



OPEN

A multi-dimensional decision framework based on the XGBoost algorithm and the constrained parametric approach

Xinxin Wang¹, BingBing Zhang¹, Zeshui Xu¹, Ming Li¹ & Marinko Skare^{2,3}✉

With the wide access to data and advanced technologies, organizations and firms prefer to use data-based and interpretable analytics to deal with uncertain and cognitive decision-making problems. In this regard, this study considers quantitative data and qualitative variables, to propose a multi-dimensional decision framework based on the nested probabilistic linguistic term sets. Under the framework, XGBoost algorithm, one of the machine learning methods, is conducted to capture the importance of attributes by using the historical data, and further calculate the attribute weights. The constrained parametric approach is used to establish membership functions of linguistic variables, and then get the objective probabilities in the linguistic model, so that we can obtain a scientific decision matrix. A case study concerning the ranking of bank credit is applied to present the proposed decision framework, and the process of making a rational decision. According to the comparative analysis, the proposed framework is flexible and the result is stable. Managers and policymakers determine the attribute weights by real data and choose the suitable decision method for a certain application. The framework provides an opportunity for capturing, integrating, analyzing data, and interpreting linguistic variables in the model to consider uncertain and cognitive decision at the both theoretical and practical levels.

Keywords Multi-dimensional decision, XGBoost algorithm, Nested probabilistic linguistic term sets, Constrained parametric approach, Bank credit

With the increasing complexity of decision-making problems, organizations and firms have been stuck in uncertainty and interpretability regarding the proper ways for translating prescriptive analytics use into organizational value¹. On the one hand, organizations highlight the significance of making rational decisions depending on objective information with the wide access to data and affordable technologies². On the other hand, managers and users prefer to deal with the descriptive models that provide meaningful and interpretive insights to support decision³. Therefore, decision-making using both quantitative data and qualitative variables has increased attention between the management science and operations research society to attack this type of problem⁴. Although a wealth of research achievements has been investigated considering machine learning, data analytics, and operation research methods^{5,6}, there are still some challenges in balancing objective data with interpretability under a multi-dimensional decision environment:

How to determine a set of attribute weights objectively in the multi-dimensional decision framework? Practical decision-making problems gradually require accurate evaluations from the multiple perspectives in most cases, therefore researchers have established relevant models for this purpose. Inspired by probabilistic linguistic term set (PLTS)⁷, the nested probabilistic linguistic term set (NPLTS) is an improved linguistic model that is suitable to describe uncertain and nested information^{8,9}. The model provides four decision-making scenarios based on the type of outer and inner linguistic variables, and has been widely applied in government investment¹⁰, communication technology¹¹, and offshore energy option¹². Attribute weight is an essential component of a decision-making framework. Ignoring rules or features of attributes would mislead the real weight, and impact the final decision to some extent¹³. Therefore, obtaining the attribute weight objectively is important to make a scientific decision, and exploring the feature attribute weights has become a popular trend¹⁴. Recent studies have been committed to developing this research direction, and proposed a series of methods to calculate the attribute

¹Business School, Sichuan University, Chengdu 610064, China. ²Juraj Dobrila University Pula, Fac Econ & Tourism Dr Mijo Mirkovic, Preradoviceva 1-1, 52100 Pula, Croatia. ³University of Economics and Human Sciences in Warsaw, Warsaw, Poland. ✉email: mskare@unipu.hr

weights. These include data-driven approaches such as the Analytic Hierarchy Process¹⁵, and consistency-driven techniques like personalized individual semantics¹⁶. Additionally, the probabilistic linguistic-Criteria Importance Through Intercriteria Correlation (CITIC) method¹⁷ and the extended entropy weight method for processing heterogeneous information in group decision-making¹⁸ have been proposed. Other approaches, such as deviation maximization combined with TOPSIS for temporal preferences¹⁹, as well as bidirectional projection and fuzzy entropy for handling unknown attribute weights²⁰, have further enriched the methodological landscape in this field. However, considering a multi-dimensional and cognitive decision-making problem, there is less research on determining the attribute weight based on historical data by using machine-learning methods, such as XGBoost algorithm, and decision tree algorithm, especially under the nested probabilistic linguistic decision environment. Table 1 presents some recent publications employing various attribute weighting methods.

How to portray the linguistic variables scientifically in the multi-dimensional decision framework? Linguistic model focused on the shape of a fuzzy set enhances the interpretability of the decision framework, while building the model that performs automatic reasoning is an increasingly important process as a necessary part of gathering data. Due to that any data collected from the real world comes with some degree of uncertainty, it is therefore crucial to design approaches that preserve the uncertainty in the modelling phase. Currently, fuzzy approaches have been popular to combine group of intervals representing the uncertainty, such as the interval approach²¹, the enhanced interval approach²², and the interval agreement approach²³. Constrained parametric approach (CPA) is a novel method to aggregate data instances modelled in a way that preserves the shape used to model individual opinions, enhancing the interpretability of the produced models²⁴. Considering that the CPA provides a way to choose the suitable shape, such triangular membership functions, to guarantee the representation of the acceptable instances, it has been applied to a case study involving combining data gathered from surveys. However, regarding the cognitive and uncertain decision-making problem, less literature considers interpretable membership functions to establish a linguistic model based on historical data, such as the CPA. It is a challenge to balance objective data with interpretable model under the nested probabilistic linguistic environment.

How to calculate the probabilities of cognitive information in the multi-dimensional decision framework? Probabilities in the linguistic model play an important part that affect the final decision directly, mainly representing confidence, ratio, weight, and proportion. If the evaluation information is provided by experts, the probabilities are relatively subjective. For example, a new-type smart city development evaluation problem is solved by using the probabilistic linguistic q-rung orthopair fuzzy set³¹. The selection of children English educational organization is considered using the two decision-making methods with probabilistic linguistic information³². Recent studies have preferred to use objective information to describe probabilities of linguistic variables. As for online information reviews, crawling comments and transforming to values is a popular way to obtain the probabilities, such as online learning platforms³¹, online doctor recommendation system³³, and the prioritization of improvements in hotel services³⁴. However, less study both considers the shape of the linguistic variable and the objective data, and it is still an open question to determine the probabilities in the linguistic model for multi-dimensional decision-making problems. In summary, Table 2 lists some recent publications using different models on the uncertain decision-making framework. As demonstrated, the Nested Probabilistic Linguistic Term Set (NPLTS) offers several key advantages over other fuzzy linguistic term sets, particularly in its ability to handle both quantitative and qualitative data, making it more versatile compared to models like Probabilistic Linguistic Term Set (PLTS) and Hesitant Fuzzy Linguistic Term Set with the Quality Function Deployment technique (HFLTS-QFD), which focus primarily on qualitative features. Additionally, NPLTS supports a multi-dimensional framework, a capability lacking in models such as PLTS and Double Hierarchy Hesitant Fuzzy Linguistic Term Set (DHHFLTS), allowing it to process multiple layers of information for more comprehensive decision-making. Moreover, NPLTS can incorporate weighting factors to reflect the varying importance of expert opinions or alternatives, a feature not available in simpler models like DHHFLTS and the Stimulus-Organism-Response framework (SOR). These strengths make NPLTS particularly well-suited for complex decision-making environments, such as ranking bank credit or evaluating multi-criteria scenarios, where nuanced data integration and expert consensus are crucial.

Aiming at these challenges both considering quantitative data and considering qualitative variables under the complex and uncertain environment, this study proposes a multi-dimensional decision framework combined with machine learning algorithm and constrained parametric approach, to calculate the attribute weights and probabilities in the linguistic model, respectively. Moreover, we apply the proposed framework to the ranking of bank credit, explore the robustness of the framework, and reveal the rules of attribute weights. To achieve these goals, the main contributions of this study are listed as follows:

Reference	Method
This paper	XGBoost algorithm
Ren, Zhu, Ren, & Ding (2022) ¹⁵	data-driven Analytic Hierarchy Process
Fan, Liang, Dong, & Pedrycz (2022) ¹⁶	consistency-driven personalized individual semantics
Zhang, Wu, Ma, Liu, & Wu (2022) ¹⁷	probabilistic linguistic-Criteria Importance Through Intercriteria Correlation
Ye et al. (2023) ¹⁸	Entropy weight for heterogeneous information
Chen, Tong, Gao, & Chen (2024) ¹⁹	Deviation maximization method
Fu & Xiao (2024) ²⁰	Bidirectional projection and fuzzy entropy

Table 1. Recent publications that use different attribute weighting methods. Sources: Authors’ own research.

Reference	Model	Application	Quantitative data	Qualitative feature	Multi-dimensional	Weighting factor
This paper	NPLTS	Ranking bank credit	√	√	√	√
Lin, Huang, Xu, & Chen (2020) ²⁵	PLTS	Evaluating IoT platforms	×	√	×	√
Finger & Lima (2022) ²⁶	HFLTS-QFD	Formulating supplier programs	×	√	×	√
Gou, Xu, & Zhou (2021) ²⁷	DHHFLTS	Diagnosis of lung cancer	×	√	√	×
Ma, Qin, Martinez, & Pedrycz (2023) ²⁸	IT2FS	Recommending tourist attractions	×	√	×	√
Liu, Liu, & Jiang (2022) ²⁹	SOR	Social commerce	√	×	√	×
Yuan et al. (2023) ³⁰	ZRO	Public-Private-Partnership	√	×	×	√

Table 2. Recent publications on the decision-making framework. Note. The full names of the abbreviations in Table 2 are: NPLTS-nested probabilistic linguistic term set, PLTS-probabilistic linguistic term set, HFLTS-QFD-hesitant fuzzy linguistic term set with the quality function deployment technique, DHHFLTS-double hierarchy hesitant fuzzy linguistic term set, IT2FS-interval type-2 fuzzy set, SOR-stimulus-organism-response framework, ZRO- Z-number based real option model. Sources: Authors' own research.

(1) To address the first issue, we establish a multi-dimensional decision-making framework using the NPLTSs. Under the framework, we use the XGBoost algorithm to capture the importance of attributes by using the historical data, and further determine the attribute weights. This process obtains the appropriate and objective weights by leveraging large volumes of data to recognize the features and their importance degrees.

(2) To address the second issue, we conduct the constrained parametric approach (CPA) to construct triangular membership functions of linguistic variables in the linguistic model, also depending on the prior data. In this way, the model not only controls the uncertainty from objective data, but also guarantees the interpretability.

(3) To address the last issue, we calculate the probabilities in the linguistic model using the historical data and obtained membership functions. In addition, we apply the decision-making framework to the ranking of bank credit, and verify the robustness considering attribute weights and different decision-making methods by comparative analysis.

The remainder of this study is organized as follows: Section "Preliminaries" briefly reviews the relevant knowledge. Section "A multi-dimensional decision-making framework" proposes a multi-dimensional decision framework based on the XGBoost algorithm and the constrained parametric approach under the nested probabilistic linguistic environment. A case study considering the ranking of bank credit is presented in Section "Case study: the ranking of bank credit". In Section "Discussions", discussions and implications are made from theoretical and practical aspects. Section "Conclusions" ends the study with some conclusions.

Preliminaries

This section reviews the relevant knowledge of the nested probabilistic linguistic term set (NPLTS), and the Constrained Parametric Approach (CPA) for modelling uncertain data.

Nested probabilistic linguistic term set

To better describe multi-dimensional and uncertain information that conforms to human cognition, a nested probabilistic linguistic term set (NPLTS) was established by a set of nested, ordered, and continuous linguistic terms with the corresponding probabilities, which could represent the confidence coefficient, proportion, importance, weight or consistency of each linguistic term.

Let an outer linguistic term set be $S_O = \{s_\alpha | \alpha = -\tau, \dots, -1, 0, 1, \dots, \tau\}$, and the corresponding inner linguistic term set be $S_I = \{n_\beta | \beta = -\varsigma, \dots, -1, 0, 1, \dots, \varsigma\}$. The nested linguistic term set is defined as $S_N = \{s_\alpha \{n_\beta\} | \alpha = -\tau, \dots, -1, 0, 1, \dots, \tau; \beta = -\varsigma, \dots, -1, 0, 1, \dots, \varsigma\}$, where $s_\alpha \{n_\beta\}$ is called the nested linguistic term⁹. A mapping function from a given set X to a subset of S_N , the NPLTS is denoted as $P_{S_N} = \{< x_i, p_{S_N}(x_i) > | x_i \in X\}$, where $p_{S_N}(x_i)$ is the set of several NPLTSs in S_N , expressed as Wang, Xu, Wen & Li⁸:

$$p_{S_N}(x_i) = \left\{ \begin{array}{l} s_{\alpha(k)}(p_{s(k)}) \{n_{\beta(l)}(p_{n(l)})\} (x_i) s_{\alpha(k)} \{n_{\beta(l)}\} \in S_N, p_{s(k)} > 0, p_{n(l)} > 0, \\ k = 1, 2, \dots, \#s_\alpha(x_i), l = 1, 2, \dots, \#n_\beta(x_i), \\ \alpha = -\tau, \dots, -1, 0, 1, \dots, \tau, \beta = -\varsigma, \dots, -1, 0, 1, \dots, \varsigma \end{array} \right\} \quad (1)$$

where $\#s_\alpha(x_i)$ is the number of the outer linguistic term elements, $\#n_\beta(x_i)$ is the number of the inner linguistic term elements, and the probabilities satisfy $\theta_P(\zeta)$.

According to the types of the linguistic variables in the outer linguistic term set and the inner linguistic term set, respectively, the NPLTS faces four scenarios for information fusion, and Wang et al.¹² proposed the transformation functions as follows:

$f^{(1)}$ in Case 1:

$$f^{(1)} : [-\tau, \tau] \times [-\varsigma, \varsigma] \rightarrow [0, 1]$$

$$f^{(1)} \left(P_{S_N}^{(k,l)} \right) = f \left(S_{\alpha(k)} \left(p_{s(k)} \right) \left\{ n_{\beta(l)} \left(p_{n(l)} \right) \right\} \right) = p_{s(k)} \left(\frac{\tau + \alpha(k)}{4\tau} + \frac{\beta(l) \times p_{n(l)} + \varsigma}{4\varsigma} \right) = \gamma \quad (2)$$

$f^{(2)}$ in Case 2:

$$f^{(2)} : [-\tau, \tau] \times [-\varsigma, \varsigma] \rightarrow [0, 1]$$

$$f^{(2)} \left(P_{S_N}^{(k,l)} \right) = f \left(S_{\alpha(k)} \left(p_{s(k)} \right) \left\{ n_{\beta(l)} \left(p_{n(l)} \right) \right\} \right) = p_{s(k)} \left(\frac{\tau + \alpha(k)}{4\tau} + \frac{\text{len}(n_{\beta(l)}) \times p_{n(l)} + 1}{4} \right) = \gamma \quad (3)$$

$f^{(3)}$ in Case 3:

$$f^{(3)} : [-\tau, \tau] \times [-\varsigma, \varsigma] \rightarrow [0, 1]$$

$$f^{(3)} \left(P_{S_N}^{(k,l)} \right) = f \left(S_{\alpha(k)} \left(p_{s(k)} \right) \left\{ n_{\beta(l)} \left(p_{n(l)} \right) \right\} \right) = p_{s(k)} \left(\frac{1 + \text{len}(s_{\alpha(k)})}{4} + \frac{\beta(l) \times p_{n(l)} + \varsigma}{4\varsigma} \right) = \gamma \quad (4)$$

$f^{(4)}$ in Case 4:

$$f^{(4)} : [-\tau, \tau] \times [-\varsigma, \varsigma] \rightarrow [0, 1]$$

$$f^{(4)} \left(P_{S_N}^{(k,l)} \right) = f \left(S_{\alpha(k)} \left(p_{s(k)} \right) \left\{ n_{\beta(l)} \left(p_{n(l)} \right) \right\} \right) = p_{s(k)} \left(\frac{1 + \text{len}(s_{\alpha(k)})}{4} + \frac{\text{len}(n_{\beta(l)}) \times p_{n(l)} + 1}{4} \right) = \gamma \quad (5)$$

where $\text{len}(s_{\alpha(k)})$ and $\text{len}(n_{\beta(l)})$ represent the positive length of elements in the outer linguistic term set and the inner linguistic term set, respectively. In this paper, the application scenario of the case study conforms to Case 3.

Constrained parametric approach

Constrained Parametric Approach (CPA) was proposed by D'Alterio et al.²⁴, to aggregate data instance modelled in a way that preserves the shape used to model individual opinions.

Let $C = \{C_1, C_2, \dots, C_n\}$ ($n \geq 2$) be a set of attributes, in this study, we use the CPA to construct the triangular membership function of each attribute for the reason that the triangular membership function not only has good flexibility and effectively suppresses the noise, but also improves the control accuracy and reduces the calculation amount of the fuzzy control system, so as to the size of the fuzzy number can be adjusted according to the actual situation, and make the system more stable, accurate, and reliable. For each attribute C_j , the triangular membership function has three parameters (a, b, c), which represents the start point, the peak point, and the end point, respectively. The constraints of each triangular membership function of C_j satisfy the following requirements (D'Alterio et al., 2022):

(1) P_0 : the membership function of C_j is in the following form to ensure that it is an actual triangle:

$$P_0 : \mu_{C_j}(x) = \begin{cases} \frac{x-a}{b-a}, & x \in [a, b] \\ \frac{b-x}{c-b}, & x \in [b, c] \\ 0, & x < a \vee x > c \end{cases} \quad (6)$$

(2) P_1 : check that the start point is with the uncertainty ranges, represented as $a \in [a_{\min}, a_{\max}]$;

(3) P_2 : check that the peak point is with the uncertainty ranges, represented as $b \in [b_{\min}, b_{\max}]$;

(4) P_3 : check that the end point is with the uncertainty ranges, represented as $c \in [c_{\min}, c_{\max}]$.

After calculating above four parameters, the triangular membership functions of attributes are modelled through parametric fuzzy sets according to the available data set. In this study, we use the CPA to complete the NPLTSs by obtaining the probabilities of the outer linguistic term sets, and the inner linguistic term sets.

A multi-dimensional decision-making framework

To make a rational and scientific decision, we propose a novel decision-making framework with NPLTSs due to optimizing the attribute weights by the XGBoost algorithm and describing cognitive information by the CPA. The framework can be applied to various fields in practice. Figure 1 shows the flow chart of the NPLTSs-based multi-dimensional decision-making framework.

The decision-making framework under nested probabilistic linguistic environment

Let $A = (A_1, A_2, \dots, A_m)$ ($m \geq 2$) be a set of alternatives, $C = (C_1, C_2, \dots, C_n)$ ($n \geq 2$) be a set of attributes, $w = (w_1, w_2, \dots, w_n)$ ($n \geq 2$) be a set of weights with respect to attributes with $w_j \geq 0$ and $\sum_{j=1}^n w_j = 1$. To comprehensively describe the alternatives A_i ($i = 1, 2, \dots, m$) with the attributes C_j ($j = 1, 2, \dots, n$) in a multi-dimensional and uncertain decision-making problem, nested probabilistic linguistic sets (NPLTSs) are used to evaluate the characteristics of alternatives by inner and outer linguistic terms, and the corresponding probability represents preferences or hesitations. According to the nested

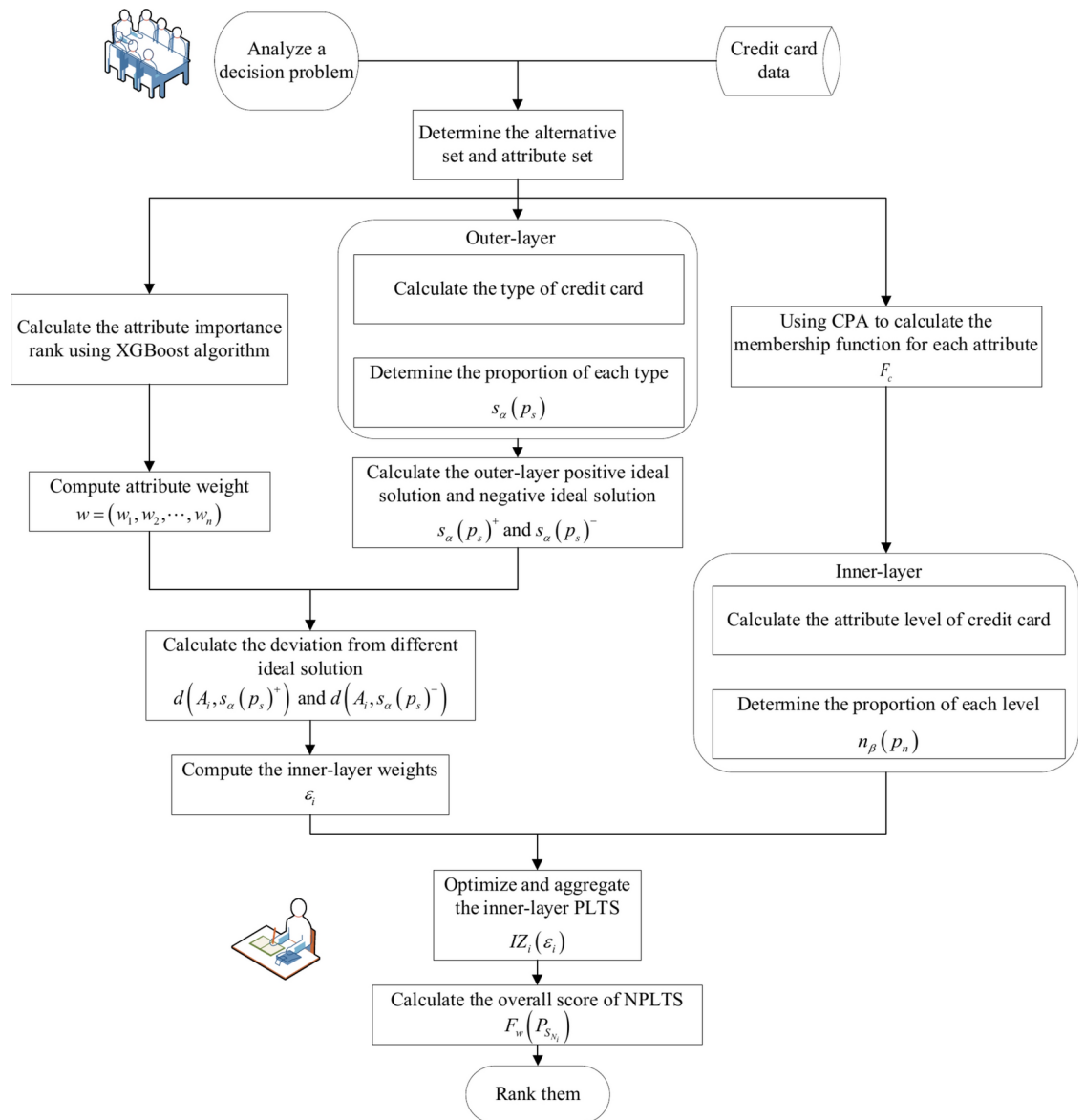


Fig. 1. The flow chart of the NPLTS-based multi-dimensional decision-making framework.

probabilistic linguistic model, $S_O = \{s_\alpha | \alpha = -\tau, \dots, -1, 0, 1, \dots, \tau\}$ is the outer linguistic term set, $S_I = \{n_\beta | \beta = -\varsigma, \dots, -1, 0, 1, \dots, \varsigma\}$ is the inner linguistic term set. NPLTSs are constructed to evaluate the alternatives A_i ($i = 1, 2, \dots, m$) with respect to the attributes C_j ($j = 1, 2, \dots, n$), and then the corresponding NPLTS decision matrix $B_N = P_{S_{N_{ij}}}$ ($i = 1, 2, \dots, m; j = 1, 2, \dots, n$) is shown as follows:

$$B_N = \begin{bmatrix} P_{S_{N_{11}}} & P_{S_{N_{12}}} & \cdots & P_{S_{N_{1j}}} & \cdots & P_{S_{N_{1n}}} \\ P_{S_{N_{21}}} & P_{S_{N_{22}}} & \cdots & P_{S_{N_{2j}}} & \cdots & P_{S_{N_{2n}}} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ P_{S_{N_{i1}}} & P_{S_{N_{i2}}} & \cdots & P_{S_{N_{ij}}} & \cdots & P_{S_{N_{in}}} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ P_{S_{N_{m1}}} & P_{S_{N_{m2}}} & \cdots & P_{S_{N_{mj}}} & \cdots & P_{S_{N_{mn}}} \end{bmatrix} \quad (7)$$

where $P_{S_{N_{ij}}}$ represents the NPLTS of the i -th alternative with respect to the j -th attribute. In the following, we go on two procedures to obtain the attribute weights and probabilities of inner linguistic terms, and outer linguistic terms, respectively.

Determining attribute weights using the XGBoost algorithm

XGBoost (eXtreme Gradient Boosting), a scalable machine learning system for tree boosting, is widely applied to sales forecasting, network text classification, and risk prediction^{35,36}. In the multi-dimensional decision-making framework, we use the algorithm to determine the attribute weights. For the test data of the alternative $A = (A_1, A_2, \dots, A_m)$ ($m \geq 2$), there are classification labels of all alternatives based on the attribute $C = (C_1, C_2, \dots, C_n)$ ($n \geq 2$). After the algorithm processing with the input data, including attributes information, the importance of each attribute that is a vector of weight $w = (w_1, w_2, \dots, w_n)$ ($n \geq 2$) will be obtained by the optimized distributed gradient enhancement library. To realize the process clearly in such a decision-making environment, the following describes the XGBoost algorithm:

Step1. Split data set. Splitting the data set into a train set and a test set, and then using the train set to train the model, and using the test set to evaluate the model's performance on previously unseen data. Here, we split the train set and the test set in 4:1.

Step2. Apply Gridsearch to optimize the parameters. Grid search is an exhaustive search method that searches for the optimal hyperparameter by traversing all possible combinations of hyperparameters. We use the commonly used the GridSearchCV package to complete the grid search process and find the optimal parameter combination.

Step3. Calculate importance. Typically, the importance provides a score that indicates the usefulness or value of each feature in the enhanced decision tree within the built model. The more attributes that use the decision tree to make key decisions, the higher their relative importance. This importance is explicitly calculated for each attribute in the data set, allowing the ranking of attributes to be compared with each other. We calculate attribute importance using the functions of the xgboost package.

Obtaining NPLTSs using the constrained parametric approach

To complete NPLTSs scientifically, we obtain the outer probabilities and the inner probabilities in each NPLTS, respectively, using the data set and the triangular membership function of each attribute established by the CPA. For a decision-making problem, let a set of alternatives be $A = (A_1, A_2, \dots, A_m)$ ($m \geq 2$), and a set of attributes be $C = (C_1, C_2, \dots, C_n)$ ($n \geq 2$). According to the definition of NPLTS, the outer probabilistic linguistic term set is $s_{\alpha_{ij}}(p_{s_{ij}}) = \{s_{\alpha(k)_{ij}}(p_{s(k)_{ij}}) | i = 1, 2, \dots, m; j = 1, 2, \dots, n; k = 1, 2, \dots, \#s_{\alpha}\}$ to describe the evaluation information with respect to attributes and outer linguistic terms. The inner probabilistic linguistic term set is $n_{\beta_{kl}}(p_{n_{kl}}) = \{n_{\beta(l)_k}(p_{n(l)_k}) | k = 1, 2, \dots, \#s_{\alpha}; l = 1, 2, \dots, \#n_{\beta}\}$ used to describe the evaluation information considering both the outer linguistic terms and the inner linguistic terms.

For the element $n_{\beta(l)_k}(p_{n(l)_k})$ of the linguistic inner term set,

$$n_{\beta(l)_k}(p_{n(l)_k}) = \sum_{i=1}^m \sum_{j=1}^n n_{\beta(l)_{ijk}}(p_{n(l)_{ijk}}), i = 1, 2, \dots, m; j = 1, 2, \dots, n; k = 1, 2, \dots, \#s_{\alpha} \quad (8)$$

The first step is to calculate the outer linguistic probabilities. According to the historical data, we collect the prior data of each alternative A_i with respect to the attribute C_j in terms of the outer linguistic term $s_{\alpha(k)}$, represented as $Data_{ijk} = \{g_{ijk} | i = 1, 2, \dots, m; j = 1, 2, \dots, n; k = 1, 2, \dots, \#s_{\alpha}\}$. At the same time, the feature is extracted denoted as $feature_{ijk}$ ($i = 1, 2, \dots, m; j = 1, 2, \dots, n; k = 1, 2, \dots, \#s_{\alpha}$). Therefore, the probability $p_{s(k)_{ij}}$ in the outer probabilistic linguistic term set $s_{\alpha_{ij}}(p_{s_{ij}})$ is calculated in Eq. (9). Table 3 provides the framework of the outer-layer decision matrix $R_O = (s_{\alpha_{ij}}(p_{s_{ij}}))_{m \times n}$.

$$p_{s(k)_{ij}} = feature_{ijk} / \sum_{k=1}^{\#s_{\alpha}} feature_{ijk}, i = 1, 2, \dots, m; j = 1, 2, \dots, n; k = 1, 2, \dots, \#s_{\alpha} \quad (9)$$

The second step is to calculate the inner linguistic probabilities. According to the CPA, we obtain the membership functions of each attribute with respect to the inner linguistic term, denoted as MF_{jl} . Put prior data $Data_{ijk}$ into the membership functions MF_{jl} , then we get the interval values $[a_{j1}, b_{j1}], [a_{j2}, b_{j2}], \dots, [a_{j\#n_{\beta}}, b_{j\#n_{\beta}}]$, and determine the inner linguistic term $n_{\beta_{kl}}$. Let $N_{n_{\beta_{ijkl}}}$ be the number of the inner linguistic terms. Therefore,

	C_1	C_2	\dots	C_n
A_1	$s_{\alpha_{11}}(p_{s_{11}})$	$s_{\alpha_{12}}(p_{s_{12}})$	\dots	$s_{\alpha_{1n}}(p_{s_{1n}})$
A_2	$s_{\alpha_{21}}(p_{s_{21}})$	$s_{\alpha_{22}}(p_{s_{22}})$	\dots	$s_{\alpha_{2n}}(p_{s_{2n}})$
\dots	\dots	\dots	\ddots	\dots
A_m	$s_{\alpha_{m1}}(p_{s_{m1}})$	$s_{\alpha_{m2}}(p_{s_{m2}})$	\dots	$s_{\alpha_{mn}}(p_{s_{mn}})$

Table 3. The framework of the outer-layer decision matrix.

	n_{β_1}	n_{β_2}	\dots	$n_{\beta_{\#n_\beta}}$
s_{α_1}	$n_{\beta_{11}}(p_{n_{11}})$	$n_{\beta_{12}}(p_{n_{12}})$	\dots	$n_{\beta_{1,\#n_\beta}}(p_{n_{1,\#n_\beta}})$
s_{α_2}	$n_{\beta_{21}}(p_{n_{21}})$	$n_{\beta_{22}}(p_{n_{22}})$	\dots	$n_{\beta_{2,\#n_\beta}}(p_{n_{2,\#n_\beta}})$
\dots	\dots	\dots	\ddots	\dots
$s_{\alpha_{\#s_\alpha}}$	$n_{\beta_{\#s_\alpha,1}}(p_{n_{\#s_\alpha,1}})$	$n_{\beta_{\#s_\alpha,2}}(p_{n_{\#s_\alpha,2}})$	\dots	$n_{\beta_{\#s_\alpha,\#n_\beta}}(p_{n_{\#s_\alpha,\#n_\beta}})$

Table 4. The framework of the inner-layer decision matrix.

	C_1	C_2	\dots	C_n
A_1	$PS_{N_{11}}$	$PS_{N_{12}}$	\dots	$PS_{N_{1n}}$
A_2	$PS_{N_{21}}$	$PS_{N_{22}}$	\dots	$PS_{N_{2n}}$
\dots	\dots	\dots	\ddots	\dots
A_m	$PS_{N_{m1}}$	$PS_{N_{m2}}$	\dots	$PS_{N_{mn}}$

Table 5. The framework of the NPLTS-based decision matrix.

the probability $p_{n(l)_{ijk}}$ in the inner probabilistic linguistic term set $n_{\beta_{kl}}(p_{n_{kl}})$ is calculated in Eq. (10). Table 4 provides the framework of the inner-layer decision matrix $R_I = (n_{\beta_{kl}}(p_{n_{kl}}))_{\#s_\alpha \times \#n_\beta}$.

$$p_{n(l)_{ijk}} = N_{n_{\beta_{ijkl}}} / \sum_{l=1}^{\#n_\beta} N_{n_{\beta_{ijkl}}}, i = 1, 2, \dots, m; j = 1, 2, \dots, n; k = 1, 2, \dots, \#s_\alpha; l = 1, 2, \dots, \#n_\beta \quad (10)$$

The last step is to combine with the outer-layer decision matrix and the inner-layer decision matrix, and get the NPLTS-based decision matrix $R = (PS_{N_{ij}})_{m \times n}$, listed in Table 5. For $PS_{N_{ij}}$,

$$PS_{N_{ij}} = \left\{ \begin{array}{l} s_{\alpha(1)_{ij}}(p_{s(1)_{ij}}) \left\{ n_{\beta(1)_{ij1}}(p_{n(1)_{ij1}}), \dots, n_{\beta(\#n_\beta)_{ij1}}(p_{n(\#n_\beta)_{ij1}}) \right\}, \\ s_{\alpha(2)_{ij}}(p_{s(2)_{ij}}) \left\{ n_{\beta(1)_{ij2}}(p_{n(1)_{ij2}}), \dots, n_{\beta(\#n_\beta)_{ij2}}(p_{n(\#n_\beta)_{ij2}}) \right\}, \\ \dots, \\ s_{\alpha(\#s_\alpha)_{ij}}(p_{s(\#s_\alpha)_{ij}}) \left\{ n_{\beta(\#s_\alpha)_{ij\#s_\alpha}}(p_{n(\#s_\alpha)_{ij\#s_\alpha}}), \dots, n_{\beta(\#n_\beta)_{ij\#s_\alpha}}(p_{n(\#n_\beta)_{ij\#s_\alpha}}) \right\} \end{array} \right\} \quad (11)$$

Deriving alternative rankings via NPLTS-based TOPSIS

According to the definition of NPLTS, the outer probabilistic linguistic term set is $s_{\alpha_{ij}}(p_{s_{ij}}) = \{s_{\alpha(k)_{ij}}(p_{s(k)_{ij}}) | i = 1, 2, \dots, m; j = 1, 2, \dots, n; k = 1, 2, \dots, \#s_\alpha\}$. For each attribute C_j , calculate the outer-layer positive ideal solution and negative ideal solution $s_\alpha(p_s)^+$ and $s_\alpha(p_s)^-$ as follows:

$$\begin{aligned} s_\alpha(p_s)^+ &= \{S_{\alpha_1}(p_{S_1})^+, S_{\alpha_2}(p_{S_2})^+, \dots, S_{\alpha_n}(p_{S_n})^+\} \\ s_\alpha(p_s)^- &= \{S_{\alpha_1}(p_{S_1})^-, S_{\alpha_2}(p_{S_2})^-, \dots, S_{\alpha_n}(p_{S_n})^-\} \end{aligned} \quad (12)$$

$$\begin{aligned} \text{where} \quad S_{\alpha_j}(p_{S_j})^+ &= \left\{ (S_{\alpha_j}^{(k)})^+ | k = 1, 2, \dots, \#S_{\alpha_{ij}}(PS_{ij}) \right\} \\ , \quad (S_{\alpha_j}^{(k)})^+ &= S_{(k)} \left(\max_i \{p_{ij}^{(k)}\} \right), j = 1, 2, \dots, n \\ S_{\alpha_j}(p_{S_j})^- &= \left\{ (S_{\alpha_j}^{(k)})^- | k = 1, 2, \dots, \#S_{\alpha_{ij}}(PS_{ij}) \right\}, (S_{\alpha_j}^{(k)})^- = S_{(k)} \left(\min_i \{p_{ij}^{(k)}\} \right), j = 1, 2, \dots, n. \end{aligned}$$

Subsequently, incorporating the attribute weights $w = (w_1, w_2, \dots, w_n)$ derived from the XGBoost algorithm, the deviation between each alternative and the positive ideal solution is calculated:

$$d(A_i, s_\alpha(p_s)^+) = w_j d(s_{\alpha_{ij}}(p_{s_{ij}}), s_\alpha(p_s)^+) = w_j \sqrt{\frac{1}{\#s_{\alpha_{ij}}(p_{s_{ij}})} \sum_{k=1}^{\#s_{\alpha_{ij}}(p_{s_{ij}})} \left(p_{s_{ij}}^{(k)} - (p_{s_{ij}}^{(k)})^+\right)^2} \quad (13)$$

and the deviation between each alternative and the negative ideal solution is

$$d(A_i, s_\alpha(p_s)^-) = w_j d(s_{\alpha_{ij}}(p_{s_{ij}}), s_\alpha(p_s)^-) = w_j \sqrt{\frac{1}{\#s_{\alpha_{ij}}(p_{s_{ij}})} \sum_{k=1}^{\#s_{\alpha_{ij}}(p_{s_{ij}})} \left(p_{s_{ij}}^{(k)} - (p_{s_{ij}}^{(k)})^-\right)^2} \quad (14)$$

The smaller the deviation $d(A_i, s_\alpha(p_s)^+)$, the better the alternative A_i , and the larger the deviation $d(A_i, s_\alpha(p_s)^-)$, the better the alternative A_i . Let

$$d_{min}(A_i, s_\alpha(p_s)^+) = \min_{1 \leq i \leq m} d(A_i, s_\alpha(p_s)^+) \quad (15)$$

be the smallest deviation between the alternative and the positive ideal solution,

$$d_{max}(A_i, s_\alpha(p_s)^-) = \min_{1 \leq i \leq m} d(A_i, s_\alpha(p_s)^-) \quad (16)$$

be the largest deviation between the alternative and the negative ideal solution. According to the improved closeness coefficient method³⁷, the inner weights are calculated as follows:

$$\varepsilon_i = \omega_{CI_i} \left(\frac{d(A_i, s_\alpha(p_s)^-)}{d_{max}(A_i, s_\alpha(p_s)^-)} + \frac{d(A_i, s_\alpha(p_s)^+)}{d_{min}(A_i, s_\alpha(p_s)^+)} \right) / \left(\sum_{i=1}^m \left(\frac{d(A_i, s_\alpha(p_s)^-)}{d_{max}(A_i, s_\alpha(p_s)^-)} + \frac{d(A_i, s_\alpha(p_s)^+)}{d_{min}(A_i, s_\alpha(p_s)^+)} \right) \right) \quad (17)$$

It is noted that the inner weight is a vector, that is $\varepsilon_i = (\varepsilon_{i(1)}, \varepsilon_{i(2)}, \dots, \varepsilon_{i(\#n_\beta)})^T$. And the inner-layer score $I Z_i$ is expressed as:

$$\begin{aligned} I Z_i(\varepsilon_i) &= \varepsilon_{i(1)} n_{\beta(i)}(p_{n(i)}) \oplus \varepsilon_{i(2)} n_{\beta(i)}(p_{n(i)}) \oplus \dots \oplus \varepsilon_{i(\#n_\beta)} n_{\beta(i)}(p_{n(i)}) \# \\ &= \varepsilon_{i(1)} \bigcup_{n_{\beta(i)(1)}(p_{n(i)(1)}) \in n_{\beta(i)}(p_{n(i)})} \{n_{\beta(i)(1)}(p_{n(i)(1)})\} \\ &\quad \oplus \varepsilon_{i(2)} \bigcup_{n_{\beta(i)(2)}(p_{n(i)(2)}) \in n_{\beta(i)}(p_{n(i)})} \{n_{\beta(i)(2)}(p_{n(i)(2)})\} \oplus \dots \\ &\quad \oplus \varepsilon_{i(\#n_\beta)} \bigcup_{n_{\beta(i)(\#n_\beta)}(p_{n(i)(\#n_\beta)}) \in n_{\beta(i)}(p_{n(i)})} \{n_{\beta(i)(\#n_\beta)}(p_{n(i)(\#n_\beta)})\} \end{aligned} \quad (18)$$

where $i = 1, 2, \dots, m$.

Finally, the whole score $F_w(P_{S_{N_i}})$ of each alternative can be obtained with the inner weight $\varepsilon_i = (\varepsilon_{i(1)}, \varepsilon_{i(2)}, \dots, \varepsilon_{i(\#n_\beta)})^T$ by the following steps:

The outer-layers's score of $s_\alpha(p_s)$ is

$$OE(s_\alpha(p_s)) = s_{\bar{\alpha}}, \quad (19)$$

where $\bar{\alpha} = \frac{\sum_{k=1}^{\#s_\alpha} \alpha(k) p_{s(k)}}{\sum_{k=1}^{\#s_\alpha} p_{s(k)}}$, and the inner-layer's score of $n_\beta(p_n)$ is

$$IE(n_\beta(p_n)) = \left\{ \bar{p}_{n(k)}^{(l)} \mid k = 1, 2, \dots, \#s_{\alpha(k)}, l = 1, 2, \dots, \#n_{\beta(l)} \right\} \quad (20)$$

where $\bar{p}_{n(k)}^{(l)} = ([\bar{\alpha}] - \bar{\alpha}) \times n_{(\lfloor \bar{\alpha} \rfloor)}(p_{n(k)}^{(l)}) + (\bar{\alpha} - [\bar{\alpha}]) \times n_{(\lceil \bar{\alpha} \rceil)}(p_{n(k)}^{(l)})$, $[\bar{\alpha}]$ means the smallest integer greater than $\bar{\alpha}$, and $\lfloor \bar{\alpha} \rfloor$ means the greatest integer less than $\bar{\alpha}$. Then, the whole score of $F_w(P_{S_{N_i}})$ is

$$F_w(P_{S_{N_i}}) = \frac{\sum_{l=1}^{\#n_\beta} \bar{p}_{n(k)}^{(l)}}{\#n_\beta} \quad (21)$$

The comparison laws of any two NPLTS, $P_{S_{N_1}}$ and $P_{S_{N_2}}$, can be presented as follows:

- (1) If $F_w(P_{S_{N_1}}) > F_w(P_{S_{N_2}})$, then $P_{S_{N_1}} \succ P_{S_{N_2}}$
- (2) If $F_w(P_{S_{N_1}}) = F_w(P_{S_{N_2}})$, then $P_{S_{N_1}} \prec P_{S_{N_2}}$
- (3) If $F_w(P_{S_{N_1}}) = F_w(P_{S_{N_2}})$, then further compared with deviation function. The deviation degree of NPLTS is

$$\sigma(P_{S_N}) = \left(\frac{\sum_{l=1}^{\#n_\beta} (p_{n(k)}^{(l)} - F_w(P_{S_N}))^2}{\#n_\beta} \right)^{1/2} \quad (22)$$

Therefore, $\sigma(P_{S_{N_1}}) > \sigma(P_{S_{N_2}})$, then $P_{S_{N_1}} \prec P_{S_{N_2}}$; if $\sigma(P_{S_{N_1}}) < \sigma(P_{S_{N_2}})$, then $P_{S_{N_1}} \succ P_{S_{N_2}}$, and if $\sigma(P_{S_{N_1}}) = \sigma(P_{S_{N_2}})$, then $P_{S_{N_1}} \sim P_{S_{N_2}}$.

Hence, we can rank the alternatives and select the best one. In particular, the comparison rule between the A_i and the A_j ($i, j = 1, 2, \dots, m; i \neq j$) is that.

- (4) If $F_w(P_{S_{N_i}}) > F_w(P_{S_{N_j}})$, then $A_i > A_j$
- (5) If $F_w(P_{S_{N_i}}) < F_w(P_{S_{N_j}})$, then $A_i < A_j$
- (6) If $F_w(P_{S_{N_i}}) = F_w(P_{S_{N_j}})$, further compare $\sigma(P_{S_{N_i}})$ and $\sigma(P_{S_{N_j}})$:
 - a) $\sigma(P_{S_{N_i}}) > \sigma(P_{S_{N_j}})$, then $A_i < A_j$
 - b) $\sigma(P_{S_{N_i}}) < \sigma(P_{S_{N_j}})$, then $A_i > A_j$
 - c) $\sigma(P_{S_{N_i}}) = \sigma(P_{S_{N_j}})$, then $A_i \sim A_j$

Case study: the ranking of bank credit

In this section, a case study on ranking the bank credit is presented to choose the safest banking institution by the proposed multi-dimensional decision-making framework based on the XGBoost algorithm and constraint parametric approach.

Case description

Bank credit has become an essential part of the considerable factors for people to choose financial service institution, especially after the 2008 global financial crisis³⁸. Modern payment systems provide a broad range of services to the users of the system, and it is one of the most advanced forms of payment³⁹. However, bank credit still suffers from some fraud problems, including card-present and online fraud. Many fraud detection systems are prone to difficulties and challenges considering overlapping data, imbalanced data, and adaptability⁴⁰. Accordingly, for individual or organizational level, choosing the optimal financial institution could decrease the risk of financial fraud, and security defense technologies are promising solutions for mid-term and long-term objectives to avoid financial fraud. In this study, we consider four alternative banks, denoted as A_1, A_2, A_3, A_4 .

Numerous factors help determine whether a transaction involves credit card fraud. In this study, we focus on four key aspects: (1) The number of replacing the device (C_1), which can indicate attempts to access the account from different or unauthorized devices; (2) The number of payment failures (C_2), which may suggest unauthorized usage or testing of stolen card information; (3) The number of IP addresses changed (C_3), as frequent changes may signal an effort to disguise the user's location; and (4) The number of IP country changed (C_4), as sudden cross-border shifts can be a red flag for potential fraud. These four variables are chosen because they capture crucial elements of suspicious network and device behavior, offering early indicators for detecting fraudulent activities. These four variables are chosen because they capture vital elements of suspicious network and device behavior, offering early indicators for detecting fraudulent activities. Financial institutions commonly use the location of online users in their fraud detection algorithms, and by analyzing the changes in IP addresses and countries⁴¹, we can enhance the effectiveness of detecting such fraudulent behavior. In general, financial systems evaluate the above-mentioned attributes from five aspects, i.e., food, reside, medical, education, and transportation. The outer linguistic term set is $S_0 = \{s_{-2} = \text{food}, s_{-1} = \text{reside}, s_0 = \text{medical}, s_1 = \text{education}, s_2 = \text{transportation}\}$ and the inner linguistic term set is given as $S_I = \{n_{-1} = \text{low}, n_0 = \text{medium}, n_1 = \text{high}\}$.

Solve the case

In the following, we solve the decision-making case by the proposed framework described in Section "A multi-dimensional decision-making framework".

Step 1. According to a set of publicly available data from a book *Python Big Data Analysis and Machine Learning Business Case Practice*⁴², we calculate the attribute weights as $w = (0.333, 0.238, 0.143, 0.286)$. Figure 2 shows feature importance of four attributes by using the XGBoost algorithm.

Sources: Authors' own research.

Step 2. Calculate the outer probabilistic linguistic term sets by Eq. (8), and obtain the outer-layer decision matrix, listed in Table 6.

Step 3. Calculate the inner probabilistic linguistic term sets. First, we construct the triangular membership functions of four attributes by using CPA, formed as:

$$P_0 : \mu_C(x) = \begin{cases} \frac{x-a}{b-a}, & x \in [a, b] \\ \frac{b-x}{c-b}, & x \in [b, c] \\ 0, & x < a \vee x > c \end{cases}$$

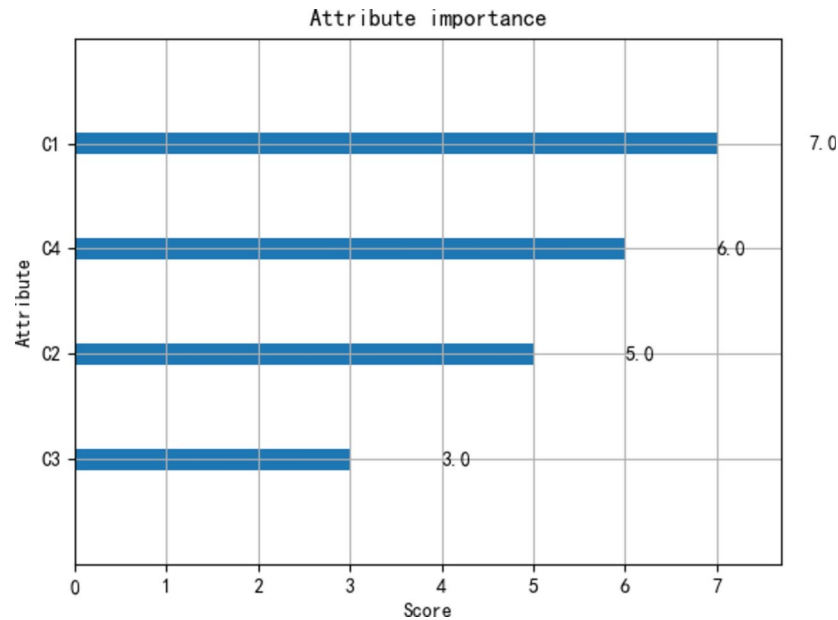


Fig. 2. Feature importance of four attributes by using XGBoost algorithm.

	C ₁	C ₂	C ₃	C ₄
A ₁	$\left\{ \begin{array}{l} s_{-2} (0.1886) \\ s_{-1} (0.1991) \\ s_0 (0.2043) \\ s_1 (0.2032) \\ s_2 (0.2047) \end{array} \right\}$	$\left\{ \begin{array}{l} s_{-2} (0.2173) \\ s_{-1} (0.1989) \\ s_0 (0.1745) \\ s_1 (0.2108) \\ s_2 (0.1985) \end{array} \right\}$	$\left\{ \begin{array}{l} s_{-2} (0.2041) \\ s_{-1} (0.1881) \\ s_0 (0.1962) \\ s_1 (0.2070) \\ s_2 (0.2046) \end{array} \right\}$	$\left\{ \begin{array}{l} s_{-2} (0.2021) \\ s_{-1} (0.2011) \\ s_0 (0.1916) \\ s_1 (0.2001) \\ s_2 (0.2052) \end{array} \right\}$
A ₂	$\left\{ \begin{array}{l} s_{-2} (0.1750) \\ s_{-1} (0.1742) \\ s_0 (0.2222) \\ s_1 (0.2510) \\ s_2 (0.1776) \end{array} \right\}$	$\left\{ \begin{array}{l} s_{-2} (0.2018) \\ s_{-1} (0.1897) \\ s_0 (0.1866) \\ s_1 (0.2187) \\ s_2 (0.2031) \end{array} \right\}$	$\left\{ \begin{array}{l} s_{-2} (0.2220) \\ s_{-1} (0.2106) \\ s_0 (0.1722) \\ s_1 (0.1797) \\ s_2 (0.2155) \end{array} \right\}$	$\left\{ \begin{array}{l} s_{-2} (0.1960) \\ s_{-1} (0.2255) \\ s_0 (0.2049) \\ s_1 (0.1996) \\ s_2 (0.1740) \end{array} \right\}$
A ₃	$\left\{ \begin{array}{l} s_{-2} (0.1761) \\ s_{-1} (0.2157) \\ s_0 (0.1748) \\ s_1 (0.1893) \\ s_2 (0.2441) \end{array} \right\}$	$\left\{ \begin{array}{l} s_{-2} (0.1896) \\ s_{-1} (0.2137) \\ s_0 (0.2038) \\ s_1 (0.1915) \\ s_2 (0.2013) \end{array} \right\}$	$\left\{ \begin{array}{l} s_{-2} (0.2090) \\ s_{-1} (0.2166) \\ s_0 (0.2194) \\ s_1 (0.1801) \\ s_2 (0.1750) \end{array} \right\}$	$\left\{ \begin{array}{l} s_{-2} (0.2195) \\ s_{-1} (0.2135) \\ s_0 (0.2020) \\ s_1 (0.1925) \\ s_2 (0.1725) \end{array} \right\}$
A ₄	$\left\{ \begin{array}{l} s_{-2} (0.1845) \\ s_{-1} (0.1843) \\ s_0 (0.2038) \\ s_1 (0.2029) \\ s_2 (0.2245) \end{array} \right\}$	$\left\{ \begin{array}{l} s_{-2} (0.2031) \\ s_{-1} (0.1926) \\ s_0 (0.1866) \\ s_1 (0.1995) \\ s_2 (0.2183) \end{array} \right\}$	$\left\{ \begin{array}{l} s_{-2} (0.1948) \\ s_{-1} (0.2247) \\ s_0 (0.1996) \\ s_1 (0.1933) \\ s_2 (0.1876) \end{array} \right\}$	$\left\{ \begin{array}{l} s_{-2} (0.1694) \\ s_{-1} (0.2052) \\ s_0 (0.2206) \\ s_1 (0.2112) \\ s_2 (0.1936) \end{array} \right\}$
A ₅	$\left\{ \begin{array}{l} s_{-2} (0.1837) \\ s_{-1} (0.1897) \\ s_0 (0.2200) \\ s_1 (0.2063) \\ s_2 (0.2004) \end{array} \right\}$	$\left\{ \begin{array}{l} s_{-2} (0.1815) \\ s_{-1} (0.1871) \\ s_0 (0.2152) \\ s_1 (0.2080) \\ s_2 (0.2081) \end{array} \right\}$	$\left\{ \begin{array}{l} s_{-2} (0.1971) \\ s_{-1} (0.1913) \\ s_0 (0.1863) \\ s_1 (0.2119) \\ s_2 (0.2134) \end{array} \right\}$	$\left\{ \begin{array}{l} s_{-2} (0.1853) \\ s_{-1} (0.2205) \\ s_0 (0.2143) \\ s_1 (0.1926) \\ s_2 (0.1873) \end{array} \right\}$

Table 6. The outer-layer decision matrix. Sources: Authors’ own research.

Attribute	C ₁			C ₂			C ₃			C ₄		
Parameter	a	b	c	a	b	c	a	b	c	a	b	c
Low	[0,0]	[0,0]	[2.874, 4.005]	[0,0]	[0,0]	[5.051, 6.461]	[0,0]	[0,0]	[3.948, 5.364]	[0,0]	[0,0]	[2.061, 2.873]
Medium	[1.005, 1.411]	[4.005, 4.411]	[6.005, 6.411]	[1.992, 3.117]	[5.992, 7.117]	[9.992, 11.117]	[1.596, 2.411]	[5.596, 6.411]	[8.596, 9.411]	[1.034, 1.467]	[3.034, 3.467]	[4.758, 4.866]
High	[4.218, 4.717]	[7,7]	[7,7]	[7.029, 8.129]	[12,12]	[12,12]	[5.21, 7.018]	[10, 10]	[10, 10]	[3.04, 3.539]	[5,5]	[5,5]

Table 7. The parameters of membership functions of inner linguistic terms with respect to four attributes. Sources: Authors’ own research.

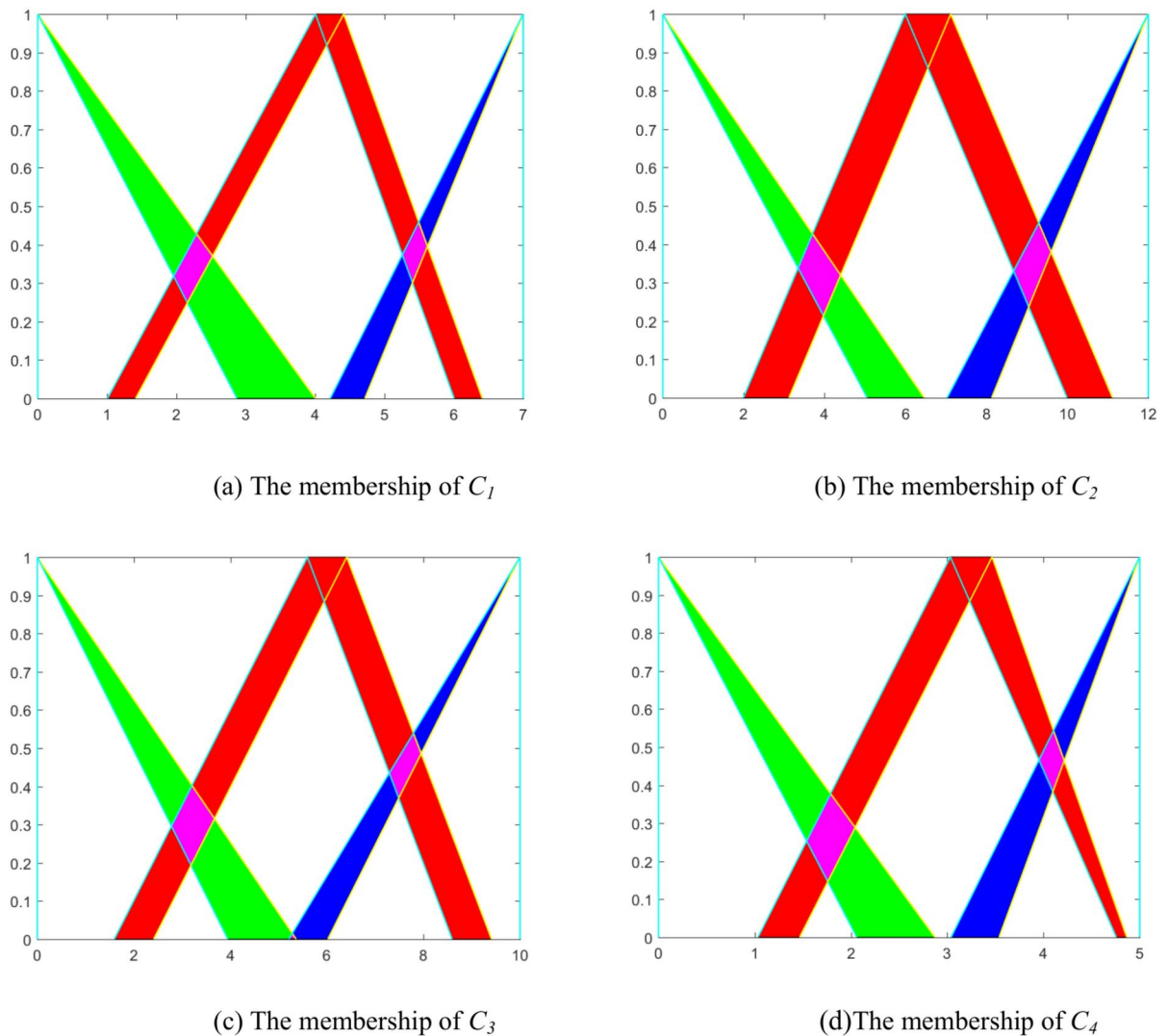


Fig. 3. The membership functions of four attributes. (a) The membership of C_1 , (b) The membership of C_2 , (c) The membership of C_3 , (d) The membership of C_4 .

The universe of discourse (UOD) of each attribute is the interval $[0,7]$, the interval $[0,12]$, the interval $[0,10]$, and the interval $[0,5]$, respectively. Table 7 lists the parameters of triangular membership functions of each attribute.

Figure 3 shows the membership functions of each attribute. Note that the green area, the red area, and the blue area represent the linguistic information “low”, “medium”, and “high”, respectively. Light blue means the left border, yellow means the right border for each inner linguistic term, and pink area is the overlap region. Considering the number of payment failures C_2 , we take Fig. 3(b) as an example, two pink areas refer to the range of payment failures around between 3 and 5, and between 9 and 10, respectively. Specifically, when the number of payment failures is 3, the membership degree of linguistic term “low” is between 0.3 and 0.4, while the membership degree of linguistic term “medium” is between around 0.2 and 0.3, which both cover 0.3. It means that this situation can be described by two membership functions, and the similar logic is suitable for other points in the pink areas.

Sources: Authors’ own research.

Next, we calculate the probabilities in the inner linguistic term sets by Eq. (9), and obtain the inner-layer decision matrix, listed in Table 8.

Step 4. We use the NPLTS-based TOPSIS (Technique for Order Preference by Similarity to an Ideal Solution) method with two types of information fusion, as detailed in Section “Deriving alternative rankings via NPLTS-based TOPSIS” and outlined in Eqs. (12) to (22), to obtain the final ranking of bank credit, and Table 9 lists the final assessment values and results.

According to the final values, the rankings of these alternatives are $A_3 \succ A_4 \succ A_1 \succ A_5 \succ A_2$ and $A_3 \succ A_4 \succ A_1 \succ A_2 \succ A_5$, respectively, and the results indicate that A_3 is the best one considering the bank credit.

	s_{-2}	s_{-1}	s_0	s_1	s_2
A_1	$\left\{ \begin{array}{l} n_{-1} (0.4892) \\ n_0 (0.3869) \\ n_1 (0.1238) \end{array} \right\}$	$\left\{ \begin{array}{l} n_{-1} (0.4786) \\ n_0 (0.4345) \\ n_1 (0.0870) \end{array} \right\}$	$\left\{ \begin{array}{l} n_{-1} (0.4678) \\ n_0 (0.4393) \\ n_1 (0.0929) \end{array} \right\}$	$\left\{ \begin{array}{l} n_{-1} (0.4333) \\ n_0 (0.4679) \\ n_1 (0.0988) \end{array} \right\}$	$\left\{ \begin{array}{l} n_{-1} (0.4274) \\ n_0 (0.4702) \\ n_1 (0.1024) \end{array} \right\}$
A_2	$\left\{ \begin{array}{l} n_{-1} (0.4953) \\ n_0 (0.3964) \\ n_1 (0.1083) \end{array} \right\}$	$\left\{ \begin{array}{l} n_{-1} (0.3929) \\ n_0 (0.5048) \\ n_1 (0.1024) \end{array} \right\}$	$\left\{ \begin{array}{l} n_{-1} (0.4357) \\ n_0 (0.4702) \\ n_1 (0.0941) \end{array} \right\}$	$\left\{ \begin{array}{l} n_{-1} (0.4857) \\ n_0 (0.4107) \\ n_1 (0.1036) \end{array} \right\}$	$\left\{ \begin{array}{l} n_{-1} (0.5143) \\ n_0 (0.3845) \\ n_1 (0.1012) \end{array} \right\}$
A_3	$\left\{ \begin{array}{l} n_{-1} (0.4083) \\ n_0 (0.4797) \\ n_1 (0.1119) \end{array} \right\}$	$\left\{ \begin{array}{l} n_{-1} (0.4690) \\ n_0 (0.4084) \\ n_1 (0.1226) \end{array} \right\}$	$\left\{ \begin{array}{l} n_{-1} (0.3964) \\ n_0 (0.4870) \\ n_1 (0.1167) \end{array} \right\}$	$\left\{ \begin{array}{l} n_{-1} (0.4000) \\ n_0 (0.4988) \\ n_1 (0.1012) \end{array} \right\}$	$\left\{ \begin{array}{l} n_{-1} (0.4762) \\ n_0 (0.4191) \\ n_1 (0.1048) \end{array} \right\}$
A_4	$\left\{ \begin{array}{l} n_{-1} (0.4488) \\ n_0 (0.4202) \\ n_1 (0.1310) \end{array} \right\}$	$\left\{ \begin{array}{l} n_{-1} (0.4929) \\ n_0 (0.4369) \\ n_1 (0.0702) \end{array} \right\}$	$\left\{ \begin{array}{l} n_{-1} (0.4440) \\ n_0 (0.4405) \\ n_1 (0.1155) \end{array} \right\}$	$\left\{ \begin{array}{l} n_{-1} (0.4036) \\ n_0 (0.4691) \\ n_1 (0.1274) \end{array} \right\}$	$\left\{ \begin{array}{l} n_{-1} (0.4250) \\ n_0 (0.4666) \\ n_1 (0.1084) \end{array} \right\}$
A_5	$\left\{ \begin{array}{l} n_{-1} (0.4429) \\ n_0 (0.4964) \\ n_1 (0.0607) \end{array} \right\}$	$\left\{ \begin{array}{l} n_{-1} (0.4703) \\ n_0 (0.3988) \\ n_1 (0.1310) \end{array} \right\}$	$\left\{ \begin{array}{l} n_{-1} (0.4988) \\ n_0 (0.4144) \\ n_1 (0.0869) \end{array} \right\}$	$\left\{ \begin{array}{l} n_{-1} (0.4821) \\ n_0 (0.3822) \\ n_1 (0.1357) \end{array} \right\}$	$\left\{ \begin{array}{l} n_{-1} (0.4785) \\ n_0 (0.3929) \\ n_1 (0.1286) \end{array} \right\}$

Table 8. The standardized inner-layer decision matrix. Sources: Authors’ own research.

	Expectation and variance (Wang et al., 2019)		Transformation function (Wang, Li, et al., 2022)	
	$\sigma (A_i)$	Rank	$s (A_i)$	Rank
A_1	0.1607	3	0.1791	3
A_2	0.1644	5	0.0879	4
A_3	0.1551	1	1	1
A_4	0.1571	2	0.7242	2
A_5	0.1629	4	0	5

Table 9. The rankings of alternatives by using the NPLTS-based TOPIS method with two fusion ways. Sources: Authors’ own research.

	Expectation and variance ⁹		Transformation function ¹²	
	$\sigma (A_i)$	Rank	$s (A_i)$	Rank
A_1	0.1611	4	0.1759	3
A_2	0.1644	5	0.0815	4
A_3	0.1550	1	1	1
A_4	0.1573	2	0.7076	2
A_5	0.1610	3	0	5

Table 10. The rankings of alternatives with the average weight vector. Sources: Authors’ own research.

Comparative analysis

To show the effectiveness and stability of the proposed decision-making framework, we conduct the comparative analysis from two aspects that may affect the final ranking, and try to find rules of the weight factor under different situations.

- The impact of attribute weights in the NPLTSs-based decision-making framework.
- (1) In the case study, we obtain the attribute weights by using the XGBoost algorithm. In the following, we compare the ranking when the attribute weight is the average weight vector, i.e., $w = (0.25, 0.25, 0.25, 0.25)$. Choosing the same decision-making method with information fusion by expectation and variance, and transformation function, respectively, we get the rankings $A_3 \succ A_4 \succ A_5 \succ A_1 \succ A_2$ and $A_3 \succ A_4 \succ A_1 \succ A_2 \succ A_5$, as shown in Table 10. At this time, the best alternative is A_3 as well as the result in the case study. However, the assessed values of alternatives are changed by two ways, especially when using the method of expectation and variance, the rankings are different from the original one. Therefore, the attribute weights would effect on the rankings, and it is a key point to get the weights objectively.
- (2) To explore whether the fluctuations of the attribute weights have effect on rankings, we consider the noise of the weight for each attribute. Let each attribute weight in $w = (0.333, 0.238, 0.143, 0.286)$ be with noise belonging to $N(0.25, 0.05)$, and the sum of all weights also equals to 1. After 1000 times, the final assessed values are calculated based on the NPLTSs-based TOPSIS method. Table 11 lists part of 1000 sets of weight.

Number	C_1	C_2	C_3	C_4
1	0.264	0.326	0.131	0.279
2	0.281	0.195	0.241	0.282
3	0.306	0.277	0.130	0.287
...
1000	0.382	0.180	0.198	0.240

Table 11. Part of the simulated weights with noise belonging to $N(0.25, 0.05)$. Sources: Authors’ own research.

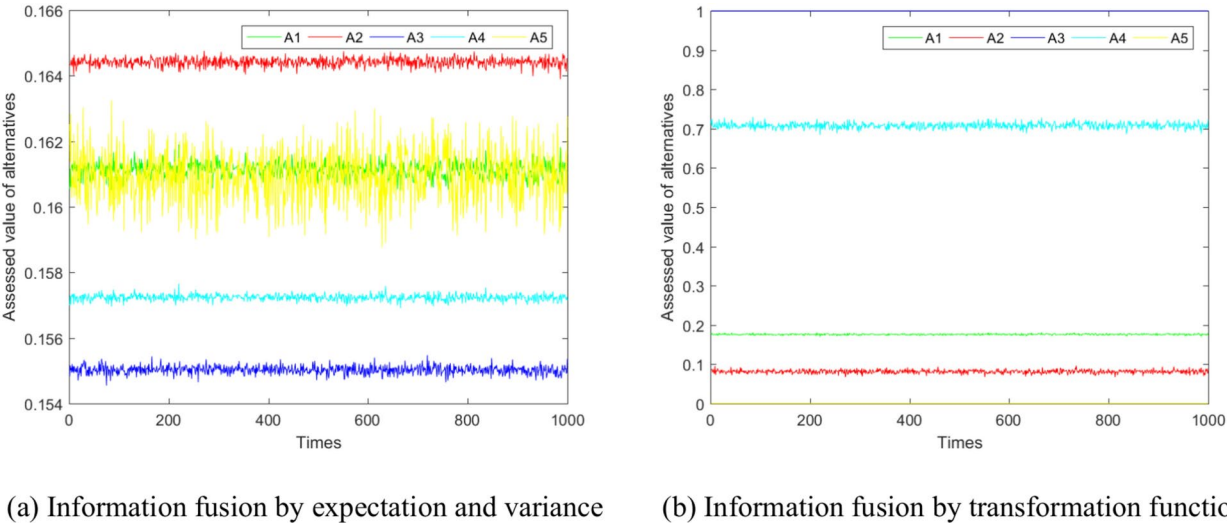


Fig. 4. The final assessed values considering the noise obey the normal distribution. (a) Information fusion by expectation and variance (b) Information fusion by transformation function.

Number	C_1	C_2	C_3	C_4
1	0	0.555	0.343	0.102
2	0	0.183	0.677	0.140
...
100	0	0.593	0.305	0.102
101	0.001	0.319	0.003	0.677
102	0.001	0.218	0.676	0.105
103	0.001	0.237	0.300	0.462
...
100,100	1	0	0	0

Table 12. Part of the simulated weights for each attribute. Sources: Authors’ own research.

Sources: Authors’ own research.

Figure 4 shows the assessed values of alternatives considering the noise obey the normal distribution by two ways. In Fig. 4a the results show that $\sigma(A_3)$ fluctuates around between 0.154 and 0.156; $\sigma(A_4)$ fluctuates around between 0.156 and 0.158; $\sigma(A_1)$ fluctuates around between 0.160 and 0.162; $\sigma(A_5)$ fluctuates roughly between 0.158 and 0.164; $\sigma(A_2)$ fluctuates around between 0.164 and 0.166; and thus the ranking relationship is $A_3 > A_4 > A_1 > A_5 > A_2$ or $A_3 > A_4 > A_5 > A_1 > A_2$. In Fig. 4b, the assessed values of alternative are more stable that fluctuate slightly, and the ranking is always $A_3 > A_4 > A_1 > A_5 > A_2$. The results by two ways indicate that different fusions may cause a little variation of the final values. As a result, when the attribute weights change a little, the ranking in the case is relatively stable.

(3) To explore the relationship between the importance of each attribute and the ranking, we consider the impact of individual attribute weight. Let each attribute weight $w_j (j = 1, 2, 3, 4)$ randomly generate 1000 sets with step size 0.001, and belongs to $[0,1]$ obeying to a standard normal distribution, and arranging in ascending order. At the same time, other three weights satisfy the standard normal distribution that the sum of all weights

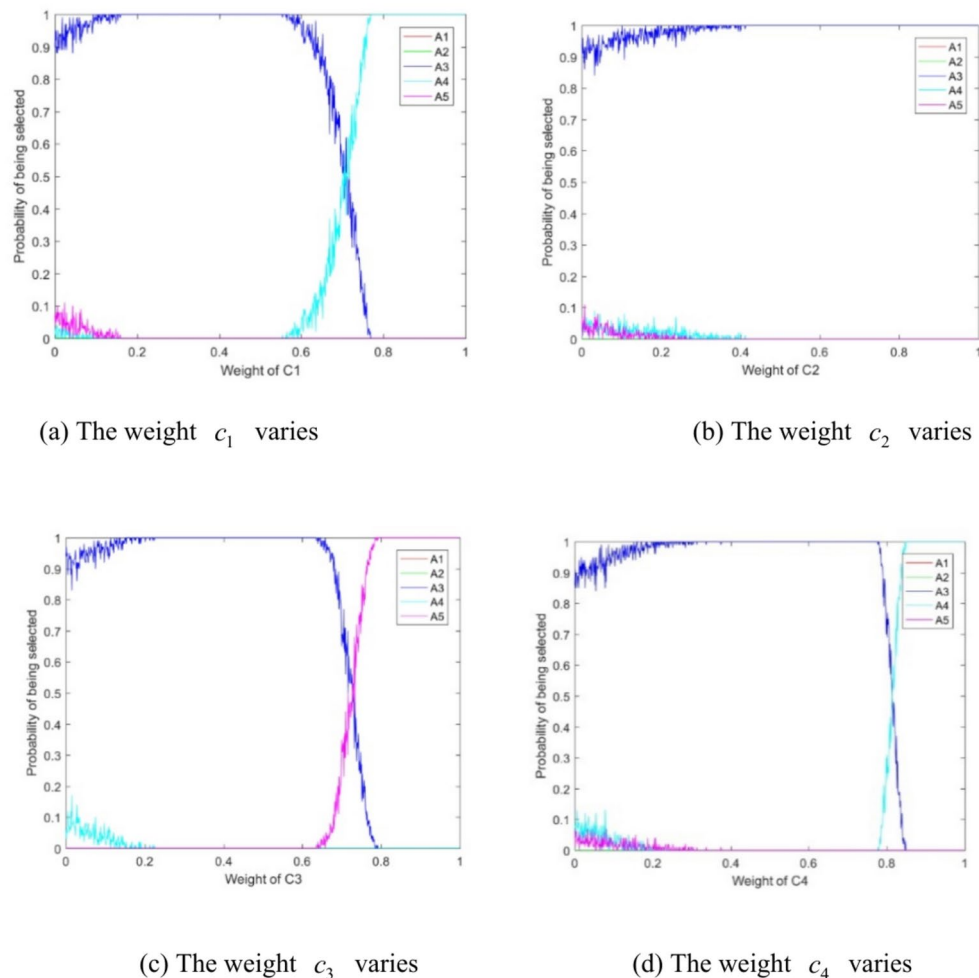


Fig. 5. The probability of being best of each alternative when the importance of each attribute varies, (a) The weight c_1 varies (b) The weight c_2 varies, (c) The weight c_3 varies (d) The weight c_4 varies.

equals to 1. Through the same decision-making method, the rankings in the case study are conducted in terms of different importance degrees of each attribute. Table 12 lists part of the 100,100 sets of weights under the above situation.

Figure 5 shows the probability of being the best of each alternative when the importance of each attribute varies. In Fig. 5a, when the weight c_1 increases from 0 to 1, the impacts of ranking on A_3 , A_5 and A_4 are bigger. To be specific, A_3 , A_5 and A_4 are the best alternative with a certain probability at the beginning, especially for A_3 . When the weight c_1 increases to 0.2, A_3 is the best alternative until the weight c_1 increases to around 0.6. At this time, A_4 is more likely to be the best one with the weight c_1 continues to increase. The phenomenon indicates that the alternative A_4 is very sensitive to the weight c_1 , i.e., the number of replacing the device.

In Fig. 5b, when the weight c_2 increases from 0 to 1, there is little effect of ranking on A_3 , A_5 and A_4 as well. As we can see, the beginning trend is similar to the weight c_1 varies that A_3 , A_5 and A_4 are possible to be the best alternative, and A_3 is with the highest probability. When the weight c_2 increases to around 0.25, A_5 loses competitiveness, and A_4 is the same situation when the weight c_2 increases to around 0.4. When the weight c_2 is larger than 0.4, A_3 is always the best one no matter what other attribute weights vary. The result shows that the weight c_2 , i.e., the number of payment failures, is a critical factor for the alternative A_3 that could enhance the competitiveness.

In Fig. 5c, when the weight c_3 increases from 0 to 1, there are obvious impacts of the best alternative on A_3 , A_5 and A_4 , especially for A_3 and A_5 . At the beginning, A_3 and A_4 are likely to be the best alternative, and A_3 is also the significant one. When the weight c_3 increases to around 0.2, A_3 holds the dominant space and A_4 withdraws from being the best alternative. When the weight c_3 increases to around 0.6, A_5 starts to be the best one, and the probability is larger when the weight c_3 continues to increase. When the weight c_3 is more than 0.8, A_5 is always the best alternative. This phenomenon indicates that the alternative A_5 is very sensitive to the weight c_3 , i.e., the number of IP addresses changed.

In Fig. 5d, when the weight c_4 increases from 0 to 1, the impacts of ranking on A_3 , A_5 and A_4 are more obvious than other alternatives. Specifically, A_3 , A_5 and A_4 have possible to be the first rank at the beginning, especially for A_3 . When the weight c_4 increases to around 0.2, A_3 is the best alternative until it increases to

around 0.78. At this time, A_4 is more likely to be the best one, and the probability becomes larger with the weight c_4 continues to increase. The phenomenon indicates that the weight c_4 i.e., the number of IP country changes, has a significant influence on the alternative A_4 , which is similar to the rules of the weight c_1 . The most difference is that A_4 is the best alternative when the weight c_4 reaches a larger value than the weight c_1 .

When each attribute weight increases gradually, we count the ranking times of each alternative to reveal the rule relationship between attribute weights and the ranking. Figure 6 presents the statistical ranking results of alternatives when the importance of each attribute varies.

Sources: Authors' own research. When the weight c_1 increases in Fig. 6a, A_1 always ranks the third or the fourth place, and the frequency with the third ranking is highest. A_2 almost ranks the fourth or the fifth place, among which the frequency with the fifth ranking is close to the frequency with the fourth ranking. A_3 almost ranks the first or the second place that the first rank is higher one. A_4 has three kinds of rankings with the larger proportion, where a second-place ranking is the highest, followed by the first-place, and the second-place. A_5 appears in all situations, and the highest one is with a fifth-place ranking, followed by the fourth-place, the third-place, and the second-place.

When the weight c_2 increases in Fig. 6b, A_1 basically ranks the third or the fourth place, and the highest frequency belongs to a fourth-place ranking, which is similar to the rule of the weight c_1 . A_2 almost ranks the fourth or the fifth place, which is far higher than the frequency of the fifth-place. A_3 almost ranks the first or the second place that being the best alternative has an absolute advantage. A_4 has all kinds of rankings, where the highest frequency is a second-place ranking. The ranking of A_5 appears in all possible situations, but the frequency with the fourth ranking is significantly higher than other four rankings. Therefore, the simulations show that all rankings tend to be stable that is $A_3 > A_4 > A_1 > A_5 > A_2$.

When the weight c_3 increases in Fig. 6c, A_1 almost ranks the third or the fourth place, and the highest frequency is with a third-place ranking, higher than that of the other ranking. A_2 ranks the fourth or the fifth

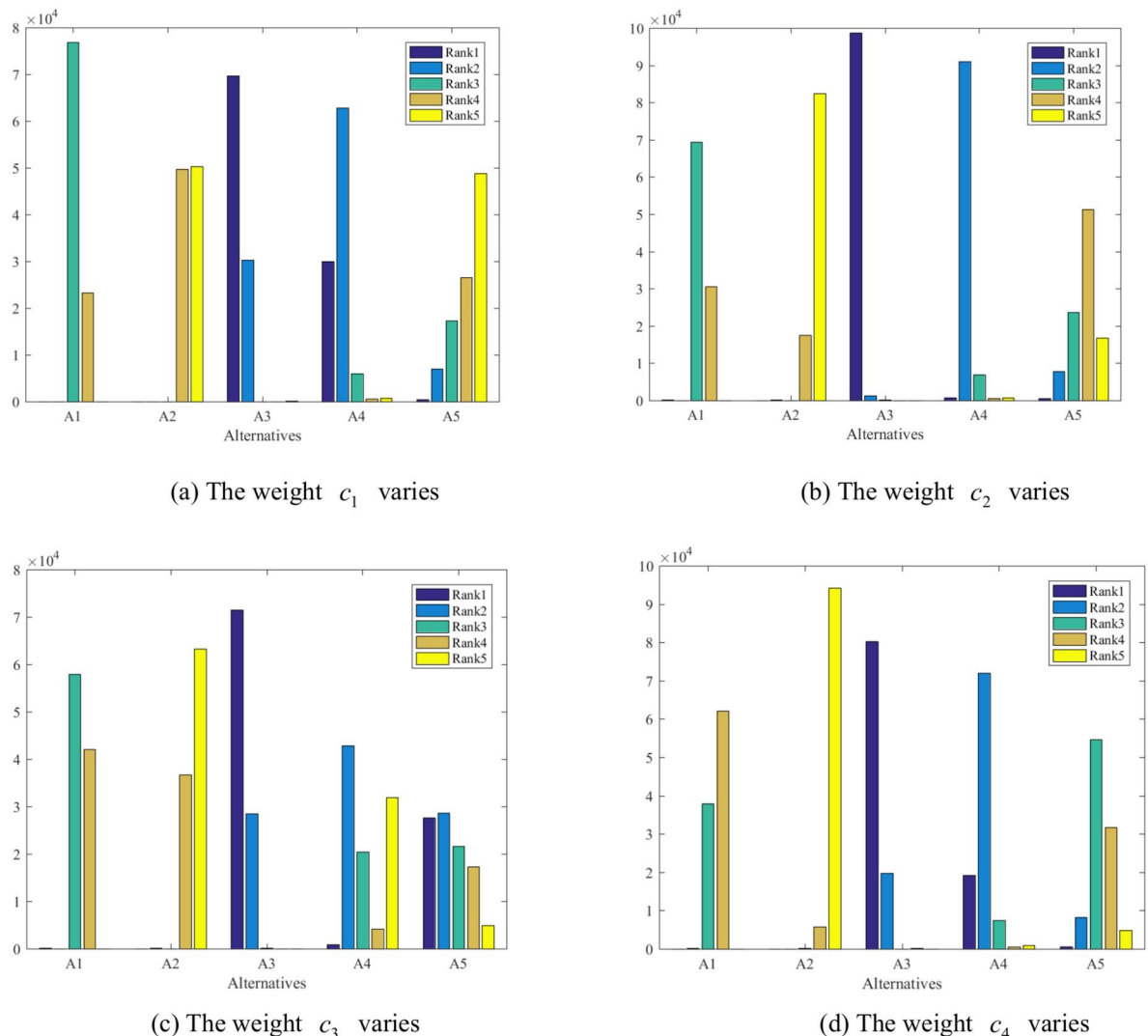


Fig. 6. The statistical results of rankings in terms of the different importance of each attribute. (a) The weight c_1 varies (b) The weight c_2 varies, (c) The weight c_3 varies (d) The weight c_4 varies.

TOPSIS with expectation and variance			TOPSIS with transformation function		TODIM with transformation function		PROMETHEE with transformation function	
	$\sigma(A_i)$	Rank	$s(A_i)$	Rank	$\varepsilon(A_i)$	Rank	$Phi(A_i)$	Rank
A_1	0.1607	3	0.1791	3	0.3141	3	-0.0128	3
A_2	0.1644	5	0.0879	4	0.1446	4	-0.0187	4
A_3	0.1551	1	1	1	1	1	0.0351	1
A_4	0.1571	2	0.7242	2	0.7666	2	0.0195	2
A_5	0.1629	4	0	5	0	5	-0.0232	5

Table 13. The rankings by different decision-making methods. Sources: Authors’ own research.

place with higher frequency. A_3 has the first or the second ranking that the frequency with the first-place ranking is highest. A_4 holds five rankings from the first to the last ranking, where the frequency with the second-place ranking is highest. A_5 also has all possible rankings, and the highest frequency is with the second-place ranking, which has a little gap compared to the first-place.

When the weight c_4 increases in Fig. 6d, A_1 basically ranks the third or the fourth place, and the highest frequency is with a fourth-place ranking. A_2 almost ranks the fourth or the fifth place, among which the frequency with the fifth ranking is far higher than the fourth ranking. A_3 almost ranks the first or the second place that being the first seems to be an obvious trend. The rankings of A_4 and A_5 appear in all possible situations, where the highest frequencies are the second-place ranking and the third-place ranking, respectively. The results indicate that all rankings tend to be stable that $A_3 > A_4 > A_5 > A_1 > A_2$.

In general, the rankings of A_1 , A_2 and A_3 are relatively stable. No matter how attribute weights vary, A_1 is either the third or the fourth one, A_2 is either the fourth or the last one, and A_3 is the first or the second one. In terms of each attribute, the weight c_1 , i.e., the number of replacing the device, mainly affects the fourth and the fifth order. The weight c_2 , i.e., the number of payment failures, has little impact on the rankings, and the stable ranking is $A_3 > A_4 > A_1 > A_5 > A_2$. The weight c_3 , i.e., the number of IP addresses changed, mainly affects the second and the fourth order, and the weight c_4 , i.e., the number of IP country changed, does not impact the stable ranking that is $A_3 > A_4 > A_5 > A_1 > A_2$.

The impact of decision-making methods in the NPLTSs-based decision-making framework. Under the proposed decision-making framework, we choose four kinds of decision-making methods to deal with the case study.

Table 13 lists the final values and rankings by using these methods. TOPSIS, TODIM (Interactive Multi-criterion Decision-making), and PROMETHEE (Preference Ranking Organization Method for Enrichment Evaluation) are classical decision-making methods to deal with selecting the most satisfied alternative with the specific characteristics. The results show that the ranking is $A_3 \succ A_4 \succ A_1 \succ A_2 \succ A_5$ by three methods with transformation function, while the ranking is $A_3 \succ A_4 \succ A_1 \succ A_5 \succ A_2$ by using TOPSIS with expectation and variance functions. In sum, the best alternative is A_3 by using these four methods, and the result indicates that the decision-making framework is flexible and reliable.

Discussions

Considering the uncertainty and the multi-dimensional structure in the complex and data-driven decision – making problems, this study proposes a cognitive and scientific decision-making framework combining the XGBoost algorithm with the constrained parametric approach (CPA) under the nested probabilistic linguistic environment. In the following, we discuss the framework that provides the objective and explicable information and implications for policy makers and managers.

The proposed decision-making framework provides a way to deal with complex problem based on historical data and interpretable model. On the one hand, linguistic terms are natural and efficient as the preference modeling tools in the elicitation of uncertain assessments²⁸, since language is the closest form to express people’s cognitive information managed flexible linguistic expression⁴³. On the other hand, data-driven techniques make up for the disadvantage of linguistic variables not being operational without detailed quantification, and have shown promising results in the analysis and reliable decision-making processes in various fields^{44,45}. Recent studies have focused on establishing suitable linguistic models and using machine learning based on data to solve real problems applied in various fields, such as location planning of electric vehicle charging stations⁴⁶, early detection of Alzheimer’s disease⁴⁷, and risk assessment of Chinese enterprises’ overseas mergers and acquisitions⁴⁸. Compared with other advanced decision-making framework, the proposed framework has the following characteristics:

- (1) Under the nested probabilistic linguistic decision-making environment, the model is not limited to the traditional structure of a decision-making problem. Depending on the inner-layer and outer-layer with two types, ordinal or nominal linguistic terms, decision-making issues related to the preference degrees or features can be solved in the proposed framework, as well as discrimination, and optimization problems.
- (2) In the proposed framework, depending on the data, a machine learning method, i.e., the XGBoost algorithm, is applied to determine the attribute weights, which provides a scientific way to indicate the importance of attributes for decision makers. Moreover, the probabilities in the linguistic model are calculated by the CPA, combined with fuzzy theory and data-based method. In this way, the established framework has features of objectiveness and interpretability.

(3) There is no specific decision-making method under the proposed framework. Each decision-making method has its own characteristics, thus, the framework is flexible and rational to adapt different problems, such as behavior decisions, cognitive decisions, and uncertain decisions. In addition, a data-based process is an essential part in the framework, which guarantees the objectivity principle. In this sense, the proposed framework not only contributes to the linguistic model with machine learning under uncertainty and interpretability, but also enriches the multi-dimensional cognitive decision-making.

(4) The proposed framework presents one of the decision-making problems, the ranking of bank credit. With the emergence of artificial intelligence, fraud detection is one of the utmost concerns for investors and financial authorities⁴⁹, and related technologies have been deployed for the management of corporate financial risk⁵⁰. In the case study, we explore the impact of attribute weights considering the noises and rules, and different decision-making methods on each alternative. After comparative analysis of attribute weights, the importance of each attribute for alternatives is portrayed and the effect is presented clearly. The phenomenon provides the basis that managers or decision makers could set a plan to adjust the impact of each attribute on alternatives according to the rules of attribute weights. In addition, the results indicate that the proposed framework has good stability and effectiveness.

Conclusions

At a genuine numeral network age with the driving of high and new technology, decision-making framework not only focuses on the most satisfied selection objectively driven by big data, but also the interpretability depending on the cognitive information and dealing with the uncertainty. Under such a complex environment, this study proposes a cognitive decision-making framework combining the XGBoost algorithm with the constrained parametric approach (CPA) using the nested probabilistic linguistic term sets (NPLTSs). We firstly provide a flexible linguistic model for decision makers to describe multi-dimensional and uncertain information that may belong to preference degree or nominal feature. And then, we obtain the complete nested probabilistic linguistic information combined with the CPA approach and the historical data. After that, the XGBoost algorithm is applied to determine the attribute weights, presenting the importance of each attribute scientifically. Finally, a case study concerning the ranking of bank credit is given to demonstrate the flexible applicability of the proposed decision-making framework. Besides, the comparative analysis explores the impact of each attribute weight on ranking, and reveals the rules between attributes and alternatives. The results of the case study show the good smoothness and the stability that the best one is constant by using two different fusion ways. However, it is important to acknowledge that the considered attributes are limited by data accessibility constraints, particularly the four selected attributes, which may not fully capture all relevant factors. In general, this work provides an objective and cognitive decision-making framework to deal with multi-dimensional and uncertain information, and presents the reliable process from the perspectives of theory and application, respectively. Policy makers and managers can choose the suitable decision-making method to solve a specific problem using the proposed decision-making framework. In the future, we will continue to focus on cognitive decision-making with NPLTSs in prescriptive analytics, artificial intelligence and big data analysis.

Data availability

The datasets generated and analyzed during the current study are based on publicly available data from the book *Python Big Data Analysis and Machine Learning Business Case Practice*.

Received: 7 April 2024; Accepted: 16 January 2025

Published online: 05 February 2025

References

1. Akter, S., Michael, K., Uddin, M. R., McCarthy, G. & Rahman, M. Transforming business using digital innovations: the application of AI, blockchain, cloud and data analytics. *Ann. Oper. Res.* **308**, 7–39 (2022).
2. Chukwuma, U., Gebremedhin, K. G. & Uyeh, D. D. Imagining AI-driven decision making for managing farming in developing and emerging economies. *Comput. Electron. Agric.* **221**, 108946 (2024).
3. Luckyardi, S., Rahayu, A., Adiwibowo, L. & Hurriyati, R. Information technology in evolutionary strategic management: decision support system in smart university. *J. Eng. Sci. Technol.* **18**, 1561–1569 (2023).
4. Dias, L. C., Passeira, C., Malca, J. & Freire, F. Integrating life-cycle assessment and multi-criteria decision analysis to compare alternative biodiesel chains. *Ann. Oper. Res.* **312**, 1359–1374 (2022).
5. Chae, B., Sheu, C. & Park, E. O. The value of data, machine learning, and deep learning in restaurant demand forecasting: Insights and lessons learned from a large restaurant chain. *Decis. Support Syst.* **184**, 114291 (2024).
6. Ulhe, P. P. et al. Flexibility management and decision making in cyber-physical systems utilizing digital lean principles with Brain-inspired computing pattern recognition in Industry 4.0. *Int. J. Comput. Integr. Manuf.* **37**, 708–725 (2024).
7. Pang, Q., Wang, H. & Xu, Z. Probabilistic linguistic linguistic term sets in multi-attribute group decision making. *Inf. Sci.* **369**, 128–143 (2016).
8. Wang, X., Xu, Z., Wen, Q. & Li, H. A multidimensional decision with nested probabilistic linguistic term sets and its application in corporate investment. *Econ. Res.-Ekon. Istraz.* **34**, 3382–3400 (2021).
9. Wang, X., Xu, Z. & Gou, X. Nested probabilistic-numerical linguistic term sets in two-stage multi-attribute group decision making. *Appl. Intell.* **49**, 2582–2602 (2019).
10. Wang, X., Xu, Z. & Li, H. Evaluation of government investment using nested probabilistic linguistic preference relations based on graph theory. *Technol. Econ. Dev. Econ.* (2022).
11. Wang, X., Xu, Z., Gou, X. & Trajkovic, L. Tracking a maneuvering target by multiple sensors using extended kalman filter with nested probabilistic-numerical linguistic information. *IEEE Trans. Fuzzy Syst.* **28**, 346–360 (2020).
12. Wang, X., Li, Y., Xu, Z. & Luo, Y. Nested information representation of multi-dimensional decision: An improved PROMETHEE method based on NPLTSs. *Inf. Sci.* **607**, 1224–1244 (2022).
13. Gao, Y. et al. Mechanical equipment health management method based on improved intuitionistic fuzzy entropy and case reasoning technology. *Eng. Appl. Artif. Intell.* **116**, 105372 (2022).

14. Zhu, H. et al. A structure optimization method for extended belief-rule-based classification system. *Knowl Based Syst.* **203**, 106096 (2020).
15. Ren, P., Zhu, B., Ren, L. & Ding, N. Online choice decision support for consumers: Data-driven analytic hierarchy process based on reviews and feedback. *J. Oper. Res. Soc.* **74**, 2227–2240 (2023).
16. Fan, S., Liang, H., Dong, Y. & Pedrycz, W. A personalized individual semantics-based multi-attribute group decision making approach with flexible linguistic expression. *EXPERT Syst. Appl.* **192**, 116392 (2022).
17. Zhang, S., Wu, Z., Ma, Z., Liu, X. & Wu, J. Wasserstein distance-based probabilistic linguistic TODIM method with application to the evaluation of sustainable rural tourism potential. *Econ. Res.-Ekon. Istraz.* **35**, 409–437 (2022).
18. Ye, J. et al. A Novel Diversified Attribute Group Decision-Making Method Over Multisource Heterogeneous Fuzzy Decision Systems With Its Application to Gout Diagnosis. *IEEE Trans. Fuzzy Syst.* **31**, 1780–1794 (2023).
19. Chen, B., Tong, R., Gao, X. & Chen, Y. A novel dynamic decision-making method: Addressing the complexity of attribute weight and time weight. *J. Comput. Sci.* **77**, 102228 (2024).
20. Fu, S. & Xiao, Y. Study on venture capital multi-attribute group decision-making based on improved Hamming-Hausdorff distance and weighted bidirectional projection. *Biomed. Signal Process. Control* **91**, 105985 (2024).
21. Akyurt, I. Z., Pamucar, D., Deveci, M., Kalan, O. & Kuvvetli, Y. A flight base selection for flight academy using a rough macbeth and rafsi based decision-making analysis. *IEEE Trans. Eng. Manag.* **71**, 258–273 (2024).
22. Li, C., Yi, J., Wang, H., Zhang, G. & Li, J. Interval data driven construction of shadowed sets with application to linguistic word modelling. *Inf. Sci.* **507**, 503–521 (2020).
23. Deveci, M., Pekaslan, D. & Canitez, F. The assessment of smart city projects using zSlice type-2 fuzzy sets based interval agreement method. *Sustain. Cities Soc.* **53**, 101889 (2020).
24. D'Alterio, P., Garibaldi, J. & Wagner, C. A Constrained Parametric Approach for Modeling Uncertain Data. *IEEE Trans. FUZZY Syst.* **30**, 3967–3978 (2022).
25. Lin, M., Huang, C., Xu, Z. & Chen, R. Evaluating IoT Platforms Using Integrated Probabilistic Linguistic MCDM Method. *IEEE Internet Things J.* **7**, 11195–11208 (2020).
26. Finger, G. S. W. & Lima-Junior, F. R. A hesitant fuzzy linguistic QFD approach for formulating sustainable supplier development programs. *Int. J. Prod. Econ.* **247**, 108428 (2022).
27. Gou, X., Xu, Z. & Zhou, W. Interval consistency repairing method for double hierarchy hesitant fuzzy linguistic preference relation and application in the diagnosis of lung cancer. *Econ. Res.-Ekon. Istraz.* **34**, 1–20 (2021).
28. Ma, X., Qin, J., Martinez, L. & Pedrycz, W. A linguistic information granulation model based on best-worst method in decision making problems. *Inf. FUSION* **89**, 210–227 (2023).
29. Liu, F., Liu, S. & Jiang, G. Consumers' decision-making process in redeeming and sharing behaviors toward app-based mobile coupons in social commerce. *Int. J. Inf. Manag.* **67**, 102550 (2022).
30. Yuan, J., Lu, W., Ding, H., Liu, J. & Mahmoudi, A. A novel Z-number based real option (ZRO) model under uncertainty: application in public-private-partnership refinancing value evaluation. *Expert Syst. Appl.* **213**, 118808 (2023).
31. Su, W., Luo, D., Zhang, C. & Zeng, S. Evaluation of online learning platforms based on probabilistic linguistic term sets with self-confidence multiple attribute group decision making method. *EXPERT Syst. Appl.* **208**, 118153 (2022).
32. Lin, M., Chen, Z., Xu, Z., Gou, X. & Herrera, F. Score function based on concentration degree for probabilistic linguistic term sets: An application to TOPSIS and VIKOR. *Inf. Sci.* **551**, 270–290 (2021).
33. Chen, X., Wang, H. & Li, X. Doctor recommendation under probabilistic linguistic environment considering patient's risk preference. *Ann. Oper. Res.* <https://doi.org/10.1007/s10479-022-04843-9> (2022).
34. Zhang, C., Xu, Z., Gou, X. & Chen, S. An online reviews-driven method for the prioritization of improvements in hotel services. *Tour. Manag.* **87**, 104382 (2021).
35. Shi, X., Wong, Y. D., Li, M. Z.-F., Palanisamy, C. & Chai, C. A feature learning approach based on XGBoost for driving assessment and risk prediction. *Accid. Anal. Prev.* **129**, 170–179 (2019).
36. Zhou, J. et al. Estimation of the TBM advance rate under hard rock conditions using XGBoost and Bayesian optimization. *Undergr. SPACE* **6**, 506–515 (2021).
37. Zavadskas, E. K., Turskis, Z., Volvaciovas, R. & Kildiene, S. Multi-criteria assessment model of technologies. *Stud. Inform. Control* **22**, 249–258 (2013).
38. Wellalage, N. H., Thrikawala, S. & Ghardallou, W. Political connections, family ownership and access to bank credit. *FINANCE Res. Lett.* <https://doi.org/10.1016/j.frl.2022.103347> (2022).
39. Krivko, M. A hybrid model for plastic card fraud detection systems. *Expert Syst. Appl.* **37**, 6070–6076 (2010).
40. Alonso-Robisco, A. & Carbó, J. M. Can machine learning models save capital for banks? Evidence from a Spanish credit portfolio. *Int. Rev. Financ. Anal.* **84**, 102372 (2022).
41. Dan, O., Parikh, V. & Davison, B. D. Improving IP Geolocation using Query Logs. in *Proceedings of the Ninth ACM International Conference on Web Search and Data Mining* 347–356 (Association for Computing Machinery, New York, NY, USA, 2016). <https://doi.org/10.1145/2835776.2835820>.
42. Wang, Y. & Qian, Y. *Python Big Data Analysis and Machine Learning Business Case Practice* (China Machine Press, 2020).
43. Zhang, H. et al. Managing flexible linguistic expression and ordinal classification-based consensus in large-scale multi-attribute group decision making. *Ann. Oper. Res.* <https://doi.org/10.1007/s10479-022-04687-3> (2022).
44. Ahmed, F. & Kim, K.-Y. Recursive approach to combine expert knowledge and data-driven RSW weldability certification decision making process. *Robot. Comput.-Integr. Manuf.* **79**, 102428 (2023).
45. Yang, F. X., Li, Y., Li, X. & Yuan, J. The beauty premium of tour guides in the customer decision-making process: An AI-based big data analysis. *Tour. Manag.* **93**, 104575 (2022).
46. Wei, G. et al. Algorithms for probabilistic uncertain linguistic multiple attribute group decision making based on the GRA and CRITIC method: application to location planning of electric vehicle charging stations. *Econ. Res.-Ekon. Istraz.* **33**, 828–846 (2020).
47. Adhikari, S. et al. Exploiting linguistic information from Nepali transcripts for early detection of Alzheimer's disease using natural language processing and machine learning techniques. *Int. J. Hum.-Comput. Stud.* **160**, 102761 (2022).
48. Lan, J. et al. CODAS methods for multiple attribute group decision making with interval-valued bipolar uncertain linguistic information and their application to risk assessment of Chinese enterprises' overseas mergers and acquisitions. *Econ. Res.-Ekon. Istraz.* **34**, 3166–3182 (2021).
49. Bernard, P. et al. A financial fraud detection indicator for investors: an IDEa. *Ann. Oper. Res.* **313**, 809–832 (2022).
50. Lin, K. & Gao, Y. Model interpretability of financial fraud detection by group SHAP. *Expert Syst. Appl.* **210**, 118354 (2022).

Author contributions

X.X W: Data curation; methodology; formal analysis; writing; B.B Z: Methodology; writing-review and editing; data curation; Z.S X: Formal analysis; writing-review and editing; M L: study design; writing-review and editing; investigation; M S: Interpretation; writing-review and editing.

Funding

This work was supported by the National Natural Science Foundation of China under Grant Nos. 72101168, 72271173; and China Postdoctoral Science Foundation under Grant No. 2021M692259.

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to M.S.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025