questionnaire collection. We extract UP keywords by TF-IDF algorithm from each product based on the reviews of fresh fruits on the Jingdong (JD) e-commerce platform, and each product is regarded as a decision maker or individual, which overcomes the disadvantage of fewer decision makers, and the UPs degree of fresh fruit is structured which is mined from big data, making the decision more objective.
- For the LSGDM process, the double-restricted Louvain model is proposed to handle the community clustering with a similarity relationship of individuals. Moreover, we design a consensus reaching process (CRP) model to accelerate reaching large-scale group (LSG) consensus with a feedback mechanism of minimizing consensus and minimizing cost. At the same time, social network analysis (SNA) is used to assign decision maker weights.
- Consensus reaching process. The independence of individuals should not be ignored by the average thinking, LSG consensus should be achieved more objectively which is necessary to ensure a high level of group consensus. It is worth mentioning that in CRP, this paper puts more emphasis on the adjustment of decision makers' UPs, instead of only considering the adjustment of discordant subgroups. A particle swarm optimization algorithm is designed to solve maximum consensus at minimum cost in combination with adjusted weights assigned by similarity relationship network.
- The assignment of weights to objects directly affects the rationality of the LSGDM outcomes. In this paper, we adopt four fuzzy neighborhood operators based on fuzzy covering with three different t -norms and R -implicators, and two fuzzy covering-based rough set models, which are fuzzy β -neighborhood and fuzzy complementary β -neighborhood as weighting methods in objects.
- Fuzzy TOPSIS is applied to LSGDM problems combining fuzzy covering-based rough set models for data that cannot be reasonably aggregated, which can deal with multiple preference ambiguities. At the same time, VIKOR method is used as a comparison to prove its effectiveness in LSGDM problem.

This paper is structured as follows. Section 2 conducts a brief literature review on LSGDM and fuzzy covering-based rough set models. In Section 3, we recall some basic definitions and conclusions used in this paper. This section introduces the concepts of fuzzy coverings and fuzzy neighborhood operators. Section 4 provides the detailed

theoretical process of LSGDM, including the solution for maximum consensus at minimum cost and decision-making with fuzzy covering-based rough set models. In Section 5, the proposed LSGDM method is implemented using online review data on fresh fruit and managerial implications and marketing strategies are discussed. Moreover, the intelligent decision-making platform is introduced. In Section 6, the comparative experiments are carried out, verifying the validity and stability of the model. Finally, Section 7 concludes and proposes future work plans. The nomenclature is presented in Table 1.

Table 1 Nomenclature

Nomenclature	
<i>Abbreviations</i>	
CRP	Consensus reaching process
CUP	Consensual user preference
DM	decision-making
LSG	Large-scale group
LSGDM	Large-scale group decision-making
PFS	Preferences fuzzy set
PSO	particle swarm optimization
SNA	Social network analysis
SRN	Similarity relationship network
TF-IDF	Term frequency-inverse document frequency
TOPSIS	Technique for order preference by similarity to ideal solution
UP	User preference
VIKOR	Vise Kriterijumski Optimizacioni Racun

2 Literature review

This section briefly reviews the relevant literature on large-scale group decision-making methods, fuzzy coverings, and fuzzy neighborhood operators.

2.1 A brief review of large-scale group decision-making method

Group decision-making (DM) refers to the process in which two or more decision makers jointly reach a consensus on a set of alternatives through DM, and the final DM conclusion is required to be completed by the group through deliberation or negotiation. With the complexity of DM environment and the increase of information involved, it has become a trend for a large number of decision makers to participate in DM process. When the number of decision makers involved in the decision problem is at least 20 [14], it can be regarded as LSGDM. The primary objective of LSGDM is to derive a collective decision that achieves an acceptable level of consensus among a large number of decision makers while accommodating heterogeneous preferences [15]. Compared with conventional group decision-making approaches that are typically designed for relatively small groups, LSGDM is better suited to handling large-scale heterogeneous

opinions and facilitating the efficient derivation of a broadly accepted collective solution through consensus-reaching processes [16]. With further study of LSGDM, it is mainly divided into the following processes, preference information processing, group clustering, consensus reaching process and decision-making process.

In DM problems, the outcome of the decision is usually represented by an exact number. In fact, as DM environment becomes more and more complex, experts may come from different fields and have different characteristics, which causes the results of decision makers' DM to be often uncertain. Therefore, in order to make the preference information set contain more valuable and accurate communication, it is necessary to adopt an appropriate information set expression method. The expression of preference information uses various forms by expert questionnaire [17]. With the increasing complexity of network relations, studies gradually pay attention to the credibility of decision makers in network relations [18]. Making the DM results more reliable by adding a trust function about decision makers in the preference relationship is more reliable [19]. Based on the literature review, we observed two problems in the current preference degree acquisition. First, researchers tend to focus on the construction form of the preference set while ignoring the way to obtain the preference. This paper focuses on how to obtain the DM preferences of LSG objectively and in large-scale. At the same time, this paper constructs a general process system to extract the preferences of decision makers based on online reviews. Secondly, the obtained reliability of decision makers only serves the next stage of cluster analysis, which is also used by this study as a decision maker's adjustment weight in consensus reaching process. Therefore, the present study aims to construct an objective large-scale decision makers' preference set.

As the number of DM participants increases, the DM process is more complex and time-consuming. LSGDM divides large groups into several subgroup, according to the similarity between interest preferences, geographical location, organization and other characteristics, to speed up DM efficiency. Group clustering is based on decision information, and various clustering algorithms are widely used. The main group clustering methods at this stage are completely partitioned clustering methods, and overlapping clustering methods [20]. Large group clustering based on subgroup detection algorithm by social network is also a common implementation. For the evolution of consensus network, Louvain algorithm can drive the community discovery algorithm to divide the entire group into multiple subgroups, which ensures the independence and strong cohesion of the subgroups [21], moreover, the dimension of LSGDM is reduced in community networks. In this paper, the Louvain algorithm combined with SNA is used to complete LSG clustering and has been proven to be a suitable clustering method for LSG preference set.

As conflicts among decision makers are common in LSGDM, a consensus-reaching process (CRP) is typically required to reconcile divergent opinions and ensure the acceptability of the final decision. Therefore, it is a necessary process for LSGDM to measure the degree of consensus and achieve discordant adjustment of decision makers to complete LSG consensus. CRP is an iterative process in which the group consensus level is assessed at each round, and a feedback mechanism is activated when the predefined consensus threshold is not met. As a key component of CRP, the feedback

mechanism guides the adjustment of divergent preferences, thereby reducing conflicts and improving consensus efficiency. At this stage, effective feedback mechanisms are often used. Wang et al. [22] proposed a three-way decision-based feedback mechanism that classifies decision makers according to their conflict levels and social-network relationships, thereby providing targeted feedback strategies. Ding et al. [23] integrated overlapping community clustering with a feedback mechanism to manage interdependent subgroups in social-network LSGDM. Li et al. [24] developed a dynamic feedback mechanism that considers communication sequences to reduce redundant interactions and improve consensus efficiency in multilevel LSGDM. However, scholars focus on the adjustment of subgroups in the discordant adjustment of CRP, but often ignore the decision makers in the discordant subgroups. A small amount of literature has adjusted the preference of decision makers, still the proportion of preference adjustment is based on the same weight. This phenomenon will cause injustice for decision makers to modify the proportion on preference degree. Therefore, this paper pays attention to decision makers' preference adjustment and uses SNA to calculate their adjustment. When the CRP is completed, the consensus of the LSG is reached. In the following decision-making process, the decision result of LSG can be obtained by weighted average according to subgroup preference scores, in which the ranking result of the alternatives is the LSGDM solution with high consensus and high consistency [20]. TOPSIS has been successfully applied in many industries and achieved satisfactory results on LSGDM problem [25]. Nevertheless, as the composition of information becomes more and more complex, the traditional TOPSIS method has limitations in applying fuzzy environments. we propose a new TOPSIS method for data that cannot be reasonably aggregated, which can deal with multiple preference ambiguities to improve the effectiveness of decision-making outcomes.

2.2 A brief review of fuzzy covering-based rough set models

In this subsection, we briefly review the development and extension of fuzzy-covering rough sets and fuzzy β -covering rough sets and their application to decision-making, which will enable us to apply these methods to LSGDM properly.

Pawlak [26] put forward a new uncertain tool, rough set theory (RST), for dealing with uncertain information in the real world. It proposes a systematic method for object classification by using an indiscernibility relation. However, when this theory is applied to practical problems, studies observe that the model based on the equivalence relation proposed by Pawlak is used to characterize a partition formed by classes of indiscernible objects which causes the limited applications [27]. By extending the partition to a covering, Zakowski [28] firstly proposed the notion of covering-based rough set (CRS). In subsequent studies, scholars also explored other forms of CRS models. In particular, Pomykala [29, 30] defined two dual approximation operators through the concepts of neighborhood and granularity. Yao [31] put forward a novel concept named neighborhood-related CRS model.

To be able to handle a wider range of data types, Dubois and Prade [32] first introduced the notion of fuzzy rough sets. Through the efforts of scholars, the original definition of fuzzy covering was gradually summarized and proposed [33, 34]. By using fuzzy β -neighborhoods, Ma [35] put forward two kinds of coverings-based

fuzzy rough sets, named fuzzy β -covering-based rough sets, which are generalizations of fuzzy covering-based rough sets when β is 1. Yang and Hu [36] explored the fuzzy complementary β -neighborhood, and further put forward its related properties. Furthermore, another critical generalization of the fuzzy rough set [37] is based on fuzzy neighborhood operators. D'eer et al. [38] proposed the t -norm-based fuzzy neighborhood operators on fuzzy coverings, their inspiration came from the four classical neighborhood operators of Yao and Yao [39].

The advantages of fuzzy set theory are that it captures complex social phenomena and retains more data information. And the success of rough set theory is that it can analyze the facts hidden in data, does not require any prior knowledge beyond the data sets, and can attain a minimal knowledge representation. And covering-based rough set is an important extension of the classical rough set and an important theoretical method for dealing with incomplete information system data. Fuzzy set and rough set are complementary and closely related, and the fuzzy covering-based rough set models generated by the combination have been proven to solve ambiguity and uncertainty problems in DM problems effectively. However, in practical problems, the models have limitations, that is often impossible to select the optimal DM result based on the membership. Then the quoted parameter β replaces 1 in the fuzzy covering [35], and the fuzzy β -covering-based rough set models are obtained, which can evaluate the importance of all objects. And the notion of fuzzy β -coverings greatly expands the application scopes of fuzzy rough set theory. Then lots of investigations about fuzzy covering-based rough sets and their application to solve practical problems in the real world are proposed have been proposed, which can be combined with models such as TOPSIS to get the optimal solution in multi-criteria DM that proved to be effective, and the application is more and more widely [25].

Inspired by fuzzy covering-based rough set models on multi-attribute decision-making problems, this paper generalizes them to LSGDM. Thus, we research that fuzzy covering-based rough set models to get the optimal preference according to approximate precision, and use fuzzy TOPSIS for the importance score of preference to decision-making.

3 Preliminaries

This section presents a concise retrospect to several fundamental notions and related results.

3.1 Fuzzy set

In the following definition, we will introduce some basic definitions of fuzzy sets.

Definition 1 [40] Let U be a non-empty finite universe. A fuzzy set A , or rather a fuzzy subset A of U , is defined by a function assigning to each element z of U a value $A(z) \in [0, 1]$. We denote by $\mathcal{F}(U)$ the family of all fuzzy subsets of U , i.e., the set of all functions from U to $[0, 1]$, and call it the fuzzy power set of U .

For any $A, B \in \mathcal{F}(U)$, we say that A is contained in B , denoted by $A \subseteq B$, if $A(z) \leq B(z)$ for all $z \in U$, and we say that $A = B$ if and only if $A \subseteq B$ and $B \subseteq A$.

Given $A, B \in \mathcal{F}(U)$, the union of A and B , denoted as $A \cup B$, is defined by $(A \cup B)(z) = A(z) \vee B(z)$ for all $z \in U$. The intersection of A and B , denoted as $A \cap B$, is given by $(A \cap B)(z) = A(z) \wedge B(z)$ for all $z \in U$. The complement of A , denoted as A^c , is defined by $A^c(z) = 1 - A(z)$ for all $z \in U$.

3.2 Fuzzy logical operator

Next, we recall some concepts related to fuzzy logical operator.

Definition 2 [41] A binary operator $\mathcal{T} : [0, 1]^2 \rightarrow [0, 1]$ is called a t -norm, if it verifies the property of commutative, associative and non-decreasing in each variable with a neutral element 1.

There are three common types of continuous t -norm:

- (1) Minimum t -norm: $\mathcal{T}_M(u, v) = \min(u, v)$;
- (2) Product t -norm: $\mathcal{T}_P(u, v) = u \times v$;
- (3) Łukasiewicz t -norm: $\mathcal{T}_L(u, v) = \max(u + v - 1, 0)$.

Definition 3 [41] A binary operator $\mathcal{I} : [0, 1]^2 \rightarrow [0, 1]$ is called a fuzzy implicator, if it verifies $\mathcal{I}(0, 0) = \mathcal{I}(0, 1) = \mathcal{I}(1, 1) = 1$, $\mathcal{I}(1, 0) = 0$ and $\mathcal{I}$ is monotonically decreasing with respect to the first argument and monotonically increasing with respect to the second argument.

One of the fuzzy implicator operators is the R -implicator, which can be induced from continuous t -norm by

$$\mathcal{I}_R(u, v) = \sup\{z \in [0, 1] \mid \mathcal{T}(u, z) \leq v\}.$$

Example 1 The common continuous t -norms and their R -implicators are as follows in table,

Table 2 The common continuous t -norms and their R -implicators

t -norms	R -implicators
$\mathcal{T}_M(u, v) = \min(u, v)$	$\mathcal{I}_M(u, v) = \begin{cases} 1, & u \leq v \\ v, & u > v \end{cases}$
$\mathcal{T}_P(u, v) = u \times v$	$\mathcal{I}_P(u, v) = \begin{cases} 1, & u \leq v \\ \frac{v}{u}, & u > v \end{cases}$
$\mathcal{T}_L(u, v) = \max(u + v - 1, 0)$	$\mathcal{I}_L(u, v) = \min(1 - u + v, 1)$

3.3 Fuzzy covering-based rough set

In this part, some basic concepts about fuzzy covering, fuzzy minimal description and fuzzy maximal description are given.

Definition 4 [42] A collection $\mathbb{C} = \{C_1, C_2, \dots, C_m\}$ with $C_i \in \mathcal{F}(U)$ ($i = 1, 2, \dots, m$) is called a fuzzy covering, if there exists a $C_i \in \mathbb{C}$ such that $C_i(x) = 1$ for each $x \in U$. Here, $(U, \mathbb{C})$ is defined as a fuzzy covering approximation space (FCAS).

In what follows, if not otherwise specified, the universe U is always finite and nonempty.

Definition 5 [38] Let $(U, \mathbb{C})$ be an FCAS. For each $x \in U$, we define the fuzzy neighborhood system of x as

$$\mathcal{C}(\mathbb{C}, x) = \{C_i \in \mathbb{C} \mid C_i(x) > 0, i = 1, 2, \dots, m\}.$$

It can be easy to obtain that $\mathcal{C}(\mathbb{C}, x)$ is not empty, since there always exists a set $C_i \in \mathbb{C}$ such that $C_i(x) = 1$.

Definition 6 [38] Let $(U, \mathbb{C})$ be an FCAS. For each $x \in U$,

$$\text{md}(\mathbb{C}, x) = \{C \in \mathcal{C}(\mathbb{C}, x) \mid (\forall S \in \mathcal{C}(\mathbb{C}, x)) \wedge (S(x) = C(x)) \wedge (S \subseteq C) \Rightarrow S = C\},$$

is called the fuzzy minimal description of x ;

$$\text{MD}(\mathbb{C}, x) = \{C \in \mathcal{C}(\mathbb{C}, x) \mid (\forall S \in \mathcal{C}(\mathbb{C}, x)) \wedge (S(x) = C(x)) \wedge (S \supseteq C) \Rightarrow S = C\},$$

is called the fuzzy maximal description of x .

It is worth reminding that $\mathcal{C}(\mathbb{C}, x)$, $\text{md}(\mathbb{C}, x)$ and $\text{MD}(\mathbb{C}, x)$ are composed of fuzzy sets, in the meantime, both $\text{md}(\mathbb{C}, x)$ and $\text{MD}(\mathbb{C}, x)$ are subsets of $\mathcal{C}(\mathbb{C}, x)$.

Definition 7 [38] Let $(U, \mathbb{C})$ be an FCAS, $\mathcal{I}$ an implicator and $\mathcal{T}$ a t -norm. $N_1^{\mathbb{C}}, N_2^{\mathbb{C}}, N_3^{\mathbb{C}}, N_4^{\mathbb{C}} : U \rightarrow \mathcal{F}(U)$ are four types of fuzzy neighborhood operators, respectively. Further, the fuzzy neighborhood $N_i^{\mathbb{C}}(x)$ ($i = 1, 2, 3, 4$) are defined as follows:

$$\begin{aligned} N_1^{\mathbb{C}}(x)(y) &= \bigwedge_{C \in \mathbb{C}} \mathcal{I}(C(x), C(y)), \\ N_2^{\mathbb{C}}(x)(y) &= \bigvee_{C \in \text{md}(\mathbb{C}, x)} \mathcal{T}(C(x), C(y)), \\ N_3^{\mathbb{C}}(x)(y) &= \bigwedge_{C \in \text{MD}(\mathbb{C}, x)} \mathcal{I}(C(x), C(y)), \\ N_4^{\mathbb{C}}(x)(y) &= \bigvee_{C \in \mathbb{C}} \mathcal{T}(C(x), C(y)). \end{aligned}$$

Definition 8 [35, 43] Let $N_i^{\mathbb{C}} (i = 1, 2, 3, 4)$ be fuzzy neighborhood operators. For each $X \in \mathcal{F}(U)$, the lower approximation $\underline{\mathbb{A}}_i(X)$ and upper approximation $\overline{\mathbb{A}}_i(X)$ of X are defined as follows:

$$\begin{aligned}\underline{\mathbb{A}}_i(X)(x) &= \bigwedge_{y \in U} [(1 - N_i^{\mathbb{C}}(x)(y)) \vee X(y)], x \in U, \\ \overline{\mathbb{A}}_i(X)(x) &= \bigvee_{y \in U} [N_i^{\mathbb{C}}(x)(y) \wedge X(y)], x \in U,\end{aligned}$$

3.4 Fuzzy β -covering approximation space

Definition 9 [35] (1) Let U be a universe set and $\mathbb{C} = \{C_1, C_2, \dots, C_m\}$, where $C_i \in \mathcal{F}(U)$ and $i = 1, 2, \dots, m$. For any $\beta \in (0, 1]$, then $\mathbb{C}$ is called a fuzzy β -covering of U if $(\bigcup_{i=1}^m C_i)(x) \geq \beta$, for all $x \in U$. And, $(U, \mathbb{C})$ is an FCAS.

(2) Let $(U, \mathbb{C})$ be an FCAS and $\mathbb{C} = \{C_1, C_2, \dots, C_m\}$ be a fuzzy β -covering of U for some $\beta \in (0, 1]$. Then $N_{\mathbb{C}(x)}^{\beta} = \cap \{C_j \in \mathbb{C} | C_j(x) \geq \beta, j = 1, 2, \dots, m\}$ is called a fuzzy β -neighborhood of x in U .

Definition 10 [35] Let $(U, \mathbb{C})$ be an FCAS and $\mathbb{C} = \{C_1, C_2, \dots, C_m\}$ be a fuzzy β -covering of U for some $\beta \in (0, 1]$. For any $X \in \mathcal{F}(U)$, the lower and upper approximations of X w.r.t. $\mathbb{C}$ are defined, respectively.

$$\begin{aligned}\underline{\mathbb{Q}}(X)(x) &= \bigwedge_{y \in U} \{(1 - N_{\mathbb{C}(x)}^{\beta}(y)) \vee X(y)\}, \\ \overline{\mathbb{Q}}(X)(x) &= \bigvee_{y \in U} \{N_{\mathbb{C}(x)}^{\beta}(y) \wedge X(y)\},\end{aligned}$$

for all $x \in U$.

Definition 11 [36] Suppose that $(U, \mathbb{C})$ is an FCAS and $\mathbb{C} = \{C_1, C_2, \dots, C_m\}$ is a fuzzy β -covering of U for some $\beta \in (0, 1]$. For any $x \in U$, define a fuzzy complementary β -neighborhood of x ,

$$M_{\mathbb{C}(x)}^{\beta}(y) = N_{\mathbb{C}(y)}^{\beta}(x),$$

for all $y \in U$.

Definition 12 [36] Assume that $(U, \mathbb{C})$ is an FCAS and $\mathbb{C} = \{C_1, C_2, \dots, C_m\}$ is a fuzzy β -covering of U for some $\beta \in (0, 1]$. For any $X \in \mathcal{F}(U)$, the lower and upper approximations of X based on a fuzzy complementary β -neighborhood $M_{\mathbb{C}(x)}$ of x in U are defined, respectively,

$$\begin{aligned}\underline{\mathbb{E}}(X)(x) &= \bigwedge_{y \in U} \{(1 - M_{\mathbb{C}(x)}^{\beta}(y)) \vee X(y)\}, \\ \overline{\mathbb{E}}(X)(x) &= \bigvee_{y \in U} \{M_{\mathbb{C}(x)}^{\beta}(y) \wedge X(y)\}.\end{aligned}$$

Definition 13 [35] Assume that $(U, \mathbb{C})$ is an FCAS and $\mathbb{C} = \{C_1, C_2, \dots, C_m\}$ is a fuzzy β -covering of U for some fuzzy β -neighborhood and $U = \{x_1, \dots, x_n\}$. Suppose

that $N_{\mathbb{C}}^{\beta} = \{N_{\mathbb{C}(x_i)}^{\beta} | x_i \in U\}$ is a fuzzy β -neighborhood system induced by $\mathbb{C}$, where $N_{\mathbb{C}(x_i)}^{\beta}(x_j) = \mu_{N_{\mathbb{C}(x_i)}^{\beta}}(x_i, x_j)$, for all $i, j = 1, 2, \dots, n$. For any $I \in \mathcal{F}(U)$, the upper and lower approximations of I w.r.t. $N_{\mathbb{C}}^{\beta}$, denoted by $\overline{N_{\mathbb{C}}^{\beta}}(I)$ and $\underline{N_{\mathbb{C}}^{\beta}}(I)$, respectively, are two FSs defined as follows:

$$\overline{N_{\mathbb{C}}^{\beta}}(I)(x_i) = \bigvee_{j=1}^n \left\{ \mu_{N_{\mathbb{C}(x_i)}^{\beta}}(x_i, x_j) \wedge \mu_I(x_j) \right\},$$

and

$$\underline{N_{\mathbb{C}}^{\beta}}(I)(x_i) = \bigwedge_{j=1}^n \left\{ \mu_{N_{\mathbb{C}(x_i)}^{\beta}}(x_i, x_j) \vee \mu_I(x_j) \right\}.$$

The operators $\overline{N_{\mathbb{C}}^{\beta}}(I), \underline{N_{\mathbb{C}}^{\beta}}(I) : \mathcal{F}(U) \rightarrow \mathcal{F}(U)$ are called the upper and lower fuzzy rough approximation operators (FRAOs) w.r.t. $N_{\mathbb{C}}^{\beta}$, respectively.

Definition 14 [36] Suppose that $(U, \mathbb{C})$ is an FCAS and $\mathbb{C} = \{C_1, C_2, \dots, C_m\}$ is a fuzzy β -covering of U for some fuzzy β -neighborhood and $U = \{x_1, \dots, x_n\}$. Suppose that $M_{\mathbb{C}}^{\beta} = \{M_{\mathbb{C}(x_i)}^{\beta} | x_i \in U\}$ is a fuzzy complementary β -neighborhood system induced by $\mathbb{C}$, where $M_{\mathbb{C}(x_i)}^{\beta}(x_j) = \mu_{M_{\mathbb{C}(x_i)}^{\beta}}(x_i, x_j)$, for all $i, j = 1, 2, \dots, n$. For any $I \in \mathcal{F}(U)$, where $I = \{\langle x_j, \mu_I(x_j) \rangle | j = 1, 2, \dots, n\}$, the upper and lower approximations of I w.r.t. $M_{\mathbb{C}}^{\beta}$, denoted by $\overline{M_{\mathbb{C}}^{\beta}}(I)$ and $\underline{M_{\mathbb{C}}^{\beta}}(I)$, respectively, are two FSs defined as follows:

$$\overline{M_{\mathbb{C}}^{\beta}}(I)(x_i) = \bigvee_{j=1}^n \left\{ \mu_{M_{\mathbb{C}(x_i)}^{\beta}}(x_i, x_j) \wedge \mu_I(x_j) \right\},$$

and

$$\underline{M_{\mathbb{C}}^{\beta}}(I)(x_i) = \bigwedge_{j=1}^n \left\{ \mu_{M_{\mathbb{C}(x_i)}^{\beta}}(x_i, x_j) \vee \mu_I(x_j) \right\}.$$

The operators $\overline{M_{\mathbb{C}}^{\beta}}(I), \underline{M_{\mathbb{C}}^{\beta}}(I) : \mathcal{F}(U) \rightarrow \mathcal{F}(U)$ are called the upper and lower FRAOs w.r.t. $M_{\mathbb{C}}^{\beta}$, respectively.

4 LSGDM method with fuzzy covering-based rough set models

This section will describe our LSGDM method and fuzzy covering-based rough set models in detail. The framework of this paper is shown in Figure 1.

- (1) Extraction of UP keywords. We collect user online review data and perform text preprocessing to extract UP keywords based on keyword frequency and TF-IDF, which detailed step is introduced in Section 5.
- (2) Construct the preference fuzzy set (PFS). The PFS is constructed according to the preference keywords and word frequency extracted from each product review that is viewed as a decision maker, i.e., an individual, preference fuzzy relations. (See Section 4.1)

- (3) Clustering and measuring the cohesion degree of subgroup. The similarity degree of decision makers is used to establish the similarity relationship network (SRN). And multiple subgroups are clustered by double-restricted Louvain model and SNA, meanwhile, the cohesion degree of subgroup is calculated to verify the clustering effect. (See Section 4.2)
- (4) Consensus reaching process (CRP). The discordant subgroups and their discordant preferences are identified. In feedback mechanism, PSO algorithm is designed to solve maximum consensus at minimum cost in combination with decision makers' weights assigned by community relation degree. And then, the LSG consensus is reached, we obtain the consensual UP fuzzy set, and the adjustment cost is calculated. (See Section 4.3)
- (5) decision-making. In Section 4.4, we use four fuzzy neighborhood operators based on fuzzy covering with three different t -norms and R -implicators, and two fuzzy covering-based rough set models which are fuzzy β -neighborhood and fuzzy complementary β -neighborhood to evaluate the weight of decision makers combined with approximate precision. And then, the fuzzy TOPSIS is used to get the result of LSGDM (see Section 4.5).

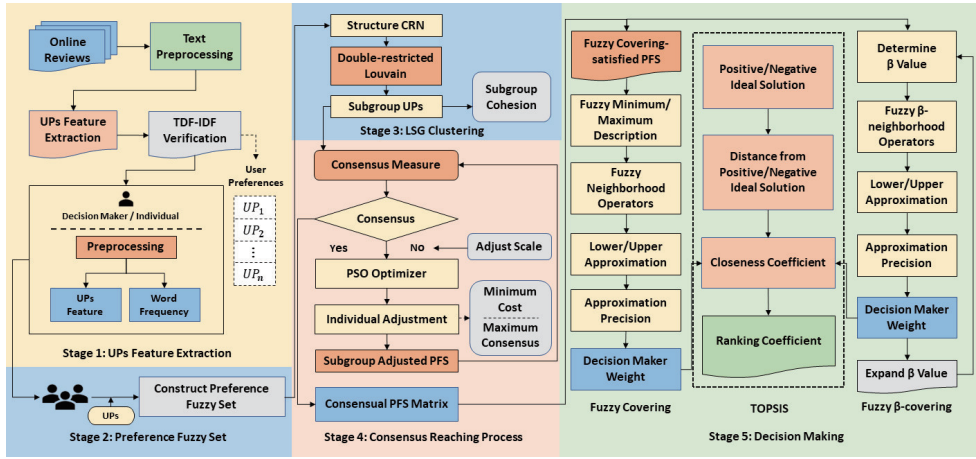


Fig. 1 The LSGDM framework for evaluating user preference by online reviews

4.1 Preference score and preference fuzzy set

In general, decision makers usually use real-valued data to represent preference scores, i.e., object scores [44]. Suppose that LSGDM problems have n preferences and m decision makers, the decision makers' scoring table is V , where v_{gc} denotes the score assigned by $\mathcal{O}_g$ to preference τ_c , that is, an object. And Table 3 represents the multi-parameter information involved in the LSGDM in this paper.

Table 3 LSGDM parameter information table

Designation	Information
Index sets	$\mathcal{N} = \{1, 2, \dots, n\},$ $\mathcal{M} = \{1, 2, \dots, m\}, \mathcal{Q} = \{1, 2, \dots, q\}$
Preference set	$P = \{\tau_c \mid c \in \mathcal{N}\}$
Decision maker set	$\mathcal{O} = \{\mathcal{O}_g \mid g \in \mathcal{M}\}$
Evaluation vector of $\mathcal{O}_g$	$\tau^{\mathcal{O}_g} = (v_{g1}, v_{g2}, \dots, v_{gn})$
Evaluation matrix	$V = (v_{gc})_{m \times n}, g \in \mathcal{M}, c \in \mathcal{N}$
Preference membership degree matrix	$F = (f_{gc})_{m \times n}, g \in \mathcal{M}, c \in \mathcal{N}$
Similarity degree	$SD^{gh}, g, h \in \mathcal{M}$
Similarity threshold	θ
Subgroup SG_κ	$SG_\kappa, \kappa \in \mathcal{Q}$
Cohesion degree of subgroup SG_κ	$\mathbb{C}(SG_\kappa)$
Consensus degree of subgroup SG_κ on preference τ_c	$\mathcal{CD}_c^{SG_\kappa}, \kappa \in \mathcal{Q}, c \in \mathcal{N}$
Consensus threshold	δ
Discordant preference set of subgroup SG_κ	$NU P^\kappa$
Discordant subgroup set	NSG
Discordant decision maker	$\mathcal{O}'_h, h \in \{1, 2, \dots, l\}$
Decision maker weight	w
Adjustment cost of discordant decision maker $\mathcal{O}'_h$ of discordant subgroup SG_κ	$\text{Cost}_h^{SG_\kappa}$
Change value of discordant decision maker $\mathcal{O}'_h$'s preference τ_j	$\Delta \tau_j^{\mathcal{O}'_h}$

Then, real-valued data transformation method is as follows, which is a normalization method to obtain dimensionless data,

$$f_{gc} = \frac{v_{gc}}{\max_{c \in \mathcal{N}} v_{gc}}, \quad (1)$$

thus, the standardized decision makers' scoring matrix is transformed into a preference membership degree matrix, where f_{gc} denotes the membership degree of preference τ_c assigned by decision maker $\mathcal{O}_g$. Now, we give an example to understand this problem.

Example 2 Suppose there have 4 user preferences (UPs) as alternatives, i.e., preferences, $P = \{\tau_c \mid c = 1, 2, 3, 4\}$, and let $\mathcal{O} = \{\mathcal{O}_g \mid g = 1, 2, 3, 4, 5\}$ be 5 decision makers to score on preferences. In addition, $\tau_c^{\mathcal{O}_g}$ indicates score value of the decision maker $\mathcal{O}_g$ w.r.t. the preference τ_c . The UP scoring table is shown in Table 4. And the standardized UP scoring table is shown in Table 5, representing the membership degree of UPs.

Table 4 Table for user scoring on preferences

$\mathcal{O}/P$	τ_1	τ_2	τ_3	τ_4
$\mathcal{O}_1$	12	4	10	15
$\mathcal{O}_2$	7	9	2	7
$\mathcal{O}_3$	6	10	28	24
$\mathcal{O}_4$	4	12	16	12
$\mathcal{O}_5$	10	6	14	23

Table 5 Table for membership degree of UPs

$\mathcal{O}/P$	τ_1	τ_2	τ_3	τ_4
$\mathcal{O}_1$	0.8000	0.2667	0.6667	1.0000
$\mathcal{O}_2$	0.7778	1.0000	0.2222	0.7778
$\mathcal{O}_3$	0.2143	0.3571	1.0000	0.8571
$\mathcal{O}_4$	0.2500	0.7500	1.0000	0.7500
$\mathcal{O}_5$	0.4348	0.2609	0.6087	1.0000

4.2 Clustering and decision maker similarity

In this subsection, the similarity degree of preference score among decision makers is calculated combined with social network analysis (SNA). Then, we propose a double-restricted Louvain algorithm that divides decision maker networks into multiple subgroup, and the definition of modularity is introduced. We use the cohesion degree of subgroups to measure the clustering effect.

4.2.1 Similarity degree within decision makers

The similarity degree of decision makers is based on preference scores and represents the relationship between them. The closer the preference score of decision makers, the higher the similarity degree between them, and vice versa. We use Definition 15 to calculate the similarity degree of decision makers for preferences, which gets the similarity relationship network (SRN), denoted as $\mathcal{SD}$ while satisfying $\mathcal{SD}^{g^h} = \mathcal{SD}^{hg}(g, h \in \mathcal{M})$, that is, the similarity degree between decision maker $\mathcal{O}_g$ and decision maker $\mathcal{O}_h$ is equal to that of $\mathcal{O}_h$ and $\mathcal{O}_g$.

And similarity degree is based on Hausdorff distance [45], which measures the degree of mismatch between two sets by measuring the distance of the point of $\mathcal{O}_g$ that is farthest from any point of $\mathcal{O}_h$. If Hausdorff distance is d , then every point of $\mathcal{O}_g$ must be within a distance d of some point of $\mathcal{O}_h$. Thus, Hausdorff distance is more suitable for measuring the similarity distance between two sets, compared with distance summation.

Let $\tau^{\mathcal{O}_g} = (\tau_c^{\mathcal{O}_g})_{1 \times n}$, $\tau^{\mathcal{O}_g}$ is a one-dimensional matrix with n vectors, including all preference scores of a decision maker. In Example 2, $\tau^{\mathcal{O}_1} = (0.8, 0.2667, 0.6667, 1)$.

Definition 15 (Decision maker similarity degree) Decision maker similarity degree by the Hausdorff distance is

$$SD^{gh} = 1 - h(\mathcal{O}_g, \mathcal{O}_h), g, h \in \{1, 2, \dots, m\},$$

where $h(\mathcal{O}_g, \mathcal{O}_h) = h(\mathcal{O}_h, \mathcal{O}_g) = \max_{a \in \tau^{\mathcal{O}_g}} \min_{b \in \tau^{\mathcal{O}_h}} |a - b|$, so that $\max\{h(\mathcal{O}_g, \mathcal{O}_h), h(\mathcal{O}_h, \mathcal{O}_g)\} = h(\mathcal{O}_g, \mathcal{O}_h)$, and $a, b \in \tau^{\mathcal{O}_g}$ represent one preference membership degree of individual $\mathcal{O}_g$.

When the similarity degree between $\mathcal{O}_g$ and $\mathcal{O}_h$ is higher than threshold θ , it is considered that there is a similarity relationship among them. Correspondingly, we mark the edge between $\mathcal{O}_g$ and $\mathcal{O}_h$, otherwise, it has no similarity relationship. Therefore, a new SRN will be constructed. We can give an example to illustrate Definition 15.

Example 3 (Continued from Example 2) According to Table 5, the decision maker similarity degree is shown in Table 6. Set the similarity threshold $\theta = 0.85$, then a new decision maker similarity relationship is shown in Table 7.

Table 6 Decision maker similarity degree

$\mathcal{O}$	$\mathcal{O}_1$	$\mathcal{O}_2$	$\mathcal{O}_3$	$\mathcal{O}_4$	$\mathcal{O}_5$
$\mathcal{O}_1$	1.0000	0.8889	0.8096	0.9167	0.8087
$\mathcal{O}_2$	0.8889	1.0000	0.8651	0.9722	0.7874
$\mathcal{O}_3$	0.8096	0.8651	1.0000	0.8929	0.7516
$\mathcal{O}_4$	0.9167	0.9722	0.8929	1.0000	0.8152
$\mathcal{O}_5$	0.8087	0.7874	0.7516	0.8152	1.0000

Table 7 Decision maker similarity relationship

$\mathcal{O}$	$\mathcal{O}_1$	$\mathcal{O}_2$	$\mathcal{O}_3$	$\mathcal{O}_4$	$\mathcal{O}_5$
$\mathcal{O}_1$	\	1	0	1	0
$\mathcal{O}_2$	1	\	1	1	0
$\mathcal{O}_3$	0	1	\	1	0
$\mathcal{O}_4$	1	1	1	\	0
$\mathcal{O}_5$	0	0	0	0	\

4.2.2 Double-restricted Louvain clustering

The Louvain [46] is an algorithm for optimizing modularity that is a better community detection method, which can perform network clustering analysis with vertices and edges. The Louvain not only can eliminate the limitation of modular partition, but

also has low time complexity, which is suitable for large-scale networks [47]. The first iteration of Louvain takes a single node as the granularity of the community, avoiding the difficulty of finding small communities when maximizing, which is easy to merge small communities, because the modularity is the characteristic of the global indicator. This advantage can make up for the shortcomings of ignoring the opinions of minority groups, and will not integrate meaningful minority opinions into large groups. Minority opinions are sometimes meaningful viewpoints that larger groups do not value [48]. Just like when multiple decision makers make decisions, most decision makers can find common problems, but often a few decision makers raise sharp questions, which may finally win the support of most of them. This is also the advantage of the Louvain in clustering.

In this subsection, Louvain is selected to cluster community subgroups according to the SRN. In Louvain, the group to which a point belongs is changed by the maximum increment of modularity. However, the increment between two points is the largest, but only a few points are associated with it in the group which will be joined, and there is more relationship with another group at this time. For example, point o has the largest modularity increment with point j , so point o will join the subgroup SG_i where point j is located. However, the associated points of point o in SG_i only have point j , but there are multiple relationship points with other subgroup SG_p . If point o joins SG_i , grouping only according to the maximum increment index seems unreasonable. Therefore, we propose a double-restricted Louvain algorithm based on the limitation of maximum increment and relationship degree, and we aim to adjust the balance between them.

In Definition 16, a new maximum increment of weighted modularization is proposed, which added a relationship index compared with the classic Louvain algorithm, it pays attention to the relationship degree in the joined group. As shown in Algorithm 1, the double-restricted Louvain algorithm is divided into three stages: initialization, modularity optimization and network cohesion.

- (1) Initial setting. Initially, each decision maker is regarded as a subgroup, and the number of subgroups is the same as the number of decision makers.
- (2) Modularity optimization phase. According to Definition 16, it needs to calculate the gain in modularity ΔQ of each decision maker and the decision maker with whom has a consensus relationship, i.e., neighboring node, and select the subgroup where the neighboring node is located which has the largest of modularity gain. If the modularity gain is greater than 0, the decision maker is put into the subgroup where the decision maker has the largest modularity gain, otherwise it will not change. Iterate the above calculation until the algorithm is stable, that is, all decision makers belong no longer change.
- (3) Subgroups coalescence phase. All decision makers in each subgroup are compressed into a super node, the relationship of all nodes in the network is used as the relationship of super nodes, and the weights between subgroups are converted into the weights of new node edges. Continue to iterate step (2) until the subgroups to which all super nodes belong do not change.

Definition 16 (Modularity increment) When node o moves from its subgroup to the subgroup of its neighbor node j , the change of the incremental modularity is defined as follows,

$$\begin{aligned}\Delta Q &= W_{limit} \times \Delta Q' \\ &= \omega_{o,in} \times \left[\frac{\sum_{in} + \omega_{o,in}}{2v} - \left(\frac{\sum_{tot} + \omega_o}{2v} \right)^2 \right] - \left[\frac{\sum_{in}}{2v} - \left(\frac{\sum_{tot}}{2v} \right)^2 - \left(\frac{\omega_o}{2v} \right)^2 \right],\end{aligned}$$

where W_{limit} represents the relationship degree with the new subgroup, $\sum_{in}$ is the sum of the weights of links in subgroup, $\sum_{tot}$ is the sum of the weights of links associated to the nodes which in the subgroup, ω_o is the sum of the weights of links associated to node o , $\omega_{o,in}$ is the sum of the weights of links from node o to nodes which in the subgroup, v is the sum of the weights of all links in the group.

4.2.3 Cohesion of subgroups

The subgroup cohesion degree can measure the clustering effect of clustering model. We give the calculation of preference membership degree and cohesion degree of subgroups, calculated by Definitions 17 and 18, respectively.

Definition 17 [8] (Subgroup preference membership degree) The large-scale group $LG = \{SG_1, SG_2, \dots, SG_q\}$, is divided into multiple subgroups, the κ th subgroup is denoted as $SG_\kappa (\kappa \in \mathcal{Q})$. The membership degree of subgroup on preferences is τ^{SG_κ} ,

$$\begin{aligned}\tau^{SG_\kappa} &= (\tau_c^{SG_\kappa})_{1 \times n}, \\ \tau_c^{SG_\kappa} &= \frac{1}{p^\kappa} \sum_{g=1}^{p^\kappa} \tau_c^{\mathcal{O}_g^{SG_\kappa}},\end{aligned}$$

where $\mathcal{O}_g^{SG_\kappa} \in SG_\kappa, \kappa \in \mathcal{Q}, c \in \mathcal{N}$, p^κ represents that the subgroup SG_κ contains p decision makers.

Definition 18 [8] (Subgroup cohesion degree) The cohesion degree of subgroup is defined as

$$\mathbb{C}(SG_\kappa) = 1 - \frac{1}{n \cdot p^\kappa} \sum_{c=1}^n \sum_{g=1}^{p^\kappa} |\tau_c^{SG_\kappa} - \tau_c^{\mathcal{O}_g^{SG_\kappa}}|, \kappa \in \mathcal{Q}, c \in \mathcal{N}.$$

The average cohesion degree $\overline{\mathbb{C}}$ of the subgroup is defined as

$$\overline{\mathbb{C}} = \frac{1}{q} \sum_{\kappa=1}^q \mathbb{C}(SG_\kappa).$$

4.3 Consensus reaching process

In order to further realize LSGDM, it is necessary to complete the consensus reaching process (CRP) which can effectively unify the opinions of large group and accelerate LSGDM. Next, each subgroup will be regarded as a super decision maker for the next step of CRP analysis. The CRP is divided into two stages.

Algorithm 1 Louvain clustering method

Input: consensus relationship list $\mathcal{ST}$ is consisted of consensus relationship pairs $\langle source, target \rangle$.

Output: the clustering list $\mathcal{S}$.

```
1: Phase I: initialization
2: for  $tuple$  in  $\mathcal{ST}$  do
3:    $\mathcal{G}[node_i][node_j] = \mathcal{G}[node_j][node_i] = edge = 1$ 
4:    $\mathcal{S}[node_i] = node_i$ 
5: end for
6: return  $\mathcal{G}, \mathcal{S}$ 
7: Phase II: modularity optimization
8: #  $can\_stop$  marks the end of the iteration
9:  $can\_stop = \mathbf{True}$ 
10: while  $can\_stop$  do
11:    $can\_stop = \mathbf{False}$ 
12:   for  $i$  in  $\mathcal{G}$  do
13:     for  $j$  in  $\mathcal{G}[node_i]$  do
14:       compute: the modularity  $\Delta Q_{ij}$ 
15:     end for
16:     if  $\Delta Q_{ij} > 0$  and  $\mathcal{S}[node_i] \neq \mathcal{S}[node_j]$  then
17:        $can\_stop = \mathbf{True}$ 
18:       return  $j \leftarrow \mathbf{Max}(\Delta Q)$ 
19:     end if
20:      $\mathcal{S}[node_i] = \mathcal{S}[node_j]$ 
21:   end for
22: end while
23: Phase III: subgroup network cohesion
24: # Repeat Phase I
25: for  $i, j$  in  $SG$  do
26:    $\mathcal{G}[SG_i][SG_j] = \mathcal{G}[SG_j][SG_i] = \frac{\sum_{i,j} edge}{2}$ 
27:    $\mathcal{S}[SG_i] = SG_i$ 
28: end for
29: return  $\mathcal{G}, \mathcal{S}$ 
30:  $can\_stop = \mathbf{True}$ 
31: while  $can\_stop$  do
32:    $can\_stop = \mathbf{False}$ 
33:   Repeat Phase II
34:   if  $\mathbf{Sum}(\Delta Q_{ij}) = 0$  then
35:      $can\_stop = \mathbf{True}$ 
36:   end if
37: end while
38: return  $\mathcal{S}$ 
```

- (1) Consensus measurement. The consensus degree of subgroups is calculated, which is used to find discordant subgroups and the discordant preference of decision makers in them.
- (2) Feedback mechanism, which is based on low-cost and high-consensus strategy. According to the cost minimization while consensus maximization, we pay attention to the adjustment of decision maker's preference membership degree in the discordant subgroup. To minimize costs, a high consensus is obtained at a low cost. That is to say, we reach the minimum consensus threshold to obtain qualification to continue large-scale group decision-making. Based on different adjustment scales, we use the Particle Swarm Optimization (PSO) algorithm to complete the feedback mechanism strategy optimization. When we reach the subgroup minimum consensus threshold, then, CRP in LSG is completed.

4.3.1 Consensus measurement

The consensus degree of subgroup is measured via the membership consensus degree of each subgroup with other subgroups by Definition 19. Afterward, the LSG consensus degree, the average consensus degree of subgroup for each preference, denoted as $\mathcal{MCD}$ is obtained. The discordant subgroup NSG is identified by consensus threshold according to Definition 20.

Definition 19 [8] (Subgroup consensus degree) The consensus degree of subgroup on each preference is

$$\mathcal{CD}_c^{SG_\kappa} = 1 - \frac{1}{q-1} \sum_{\vartheta=1, \vartheta \neq \kappa}^q |\tau_c^{SG_\kappa} - \tau_c^{SG_\vartheta}|, \kappa, \vartheta \in \mathcal{Q}.$$

The consensus degree of subgroup is

$$\mathcal{MCD}^{SG_\kappa} = \frac{1}{n} \sum_{c=1}^n \mathcal{CD}_c^{SG_\kappa}. \quad (2)$$

LSG consensus degree is,

$$\mathcal{MCD} = \frac{1}{q} \sum_{\kappa=1}^q \mathcal{MCD}^{SG_\kappa}.$$

Definition 20 (Discordant measurement) The discordant subgroups are identified by the minimum consensus threshold δ ,

$$\mathcal{MCD}^{SG_\kappa} < \delta \implies SG_\kappa \in NSG,$$

where the set of discordant subgroups is NSG .

Next, the discordant preference set denoted as NUP^κ in the discordant subgroup SG_κ is found out,

$$\mathcal{CD}_c^{SG_\kappa} < \delta \implies \tau_c^{SG_\kappa} \in NUP^\kappa, SG_\kappa \in NSG,$$

where $\tau_c^{SG_\kappa}$ is the c th preference of SG_κ .

4.3.2 Feedback mechanism of maximum consensus at minimum cost strategy

When there is discordant behavior in the LSG consensus, in order to obtain the LSGDM result, it is necessary to encourage discordant subgroups to modify the preference score for achieving LSG consensus. We focus on the adjustment of preferences of decision makers in discordant subgroups, instead of methods that only attention to subgroup adjustment and ignore individuals.

In the feedback mechanism, we complete CRP by the optimization of low-cost and high-consensus strategy based on the PSO algorithm. CRP has multiple rounds of consensus testing. Smaller adjustment scales are allowed at the beginning. If consensus cannot be reached, we will expand the scale. Under the permission of different adjustment scale settings, the lowest adjustment cost is obtained through PSO, and then particles that meet the consensus conditions will be selected. With different adjustment scales and different adjustment plans, situations arise that have the same cost but different consensus levels. The feedback mechanism is as follows.

Step 1: Adjustment cost of decision maker. In order to reach the consensus threshold of the LSG, if all discordant decision makers are required to adjust the preference score membership degree in the same proportion, which will inevitably cause some decision makers to be dissatisfied. We use the status of decision makers in the LSG community network to deal with this situation. In reality, decisionmakers with status will have greater say. The adjustment cost is obtained according to Definition 22, where similarity degree is used as the weight of decision makers' adjustment cost. Decision makers with a higher degree of SRN have greater adjustment costs. the discordant preferences in the discordant subgroup are identified denoted as $\zeta (\zeta \in \{1, 2, \dots, a\})$.

Definition 21 (Decision makers' adjustment weight)

$$w_{\mathcal{O}'_h} = \frac{1}{m} \sum_g^m \mathcal{SD}^{hg},$$

where $\mathcal{O}'_h$ is the decision maker in the discordant subgroup.

Definition 22 (Adjustment cost) The adjustment cost $Cost_h^{SG_\kappa}$ of discordant decision maker $\mathcal{O}'_h$ of the discordant subgroup SG_κ is

$$Cost_h^{SG_\kappa} = \sum_{\zeta=1}^a \Delta_{\tau_\zeta}^{\mathcal{O}'_h} \times w_{\mathcal{O}'_h}, \mathcal{O}'_h \in SG_\kappa,$$

where a is the number of the discordant preferences of the discordant subgroup. The total adjustment cost of discordant subgroup SG_κ is

$$Cost^{SG_\kappa} = \sum_{h=1}^l Cost_h^{SG_\kappa}. \quad (3)$$

Step 2: Feedback mechanism. The low-cost and high-consensus feedback mechanism Model 4 is established. We take cost minimization as the main problem. Given

the upper and lower range of the adjustment of UP in the discordant subgroup's discordant preferences, the lowest adjustment cost is obtained, when the consensus level of discordant subgroup is guaranteed to reach the threshold. The pseudocode is shown in Algorithm 2.

$$\min \sum_{h=1}^l \sum_{\zeta=1}^a \Delta \tau_{\zeta}^{\mathcal{O}'_h} \cdot w_{\mathcal{O}'_h}$$

$$s.t. \begin{cases} \Delta \tau_c^{\mathcal{O}'_h} \in [lb_c^h, ub_c^h], c \in NUP, \\ \mathcal{MCD}^{SG_{\kappa}} \geq \delta \\ \max \mathcal{MCD}^{SG_{\kappa}} = \frac{1}{n} \sum_{c=1}^n \mathcal{CD}_c^{SG_{\kappa}}, c \in NUP, SG_{\kappa} \in NSG, \\ \mathcal{CD}_c^{SG_{\kappa}} = 1 - \frac{1}{q-1} \sum_{\vartheta=1, \vartheta \neq \kappa}^q |\tau_c^{SG_{\kappa}} \pm \Delta \tau_c^{SG_{\kappa}} - \tau_c^{SG_{\vartheta}}|, \kappa, \vartheta \in \mathcal{Q}, \end{cases} \quad (4)$$

Algorithm 2 The algorithm for consensus reaching process

Input: decision makers' preference membership matrix, discordant subgroups and corresponding discordant preferences, lb_c^h, ub_c^h

Output: adjusted decision makers' preference membership matrix, adjustment cost, new consensus level of discordant subgroups.

```

1: for  $i$  in  $[lb_c^h, ub_c^h]$  do
2:   for  $\kappa$  in  $NSG$  do
3:     for  $c$  in  $NUP$  do
4:       new_UP_matrix =  $\tau_c^{SG_{\kappa}} \pm \Delta \tau_c^{SG_{\kappa}}$ 
5:     end for
6:   end for
7:   return new_UP_matrix
8:   Calculate subgroup consensus level
9:   if  $\mathcal{CD}$  all  $\geq \delta$  then
10:    Calculate adjustment cost
11:    return  $\Delta \tau$  matrix, adjustment cost, subgroup consensus level
12:   else
13:    Continue
14:   end if
15: end for
16: Minimum adjustment cost when reaching consensus level
17: return  $\Delta \tau$  matrix, adjustment cost, new consensus level

```

4.3.3 Optimization of different adjustment scale strategy based on the PSO algorithm

Finding the optimal solution of a bi-level programming model has been proved to be a NP-hard problem, and even linear bi-level programming is strongly NP-hard [49]. Finding the optimal solution of a bi-level programming model has been proved to be a NP-hard problem, and even linear bi-level programming is strongly NP-hard. PSO is a group-based search method. This heuristic method makes PSO effective in finding the global optimal solution [50], especially in high-dimensional and complex search spaces. We use the PSO algorithm to solve the optimal solution for low-cost and high-consensus under different adjustment scale strategy.

In PSO, the velocity $v_p(t)$ and position $x_p(t)$ of each particle are initialized to produce a random solution. Then, in the evolutionary process, the particle updates its velocity and position by the following rule,

$$\begin{cases} v_p(t) = w \times v_p(t) + c_1 \times \text{rand}() \times (pbest_p(t) - x_p(t)) + c_2 \times \text{rand}() \times (gbest(t) - x_p(t)), \\ x_p(t) = x_p(t) + v_p(t), \end{cases}$$

where w is the inertia weight, c_1 and c_2 are the acceleration constants, which are suggested to set value 2.0 [51], $\text{rand}() \in [0, 1]$ is random value, $pbest_p(t)$ represents the best fitness at present particle, and $gbest(t)$ represents the overall best position in the t th generation. And, the inertia weight w adopts linearly decreasing weight,

$$w = \frac{(w_{\text{init}} - w_{\text{end}})(T - t)}{T} + w_{\text{end}},$$

where T is the maximum number of iterations, w_{init} and w_{end} are usually set to 0.9 and 0.4, respectively [51]. The foundational setting based on PSO are shown in Table 8, and PSO algorithm is designed as Algorithm 3.

Table 8 Fundamental setting based on PSO

	Fundamental setting
Particle swarm number	pop = particles number produced at one time
One particle dimension 1	obj = discordant decision maker number
One particle dimension 2	dim = discordant preference number
Particle dimension	$D = [\text{pop}, \text{obj}, \text{dim}]$
p th particle in r th generation	$x_{pr} = \begin{bmatrix} w_1; & x_{p1,1} & \dots & x_{p1,c} & \dots & x_{p1,n} \\ \vdots & \vdots & & \ddots & & \vdots \\ w_h; & x_{ph,1} & \dots & x_{ph,c} & \dots & x_{ph,n} \\ \vdots & \vdots & & \ddots & & \vdots \\ w_l; & x_{pl,1} & \dots & x_{pl,c} & \dots & x_{pl,n} \end{bmatrix}$
Search space	$x_{pr} \in [\text{lb}_{pr}, \text{ub}_{pr}]$
Fitness function	cost = Eq. (3)
Consensus level	CD = Eq. (2)

Algorithm 3 PSO for minimum cost on different adjustment scale

Input: Particle dimension D , search space ub, lb , maximum number T of iterations, default parameters $c1, c2, w_max, w_min$, specified parameter $V_max, range_step$.

Output: Particle and minimum cost value of different scale of adjustment range.

- 1: **step 1:** Initialize the particle swarm with random positions and velocities within the defined search space. Calculate $pbest_p(t)$ and $gbest(t)$
 - 2: **step 2:** Calculate the inertia weight w for the current iteration, and update the velocities and positions of the particles.
 - 3: **step 3:** Perform boundary checking for the updated velocities and positions.
 - 4: **if** $v_p(t) > V_max$ **then**
 - 5: $v_p(t) = \text{rand}() \times V_max$
 - 6: **end if**
 - 7: **step 4:** Calculate the fitness function of each particle, and sort the particle swarm according to the fitness value.
 - 8: $pbest_p(t) = \text{particle } x \text{ at position } 1$
 - 9: **step 5:** Update the global best position.
 - 10: **if then** $pbest_p(t) > gbest$
 - 11: $gbest = pbest_p(t)$
 - 12: **end if**
 - 13: Repeat Step 2 to Step 4 until the maximum number T of iterations is reached.
 - 14: **step 6:** Change the different scale of adjustment range, repeat Step 1 to Step 5.
 - 15: **return** Particle and minimum cost value of different scale of adjustment range.
-

4.4 Fuzzy covering-based rough set models

In this subsection, in order to evaluate the weight of decision makers, we use four fuzzy neighborhood operators based on fuzzy covering with three different t -norms and R -implicators, and two fuzzy covering-based rough set models which are fuzzy β -neighborhood and fuzzy complementary β -neighborhood that construct the lower approximation and upper approximation. Approximate precision is proposed via lower and upper approximation of six types of neighborhood-related fuzzy covering-based rough set models, which is as weight of decision makers. And in this paper, covering-based fuzzy rough set carries UP information, so the weight of decision makers obtained is based on UP. In multi-attribute decision-making, covering-based fuzzy rough set models are often used to rank the importance of objects or alternatives, which have been proven to be a good solution.

4.4.1 Fuzzy neighborhood operators based on fuzzy covering

The four fuzzy neighborhood operators based on fuzzy covering proposed by D'eer with three different t -norms and R -implicators are discussed to construct the lower approximation and upper approximation. The specific process steps are shown below.

- (1) In order to satisfy the condition of fuzzy covering by Definition 4, standardize consensual decision makers' score fuzzy set according to the following equation [44],

then we get new decision maker's preference score fuzzy set.

$$s_{gc} = \frac{f_{gc}}{\max_{g \in \mathcal{M}} f_{gc}}. \quad (5)$$

- (2) When taking the fuzzy neighborhood operators, the minimum and maximum descriptions of each preference need to be calculated first, as shown in Algorithm 4.4.1.

Algorithm 4 Fuzzy minimal description md_{x_i} and fuzzy maximal description MD_{x_i}

Input: decision makers' preference fuzzy set K of fuzzy covering $\mathcal{O} = \{\mathcal{O}_j \mid j = 1, 2, \dots, m\}$ about universe $U = \{\tau_c \mid c = 1, 2, \dots, n\}$.
Output: md_{τ_j} and MD_{τ_j} of each preference τ_j .

```

1:  $r, c = K.shape$ ;
2: # Compute the fuzzy neighborhood system of  $\tau_j$ 
3: for  $j$  in  $c$  do
4:    $K_{\tau_j} = []$ ;  $q = 1$ 
5:   for  $i$  in  $r$  do
6:     if  $K[i, j] > 0$  then
7:        $K_{\tau_j}[q, :] = K[i, :]$ 
8:        $q = q + 1$ 
9:     end if
10:  end for
11:   $r_j = K_{\tau_j}[1].shape$ 
12:  Initialize:  $K_{\tau_j}^{md} = K_{\tau_j}$ ;  $K_{\tau_j}^{MD} = K_{\tau_j}$ ;
13:  for  $s$  in  $r_j$  do
14:    for  $t$  in  $r_j$  do
15:      if  $i \neq j$  and  $(K_{\tau_j}^{md}[s, j] == K_{\tau_j}^{md}[t, j])$  and  $(sum(K_{\tau_j}^{md}[:, :] \geq K_{\tau_j}^*[t, :]) == n)$  then
16:         $sum(K_{\tau_j}^{md}[s, :] = zeros(1, n))$ 
17:      else if  $i \neq j$  and  $(K_{\tau_j}^{MD}[s, j] == K_{\tau_j}^{MD}[t, j])$  and  $(sum(K_{\tau_j}^{MD}[:, :] \leq K_{\tau_j}^*[t, :]) == n)$  then
18:         $sum(K_{\tau_j}^{MD}[s, :] = zeros(1, n))$ 
19:      end if
20:    end for
21:  end for
22:  Compute:  $md_{\tau_j} = K_{\tau_j}^{md}(any(K_{\tau_j}^*, 2), :)$ 
23:  Compute:  $MD_{\tau_j} = K_{\tau_j}^{MD}(any(K_{\tau_j}^*, 2), :)$ 
24: end for

```

- (3) The fuzzy neighborhood operators $N_1^C(x)$ to $N_4^C(x)$ are obtained by bringing the fuzzy minimum (maximum) description into Algorithm 4.4.1.

Algorithm 5 Fuzzy neighborhood operators

Input: The decision makers' preference fuzzy set K , md_{τ_j} , MD_{τ_j} of each object τ_j .

Output: Four fuzzy neighborhood operators with three different t-norms and R-implicators.

```
1:  $r, c = K.shape$ 
2: # Compute the fuzzy neighborhood operator  $N_k^C$ 
3: if  $k == 1$  then
4:   for  $i$  in  $c$  do
5:     for  $j$  in  $c$  do
6:       for  $t$  in  $r$  do
7:          $D = \mathcal{J}(K[t, \tau_i], K[t, \tau_j])$ 
8:       end for
9:        $N_k^C(\tau_i)(\tau_j) = \min(D)$ 
10:    end for
11:  end for
12: else if  $k == 2$  then
13:   for  $i$  in  $c$  do
14:      $r_m = md_{\tau_i}.shape[1]$ 
15:     for  $j$  in  $c$  do
16:       for  $t$  in  $r_m$  do
17:          $D = \mathcal{T}(md_{\tau_i}[t, \tau_i], md_{\tau_i}[t, \tau_j])$ 
18:       end for
19:        $N_k^C(\tau_i)(\tau_j) = \max(D)$ 
20:     end for
21:   end for
22: else if  $k == 3$  then
23:   for  $i$  in  $c$  do
24:      $r_M = MD_{\tau_i}.shape[1]$ 
25:     for  $j$  in  $c$  do
26:       for  $t$  in  $r_M$  do
27:          $D = \mathcal{J}(MD_{\tau_i}[t, \tau_i], MD_{\tau_i}[t, \tau_j])$ 
28:       end for
29:        $N_k^C(\tau_i)(\tau_j) = \min(D)$ 
30:     end for
31:   end for
32: else if  $k == 4$  then
33:   for  $i$  in  $c$  do
34:     for  $j$  in  $c$  do
35:       for  $t$  in  $r$  do
36:          $D = \mathcal{T}(K(t, \tau_i), K(t, \tau_j))$ 
37:       end for
38:        $N_k^C(\tau_i)(\tau_j) = \max(D)$ 
39:     end for
40:   end for
41: end if
```

4.4.2 Fuzzy β -neighborhood and complementary β -neighborhood operators

In the calculation of fuzzy β -neighborhood operator and fuzzy complementary β -neighborhood operator, we use consensual decision makers' preference fuzzy set, and the specific process steps are shown below.

- (1) Set the value of β .
- (2) By Definitions 9 and 11, the fuzzy β -neighborhood operator and fuzzy complementary β -neighborhood operator are obtained, respectively.

4.4.3 Fuzzy covering-based lower approximation and upper approximation

The lower approximation and upper approximation constitute the fuzzy covering-based rough set model. We use four fuzzy neighborhood, fuzzy β -neighborhood and complementary β -neighborhood operators to construct lower and upper approximations separately. And 14 upper (lower) approximation matrices will be obtained.

4.5 LSGDM on evaluation of preferences

In the subsection, approximate precision is used to obtain the weight score of preferences. Then, we use the fuzzy TOPSIS method to get the importance ranking of preferences, which is the result of LSGDM.

4.5.1 Approximate precision

We propose the concept of approximate precision combined with the fuzzy covering-based rough set models. Approximate precision $\mathcal{P}$ is defined as the ratio of lower approximation to upper approximation. The approximate precision of fuzzy neighborhood operators is

$$\mathcal{P}_i^j(C_k) = \frac{|\underline{\mathbb{A}}_i^j(C_k)|}{|\overline{\mathbb{A}}_i^j(C_k)|}, k \in 0, 1, \dots, n,$$

where $i = 1, 2, 3, 4$ and $j = 1, 2, 3$ represent four fuzzy neighborhood operators and three t -norms or R -implicators, and $\mathbb{C} = \{C_1, C_2, \dots, C_m\}$ is an fuzzy β -covering of U .

Then, the decision makers weight vector $\mathcal{W}$ about approximate precision $\mathcal{P}$ is defined as follows,

$$\mathcal{W}_{\mathcal{P}_i^j} = \frac{\mathcal{P}_i^j(C_k)}{\sum_{k=0}^l \mathcal{P}_i^j(C_k)}. \quad (6)$$

4.5.2 Fuzzy TOPSIS model

The fuzzy TOPSIS method is used to calculate the ranking coefficient of preferences, and finally we obtain the preference importance of LSG. The specific process steps are shown below.

- (1) TOPSIS is used to confirm the order of preferences in virtue of the positive ideal solution and the negative ideal solution [52]. Let ψ^+ and ψ^- be defined as,

$$\begin{cases} \psi^+ = [\vee\{\psi_1^1, \psi_1^2, \dots, \psi_1^n\}, \vee\{\psi_2^1, \psi_2^2, \dots, \psi_2^n\}, \dots, \vee\{\psi_m^1, \psi_m^2, \dots, \psi_m^n\}] \\ \quad = [\psi_1^+, \psi_2^+, \dots, \psi_m^+], \\ \psi^- = [\wedge\{\psi_1^1, \psi_1^2, \dots, \psi_1^n\}, \wedge\{\psi_2^1, \psi_2^2, \dots, \psi_2^n\}, \dots, \wedge\{\psi_m^1, \psi_m^2, \dots, \psi_m^n\}] \\ \quad = [\psi_1^-, \psi_2^-, \dots, \psi_m^-], \end{cases}$$

where n is the number of preferences, and m is the number of decision makers.

- (2) The distances of each decision makers from the positive ideal solution and the negative ideal solution are as follows,

$$\begin{cases} D_c^+ = \|\psi^+ - \psi^c\| = \psi^+ - \psi^c, \\ D_c^- = \|\psi^c - \psi^-\| = \psi^c - \psi^-, \end{cases}$$

where $c \in \{1, 2, \dots, n\}$.

- (3) Then, the closeness coefficient of preferences can be defined combined with the D^+ and D^- , as follows,

$$\mathcal{R} = \frac{\sum_{i=1}^m \mathcal{W}_c D_{c,i}^-}{\max_{1 \leq c \leq n} (\sum_{i=1}^m \mathcal{W}_c D_{c,i}^-)} - \frac{\sum_{i=1}^m \mathcal{W}_c D_{c,i}^+}{\min_{1 \leq c \leq n} (\sum_{i=1}^m \mathcal{W}_c D_{c,i}^+)}.$$

- (4) Finally, we rank the importance of preferences according to the ranking closeness, if $\mathcal{R}_c > \mathcal{R}_d$, the importance of c th preferences is higher than the d th preferences in LSG, recorded as $C_c \succ C_d$. It means that the larger the value, the higher the ranking.

5 An online reviews-driven LSGDM for evaluating UP of fresh fruit on e-commerce platform

This section uses the LSGDM method with fuzzy covering-based rough set models and fuzzy TOPSIS to evaluate the UPs on JD platform for purchasing fresh fruit. In Section 5.1 describes the extraction process of UP feature keywords and the data preprocessing process in detail. Section 5.2 introduces the construction of UP fuzzy set and the clustering results. Section 5.3 is the consensus reaching process and uses PSO to solve maximum consensus at minimum cost to adjust the discordant decision makers and subgroups. The final decision-making process based on large-scale group is presented in Section 5.4. Finally, Section 5.5 implements the resulting approach in a platform-based intelligent decision-making, enabling interactive question answering and decision-making support.

5.1 User preference feature

User online reviews effectively convey UPs and concerns that are characterized by high volume and low value density, which prevent the preference information of most

users from being quickly summarized. Moreover, unstructured text cannot be directly applied to decision analysis. Therefore, when analyzing unstructured and massive online text reviews, the most important process is to convert text into structured data that is easy to analyze, to facilitate further analysis of user preference features. In this section, empirical analysis is used to conduct information classification and preference research on user online reviews on e-commerce platforms.

The JD e-commerce platform not only has its shop, but also many merchants and suppliers of agricultural products have registered. New shopping methods such as live streaming, group buying, etc., make the development of the platform more diversified, which give play to increase sales volume for seller of fresh fruit. However, it is found that the online reviews of the platform products can reflect the user's purchase experience, which effectively conveys users' emotions. Compared with positive reviews, negative reviews focus more on the outstanding problems [53], which better reflect UPs and users' unsatisfactory demands. Therefore, it is representative and practical to analyze the negative reviews to explore the user needs of consumers to buy fresh fruit on JD. The product information of fresh fruits is obtained based on JD, which includes product id, user id, product title, comment time, negative review text, product description information and store id.

From the data collected, there is a large amount of meaningless information. In addition, the low volume products have the problem of too few reviews, which will not be enough to extract UP features and will be eliminated. (1) Remove meaningless data. Remove empty data, system default reviews, meaningless characters, non-Chinese characters, and repeated reviews. (2) Filter data. Complete the intercept according to the time. Reviews will be removed which is under two characters. Product reviews greater than 100 will be retained. (3) Remove invalid characters. Invalid comment content can be effectively removed, such as emoticons and so on. Finally, the number of products retained is 181, and the effective reviews are 48,601. Next, user preference feature keywords are extracted further.

The product review data can reflect the key points that users pay attention to when buying fresh fruit online, such as logistics, service, etc., which are the UPs characteristics. That user reviews are converted from unstructured text to structured text is the primary task in text preprocessing. UP keywords are extracted and verified, respectively.

First, feature extraction. The UP keywords and word frequency, denoted as a tuple pair $kf = \langle kw, fre \rangle$, are extracted from the effective reviews filtered, where kw is the keyword and fre is the word frequency. The extraction flow chart of UP keywords is shown in Figure 2, and Algorithm 6 is the process of text preprocessing.

- (1) Merge entries. Merge all user reviews C into one entry information C^{all} .
- (2) The Jieba splits words. Python Jieba package is used to split C^{all} that gets C^{jieba} .
- (3) Remove the stop words. The invalid words and characters are removed by referring to the stop words dictionary (the sentiment analysis Chinese glossary on CNKI).
- (4) Count word frequency. Word frequency statistics are performed for the selected keywords and a list of $kfs = \{kf_1, kf_2, \dots, kf_z\}$ is got, where kfs is composed of z tuple pair $\langle kw, fre \rangle$.
- (5) Sort. These tuples z are sorted according to the number of keyword frequencies.

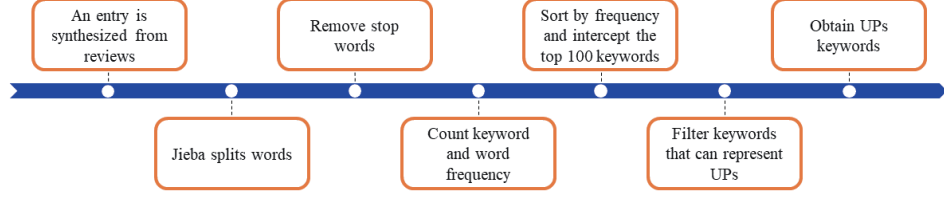


Fig. 2 The extraction flow chart of UP keywords

We select the keywords which can represent UPs from the top 100 words by word frequency, as a total of 25 UP keywords, i.e., $P = \{UP_1, UP_2, \dots, UP_{25}\}$. The proportions of preference word frequency can be obtained in the total word frequency, as shown in Table 9.

Algorithm 6 User preference keyword extraction method

Input: online review data C after data preprocessing

Output: the tuples kfs of UP keywords and word frequency sorted by word frequency

```

1:  $C^{all} \rightarrow \text{String}$ ,  $C^{jieba} \rightarrow \text{List}$ ,  $C^{words} \rightarrow \text{List}$ 
2: for  $comment$  in  $C$  do
3:    $C^{all} \cdot \text{join}(comment)$ 
4: end for
5: return  $C^{all}$ 
6: for  $string$  in  $C^{all}$  do
7:   apply Jieba method
8: end for
9: return  $C^{jieba}$ 
10: for  $word$  in  $C^{jieba}$  do
11:   apply stop words dictionary
12: end for
13: return  $C^{words}$ 
14: for  $word$  in  $C^{words}$  do
15:   compute: words frequency
16: end for
17: return user preference keywords and word frequency,  $kfs = \{\langle kw_1, fre_1 \rangle, \langle kw_2, fre_2 \rangle, \dots, \langle kw_z, fre_z \rangle\}$ 
18: sorted  $kfs$  with  $fre$  as the key value
19: return  $kfs$ 
  
```

Second, feature verification. This subsection verifies whether the 25 UP features are important concerns of users. TF-IDF model is used to verify the importance of selected preference keywords in Table 9, and the process is shown in Algorithm 7.

- (1) Split words by Jieba and remove the stop words. Each good review is split by Jieba package, then the stop words are removed to obtain the list C^{pre} .

- (2) Convert word to vector. The obtained keywords are processed by using *CountVectorizer* method under Python's *scikit-learn* library for text vectorization, and the vectorization matrix *word2vec* is obtained. The primary purpose is to convert the preference words into a series of vectors that machines can recognize, which is the pretreatment process of TF-IDF.
- (3) TF-IDF model. The *TfidfTransformer* method under the same library is used for TF-IDF modeling, and the weights of different word vectors in each entry are calculated which makes the word vector weight matrix *weight_matrix* is obtained.
- (4) Calculate keyword weight. The weight of TF-IDF in each item by TF-IDF is added to get the weight of the word vector as the keyword weight and construct tuple pairs $\langle kw, wgt \rangle$ of keywords and weights.
- (5) Sort. The keywords are sorted to obtain *kts* composed of z tuple pairs by the keyword weight. The cloud map of weighted words is generated according to the top 100 keywords. Table 9 is shown the ranking number of keyword frequency and TF-IDF weight, respectively.

It can be found that the ranking of keyword importance fluctuates little from the ranking of the two groups of keywords. It is worth mentioning that the rankings of keyword importance “too small” and “not ripe” increases significantly in TF-IDF. In summary, the selected preference keywords are also highly important in the TF-IDF algorithm, and all of them are in the top 100, indicating that the 25 keywords selected in Table 9 have high credibility.

Algorithm 7 Preference keywords verification method

Input: online review data C after data preprocessing

Output: TF-IDF's preference keywords and weight ranking

$kts = \{\langle kw_1, wgt_1 \rangle, \langle kw_2, wgt_2 \rangle, \dots, \langle kw_z, wgt_z \rangle\}$

```

1: for comment in  $C$  do
2:   apply Jieba and stop words method
3: end for
4: return  $C^{pre} \rightarrow$  List
5: for word in  $C^{pre}$  do
6:    $word2vec \leftarrow$  word vector transformation
7:    $weight\_matrix \leftarrow$  TF-IDF algorithm
8: end for
9: return  $weight\_matrix$ 
10: sum the weight of keywords
11: sort by keyword weight
12: return the list  $kts$  included keyword and ranking of weight

```

5.2 User preference fuzzy set and Louvain clustering

This subsection is to construct user preference fuzzy set (PFS) based on user preference keywords and word frequency, then complete Louvain clustering.

Table 9 User preference information, keywords frequency and TF-IDF weight ranking

UPs	UP keywords	Frequency	Proportion	Ranking	
				Frequency	TF-IDF weight
UP_1	bad fruit	1954	3.99%	33	26
UP_2	customer service	4514	11.12%	6	3
UP_3	after-sale	1448	2.96%	44	38
UP_4	flavor	2484	5.08%	24	18
UP_5	taste	1611	3.29%	41	31
UP_6	size	1172	2.40%	64	48
UP_7	arrival	940	1.92%	83	73
UP_8	delivery	1182	5.05%	25	22
UP_9	very small	978	2.00%	79	59
UP_{10}	moisture	1345	2.75%	51	27
UP_{11}	quality	1415	2.89%	47	43
UP_{12}	store	4703	9.23%	10	7
UP_{13}	logistics	2471	2.42%	63	52
UP_{14}	JD	5437	9.61%	8	4
UP_{15}	return	882	1.80%	95	80
UP_{16}	fresh	3107	6.35%	16	10
UP_{17}	express	2816	5.76%	21	16
UP_{18}	box	937	1.92%	85	62
UP_{19}	unpalatable	2629	5.37%	23	12
UP_{20}	package	1705	3.49%	38	33
UP_{21}	not sweet	1416	2.89%	46	30
UP_{22}	not ripe	895	1.83%	94	49
UP_{23}	refund	937	1.92%	84	77
UP_{24}	level	1004	2.05%	76	58
UP_{25}	too small	933	1.91%	86	53

5.2.1 User preference fuzzy set

In Section 4, we get the review information of 181 fresh fruit products through data pre-processing. We regard 181 products as independent individuals $\mathcal{O} = \{\mathcal{O}_1, \mathcal{O}_2, \dots, \mathcal{O}_m\}$, i.e., decision makers. And the frequency of the preference keyword which is obtained from the individual review information, is standardized as the membership degree of the individual preference features. The selected preference features are recorded as $\{\tau_c | c = 1, 2, \dots, n\} \in P$. The steps of constructing PFS are as follows.

- (1) Multiple reviews of the one product are integrated into one entry, thus 181 reviews can be obtained, and text preprocessing (word segmentation and removal of stop words) is performed on each entry.
- (2) Calculate the word frequency of the preference keyword contained in the 25 preference features.
- (3) According to Eq. (1), the membership degree of each individual to each preference feature is calculated.
- (4) Construct the PFS matrix of individual preference features.

The same as the method in Algorithm 6, carry out text preprocessing on each individual reviews, realize word segmentation and remove stop words. Next, that individual preference keywords kw and word frequency fre are counted is denoted as $\mathcal{O}_{81}$ to $\mathcal{O}_{90}$ in Table 10. And the PFS matrix is constructed, shown in Table 11.

Table 10 Keywords and word frequency of individuals

$\mathcal{O}$	UP_1	UP_2	UP_3	UP_4	UP_5	UP_6	UP_7	UP_8	...
81	0	5	0	5	5	5	4	9	...
82	16	21	4	11	8	5	7	3	...
83	2	8	0	12	7	3	1	4	...
84	4	5	7	13	7	10	3	5	...
85	3	3	3	5	4	4	1	6	...
86	6	14	2	8	5	3	3	8	...
87	46	62	18	30	51	18	10	28	...
88	7	23	4	12	1	3	4	13	...
89	2	5	4	10	4	2	0	2	...
90	23	10	15	7	0	5	0	2	...

Table 11 The membership degree of preferences in individuals

$\mathcal{O}$	UP_1	UP_2	UP_3	UP_4	UP_5	UP_6	UP_7	UP_8	...
81	0.0000	0.2381	0.0000	0.2381	0.2381	0.2381	0.1905	0.4286	...
82	0.2857	0.3750	0.0714	0.1964	0.1429	0.0893	0.1250	0.0536	...
83	0.0541	0.2162	0.0000	0.3243	0.1892	0.0811	0.0270	0.1081	...
84	0.1481	0.1852	0.2593	0.4815	0.2593	0.3704	0.1111	0.1852	...
85	0.1111	0.1111	0.1111	0.1852	0.1481	0.1481	0.0370	0.2222	...
86	0.3000	0.7000	0.1000	0.4000	0.2500	0.1500	0.1500	0.4000	...
87	0.4107	0.5536	0.1607	0.2679	0.4554	0.1607	0.0893	0.2500	...
88	0.3043	1.0000	0.1739	0.5217	0.0435	0.1304	0.1739	0.5652	...
89	0.1429	0.3571	0.2857	0.7143	0.2857	0.1429	0.0000	0.1429	...
90	0.7667	0.3333	0.5000	0.2333	0.0000	0.1667	0.0000	0.0667	...

5.2.2 Louvain clustering

The degree of preference consensus among individuals is calculated by social network analysis (SNA). And Louvain algorithm cluster individuals with social network similarity.

Definition 15 is used to calculate the consensus degree among 181 individuals, which gets a similarity degree matrix $\mathcal{SD}$ with a size of 181×181 , that is, the individual similarity relationship network (SRN). It believes that when the similarity degree is greater than threshold θ , there is a new SRN between individuals. Let the similarity threshold $\theta = 0.91$ which is strictly required, and if the similarity degree (SD) is higher than 0.91 among individual $\mathcal{O}_g$ and $\mathcal{O}_h$, then it is considered that there is a preference

relationship among them. Correspondingly, we mark the $\langle g, h \rangle$ with existence of preference relationship that similarity degree as 1 in the SRN matrix, otherwise it is 0. Therefore, a new SRN matrix denoted as $\mathcal{SD}'$ will be constructed, which is visualized using Gephi software, as shown in Figure 3, lines between individuals represent they have the similarity relationship. The cluster's marks and the number of individuals in the cluster are also indicated. And, individuals of the same color are the same subgroup. By Louvain community clustering, the LSG is clustered into 4 subgroups, and the number of individuals included is 78, 53, 46, and 4 respectively.

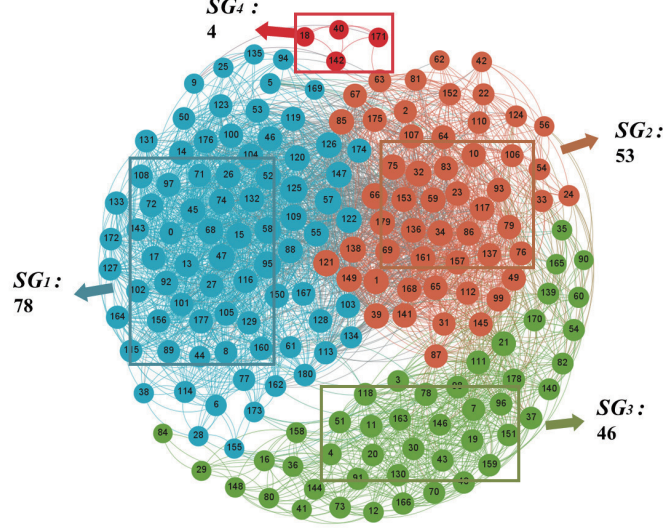


Fig. 3 Individual consensus relationship network

5.2.3 Cohesion of subgroups

According to Definition 17, the membership degree of four subgroups to 25 preference features can be calculated, we get the subgroup preference matrix that the size is 4×25 . And the degree of cohesion of the subgroups is obtained, as shown in Table 12. Then the average cohesion of subgroups $\bar{C}$ is cohesion of the LSG, which reaches more than 0.85, which means that the clustering effect on preferences is better.

Table 12 Subgroup cohesion degree

	SG_1	SG_2	SG_3	SG_4	LSG
$\bar{C}$	0.8443	0.8593	0.8699	0.8576	0.8578

5.3 Consensus reaching process

5.3.1 Consensus measurement

By Definition 19, consensus degree of subgroups is measured, meanwhile, the LSG consensus degree is obtained. Next, suppose the minimum consensus threshold δ is 0.92, the discordant subgroup and its discordant preference features of it are identified by Definition 20, which is shown in Table 13. It can see SG_3 did not reach the consensus threshold, and further analyzes the preferences consensus degree of SG_3 , as shown in Table 14, the discordant preferences are $\{UP_1, UP_2, UP_4, UP_8, UP_{10}, UP_{12}, UP_{14}, UP_{16}, UP_{17}, UP_{19}, UP_{21}\}$ that is 10 total, which marked in bold in table.

Table 13 Subgroup consensus degree

	SG_1	SG_2	SG_3	SG_4	LSG
CD	0.9263	0.9444	0.9189	0.9223	0.9280

Table 14 The preference consensus degree of discordant subgroup

	UP	CD	UP	CD	UP	CD	UP	CD	UP	CD
SG_3	1	0.9117	2	0.8413	3	0.9645	4	0.8635	5	0.9334
	6	0.9599	7	0.9422	8	0.9060	9	0.9809	10	0.8571
	11	0.9511	12	0.8301	13	0.9523	14	0.8148	15	0.9574
	16	0.9183	17	0.8564	18	0.9503	19	0.8510	20	0.9730
	21	0.9085	22	0.9526	23	0.9613	24	0.9626	25	0.9722

5.3.2 Feedback mechanism

In feedback mechanism, we design PSO algorithm to solve maximum consensus at minimum cost in combination with adjusted weights assigned by community relation degree, and get the change value and after adjusted value of each discordant preference of discordant subgroup SG_3 according to Model 4, and calculate the cost of discordant subgroup.

The results are shown in Table 15, $\Delta_{7_j}^{SG_3}$ is the adjustment value of subgroup preference, where the weight of individuals is shown in Table 16. And the subgroup's new consensus degree $\tilde{CD}$ and the new cohesion degree $\tilde{C}$ are calculated. Then, we find that the cohesion degree of SG_3 is a slight improvement, and the consensus degrees of the adjusted subgroups all reached the consensus threshold. In the end, the LSG reaches the consensus level, and we get the adjusted preference degree of individual, named consensual UP (CUP) fuzzy set.

Table 15 Adjustment value and cost of discordant preferences of discordant subgroup

Discordant subgroup	$\Delta\tau_j^{SG_\kappa}$	$\widetilde{CD}$	$\widetilde{C}$	Cost
SG_3	3.2287	0.9201	0.8717	2.6981

Table 16 Social network association degree of individual

individual	social network association degree											
$\mathcal{O}_1\text{-}\mathcal{O}_{11}$	0.8770	0.8834	0.8745	0.8656	0.8273	0.8689	0.8598	0.8525	0.8727	0.8596	0.8784	
$\mathcal{O}_{12}\text{-}\mathcal{O}_{22}$	0.8534	0.8274	0.8733	0.8640	0.8692	0.7858	0.8677	0.7942	0.8523	0.8294	0.8735	
$\mathcal{O}_{23}\text{-}\mathcal{O}_{33}$	0.8711	0.8791	0.8660	0.8611	0.8656	0.8823	0.8422	0.7570	0.8456	0.8754	0.8775	
$\mathcal{O}_{34}\text{-}\mathcal{O}_{44}$	0.8716	0.8784	0.8615	0.7952	0.8575	0.8539	0.8851	0.7872	0.783	0.8415	0.8477	
$\mathcal{O}_{45}\text{-}\mathcal{O}_{55}$	0.8842	0.8710	0.8726	0.8669	0.8443	0.8770	0.8579	0.8337	0.8705	0.8645	0.8652	
$\mathcal{O}_{56}\text{-}\mathcal{O}_{66}$	0.8768	0.8546	0.8806	0.8749	0.8771	0.8601	0.8812	0.8693	0.8659	0.8778	0.8784	
$\mathcal{O}_{67}\text{-}\mathcal{O}_{77}$	0.8784	0.8759	0.8691	0.8755	0.8453	0.8661	0.8627	0.8106	0.8736	0.8786	0.8771	
$\mathcal{O}_{78}\text{-}\mathcal{O}_{88}$	0.8689	0.8616	0.8792	0.7912	0.8719	0.8574	0.8773	0.7322	0.8744	0.8787	0.8692	
$\mathcal{O}_{89}\text{-}\mathcal{O}_{99}$	0.8714	0.8732	0.8300	0.8157	0.8684	0.8789	0.8652	0.8784	0.8572	0.8703	0.8585	
$\mathcal{O}_{100}\text{-}\mathcal{O}_{110}$	0.8754	0.8698	0.8715	0.8630	0.8761	0.8722	0.8746	0.8771	0.8763	0.8646	0.8739	
$\mathcal{O}_{111}\text{-}\mathcal{O}_{121}$	0.8744	0.8674	0.8772	0.8734	0.8614	0.8702	0.8684	0.8799	0.8662	0.8694	0.8716	
$\mathcal{O}_{122}\text{-}\mathcal{O}_{132}$	0.8782	0.8772	0.8553	0.8699	0.8792	0.8773	0.8645	0.874	0.8732	0.8378	0.8615	
$\mathcal{O}_{133}\text{-}\mathcal{O}_{143}$	0.8752	0.8673	0.8617	0.8668	0.8791	0.8772	0.8847	0.8704	0.8555	0.8834	0.8185	
$\mathcal{O}_{144}\text{-}\mathcal{O}_{154}$	0.8709	0.8004	0.8760	0.8543	0.8776	0.7702	0.8833	0.8682	0.8540	0.8730	0.8793	
$\mathcal{O}_{155}\text{-}\mathcal{O}_{165}$	0.8697	0.8576	0.8707	0.8787	0.7951	0.8454	0.8595	0.8806	0.8849	0.8556	0.8725	
$\mathcal{O}_{166}\text{-}\mathcal{O}_{176}$	0.8647	0.8300	0.8738	0.8783	0.8396	0.8701	0.8212	0.8638	0.8833	0.8741	0.8761	
$\mathcal{O}_{177}\text{-}\mathcal{O}_{181}$	0.8622	0.8665	0.8578	0.8756	0.8663							

5.4 Large-scale group decision-making on evaluation of user preference

In this subsection, to evaluate the scores of UPs for decision-making, approximate precision is proposed via six types of neighborhood-related fuzzy covering-based rough set models. And the fuzzy TOPSIS method is used to calculate the ranking coefficient to obtain the preference information that users value most in LSGDM.

5.4.1 Fuzzy covering-based rough set and fuzzy TOPSIS models

We get the fuzzy covering-based rough set models, the experimental steps are shown below.

- (1) Standardize CUP fuzzy set to satisfy the condition of fuzzy covering by Eq. (5).
- (2) Fuzzy neighborhood operators. Fuzzy covering-based maximum and minimum descriptions are calculated by Algorithm 4.4.1. Then, the fuzzy neighborhood operators $N_1^C(x)$ to $N_4^C(x)$ are obtained via Algorithm 4.4.1, a total of 12 fuzzy neighborhood operators are attained.
- (3) Fuzzy β -neighborhood and complementary β -neighborhood operators. Let β value be $\bigwedge_{i \in \mathcal{M}} (\bigvee_{j \in \mathcal{N}} C_i^j) = 0.4737$, where n is the number of UPs, and m is the number

of individuals. The fuzzy β -neighborhood operator is shown in Table 17 and the complementary β -neighborhood is calculated.

- (4) The lower approximation and upper approximation are calculated by Definition 8. We obtain 14 upper approximation tables and corresponding 14 lower approximation tables.
- (5) Approximate precision of each fuzzy covering-based rough set model is calculated.

Table 17 The fuzzy β -neighborhood operator

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
1	0.4874	0.3333	0	0.0588	0	0.0455	0	0.0642	0	0	0.0417	0.1	0	0.27	0	0	0.1	0	0	0.0333	0	0	0	0	0
2	0	0.5	0	0.0213	0	0	0	0.04	0	0	0	0.0227	0	0.1071	0	0	0.0052	0	0	0	0	0	0	0	0
3	0	0.3333	0.4783	0.0571	0	0	0	0.0667	0	0	0.0312	0.0769	0.0333	0.5143	0	0	0.0769	0	0	0.0333	0	0	0.0286	0.0312	0
4	0	0.0625	0	0.48	0	0	0	0	0	0	0.0357	0	0	0.1071	0	0	0	0	0.1613	0	0	0	0	0	0
5	0.037	0.25	0.0455	0.1429	0.4783	0	0	0.0263	0.0455	0	0.0833	0	0	0.1667	0	0.087	0.0517	0	0.1739	0.0652	0.0312	0	0	0.0714	0
6	0.2	0.54	0.0862	0.4	0.1897	0.5	0.1429	0.4	0.3103	0	0.1143	0.72	0.1897	0.6571	0	0.0862	0.2571	0.14	0.3714	0.12	0.0571	0.0571	0	0.0857	0.3448
7	0.3333	0.5556	0	0.1111	0.1111	0	0.5789	0.6842	0	0	0	0.3333	0.2105	0.1111	0.1111	0.2222	0.3333	0	0	0.2105	0	0	0.4444	0.1111	0
8	0	0.12	0	0.0256	0	0	0	0.4762	0	0	0	0.0227	0.0435	0.1111	0.0455	0.0309	0	0	0.04	0	0	0	0	0	0
9	0.58	0.54	0.12	0.48	0.3	0.5	0.16	0.42	0.52	0.22	0.14	0.72	0.2	0.8	0.22	0.4	0.64	0.14	0.42	0.12	1	0.16	0.06	0.16	0.38
10	0	0.0159	0	0.1152	0	0	0	0	0	0.5	0	0	0.066	0	0.1099	0	0.0454	0	0.1493	0	0	0	0	0	0
11	0.0679	0.3333	0	0.1455	0	0.1304	0	0	0.0417	0	0.5135	0.0435	0	0.25	0	0.0035	0.1	0	0	0.0333	0	0	0.0417	0	0
12	0	0.2218	0	0.0056	0	0	0	0	0	0	0	0.48	0	0.134	0	0	0.0052	0	0	0.0213	0	0	0	0	0
13	0	0.0689	0.0476	0.0764	0	0.0667	0.0513	0.1111	0.0667	0	0.1739	0.2799	0.4762	0.3333	0	0.0667	0.2222	0.0714	0.119	0.05	0.1282	0	0	0	0
14	0	0.0625	0	0.0213	0	0	0	0	0	0	0	0	0	0	0.4762	0	0	0	0	0.0123	0	0	0	0	0
15	0.3333	0.3333	0	0.4991	0	0.3182	0.1111	0	0.1111	0.1818	0.2727	0.9545	0	0.7727	0.6364	0.1818	0.5455	0.3182	0.3636	0.0455	0.1364	0.2222	0.2273	0.3636	0.1364
16	0	0.054	0	0.0833	0	0	0	0.0208	0	0	0.0333	0	0	0.2	0	0.48	0	0	0.0333	0	0	0	0	0	0
17	0	0.3333	0	0.0213	0	0	0	0	0	0	0	0.0952	0	0.2	0	0	0.48	0	0	0.0213	0	0	0	0	0
18	0.2069	0.6818	0	0.1818	0.125	0	0.0909	0.1765	0.0455	0	0.0417	0.7917	0.0455	0.5882	0.0588	0	0.3636	0.4828	0.0833	0.2273	0	0	0.069	0	0
19	0	0.0741	0	0.1071	0	0	0	0	0	0	0.0323	0	0	0.1071	0	0.0513	0	0	0.475	0	0	0	0	0	0
20	0.027	0.2606	0	0.0938	0	0	0	0	0	0	0.0312	0.0787	0	0.4375	0.0357	0	0.0086	0.0435	0.045	0.4783	0	0	0	0	0
21	0	0.1045	0	0.3453	0.0667	0	0	0.0263	0	0	0.0714	0.0299	0.0263	0.2	0	0.1613	0.0323	0	0.3077	0.0645	0.4865	0	0	0	0
22	0	0.2176	0	0.0656	0	0.0526	0	0.0665	0	0	0.0625	0.0598	0	0.1667	0	0	0.0572	0	0.1708	0.0455	0.0128	0.5	0	0	0
23	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
24	0.16	0.12	0.08	0.3043	0.4	0.3125	0	0.25	0.12	0	0.16	0.0435	0	0.5	0.12	0.3125	0	0	0.875	0.16	0	0	0.08	0.48	0
25	0.0679	0.3651	0.0476	0.1455	0.04	0.16	0.0476	0.16	0.1429	0	0.1143	0.5374	0	0.36	0	0.0035	0.2289	0.0541	0.08	0.0476	0.0295	0	0	0	0.52

we get the values of closeness coefficient of decision makers, add them to the weight, and calculate the ranking coefficient of UPs via the fuzzy TOPSIS method. According to 12 fuzzy neighbourhood operators based on fuzzy covering, $(\underline{A}, \overline{A})$, and fuzzy β -neighborhood, $(\underline{Q}, \overline{Q})$, and complementary β -neighborhood operators, $(\underline{E}, \overline{E})$, their rankings are shown in Table 18.

Table 18 UPs ranking of TOPSIS based on fuzzy neighborhood operators

MODEL	THE RANKING OF UPs
$(\underline{A}_1, \overline{A}_1)$	$C_2 > C_{14} > C_{12} > C_{19} > C_{16} > C_{17} > C_4 > C_8 > C_{11} > C_6 > C_9 > C_{13} > C_1 > C_{18} > C_{23} > C_5 > C_{20} > C_{21} > C_{10} > C_{24} > C_3 > C_{25} > C_{15} > C_{22} > C_7$
$(\underline{A}_2, \overline{A}_2)$	$C_2 > C_{14} > C_{12} > C_{19} > C_{16} > C_{17} > C_4 > C_8 > C_{11} > C_6 > C_9 > C_{13} > C_1 > C_{23} > C_{18} > C_5 > C_{20} > C_{21} > C_{10} > C_{24} > C_3 > C_{25} > C_{15} > C_{22} > C_7$
$(\underline{A}_3, \overline{A}_3)$	$C_2 > C_{14} > C_{12} > C_{19} > C_{16} > C_{17} > C_4 > C_8 > C_{11} > C_6 > C_9 > C_{13} > C_1 > C_{23} > C_{18} > C_5 > C_{20} > C_{21} > C_{10} > C_{24} > C_3 > C_{25} > C_{15} > C_{22} > C_7$
$(\underline{A}_4, \overline{A}_4)$	$C_2 > C_{14} > C_{12} > C_{19} > C_{16} > C_{17} > C_4 > C_8 > C_{11} > C_6 > C_9 > C_{13} > C_{18} > C_1 > C_{23} > C_5 > C_{20} > C_{21} > C_{24} > C_{10} > C_3 > C_{25} > C_{15} > C_{22} > C_7$
$(\underline{A}_5, \overline{A}_5)$	$C_2 > C_{14} > C_{12} > C_{19} > C_{16} > C_{17} > C_4 > C_8 > C_{11} > C_6 > C_9 > C_{13} > C_{18} > C_1 > C_{23} > C_5 > C_{20} > C_{21} > C_{24} > C_{10} > C_3 > C_{25} > C_{15} > C_{22} > C_7$
$(\underline{A}_6, \overline{A}_6)$	$C_2 > C_{14} > C_{12} > C_{19} > C_{16} > C_{17} > C_4 > C_8 > C_{11} > C_6 > C_9 > C_{13} > C_{18} > C_1 > C_{23} > C_5 > C_{20} > C_{21} > C_{24} > C_{10} > C_3 > C_{25} > C_{15} > C_{22} > C_7$
$(\underline{A}_7, \overline{A}_7)$	$C_2 > C_{14} > C_{12} > C_{19} > C_{16} > C_{17} > C_4 > C_8 > C_{11} > C_6 > C_9 > C_{13} > C_{18} > C_1 > C_{23} > C_5 > C_{20} > C_{21} > C_{24} > C_{10} > C_3 > C_{25} > C_{15} > C_{22} > C_7$
$(\underline{A}_8, \overline{A}_8)$	$C_2 > C_{14} > C_{12} > C_{19} > C_{16} > C_{17} > C_4 > C_8 > C_{11} > C_6 > C_9 > C_{13} > C_{18} > C_1 > C_{23} > C_5 > C_{20} > C_{21} > C_{24} > C_{10} > C_3 > C_{25} > C_{15} > C_{22} > C_7$
$(\underline{A}_9, \overline{A}_9)$	$C_2 > C_{14} > C_{12} > C_{19} > C_{16} > C_{17} > C_4 > C_8 > C_{11} > C_6 > C_9 > C_{13} > C_{18} > C_1 > C_{23} > C_5 > C_{20} > C_{21} > C_{24} > C_{10} > C_3 > C_{25} > C_{15} > C_{22} > C_7$
$(\underline{A}_{10}, \overline{A}_{10})$	$C_2 > C_{14} > C_{12} > C_{19} > C_{16} > C_{17} > C_4 > C_8 > C_{11} > C_6 > C_9 > C_{13} > C_{18} > C_1 > C_{23} > C_5 > C_{20} > C_{21} > C_{24} > C_{10} > C_3 > C_{25} > C_{15} > C_{22} > C_7$
$(\underline{A}_{11}, \overline{A}_{11})$	$C_2 > C_{14} > C_{12} > C_{19} > C_{16} > C_{17} > C_4 > C_8 > C_{11} > C_6 > C_9 > C_{13} > C_{18} > C_1 > C_{23} > C_5 > C_{20} > C_{21} > C_{24} > C_{10} > C_3 > C_{25} > C_{15} > C_{22} > C_7$
$(\underline{A}_{12}, \overline{A}_{12})$	$C_2 > C_{14} > C_{12} > C_{19} > C_{16} > C_{17} > C_4 > C_8 > C_{11} > C_6 > C_9 > C_{13} > C_{18} > C_1 > C_{23} > C_5 > C_{20} > C_{21} > C_{24} > C_{10} > C_3 > C_{25} > C_{15} > C_{22} > C_7$
$(\underline{Q}, \overline{Q})$	$C_2 > C_{14} > C_{12} > C_{19} > C_{16} > C_{17} > C_4 > C_8 > C_1 > C_5 > C_{20} > C_{21} > C_{11} > C_{10} > C_3 > C_6 > C_{13} > C_{22} > C_{24} > C_{25} > C_9 > C_{18} > C_7 > C_{15} > C_{23}$
$(\underline{E}, \overline{E})$	$C_2 > C_{14} > C_{12} > C_{19} > C_{16} > C_{17} > C_4 > C_8 > C_1 > C_5 > C_{20} > C_{21} > C_{11} > C_{10} > C_3 > C_6 > C_{13} > C_{22} > C_{24} > C_{25} > C_9 > C_{18} > C_7 > C_{15} > C_{23}$

5.4.2 User preferences analysis on LSGDM

In this subsection, a brief analysis of UPs sorting results of 4 operators of D'eer and fuzzy β -covering-based rough set models are carried out.

Among the 4 operators in D'eer, the experimental results of the first and third operators are consistent, which shows that the fuzzy covering is close to the maximum description. As shown in Figure 4, the larger the value, the higher the ranking,

where “IP” represents R -implicator induced by $\mathcal{T}_P$, the meanings of “IL” and “IM” can be obtained in a similar way. The experimental results of the second and second operators are consistent, which is represented by scatter points. “Customer service”, “JD” and “store” occupies the stable primary position, which is proof that people choose the store and shopping platform, they often have high expectations for them. This phenomenon shows that users pay high attention to the seller’s brand and shopping platform. Many shopping groups choose JD because of the platform’s advantages, which include JD’s self-operated stores and its own logistics chain. Therefore, when a large group of shopping is dissatisfied, opinions and dissatisfaction with JD platform will appear in the negative reviews. The ranking results of D’eer’s second and fourth operators and the second and fourth operators are consistent. The top 4 rankings remain unchanged, only the rankings of “express” and “fresh” are reversed. In practice, when consumers have problems purchasing fruits, they will immediately ask customer service to express their dissatisfaction. At this time, the importance of customer service will be reflected. How to effectively appease consumers and solve consumer problems to give consumers a better experience in a timely manner is an important concern for platforms and merchants. Then, in D’eer’s first and third operators and the second and fourth operators, the second-ranked preference words are “unpalatable”, “fresh” and “express”. It can reflect that users pay a lot of attention to the taste and quality when buying fresh fruits in JD.

According to the UPs ranking obtained by fuzzy β -neighborhood and fuzzy complementary β -neighborhood operators, the preference ranking of these two models is consistent, as shown in Figure 5, which the larger the value, the higher the ranking, and the ranking of the first operator of D’eer is added in this figure for comparison. Through the radar chart, it can be found that models based on fuzzy β -neighborhood and fuzzy complementary β -neighborhood operators have a high similarity with the model results of the first and third types of operators of D’eer. It is worth noting that the “quality” ranking dropped by 4 places in fuzzy β -covering-based rough set models. At the same time, the rankings of “bad fruit”, “taste”, “package” and “not sweet” have improved slightly. To summarize the experimental results, consumers pay the most attention to “customer service”, followed by “JD” and “store” when purchasing fresh fruit online, representing the store service and the protection of rights and interests of e-commerce platforms. In order to improve the sales of fresh fruits in agricultural products, there is an urgent need for:

(1) The JD e-commerce platform can enhance the user’s shopping experience by improving the service level. Strict screening of settled merchants and selection of products can improve the quality of fresh fruits. At the same time, optimize logistics and transportation routes to provide the best and fastest delivery, which not only reduces time and increases user repurchase rate, but also greatly reduces the consumption of fresh fruit. And improving after-sales service can also greatly enhance users’ loyalty to the platform.

(2) Take preservation measures. Merchants generally have their warehouses for fresh fruit storage, during this period, attention should be paid to the freshness of the fruit. After a period of time in transit, consumers will receive the fruit. In this regard, merchants can improve in two aspects. One is to take suitable fresh-keeping

measures, such as adding ice packs, etc., and the other is to reasonably ship fresh fruits according to the actual logistics and the experience of fruit ripening time. Improving the quality of fresh fruit can enhance the user’s shopping experience and increase the user’s repurchase rate.

(3) Improving customer service’s timely processing and rapid response capabilities. When consumers are dissatisfied after shopping, they will look for customer service. If customer service fails to respond in time and give a satisfactory answer, they will often give negative reviews. In order to reduce negative reviews from users and increase sales of fresh fruits, merchants should pay attention to the training of customer service personnel so that they guide customers well. And then, merchants should improve the after-sales mechanism. We know that fresh fruit has a high timeliness, and fruit rot is often caused by reasons such as the transportation time process, such a phenomenon is unavoidable. Therefore, merchants should improve the after-sales mechanism, provide timely and good solutions to the purchase problems reported by consumers, which can increase user repurchase rate. It can be refunded or reissued for lousy fruit. High store sales often have high-standard after-sales service, and it is important to improve the standard of the after-sales department.

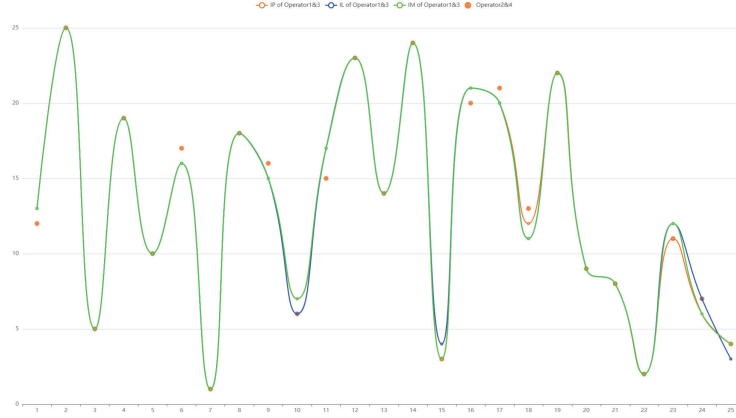


Fig. 4 The UPs ranking of first and third operator of D’eer

5.5 Platform implementation and interaction

To facilitate the solution of decision-making problems, a platform-based intelligent decision-support system was developed and deployed.¹ In this way, the proposed method is transformed from a theoretical decision-making framework into an accessible application for practical decision support. Figure 6 presents the main interface of the platform. The platform integrates the database and decision-making algorithmic modules developed in this study. It provides two main functions. First, the platform implements retrieval-augmented generation to provide more focused, reliable,

¹The platform is available at: <http://47.109.40.169>.

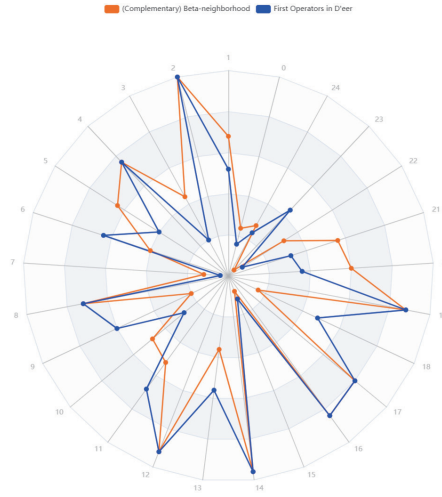


Fig. 5 The UPs ranking of fuzzy (complementary) β beta-neighborhood operator

and effective support for interactive question answering on decision-making problems. Second, the decision process is organized into a standardized two-stage workflow: data and algorithm upload or configuration, followed by decision execution. Based on the domain-specific knowledge base and the proposed algorithmic model, the platform executes the corresponding decision code and returns analytical results in a consistent format. The platform can also be extended by incorporating additional domain-specific databases or expert knowledge bases, enabling its application to decision-support tasks involving other products or service scenarios. This design improves the accessibility and reproducibility of the proposed method and enhances the efficiency of decision analysis.



Fig. 6 Intelligent decision-support platform homepage

6 Discussion

In this section, firstly, we test the robustness of clustering effect of the Louvain model. Then, multiple β values are taken for fuzzy β -covering to verify the experimental results. Finally, different decision-making methods are compared, including classical methods in LSGDM, and VIKOR model instead of TOPSIS.

6.1 Comparison of clustering models

In this section, to verify the clustering effect of our proposed double-restricted Louvain model, we compare it with the model itself and other clustering models. We choose the silhouette coefficient (SC) as the index of LSG clustering effect, including two factors, cohesion and separation, which can well reflect the clustering effect. Its value range is $[-1, 1]$, and the larger the value, the better the clustering effect.

First, we compare the double-restricted Louvain model with the classic Louvain model, and the Louvain model with an edge weight which is similarity relationship degree between individuals. As shown in Table 19, the three models are clustered into 4 subgroups, the double-restricted Louvain model has the highest cohesion, followed by Louvain with edge weight. The consensus degree of the three is basically close, but the silhouette coefficient of the improved Louvain model is obviously higher than that of the other two. And Dunn [54] evaluation value is consistent among the three, compared with other clustering models, it is mostly higher than theirs, which shows that we are wise to base it on Louvain clustering. Based on the above analysis and comparison, we find that the double-restricted Louvain model proposed by us has a better clustering effect compared with the classical model.

Then, four clustering models are compared, including k -means, fuzzy c -means, hierarchical clustering and DBSCAN. Compared with Louvain, the k -means model needs to specify the number of clusters during cluster analysis manually, and the optimal number of clusters k can be obtained by using the elbow method. The core idea of the elbow method is to use the core indicator SSE (sum of the squared errors),

$$SSE = \sum_{i=1}^k \sum_{p \in C_i} |p - m_i|^2,$$

where C_i is the i th cluster, p is the sample point in C_i , m_i is the centroid of C_i , which is the mean of all samples in C_i , and SSE is all samples clustering error, which can explain the quality of the clustering effect.

As the value of k increases, SSE will decrease rapidly and then level off. The relationship between SSE and k is an elbow shape, and the k value corresponding to the elbow is optimal clustering number of the data. We iterate the clustering 100 times to obtain the mean value of SSE . The result is shown in Figure 7, which shows that the elbow is not particularly significant. We take 2 to 5 according to the figure as the clustering number. And we set the number of clusters to 2 for clustering, according to our practical experience, it is found that the number of subgroups 2 is not reasonable, because too few subgroups will increase the difficulty of LSG consensus. Next,

we compare experiments with other cluster numbers, which show that the silhouette coefficient is around 0 and that the clustering effect is not good.

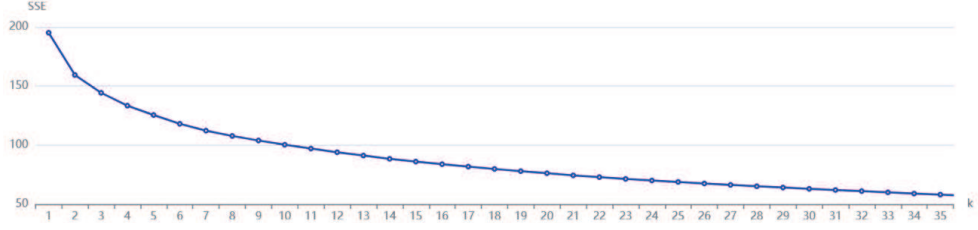


Fig. 7 SSE and profile factor of k -means

The fuzzy c-means model clustering number is 2, the cohesion is too low, the consensus between subgroups is too high, and the silhouette coefficient is close to 0, indicating that the clustering effect is not ideal. According to the results of the hierarchical clustering method, it is found that the cohesion and consensus indicators have good results, but the silhouette coefficient is low. About DBSCAN model, when set the parameter that the maximum distance between two samples exceeds 0.5, there will be many invalid values in the clustering results. In order to divide all individuals, the parameter value can only be reduced, and at this time, LSG is clustered into one class.

After overall evaluation, it is found that the double-restricted Louvain is a more suitable clustering method that fully divides subgroups while ensuring a higher silhouette coefficient.

Table 19 Comparison of clustering models

MODEL	Number of Subgroup	Cohesion	Consensus	Silhouette Coefficient	Dunn
Double-restricted Louvain	4	0.8578	0.9280	0.2420	0.0714
Classic Louvain	4	0.8562	0.9275	0.2245	0.0714
Louvain with edge weights	4	0.8576	0.9273	0.2190	0.0714
K-means	2	0.8656	0.9588	0.0659	0.0714
	3	0.8724	0.9151	0.0054	0.0675
	4	0.8746	0.8748	-0.0527	0.0675
	5	0.8811	0.8199	-0.0711	0.0675
Fuzzy c-means	2	0.8644	0.9593	0.0323	0.0714
Hierarchical clustering	4	0.8743	0.8724	-0.0707	0.0675
DBSCAN	1	-	-	-	-

6.2 Convergence and comparison of consensus model

For the consensus Model 4, this section uses adjustment ranges of different scales for non-consensus adjustments, as shown in the Figure 8 shown. As the adjustment scale increases, the cost gradually increases, but the consensus gradually stabilizes. The experimental result illustrates the effectiveness of the consensus model.

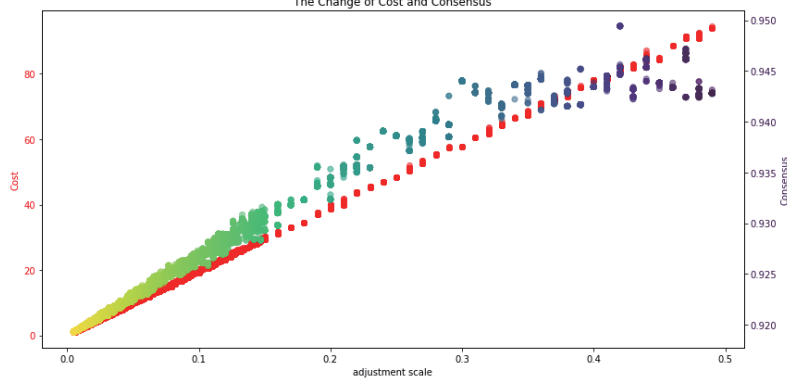


Fig. 8 Consensus and cost with different adjustment scales

If we experiment with average thinking instead of the optimization strategy, at this stage, we modify the preference scores of individuals of discordant subgroups by gradually increasing the adjustment value of preference. When the LSG reaches consensus, the iteration stops, which ensures that LSG consensus reaches the minimum threshold and the adjustment cost is minimal. At this time, the decision maker's adjustment scale is 0.0073, the consensus level of the discordant subgroup SG_3 is 0.9201, and the adjustment cost is 3.0894, which is significantly higher than our optimization strategy, thus proving the superiority of the consensus strategy proposed in this paper.

6.3 Comparison of β -covering-based fuzzy rough sets models

In fuzzy β -covering, let $\beta \in (0, 0.4737]$, we take multiple β values, $[0.01, 0.4737]$, with a step size of 0.03, and their final ranking results about fuzzy β -neighborhood and fuzzy complementary β -neighborhood whose results number is 16 individually are the same, as shown in Table 20, it is found that the rankings are consistent with those in Table 18. This proves that fuzzy β -covering-based rough set models are stable in experiment.

Table 20 Multiple β values of UPs ranking

MODEL	THE RANKING OF PREFERENCES
$(\underline{Q}, \underline{Q})$	$C_2 > C_{14} > C_{12} > C_{19} > C_{16} > C_{17} > C_4 > C_8 > C_1 > C_5 > C_{20} > C_{21} > C_{11} > C_{10} > C_3 > C_6 > C_{13} > C_{22} > C_{24} > C_{25} > C_9 > C_{18} > C_7 > C_{15} > C_{23}$
$(\underline{E}, \underline{E})$	$C_2 > C_{14} > C_{12} > C_{19} > C_{16} > C_{17} > C_4 > C_8 > C_1 > C_5 > C_{20} > C_{21} > C_{11} > C_{10} > C_3 > C_6 > C_{13} > C_{22} > C_{24} > C_{25} > C_9 > C_{18} > C_7 > C_{15} > C_{23}$

6.4 Comparison of decision-making methods

In this subsection, the results of our LSGDM model will be compared with the classical LSGDM models. Afterward, we will use the VIKOR model to verify the stability of the TOPSIS model.

6.4.1 Comparison of classic LSGDM methods

(1) The preference of LSGDM based on subgroup cardinality and adjustment cost as subgroup weight by Ji et al. [8], as shown below,

$$\tau_c^{LG} = \sum_{\kappa=0}^3 num_{\kappa} \times \omega_{\kappa} \times \tau_c^{SG_{\kappa}} (\kappa \in \mathcal{Q}, c \in \mathcal{N}),$$

let $\omega_{\kappa} = [1, 1, 1/cost, 1]$. In Figure 3, $num = [78, 53, 46, 4]$ is shown. We obtain the results of LSGDM, and sort the results to get the importance of UPs, as shown in Table 21.

Table 21 The UPs ranking of LSGDM in [8]

MODEL	THE RANKING OF UPs
[8]	$C_2 > C_{14} > C_{12} > C_{19} > C_{16} > C_{17} > C_4 > C_8 > C_1 > C_5 > C_{20} > C_{21} > C_{11} > C_3 > C_6 > C_{10} > C_{13} > C_{24} > C_{22} > C_{25} > C_9 > C_{18} > C_7 > C_{15} > C_{23}$

(2) Labella et al. [55] used subgroup cohesion $\mathbb{C}$ in Definition 18 as the importance degree $\omega^{\mathbb{C}}$ of subgroup,

$$\mathbb{C}(SG_{\kappa}) < \mathbb{C}(SG_{\sigma}) \Rightarrow w_{\kappa}^{\mathbb{C}} < w_{\sigma}^{\mathbb{C}} (\kappa, \sigma \in \mathcal{Q}).$$

The preference of LSG is $\tau^{LG} = (\tau_c^{LG})_{1 \times n}$ by

$$\tau_c^{LG} = \sum_{\kappa=0}^3 w_{\kappa} \times \tau_c^{SG_{\kappa}} (\kappa \in \mathcal{Q}, c \in \mathcal{N}).$$

We obtain the results of LSGDM, and sort the results to get the importance of UPs, as shown in Table 22.

Table 22 The UPs ranking of LSGDM in [55]

MODEL	THE RANKING OF UPs
[55]	$C_{14} > C_2 > C_{12} > C_{19} > C_{16} > C_4 > C_{17} > C_8 > C_{21} > C_{20} > C_1 > C_5 > C_{11} > C_{22} > C_3 > C_{10} > C_6 > C_{24} > C_{13} > C_9 > C_7 > C_{25} > C_{18} > C_{15} > C_{23}$

After obtaining the results of the above two groups of classic LSGDM methods, compared with our comprehensive experimental results, several UPs have small ranking fluctuations, especially with fuzzy β -covering-based rough set models. In summary, it proves that the model of covering-based fuzzy rough sets and fuzzy β -covering-based rough sets can objectively calculate the weight of objects, and have good applicability in LSGDM. In addition, multiple operators in different forms in the model cover multiple decision-making results, which can be comprehensively judged on the decision-making results.

6.4.2 Comparison of VIKOR model

In this subsection, VIKOR method, which is also a potential method for solving decision-making problems in complex systems [56], is compared with the results of TOPSIS in LSGDM. VIKOR method has two indicators, the benefit B_i and the regret value G_i . The equations [56] are as follows,

$$B_i = \sum_{j=1}^n \frac{\mathcal{W}_i^j \mathcal{D}(\psi_i, \psi^+)}{\mathcal{D}(\psi^+, \psi^-)},$$

$$G_i = \max_j \left[\frac{\mathcal{W}_i^j \mathcal{D}(\psi_i, \psi^+)}{\mathcal{D}(\psi^+, \psi^-)} \right],$$

where $\mathcal{W}$ is the weight of UPs, which is got in Eq. (6). $\mathcal{D}(X_i, X^+) = \sum |\psi_i - \psi^+|$, $\mathcal{D}(X^+, X^-) = \sum |\psi^- - \psi^+|$. Then, the ranking score $\mathcal{Z}_i$ of UPs by VIKOR is

$$\mathcal{Z}_i = \frac{\varpi (B_i - B^-)}{B^+ - B^-} + \frac{(1 - \varpi) (G_i - G^-)}{G^+ - G^-}$$

where $B^+ = \max_i (B_i)$, $B^- = \min_i (B_i)$, $G^+ = \max_i (G_i)$, $G^- = \min_i (G_i)$, $\varpi \in [0, 1]$ is the weight of maximum group utility. The importance of UPs is sorted according to $\mathcal{Z}_i$. The larger the $\mathcal{Z}_i$ value, the more important the UPs.

In the VIKOR model, we conducted a total of 12 experiments according to different operators. In each experiment, we selected 10 ϖ in $[0, 1]$, the step size is 0.1. It is found in the experiment that different ϖ are selected in each experiment, the ranking of UPs fluctuates but is small, and the experimental results are basically the same. We choose the experimental results of $\varpi = 0.5$ as shown in Table 23. In Figure 9, the ranking results of three calculation methods of one operator are stacked, the more important the greater the value. In the figure, each UP has four stacked columns. The first two stacked columns are stacked graphs of the three ranking results of VIKOR's first operators and TOPSIS's. The latter two stacked columns are thinner, representing the stacked graphs of the three ranking results of VIKOR's second operators and TOPSIS's, respectively. The two line graphs respectively represent the average ranking of all VIKOR operators and TOPSIS operators. In the legend, "V1IP" represents the VIKOR with R -implicator induced by $\mathcal{T}_P$ of first operator, "T1IP" represents the TOPSIS with that, and "V_AVE" and "T_AVE" represents the average of all rankings in VIKOR and TOPSIS respectively. According to the comparison of the experimental results of TOPSIS and VIKOR, it is found that under different forms of different operators, the ranking of UPs fluctuates by a few places. At the same time, we use Spearman correlation to test the similarity between VIKOR and TOPSIS for each ranking. According to the results in Table 24, The p value is less than 0.01, passing the significance test, the two have a high correlation, and the conclusions drawn by the two models are consistent, which proves the effectiveness of the TOPSIS and its good applicability in LSGDM.

Table 23 UPs ranking of VIKOR

MODEL	THE RANKING OF UPs
$(\bar{A}_{I_P}^1, \bar{A}_{I_P}^1)$	$C_2 > C_{14} > C_{12} > C_{17} > C_4 > C_{19} > C_{16} > C_8 > C_{13} > C_{18} > C_5 > C_{11} > C_9 > C_{21} > C_{20} > C_1 > C_{25} > C_6 > C_{10} > C_{23} > C_3 > C_{15} > C_{24} > C_{22} > C_7$
$(\bar{A}_{I_L}^1, \bar{A}_{I_L}^1)$	$C_{14} > C_2 > C_{12} > C_{19} > C_{16} > C_{17} > C_4 > C_8 > C_{11} > C_6 > C_9 > C_{13} > C_1 > C_{23} > C_{18} > C_5 > C_{20} > C_{21} > C_{24} > C_{10} > C_3 > C_{25} > C_{15} > C_{22} > C_7$
$(\bar{A}_{I_M}^1, \bar{A}_{I_M}^1)$	$C_{14} > C_2 > C_{12} > C_{17} > C_{16} > C_4 > C_{19} > C_8 > C_9 > C_{18} > C_5 > C_{13} > C_{20} > C_{21} > C_{10} > C_7 > C_{11} > C_6 > C_{23} > C_1 > C_{22} > C_{24} > C_{15} > C_3 > C_{25}$
$(\bar{A}_{I_P}^2, \bar{A}_{I_P}^2)$	$C_{14} > C_2 > C_{17} > C_{12} > C_{16} > C_{19} > C_4 > C_8 > C_9 > C_{13} > C_{25} > C_5 > C_1 > C_{18} > C_6 > C_{21} > C_{11} > C_{24} > C_{15} > C_{10} > C_{23} > C_{20} > C_{22} > C_7 > C_3$
$(\bar{A}_{I_L}^2, \bar{A}_{I_L}^2)$	$C_{14} > C_2 > C_{17} > C_{12} > C_4 > C_{16} > C_{19} > C_9 > C_8 > C_{13} > C_5 > C_{25} > C_{18} > C_{21} > C_1 > C_6 > C_{11} > C_{24} > C_{15} > C_{10} > C_{23} > C_{20} > C_{22} > C_7 > C_3$
$(\bar{A}_{I_M}^2, \bar{A}_{I_M}^2)$	$C_{14} > C_2 > C_{17} > C_{12} > C_{16} > C_{19} > C_4 > C_8 > C_9 > C_{13} > C_{25} > C_5 > C_{18} > C_1 > C_6 > C_{21} > C_{11} > C_{24} > C_{15} > C_{10} > C_{23} > C_{20} > C_{22} > C_7 > C_3$
$(\bar{A}_{I_P}^3, \bar{A}_{I_P}^3)$	$C_2 > C_{14} > C_{12} > C_{17} > C_4 > C_{19} > C_{16} > C_8 > C_{13} > C_{18} > C_5 > C_{11} > C_9 > C_{21} > C_{20} > C_1 > C_{25} > C_6 > C_{10} > C_{23} > C_3 > C_{15} > C_{24} > C_{22} > C_7$
$(\bar{A}_{I_L}^3, \bar{A}_{I_L}^3)$	$C_{14} > C_2 > C_{12} > C_{19} > C_{16} > C_{17} > C_4 > C_8 > C_{11} > C_6 > C_9 > C_{13} > C_1 > C_{23} > C_{18} > C_5 > C_{20} > C_{21} > C_{24} > C_{10} > C_3 > C_{25} > C_{15} > C_{22} > C_7$
$(\bar{A}_{I_M}^3, \bar{A}_{I_M}^3)$	$C_{14} > C_2 > C_{12} > C_{17} > C_{16} > C_4 > C_{19} > C_8 > C_9 > C_{18} > C_5 > C_{13} > C_{20} > C_{21} > C_{10} > C_7 > C_{11} > C_6 > C_{23} > C_1 > C_{22} > C_{24} > C_{15} > C_3 > C_{25}$
$(\bar{A}_{I_P}^4, \bar{A}_{I_P}^4)$	$C_{14} > C_2 > C_{17} > C_{12} > C_{16} > C_{19} > C_4 > C_8 > C_9 > C_{13} > C_{25} > C_5 > C_1 > C_{18} > C_6 > C_{21} > C_{11} > C_{24} > C_{15} > C_{10} > C_{23} > C_{20} > C_{22} > C_7 > C_3$
$(\bar{A}_{I_L}^4, \bar{A}_{I_L}^4)$	$C_{14} > C_2 > C_{17} > C_{12} > C_4 > C_{16} > C_{19} > C_9 > C_8 > C_{13} > C_5 > C_{25} > C_{18} > C_{21} > C_1 > C_6 > C_{11} > C_{24} > C_{15} > C_{10} > C_{23} > C_{20} > C_{22} > C_7 > C_3$
$(\bar{A}_{I_M}^4, \bar{A}_{I_M}^4)$	$C_{14} > C_2 > C_{17} > C_{12} > C_{16} > C_{19} > C_4 > C_8 > C_9 > C_{13} > C_{25} > C_5 > C_{18} > C_1 > C_6 > C_{21} > C_{11} > C_{24} > C_{15} > C_{10} > C_{23} > C_{20} > C_{22} > C_7 > C_3$

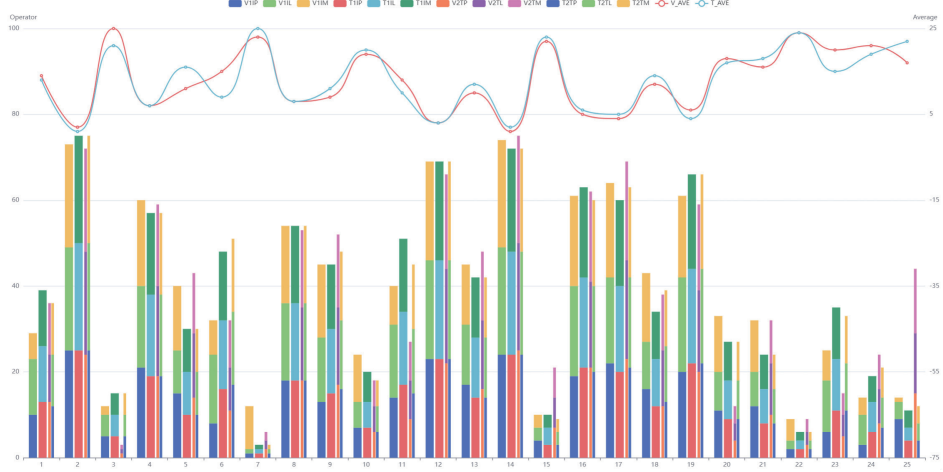


Fig. 9 Comparison of VIKOR and TOPSIS

Table 24 Spearman correlation of VIKOR and TOPSIS

	$\bar{A}_{I_P}^1$	$\bar{A}_{I_L}^1$	$\bar{A}_{I_M}^1$	$\bar{A}_{I_P}^2$	$\bar{A}_{I_L}^2$	$\bar{A}_{I_M}^2$	$\bar{A}_{I_P}^3$	$\bar{A}_{I_L}^3$	$\bar{A}_{I_M}^3$	$\bar{A}_{I_P}^4$	$\bar{A}_{I_L}^4$	$\bar{A}_{I_M}^4$
Spearman correlation	0.9108	0.9985	0.8338	0.8738	0.8646	0.8746	0.9108	0.9985	0.8338	0.8738	0.8646	0.8746
P value < 0.01	True	True	True	True	True	True	True	True	True	True	True	True

7 Conclusion and future work

The paper analyzes the online reviews of fresh fruit on e-commerce platform, studies the user preferences of consumers for buying fresh fruit. We propose an LSGDM based on consensus and fuzzy β -covering-based rough set models to determine large-scale group preferences for fresh fruit purchases. The main advantages and differences compared to existing methods are,

- (1) Different from the traditional method of finding user demands based on decision makers' ratings, the text analysis technique is used to collect more decision makers' opinions to obtain more objective user preferences instead of expert scoring, and we form the user preference extraction and build user preference fuzzy set process framework.

(2) In the clustering stage of LSGDM, SNA is used to determine individual weights, and the proposed double-restricted Louvain model has good clustering performance in graph clustering, and its advantages and feasibility are proved. In consensus reaching process, it is effective in improving consensus progress and consensus degree of LSG that PSO algorithm with the strategy of maximizing consensus and minimizing cost is used to get the optimal solution of subgroup and individual adjustment. Besides, we define the adjustment cost.

(3) Fuzzy β -covering-based rough set models have been proven effective in objectively obtaining objects' weights and can be well applied in LSGDM. The score results of UPs based on four types of fuzzy approximation operators, fuzzy β -neighborhood and fuzzy complementary β -neighborhood operators are compared. Especially for the latter, the decision results obtained are highly consistent with the classical LSGDM method. And when making the decision, we propose a weighted fuzzy TOPSIS method to obtain the preferences of LSG, which can objectively and clearly distinguish the important scores of UPs.

(4) The analysis results of LSGDM can provide marketing references for platforms and merchants to improve achievement of fresh fruits online. Merchants should improve service level, the quality of fresh fruits, and the shopping experience of consumers to increase the user purchase and repurchase rate. At the same time, the platform should play a regulatory role and work together to jointly promote the development of online sales model of agricultural products. Moreover, a platform-based intelligent decision-support system is developed and deployed, which can also incorporate other domain-specific expert databases as alternative data sources for decision support.

However, there are certain limitations here. For the sake of sales and profits, merchants give discounts to users with bad reviews to reduce negative information about products, and merchants often hire “water troops” to give positive reviews to hedge against negative reviews of products. Therefore, how to reasonably process and choose the review text of fresh fruit deserves our further study. In addition, the comment on fresh fruit has a large number of user features, objectively and effectively screening and classifying the characteristics will be our next step.

Declarations

Conflict of interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Author contributions

Xiaolei Wang: Conceptualization, Methodology, Writing–original draft. **Bin Yang:** Investigation, Writing–review & editing.

Funding

This research was supported by the National Natural Science Foundation of China (Grant no. 12101500) and the Chinese Universities Scientific Fund (Grant nos. 2452018054 and 2452022370) and the Natural Science Basic Research Program of Shaanxi (Program No. 2026JC-YBMS-0056).

References

- [1] Xie, G.: The impact of generative ai shopping assistants on e-commerce consumer motivation and behavior: Consumer-ai interaction design. *International Journal of Information Management* **86**, 102983 (2026)
- [2] Huang, Y., Gong, L., Ren, K., Zhu, N.: Determinants of continuous purchase intention on fresh food e-commerce platforms. *Frontiers in Sustainable Food Systems* **9**, 1554600 (2025)
- [3] Li, J., Kang, J.: Less stress, fewer delays: The role of sophisticated ai in mitigating decision fatigue and purchase postponement in luxury retail. *Journal of Retailing and Consumer Services* **85**, 104268 (2025)
- [4] Kang, D., Park, Y.: based measurement of customer satisfaction in mobile service: Sentiment analysis and vikor approach. *Expert Systems with Applications* **41**(4), 1041–1050 (2014)
- [5] Liang, D., Dai, Z., Wang, M.: Assessing customer satisfaction of o2o takeaway based on online reviews by integrating fuzzy comprehensive evaluation with ahp and probabilistic linguistic term sets. *Applied Soft Computing* **98**, 106847 (2021)
- [6] Wang, X., Lin, X., Spencer, M.K.: Exploring the effects of extrinsic motivation on consumer behaviors in social commerce: Revealing consumers' perceptions of social commerce benefits. *International Journal of Information Management* **45**, 163–175 (2019)
- [7] Melović, B., Šehović, D., Karadžić, V., Dabić, M., Ćirović, D.: Determinants of millennials' behavior in online shopping—implications on consumers' satisfaction and e-business development. *Technology in society* **65**, 101561 (2021)
- [8] Ji, F., Cao, Q., Li, H., Fujita, H., Liang, C., Wu, J.: An online reviews-driven large-scale group decision-making approach for evaluating user satisfaction of sharing accommodation. *Expert Systems with Applications* **213**, 118875 (2023)
- [9] Wu, P., Sun, S., Zhou, L., Yao, Y., Deveci, M.: Platforms empower: Mining online reviews for supporting consumers decisions. *Journal of Retailing and Consumer Services* **84**, 104214 (2025)
- [10] Cao, J., Xu, X., Yin, X., Pan, B.: A risky large group emergency decision-making method based on topic sentiment analysis. *Expert Systems with Applications*

- [11] Zhou, J.-L., Tu, R.-F., Xiao, H.: Large-scale group decision-making to facilitate inter-rater reliability of human-factors analysis for the railway system. *Reliability Engineering & System Safety* **228**, 108806 (2022)
- [12] Chen, Z.-S., Liu, X.-L., Chin, K.-S., Pedrycz, W., Tsui, K.-L., Skibniewski, M.J.: Online-review analysis based large-scale group decision-making for determining passenger demands and evaluating passenger satisfaction: Case study of high-speed rail system in china. *Information Fusion* **69**, 22–39 (2021)
- [13] Chen, Z.-S., Zhu, Z., Wang, Z.-J., Tsang, Y.: Fairness-aware large-scale collective opinion generation paradigm: A case study of evaluating blockchain adoption barriers in medical supply chain. *Information Sciences* **635**, 257–278 (2023)
- [14] Liao, H., Li, X., Tang, M.: How to process local and global consensus? a large-scale group decision-making model based on social network analysis with probabilistic linguistic information. *Information Sciences* **579**, 368–387 (2021)
- [15] Shen, Y., Ma, X., Deveci, M., Herrera-Viedma, E., Zhan, J.: A hybrid opinion dynamics model with leaders and followers fusing dynamic social networks in large-scale group decision-making. *Information Fusion* **116**, 102799 (2025)
- [16] Li, Y., Wang, Z., Shen, Y., Pedrycz, W.: Cost management with default risk in large-scale group decision-making: A method based on cooperative bargaining game. *European Journal of Operational Research* (2025)
- [17] Yuan, R., Wu, Z., Tu, J.: Large-scale group decision-making with incomplete fuzzy preference relations: The perspective of ordinal consistency. *Fuzzy Sets and Systems* **454**, 100–124 (2023)
- [18] Wan, Q., Xu, X., Han, J.: A dimensionality reduction method for large-scale group decision-making using tf-idf feature similarity and information loss entropy. *Applied Soft Computing* **150**, 111039 (2024)
- [19] Liu, Z., Wang, W., Liu, P.: Dynamic consensus of large group emergency decision-making under dual-trust relationship-based social network. *Information Sciences* **615**, 58–89 (2022)
- [20] Ding, R.-X., Cheng, R.-X., Li, M.-N., Yang, G.-R., Herrera-Viedma, E.: Conflict management-based consensus reaching process considering conflict relationship clustering in large-scale group decision-making problems. *Expert Systems with Applications* **238**, 122095 (2024)
- [21] Qin, J., Li, M., Liang, Y.: Minimum cost consensus model for crp-driven preference optimization analysis in large-scale group decision-making using louvain algorithm. *Information Fusion* **80**, 121–136 (2022)

- [22] Wang, Y.-H., Ding, R.-X., Ju, Y., Xu, Y., Chiclana, F.: Three-way decision-based consensus feedback mechanism considering conflict level and social network for large-scale group decision-making. *Information Fusion*, 103634 (2025)
- [23] Ding, R.-X., Yang, B., Huang, Y., Zhang, Y., Chiclana, F.: Social network-based overlapping community clustering and feedback mechanism for large-scale group decision-making. *European Journal of Operational Research* (2025)
- [24] Li, M.-N., Wang, X., Ding, R.-X., Wang, Y.-H.: Primacy effect-based dynamic feedback mechanism considering communication sequence for multilevel infiltrative large-scale group decision-making. *European Journal of Operational Research* (2025)
- [25] Li, W., Pang, B.: Covering-based-variable precision fuzzy rough set and its application in decision-making. *Information Sciences*, 123096 (2026)
- [26] Pawlak, Z.: Rough sets. *International journal of computer & information sciences* **11**, 341–356 (1982)
- [27] Zhan, J., Sun, B.: Covering-based intuitionistic fuzzy rough sets and applications in multi-attribute decision-making. *Artificial Intelligence Review* **53**(1), 671–701 (2020)
- [28] Zakowski, W.: Approximations in the space (u, π) . *Demonstratio mathematica* **16**(3), 761–770 (1983)
- [29] Pomykala, J.A.: Approximation operations in approximation space. *Bulletin of the Polish Academy of Sciences* **35**(9-10), 653–662 (1987)
- [30] Pomykała, J.: On definability in the nondeterministic information system. *Bulletin of the Polish Academy of Sciences. Mathematics* **36**(3-4), 193–210 (1988)
- [31] Yao, Y.: Relational interpretations of neighborhood operators and rough set approximation operators. *Information sciences* **111**(1-4), 239–259 (1998)
- [32] Dubois, D., Prade, H.: Rough fuzzy sets and fuzzy rough sets. *International Journal of General System* **17**(2-3), 191–209 (1990)
- [33] Li, T.-J., Leung, Y., Zhang, W.-X.: Generalized fuzzy rough approximation operators based on fuzzy coverings. *International Journal of Approximate Reasoning* **48**(3), 836–856 (2008)
- [34] Feng, T., Zhang, S.-P., Mi, J.-S.: The reduction and fusion of fuzzy covering systems based on the evidence theory. *International Journal of Approximate Reasoning* **53**(1), 87–103 (2012)
- [35] Ma, L.: Two fuzzy covering rough set models and their generalizations over fuzzy lattices. *Fuzzy Sets and Systems* **294**, 1–17 (2016)

- [36] Yang, B., Hu, B.-Q.: On some types of fuzzy covering-based rough sets. *Fuzzy Sets and Systems* **312**, 36–65 (2017)
- [37] D’eer, L., Cornelis, C.: A comprehensive study of fuzzy covering-based rough set models: Definitions, properties and interrelationships. *Fuzzy Sets and Systems* **336**, 1–26 (2018)
- [38] D’eer, L., Cornelis, C., Godo, L.: Fuzzy neighborhood operators based on fuzzy coverings. *Fuzzy Sets and Systems* **312**, 17–35 (2017)
- [39] Yao, Y., Yao, B.: Covering based rough set approximations. *Information Sciences* **200**, 91–107 (2012)
- [40] Klir, G.J., Yuan, B.: Fuzzy sets and fuzzy logic: theory and applications. Possibility Theory versus Probab. Theory **32**(2), 207–208 (1996)
- [41] Klement, E.P., Mesiar, R., Pap, E.: Triangular Norms vol. 8. Springer, Dordrecht (2013)
- [42] Deng, T., Chen, Y., Xu, W., Dai, Q.: A novel approach to fuzzy rough sets based on a fuzzy covering. *Information Sciences* **177**(11), 2308–2326 (2007)
- [43] Qi, G., Yang, B., Li, W.: Some neighborhood-related fuzzy covering-based rough set models and their applications for decision-making. *Information Sciences* **621**, 799–843 (2023)
- [44] Zhang, K., Zhan, J., Wu, W.-Z.: On multicriteria decision-making method based on a fuzzy rough set model with fuzzy *alpha*-neighborhoods. *IEEE Transactions on Fuzzy Systems* **29**(9), 2491–2505 (2020)
- [45] Huttenlocher, D.P., Klanderman, G.A., Rucklidge, W.J.: Comparing images using the hausdorff distance. *IEEE Transactions on pattern analysis and machine intelligence* **15**(9), 850–863 (1993)
- [46] Blondel, V.D., Guillaume, J.-L., Lambiotte, R., Lefebvre, E.: Fast unfolding of communities in large networks. *Journal of statistical mechanics: theory and experiment* **2008**(10), 10008 (2008)
- [47] Waltman, L., Van Eck, N.J.: A smart local moving algorithm for large-scale modularity-based community detection. *The European physical journal B* **86**, 1–14 (2013)
- [48] Xu, X.-h., Du, Z.-j., Chen, X.-h.: Consensus model for multi-criteria large-group emergency decision-making considering non-cooperative behaviors and minority opinions. *Decision Support Systems* **79**, 150–160 (2015)
- [49] Jeroslow, R.G.: The polynomial hierarchy and a simple model for competitive analysis. *Mathematical programming* **32**(2), 146–164 (1985)

- [50] Qin, J., Wang, D., Liang, Y.: Social network-driven bi-level minimum cost consensus model for large-scale group decision-making: A perspective of structural holes. *Information Sciences* **649**, 119678 (2023)
- [51] Shi, Y., *et al.*: Particle swarm optimization: developments, applications and resources. In: *Proceedings of the 2001 Congress on Evolutionary Computation* (IEEE Cat. No. 01TH8546), vol. 1, pp. 81–86 (2001). IEEE
- [52] Hwang, C.-L., Yoon, K.: *Methods for Multiple Attribute decision-making*, pp. 58–191. Springer, Berlin, Heidelberg (1981)
- [53] Mankad, S., Han, H.S., Goh, J., Gavirneni, S.: Understanding online hotel reviews through automated text analysis. *Service Science* **8**(2), 124–138 (2016)
- [54] Dunn, J.C.: A fuzzy relative of the isodata process and its use in detecting compact well-separated clusters. *Journal of Cybernetics* (1973)
- [55] Labella, Á., Rodríguez, R.M., De Tré, G., Martínez, L.: A cohesion measure for improving the weighting of experts' subgroups in large-scale group decision-making clustering methods. In: *2019 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, pp. 1–6 (2019). IEEE
- [56] Opricovic, S., Tzeng, G.-H.: Compromise solution by mcdm methods: A comparative analysis of vikor and topsis. *European journal of operational research* **156**(2), 445–455 (2004)