

Research Article

Fuzzy Boost Classifier of Decision Experts for Multicriteria Group Decision-Making

Xiao-yi Wang ^{1,2} Yi Yang,^{1,2} Yu-ting Bai ^{1,2} Jia-bin Yu,^{1,2} Zhi-yao Zhao ^{1,2}
and Xue-bo Jin ^{1,2}

¹School of Artificial Intelligence, Beijing Technology and Business University, Beijing 100048, China

²China Light Industry Key Laboratory of Industrial Internet and Big Data, Beijing Technology and Business University, Beijing 100048, China

Correspondence should be addressed to Yu-ting Bai; byting@foxmail.com

Received 6 May 2020; Revised 17 June 2020; Accepted 25 July 2020; Published 17 August 2020

Academic Editor: Lucia Valentina Gambuzza

Copyright © 2020 Xiao-yi Wang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The expert is a vital role in multicriteria decision-making, which provides source decision opinions. In the existing group decision-making activities, the selection of experts is usually conducted artificially, which relies on personal subjective experience. It has been the urgent demand for an automatic selection of experts, which can help to determine their weights for the follow-up decision calculation. In this paper, an expert classification method is proposed to solve the problem. First, the CatBoost classification algorithm is improved by integrating the 2-tuple linguistic, which can effectively extract the features of samples. Second, the framework of the expert classification is designed. The flow combines the expert resume collection, expert classification, and database update. Third, a decision-making case is analyzed for the expert selection issue. The experiment and result indicate that the proposed classifier performs better than the classic methods. The proposed classification method of the decision experts can support the automatic and intelligent operation of the decision-making activities.

1. Introduction

Decision-making is a vital activity in various management works. It is a complex process of thinking and operation [1]. Rational decision-making has become an important part of the modern management theory, in which the decision-making always involves multicriteria group decision-making [2]. The decision-making methods originate from the persons' thinking process. Therefore, the essential basis is the experts' information and knowledge. The expert here means the administrator, decision-maker, and scholar in the related field. The professional and adaption level of experts affects the decision-making greatly. How to select the appropriate decision experts has been a novel issue.

Moreover, the computer and information technology bring the automatic decision support system (DSS) [3, 4]. In the automatic decision-making programs like the DSS, the expert information is collected and stored, such as the professional title, educational background, professional

background, and research field. It provides the opportunity to select the decision experts automatically and rationally.

Specifically, the selection of experts is usually done manually, which relies on the subjective judgment of administrators. The selection of experts is not objective enough so that the results are often not persuasive. The artificial approach is unreliable and inefficient in the actual operation. Moreover, in the Internet information era, it is an available demand to automatically analyze the expert information and select the appropriate ones for the decision-making [5]. The experts should be analyzed with their research level and practical experience. Then the experts can be classified and selected according to their professional grades and the decision-making situation.

The core issue can be abstracted as the classification and evaluation of the experts. The typical classification algorithms of machine learning include naive Bayes, k -nearest neighbor (KNN), decision tree (DT), support vector machine (SVM), and logistic regression [6–9]. It develops with

the integrated learning algorithms such as random forest, AdaBoost, XgBoost, and CatBoost [10, 11]. Meanwhile, other machine learning methods are also studied for feature extraction [12–14]. The input of the classifiers may be a real number, linguistic text, and grade variable. For the fuzzy inputs, they are usually converted with the one-hot coding method [15]. The simple conversion may lose the detailed and hidden information of the categorical features [16–18], which makes it impossible to accurately and completely express the information carried by the original data.

Given the demands above, the experts should be analyzed automatically with the machine learning, in which the problems of information coding and feature recognition should be solved. This paper proposes an expert classification method from the perspective of objectivity and intelligent computing. The main work is briefly presented as follows.

- (i) The 2-tuple linguistic [19] is introduced into the integrated classifier CatBoost [20] to fully express the fuzzy features and classify the experts.
- (ii) The general framework is designed with the fuzzy classification model. It can be used to select decision experts and update the expert library.
- (iii) The ability of feature extraction and classification is tested with the one-hot encoding and traditional classifiers. Results show that CatBoost integrated with 2-tuple linguistic can effectively recognize the experts and help evaluate and select the decision expert in an automatic solution.

This paper is organized as follows. Section 2 introduces the works related to the expert classification and decision-making. Section 3 expounds on the proposed method of expert classification. Section 4 carries out a case of the expert classification for the decision-making in water environment pollution. The method and results are discussed in Section 5, and the paper is concluded finally in Section 6.

2. Related Work

2.1. Group Decision-Making and Decision Support System. The core thought of rational decision-making is evaluating the alternatives from multiple aspects and persons. The multicriteria group decision-making realizes the thought, and it will be introduced firstly in this section.

Decision-making is the process of judging a decision problem or finding a solution. In general, a satisfactory solution should be most acceptable by the entire administrators. Therefore, the group decision method is proposed to improve the traditional individual method. The group decision-making mainly uses the core theory of the multicriteria and multiattribute decision-making [21, 22], which involves the integration of the opinions of multiple experts and the weight calculation of alternative scheme attributes. In most situations, the decision opinions are given in the form of linguistic text by the group of decision experts. In order to aggregate the linguistic information, Xu [23] proposed a group decision-making method based on the

relation of language preference. Qian et al. [24] redefined some basic operations of the generalized hesitant fuzzy set and presented a group decision-making method based on the improved hesitant fuzzy set. Chen [25] extended the technique for order preference by similarity to an ideal solution (TOPSIS) to the fuzzy decision environment. Besides, the decision-making methods are extended in the neutrosophic set environment [26–28], which try to promote the information expression level. More types of group decision-making are explored, such as the methods based on granular computing [29–31]. They aim at the decision-making with multiple dimensions.

With information technology development, the different decision-making methods can be realized in the DSS. Since Gorry [32] first put forward the concept of decision-making with the software system, the DSS has been studied and developed. Up to now, DSS has been widely used, including clinical DSS in hospitals, environmental DSS in government departments, and DSS in urban water supply management [33]. The DSS mainly consists of three parts: data access and operation, model based on data, and case matching. With the development of computer and information technology, the DSS develops in various solutions [34, 35]. The DSS has provided an automatic platform to run the decision-making methods.

It can be seen that multicriteria group decision-making is built based on decision experts. The experts need to be selected to express decision opinions [36]. The existing studies on group decision-making generally focus on the integration of experts' decision opinions. However, the analysis of experts themselves is not sufficient, and the level differences among experts are not fully considered. Besides, the expert library is an important module of the DSS. Most of the systems collect and evaluate the experts in the manual solution. Meanwhile, the system pays more attention to the case library and other knowledge bases, instead of the expert library [37]. It can be concluded that the selection and evaluation of decision experts directly affect the reliability of decision results. Moreover, the modern and intelligent DSS has an urgent need for automatic analysis and classification of experts. Therefore, it is necessary to explore the automatic method of the expert classification.

2.2. Classification Methods in Machine Learning. There are many kinds of classification algorithms in machine learning. The classifiers are widely used in data mining, text recognition, and other fields.

As mentioned in Introduction, the classical algorithms include naive Bayes, KNN, DT, SVM, and logistic regression [6–9]. The integrated learning algorithms include random forest, AdaBoost, XgBoost, and CatBoost [10, 11, 20]. However, the classifiers themselves do not provide a mechanism for input features. The sample data are usually converted into numerical data by one-hot encoding before training models. The hidden information may be lost by the categorical feature conversation with one-hot encoding, which will impact the classification accuracy [16].

The CatBoost algorithm has the characteristics of high classification accuracy and fast training speed. It has been the latest classifier which is of strong learning ability. Moreover, CatBoost provides a mechanism of converting input features into numerical values. However, the conversation method of CatBoost is not suitable for expert sets with small samples, especially for the samples with a few input features. For the problem of expert classification, the integrated method represented by CatBoost can provide an effective learning technique, but it should be remolded for the input feature conversation.

For the feature extraction of input data, some fuzzy analysis methods can provide the basic technique. The classical theories include the fuzzy set and rough set [38, 39], in which the degree of information uncertainty is expressed with the probability and membership function. The fuzzy analysis methods are developed in the demand for language comprehension. The grade variables in the text can be transformed and calculated with the vague set [40] and 2-tuple linguistic [19, 41]. The related studies show that the 2-tuple linguistic is flexible and effective to express the rating grade information, of which the calculation load is lower than the complex natural language processing technique.

Significantly different from previous studies, we will propose an intelligent method to create the decision expert library. Our innovative contributions are highlighted as follows:

- (1) The automatic classification and evaluation of decision experts are mainly focused. An improved fuzzy CatBoost classifier is proposed, integrated with 2-tuple linguistic. The information of experts is expected to be extracted more soundly and deeply.
- (2) The creative flow of the expert classification and selection is designed based on the improved classifier, which can help the application of group decision-making in the software programming of the DSS.

3. Classification Method of Decision Expert

3.1. Problem Description. The existing research indicates that the group multicriteria decision method depends on the selection and aggregation of experts. It is the classification problem of decision experts. The reasonable experts should be selected considering as many characteristics of experts as possible, such as the research field, academic level, and practical experience. Also, the representation of expert information should be easy and available for computer computation. The information collection and analysis calculation should be realized with the software program. Considering the demand, the problem is abstracted and presented as follows.

The fundamental task for expert classification is the importing of the expert information. In the task, the experts are selected and classified to form different categories. First, the resumes of experts are collected, and the information will be processed uniformly. In order to meet the requirements of machine learning, the input data of experts should be

objective enough. This paper uses structured data to represent the attributes of experts. There are n experts, and the expert set is represented as E , $E = \{e_1, e_2, \dots, e_n\}$. Each expert is described with m attributes, and the attribute is represented as $x_k = \{x_k^1, x_k^2, \dots, x_k^m\}$, where k is the serial number of the experts. The attribute variable is in different forms according to the property information. One of the attribute variables is the numerical value, while others are represented with texts. The attribute variable in the text form should be converted for the unified input of the classifier. Moreover, for the elements in the expert set, they are corresponding to the category label. The category means the general evaluation of the expert level, in view of the academic level, the professional degree in the field, and the quantized achievements. The category label is represented by y_k , where $y_k = \{1, 2, 3, \dots\}$ and the number of categories can be determined in the concrete decision-making situation. The main task in the paper is to build the classifier that can model the relation between the attribute set x_k and the category label y_k . The data processing and classification method are introduced in Section 3.2.

3.2. CatBoost Algorithm Integrated with 2-Tuple Linguistic.

For the advanced integrated classification method, the CatBoost algorithm is embedded with a processing mechanism of input features. The processing replaces the original inputs of numerical codes. The main solution of CatBoost is as follows.

The attribute variable x_k^i represents the i -th dimensional input feature of the k -th training sample. The category label of the k -th sample is y_k . The conditional expected value of x_k^i under the category y_k is expressed as \hat{x}_k^i ; namely, $\hat{x}_k^i \approx E(y_k | x^i = x_k^i)$. Take $D_k = \{x_j: \sigma(j) < \sigma(k)\}$ as a training sample, and σ is a random arrangement. \hat{x}_k^i is calculated as

$$\hat{x}_k^i = \frac{\sum_{x_j \in D_k} \mathbf{1}_{x_j^i = x_k^i} \cdot y_j + ap}{\sum_{x_j \in D_k} \mathbf{1}_{x_j^i = x_k^i} + ap}, \quad (1)$$

where p is the smoothing parameter and it is usually set to the mean value of the target data set. a is the coefficient of p , and it is greater than 0. The sign $\mathbf{1}_{x_j^i = x_k^i}$ means the function of which its value is 1 if the marked variables are equal. CatBoost uses different permutations in the steps of gradient enhancement.

The classical algorithm of CatBoost above has been applied in various classification issues. However, it does not work well for the small samples with its built-in processing tool of input features, as shown in the example analysis of Section 4.2. Moreover, some inputs are in the text form with fuzzy information. Then 2-tuple linguistic is introduced as the analysis tool for the fuzzy text variable.

The variable of 2-tuple linguistic can express the grade and membership degree of an object. It is expressed in the form of (s_i, a_i) . The symbol s_i is the i -th element of the predefined grade set s . a_i is the sign transfer value: $a_i \in [-0.5, 0.5]$, indicating the deviation between the real evaluation degree and s_i .

For the transform of the common input attribute to the 2-tuple linguistic, different measures can be taken in view of the input format. If the input attribute is expressed with a defined grade set, namely, $s_i \in S$, the 2-tuple linguistic can be obtained directly with the function θ :

$$\theta: S \longrightarrow S \times [-0.5, 0.5), \quad (2)$$

$$\theta(s_i) = (s_i, 0), \quad s_i \in S. \quad (3)$$

If the input attribute is expressed with a real number $\beta \in [0, T]$, T is the number of elements in the grade evaluation set S . round is the rounding operator. β is transformed into the 2-tuple linguistic variable by the function Δ . The function Δ is defined as

$$\Delta: [0, T] \longrightarrow S \times [-0.5, 0.5), \quad (4)$$

$$\Delta_\beta = \begin{cases} s_k, & k = \text{round}(\beta), \\ a_k = \beta - k, & a_k \in [-0.5, 0.5). \end{cases} \quad (5)$$

Conversely, the 2-tuple linguistic (s_i, a_i) can be converted to β by the inverse function Δ^{-1} :

$$\Delta^{-1}: S \times [-0.5, 0.5) \longrightarrow [0, T], \quad (6)$$

$$\Delta^{-1}(s_k, a_k) = k + a_k = \beta.$$

In the proposed classifier, the built-in input feature processing tool in CatBoost is replaced with the 2-tuple linguistic. The CatBoost algorithm integrated with 2-tuple linguistic is proposed as the expert classifier. The improved method is abbreviated as 2L-CatBoost.

In the method, the inputs are preprocessed according to their formats. If the inputs are expressed with the grade set in text format, they can be transformed into 2-tuple linguistic following formulas (2) and (3). If the inputs are given with numbers, they can be transformed following formulas (4) and (5). Besides, the membership degree of a_i can be given by the calculation of semantic similarity. Then the processed inputs can be imported into the CatBoost, and the CatBoost algorithm is shown in Algorithm 1.

3.3. Automatic Flow of Expert Classification and Selection.

The experts should be analyzed and selected reliably for multicriteria group decision-making. Especially in modern DSS, the work should be operated with the software. Then an automatic flow of the expert classification and selection is designed, in which the collection and update of the expert information are also contained. The experts are classified by the 2L-CatBoost algorithm. The classification results can be the main reference for the expert selection, and the selected experts can provide decision opinions for the multicriteria decision-making activities. The flow of the expert classification and selection is shown in Figure 1.

As shown in Figure 1, the process of expert classification and selection consists of two parts, namely, the training and application. In the figure, the training part is shown with the green blocks. The application part is blue, and the shared

part is dark orange. The concrete flow can be summarized as follows:

- (1) The expert information is collected with the web crawler or artificially. The information should meet the needs as much as possible; that is, the information can well reflect the academic level and experience degree of experts in a certain field.
- (2) The information of experts needs to be stored with structured data, which can be used to train the classifier model and facilitate users to consult expert information. Therefore, if the data source contains semistructured data, they will be converted into structured data with 2-tuple linguistic.
- (3) The structured expert data are labeled with the category of the expert. Concretely, the experts are allocated with a certain degree, and the labeled expert data are stored in the expert database.
- (4) The stored data in the expert database are used as training samples. Then the 2L-CatBoost algorithm is modeled with the existing data. At this point, the basic framework of the expert classification is completed.
- (5) The expert database and classifier will be improved and evolved in the application. When the new data of experts are imported, the experts will be classified automatically with the trained 2L-CatBoost classifier. Meanwhile, the new experts can be set as the supplement of the training samples to retrain and improve the existing classifier.

The flow above is designed for the automatic building and updating of the expert library in DSS. In the process of group decision-making, the anticipant experts can be selected by the system according to the category result. The subsequent decision-making activities can be executed by other modules in the DSS. With the operation of DSS, the data in the expert library will increase, which can help to improve and evolve the system and library based on the proposed classification learning method. Then it can help the DSS to obtain the strong ability of automatic decision-making in the mode of human thinking.

4. Case Study

4.1. Data and Experiment Setting. The proposed classification method for the decision expert selection is verified with a decision-making case. In the previous studies [42, 43], we have analyzed the decision-making of water pollution governance. For the pollution of algal bloom in the urban lakes, it is necessary to monitor and predict its trends by using some parameter estimation methods [44–52] such as the recursive algorithms [18, 50, 53–58]. When it breaks or is going to break, rational decision-making should be carried out for emergency management. In order to promote the scientific and efficient governance of algal bloom, it is the first task to select and invite experts from the DSS. Then the other decision-making activities can run based on the experts' decision opinion.

```

input:  $\{(x_i, y_i)\}_{i=1}^n, I, \alpha, L, s, \text{Mode}$ 
 $\sigma_i \leftarrow$  random permutation of  $[1, n]$  for  $i = 0..s$ ;
 $S_r(i) \leftarrow 0$  for  $r = 0 \dots s, i = 1 \dots n$ ;
 $S'_r(i) \leftarrow 0$  for  $r = 1 \dots s, i = 1 \dots n, j = 1 \dots \lceil \log_2 n \rceil$ ;
for  $t \leftarrow 1$  to  $I$  do
   $\text{grad} \leftarrow \text{CalcGradient}(L, S, y)$ ;
   $\text{grad}' \leftarrow \text{CalcGradient}(L, S', y)$ ;
   $r \leftarrow \text{random}(1, s)$ ;
   $T_t \leftarrow \text{BuildTree}(\text{Mode}, \text{grad}_r, \text{grad}'_r, \sigma_r, \{x_i\}_{i=1}^n)$ ;
   $\text{leaf}_{r,i} \leftarrow \text{GetLeaf}(x_i, T_t, \sigma_r)$  for  $r = 0..s, i = 1..n$ ;
  foreach leaf  $R_j^t$  in  $T_t$  do
     $b_j^t \leftarrow -\text{avg}(\text{grad}_0(i))$  for  $i: \text{leaf}_{r,i} = j$ ;
   $S, S' \leftarrow \text{UM}(\text{Mode}, \text{leaf}, T_t, \{b_j^t\}_j, S, S', \text{grad}, \text{grad}', \{\sigma_r\}_{r=1}^s)$ 
return  $(x) = \sum_{t=1}^I \sum_j \alpha b_j^t \mathbf{1}_{\{\text{GetLeaf}(x, T_t, \text{ApplyMode})=j\}}$ 

```

ALGORITHM 1: CatBoost.

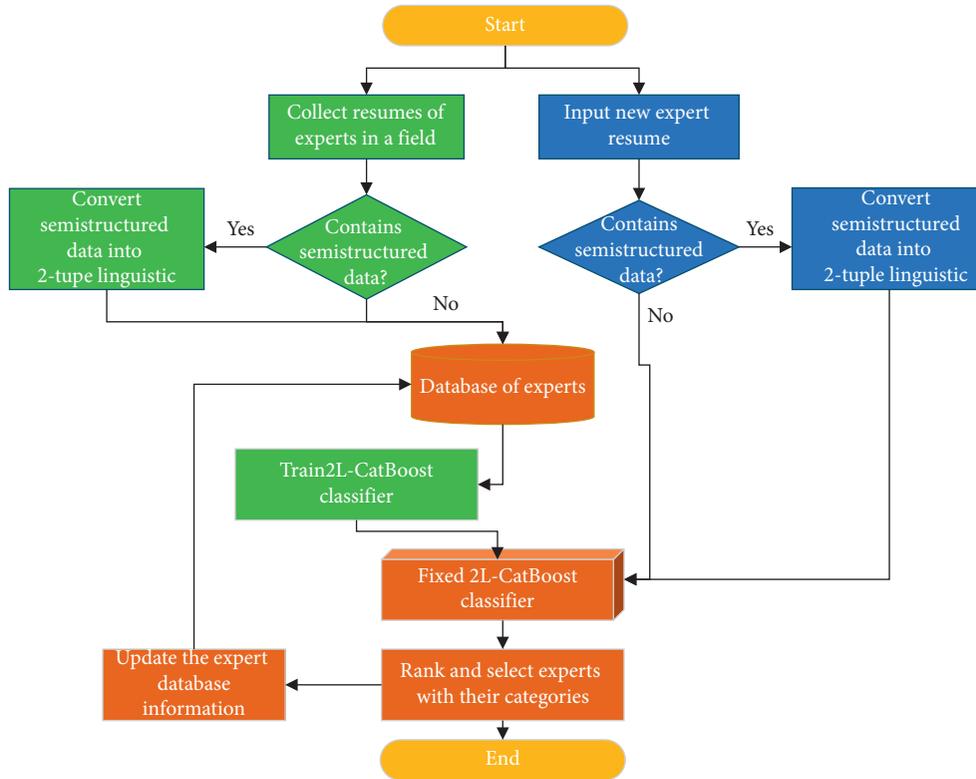


FIGURE 1: The process of expert classification and selection.

In the selection of experts, the administrators usually give priority to ones with high academic level and rich practical experience in the field. The classification and evaluation of experts will be tested in the experiment.

The basic resumes of experts are collected in previous studies. The information is from the affiliated websites and China National Knowledge Infrastructure (CNKI). The collected expert information is stored in a structured datasheet. A total of 56 experts and scholars related to algal bloom management are collected [42]. Each expert is represented with 11 attribute variables that can reflect the academic level and professional experience. The names and

meanings of attributes are shown in Table 1. The professional title is set as a categorical feature and other attributes are of a numerical value. The basic data set of experts is shown in Table 2.

The experts in Table 2 are divided into five categories. 56 samples are tagged with the category labels of I, II, III, IV, and V. Level I means the expert is the highest professional, and V means the lowest. The distribution of the expert category is shown in Figure 2. For the demand for the minimum size of samples in the classifier training, the existing data set is expanded. The Monte Carlo simulation [59] is used to expand the data set to 50 samples in each

TABLE 1: Attribute names and their representations.

No.	Expert feature factors	Optional values
f_1	Professional title	1: professor/researcher; 2: associate professor/researcher; 3: middle; 4: others
f_2	He/she is a doctoral supervisor	1: true; 2: false
f_3	Age	Real number
f_4	Highest education degree	1: doctor; 2: master
f_5	Number of papers in CNKI	Real number
f_6	Numbers of lakes and reservoirs	Real number
f_7	Number of papers about governance blooms	Real number
f_8	Number of papers about the water environment	Real number
f_9	Number of papers about chlorophyll	Real number
f_{10}	Number of papers about blue-green algae	Real number
f_{11}	Number of papers about aquatic organism	Real number

TABLE 2: Basic data set of experts in the field of algal bloom governance.

Expert	f_1	f_2	f_3	f_4	f_5	f_6	f_7	f_8	f_9	f_{10}	f_{11}
Exp1	2	2	1984	1	12	5	0	7	0	0	0
Exp2	4	2	1980	2	5	2	2	1	0	0	0
Exp3	1	1	1975	1	118	25	25	83	0	0	0
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
Exp55	2	2	1995	1	25	25	0	0	0	0	0
Exp56	1	1	1964	1	80	18	13	29	0	20	0

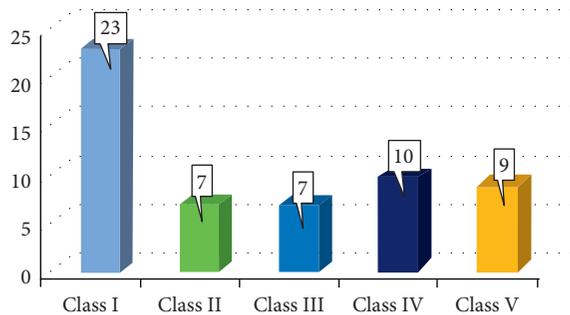


FIGURE 2: The distribution of experts in five categories.

category. Then 250 expanded samples are obtained, which are set as the training set, and the original 56 samples are used as the test set.

The training and test of the classifier are conducted in the preset environment. The concrete setting of the hardware and software is shown in Table 3. The 2L-CatBoost is trained with the expanded samples. The parameters in the training are shown in Table 4.

4.2. Result. The original 56 samples of experts are imported into the trained 2L-CatBoost classifier. The classification results are shown in Table 5, where the value in the table means the number classified into the class of columns from the class of the row. The confusion matrix of the classification result is shown in Figure 3.

From the results in Table 5, it can be seen that only 3 of the 56 test samples are misclassified. The total classification accuracy is 94.64%. In the misclassified samples, 2 experts in

class III are classified into class II; and 1 expert in class IV is classified into class III. The error can also be seen in Figure 3.

The proposed 2L-CatBoost mainly improves the classifier by processing the input with 2-tuple linguistic. Then the effect of the processing approach is tested and compared with the traditional ways, of which the one-hot encoding and the embedded feature conversation in CatBoost are set as the contrast. The samples of 56 experts are preprocessed with the three methods; then they are imported into the same CatBoost classifier. The classification accuracy of the three methods is shown in Figure 4.

For the different preprocessing methods, the proposed method obtains the best result. The result of the embedded feature conversation is the worst. It indicates that the mechanism embedded in CatBoost cannot effectively dispose of the samples in small size. Its accuracy is even worse than the one-hot encoding method.

Moreover, the classical classifiers are set as the contrast methods, including the SVM, DT, AdaBoost, random forest, naive Bayes, KNN, logistic regression, and XgBoost. For the SVM, the different kernel functions are used. The classification accuracy of the methods is shown in Figure 5. In the contrast experiments, two preprocessing approaches of the inputs are adapted, the one-hot encoding and the 2-tuple linguistic, represented with different colors in Figure 5. The results show that 2L-CatBoost has the highest classification accuracy.

5. Discussion

As the core information source of the multicriteria decision-making, the experts should be evaluated and selected rationally and automatically. In this paper, the classifier of machine learning is introduced to solve the problem. An improved fuzzy classifier 2L-CatBoost is proposed. The experiments are designed to test the performance. The proposed classifier is discussed based on the results.

For the proposed 2L-CatBoost classifier, it integrates CatBoost and the 2-tuple linguistic, which can take the advantages of both methods. For the CatBoost algorithm, it is proved to be effective in the expert classification. The proposed classifier has obtained a better result than other classifiers, including KNN, XgBoost, SVM, and naive Bayes. Besides, the training time of the 2L-CatBoost classifier is very short. In the experiment, it takes 7 iterations to obtain the

TABLE 3: Hardware and software environment for the experiment.

Software	Hardware
Anaconda 3 (2019.10)	
Python 3.7	
CatBoost: 0.22	Intel i5-7300u (2.7 Ghz~3.5 Ghz)
XgBoost: 1.0.2	

TABLE 4: The parameter settings of the CatBoost.

Parameter	Value
Logging level	Silent
Depth	5
Learning rate	0.001
Custom loss	Accuracy
Eval metric	Accuracy
Bagging temperature	0.83
Od type	Iter
RSM	0.78
Od wait	150
Metric period	400
l2-leaf reg	0
Thread count	20
Random seed	967
Iterations	7

TABLE 5: Classification result of the experts.

	Class I	Class II	Class III	Class IV	Class V
Class I	23	0	0	0	0
Class II	0	7	0	0	0
Class III	0	2	5	0	0
Class IV	0	0	1	9	0
Class V	0	0	0	0	9

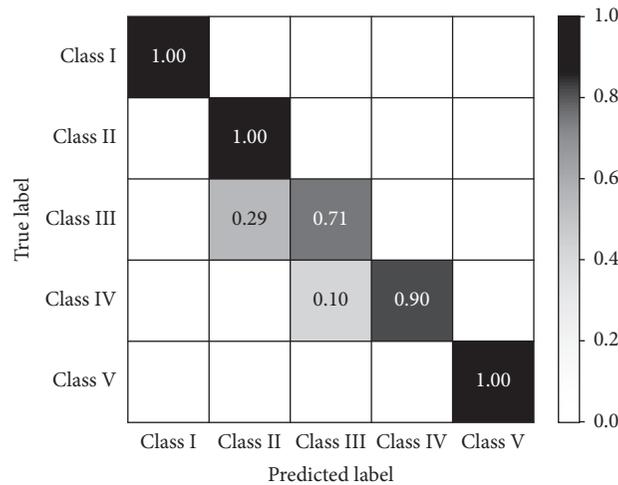


FIGURE 3: Confusion matrix of the classification results.

optimal model. The advantage makes rapid updates and involvement of the expert library in the DSS possible.

The other main contribution is the processing of fuzzy inputs. The 2L-CatBoost classifier introduces the 2-tuple linguistic for the feature conversation of the input fuzzy

information. The 2-tuple linguistic helps to increase the accuracy of the CatBoost classifier. Compared with the classical one-hot encoding method, the 2-tuple linguistic makes logistic regression, naive Bayes, and CatBoost better on the categorization. The results also show that

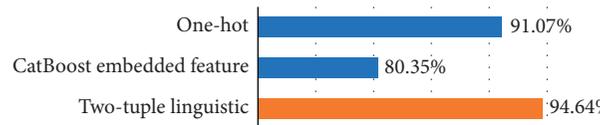


FIGURE 4: Classification accuracy of three preprocessing methods.

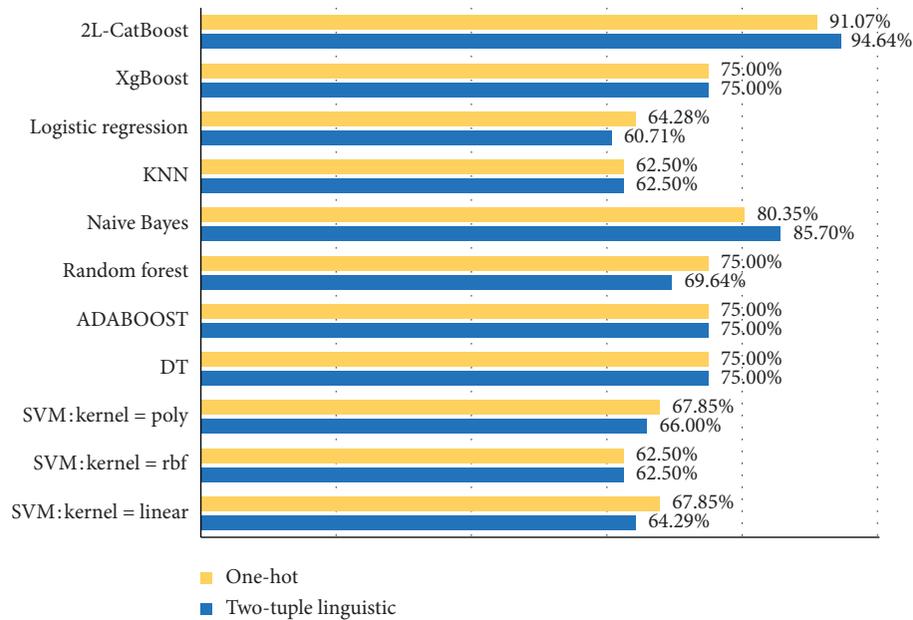


FIGURE 5: Classification accuracy of different classifiers.

not all classifiers are suitable to use 2-tuple linguistic to deal with categorical features in the case of small samples. CatBoost has been proved to be effective in the case of a large sample size, but it fails when the data size is limited with the embedded feature processing. In this case, the proposed 2L-CatBoost still performs well with the help of the 2-tuple linguistic.

In summary, the intelligent classification method of decision experts based on machine learning proposed in this paper can effectively process expert information. After the construction of the expert information database, the task of expert classification can be completed. The process is objective enough, which is helpful to promote the standardization and efficiency of the decision-making process. The selection and classification of experts will help the subsequent group decision-making with multiple approaches [60, 61].

The method proposed in this paper still has some shortcomings that need to be further improved in the follow-up work. The method has not been proved to be useful in any other data set. It is indeed difficult to adapt to all the data situations. The transferability of the method should be explored in the future. Besides, if the input data contains semistructured data, it needs the manual concertation from semistructured data to structured data, which undoubtedly increases the workload of users and reduces the efficiency. In the future, it is expected to explore the automatic way to convert the unstructured and semistructured data.

6. Conclusion

The issue of the decision source is studied for the multicriteria group decision-making. An automatic classification method of the experts is expected to be the important support for expert selection. In the solution, the improved fuzzy CatBoost classifier is proposed, integrated with the 2-tuple linguistic. It can dispose of the fuzzy input features effectively for accurate classification. Meanwhile, the general creation and update of the decision expert database are also designed. The experiment and results indicate that the proposed 2L-CatBoost is available and suitable for the expert classification with small samples and fuzzy inputs of features. In the future, more intelligent techniques can be introduced and studied to improve the fuzzy information in multicriteria decision-making, including automatic text collection and analysis, data prediction, and natural language understanding. Then the multicriteria group decision-making method will be efficient and intelligent. The proposed method in this paper can combine other estimation algorithms [62–65] to study the multicriteria decision-making problems.

Data Availability

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

Conflicts of Interest

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

Acknowledgments

This research was funded by National Social Science Fund of China (no. 19BGL184), Beijing Excellent Talent Training Support Project for Young Top-Notch Team (no. 2018000026833TD01), the National Natural Science Foundation of China (nos. 61903008 and 61673002), and the Research Foundation for Youth Scholars of Beijing Technology and Business University (no. QNJJ2020-26).

References

- [1] W. Che, *Encyclopedia of Psychological Counseling*, Zhejiang Science and Technology Press, Hangzhou, China, 2001.
- [2] M. Zeleny, "Multiple criteria decision making (MCDM): from paradigm lost to paradigm regained?" *Journal of Multi-Criteria Decision Analysis*, vol. 18, no. 1-2, pp. 77–89, 2011.
- [3] U. Cortés, M. Sánchez-Marré, L. Ceccaroni, I. R-Roda, and M. Poch, "Artificial intelligence and environmental decision support systems," *Applied Intelligence*, vol. 13, no. 1, pp. 77–91, 2000.
- [4] T. H. Payne and H. Thomas, "Computer decision support systems," *Chest*, vol. 118, no. 2, pp. 47S–52S, 2000.
- [5] J. D'Haen, P. D. Van d, D. Thorleuchter et al., "Integrating expert knowledge and multilingual web crawling data in a lead qualification system," *Decision Support Systems*, vol. 82, pp. 69–78, 2016.
- [6] N. Friedmann, D. Geigerd, and M. Goldszmidt, "Bayesian network classifiers," *Machine Learning*, vol. 29, no. 1, pp. 131–163, 1997.
- [7] P. Hart, "The condensed nearest neighbor rule (Corresp.)," *IEEE Transactions on Information Theory*, vol. 14, no. 3, pp. 515–516, 1968.
- [8] L. Breiman and J. Friedman, *Classification and Regression Trees*, Thomson Wadsworth, Belmont, CA, USA, 1984.
- [9] C. Cortes and V. Vapnik, "Support-vector networks," *Machine Learning*, vol. 20, no. 3, pp. 273–297, 1995.
- [10] T. Chen and C. Guestrin, "XGBoost: a scalable tree boosting system," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, San Francisco, CA, USA, August 2016.
- [11] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.
- [12] Y. Bai, X. Jin, X. Wang et al., "Dynamic correlation analysis method of air pollutants in spatio-temporal analysis," *International Journal of Environmental Research and Public Health*, vol. 17, no. 1, p. 360, 2020.
- [13] F. Ding, L. Lv, J. Pan, X. Wan, and X.-B. Jin, "Two-stage gradient-based iterative estimation methods for controlled autoregressive systems using the measurement data," *International Journal of Control, Automation and Systems*, vol. 18, no. 4, pp. 886–896, 2020.
- [14] Y. Bai, X. Wang, X. Jin et al., "A neuron-based Kalman filter with nonlinear autoregressive model," *Sensors*, vol. 20, no. 1, p. 299, 2020.
- [15] O. Chapelle, E. Manavoglu, and R. Rosales, "Simple and scalable response prediction for display advertising," *ACM Transactions on Intelligent Systems and Technology*, vol. 5, no. 4, pp. 1–34, 2014.
- [16] P. Rodríguez, M. A. Bautista, J. González, and S. Escalera, "Beyond one-hot encoding: lower dimensional target embedding," *Image and Vision Computing*, vol. 75, pp. 21–31, 2018.
- [17] Y. Bai, X. Wang, X. Jin, T. Su, J. Kong, and B. Zhang, "Adaptive filtering for MEMS gyroscope with dynamic noise model," *ISA Transactions*, vol. 101, pp. 430–441, 2020.
- [18] X. Zhang and F. Ding, "Hierarchical parameter and state estimation for bilinear systems," *International Journal of Systems Science*, vol. 51, no. 2, pp. 275–290, 2020.
- [19] F. Herrera and L. Martinez, "A 2-tuple fuzzy linguistic representation model for computing with words," *IEEE Transactions on Fuzzy Systems*, vol. 8, pp. 746–752, 2000.
- [20] L. Prokhorenkova, G. Gusev, A. Vorobev et al., "CatBoost: unbiased Boosting with categorical features," in *Proceedings of the 32nd Conference on Neural Information Processing Systems*, Montreal, Canada, December 2018.
- [21] P. Biswas, S. Pramanik, and B. C. Giri, "TOPSIS method for multi-attribute group decision-making under single-valued neutrosophic environment," *Neural Computing and Applications*, vol. 27, no. 3, pp. 727–737, 2016.
- [22] S. Pramanik and R. Mallick, "TODIM strategy for multi-attribute group decision making in trapezoidal neutrosophic number environment," *Complex & Intelligent Systems*, vol. 5, no. 4, pp. 379–389, 2019.
- [23] Z. Xu, "A method based on linguistic aggregation operators for group decision making with linguistic preference relations*1," *Information Sciences*, vol. 166, no. 1-4, pp. 19–30, 2004.
- [24] G. Qian, H. Wang, and X. Feng, "Generalized hesitant fuzzy sets and their application in decision support system," *Knowledge-Based Systems*, vol. 37, pp. 357–365, 2013.
- [25] C.-T. Chen, "Extensions of the TOPSIS for group decision-making under fuzzy environment," *Fuzzy Sets and Systems*, vol. 114, no. 1, pp. 1–9, 2000.
- [26] P. Biswas, S. Pramanik, and B. C. Giri, "NonLinear programming approach for single-valued neutrosophic TOPSIS method," *New Mathematics and Natural Computation*, vol. 15, no. 2, pp. 307–326, 2019.
- [27] S. Pramanik, S. Dalapati, S. Alam, F. Smarandache, and T. K. Roy, "NS-cross entropy based MAGDM under single-valued neutrosophic set environment," *Information*, vol. 9, no. 2, p. 37, 2018.
- [28] P. Biswas, S. Pramanik, and B. C. Giri, "NH-MADM strategy in neutrosophic hesitant fuzzy set environment based on extended GRA," *Informatica*, vol. 30, no. 2, pp. 213–242, 2019.
- [29] B. Sun, W. Ma, B. Li, and X. Li, "Three-way decisions approach to multiple attribute group decision making with linguistic information-based decision-theoretic rough fuzzy set," *International Journal of Approximate Reasoning*, vol. 93, pp. 424–442, 2018.
- [30] D. Liang, M. Wang, Z. Xu, and D. Liu, "Risk appetite dual hesitant fuzzy three-way decisions with TODIM," *Information Sciences*, vol. 507, pp. 585–605, 2020.
- [31] C. Zhang, D. Li, and J. Liang, "Hesitant fuzzy linguistic rough set over two universes model and its applications," *International Journal of Machine Learning and Cybernetics*, vol. 9, no. 4, pp. 577–588, 2018.
- [32] G. A. Gorry and M. S. Morton, "A framework for management information systems," *Sloan Management Review*, vol. 30, no. 3, pp. 49–61, 1989.

- [33] S. Eom and E. Kim, "A survey of decision support system applications (1995–2001)," *Journal of the Operational Research Society*, vol. 57, no. 11, pp. 1264–1278, 2006.
- [34] J. Kazak, J. van Hoof, and S. Szewranski, "Challenges in the wind turbines location process in Central Europe - the use of spatial decision support systems," *Renewable and Sustainable Energy Reviews*, vol. 76, pp. 425–433, 2017.
- [35] B. S. McIntosh, J. C. Ascough, M. Twery, J. Chew et al., "Environmental decision support systems (EDSS) development - challenges and best practices," *Environmental Modelling & Software*, vol. 26, no. 12, pp. 1389–1402, 2011.
- [36] Q. Pang, H. Wang, and Z. Xu, "Probabilistic linguistic term sets in multi-attribute group decision making," *Information Sciences*, vol. 369, pp. 128–143, 2016.
- [37] X. Zhou, J. Xiao, Z. Zhou et al., "Research and application of an intelligent decision support system," in *Proceedings of the International Conference on Computer Network, Electronic and Automation (ICCNEA)*, pp. 186–189, Xian China, September 2017.
- [38] S. Faizi, T. Rashid, W. Saifun, S. Zafar, and J. Wątróbski, "Decision making with uncertainty using hesitant fuzzy sets," *International Journal of Fuzzy Systems*, vol. 20, no. 1, pp. 93–103, 2018.
- [39] J. Zhan, Q. Liu, and T. Herawan, "A novel soft rough set: soft rough hemirings and corresponding multicriteria group decision making," *Applied Soft Computing*, vol. 54, pp. 393–402, 2017.
- [40] Y. Bai, B. Zhang, X. Wang et al., "A novel group decision-making method based on sensor data and fuzzy information," *Sensors*, vol. 16, no. 11, p. 1799, 2016.
- [41] G. Wei, "2-tuple intuitionistic fuzzy linguistic aggregation operators in multiple attribute decision making," *Iranian Journal of Fuzzy Systems*, vol. 16, no. 4, pp. 159–174, 2019.
- [42] Y. Yang, Y. Bai, X. Wang et al., "Group decision-making support for sustainable governance of algal bloom in urban lakes," *Sustainability*, vol. 12, no. 4, p. 1494, 2020.
- [43] L. Wang, T. Zhang, X. Jin et al., "An approach of recursive timing deep belief network for algal bloom forecasting," *Neural Computing and Applications*, vol. 32, no. 1, pp. 163–171, 2020.
- [44] Y. Bai, X. Jin, X. Wang et al., "Compound autoregressive network for prediction of multivariate time series," *Complexity*, vol. 2019, Article ID 9107167, 11 pages, 2019.
- [45] X. Jin, N. Yang, X. Wang et al., "Deep Hybrid model based on EMD with classification by frequency characteristics for long-term air quality prediction," *Mathematics*, vol. 8, p. 214, 2020.
- [46] X. Jin, N. Yang, X. Wang et al., "Integrated predictor based on decomposition mechanism for PM2.5 long-term prediction," *Applied Sciences*, vol. 9, no. 21, p. 4533, 2019.
- [47] X. Jin, N. Yang, X. Wang et al., "Hybrid deep learning predictor for smart agriculture sensing based on empirical mode decomposition and gated recurrent unit group model," *Sensors*, vol. 20, no. 5, p. 1334, 2020.
- [48] X. Jin, X. Yu, X. Wang et al., "Deep learning predictor for sustainable precision agriculture based on internet of things system," *Sustainability*, vol. 12, p. 1433, 2020.
- [49] F. Ding, L. Xu, D. Meng et al., "Gradient estimation algorithms for the parameter identification of bilinear systems using the auxiliary model," *Journal of Computational and Applied Mathematics*, vol. 369, Article ID 112575, 2020.
- [50] X. Zhang, F. Ding, and E. Yang, "State estimation for bilinear systems through minimizing the covariance matrix of the state estimation errors," *International Journal of Adaptive Control and Signal Processing*, vol. 33, no. 7, pp. 1157–1173, 2019.
- [51] F. Ding, G. Liu, and X. P. Liu, "Parameter estimation with scarce measurements," *Automatica*, vol. 47, no. 8, pp. 1646–1655, 2011.
- [52] Y. Wang and F. Ding, "Novel data filtering based parameter identification for multiple-input multiple-output systems using the auxiliary model," *Automatica*, vol. 71, pp. 308–313, 2016.
- [53] F. Ding, L. Qiu, and T. Chen, "Reconstruction of continuous-time systems from their non-uniformly sampled discrete-time systems," *Automatica*, vol. 45, no. 2, pp. 324–332, 2009.
- [54] X. Zhang and F. Ding, "Recursive parameter estimation and its convergence for bilinear systems," *IET Control Theory & Applications*, vol. 14, no. 5, pp. 677–688, 2020.
- [55] X. Zhang, Q. Liu, F. Ding, A. Alsaedi, and T. Hayat, "Recursive identification of bilinear time-delay systems through the redundant rule," *Journal of the Franklin Institute*, vol. 357, no. 1, pp. 726–747, 2020.
- [56] Y. Liu, F. Ding, and Y. Shi, "An efficient hierarchical identification method for general dual-rate sampled-data systems," *Automatica*, vol. 50, no. 3, pp. 962–970, 2014.
- [57] J. Ding, F. Ding, X. P. Liu, and G. Liu, "Hierarchical least squares identification for linear SISO systems with dual-rate sampled-data," *IEEE Transactions on Automatic Control*, vol. 56, no. 11, pp. 2677–2683, 2011.
- [58] F. Ding, G. Liu, and X. Liu, "Partially coupled stochastic gradient identification methods for non-uniformly sampled systems," *IEEE Transactions on Automatic Control*, vol. 55, no. 8, pp. 1976–1981, 2010.
- [59] Q. Pan and D. Dias, "An efficient reliability method combining adaptive support vector machine and Monte Carlo simulation," *Structural Safety*, vol. 67, pp. 85–95, 2017.
- [60] B. Sun, X. Zhou, and N. Lin, "Diversified binary relation-based fuzzy multigranulation rough set over two universes and application to multiple attribute group decision making," *Information Fusion*, vol. 55, pp. 91–104, 2020.
- [61] C. Zhang, D. Li, and J. Liang, "Multi-granularity three-way decisions with adjustable hesitant fuzzy linguistic multi-granulation decision-theoretic rough sets over two universes," *Information Sciences*, vol. 507, pp. 665–683, 2020.
- [62] X. Zhang, F. Ding, L. Xu, and E. Yang, "Highly computationally efficient state filter based on the delta operator," *International Journal of Adaptive Control and Signal Processing*, vol. 33, no. 6, pp. 875–889, 2019.
- [63] X. Zhang and F. Ding, "Adaptive parameter estimation for a general dynamical system with unknown states," *International Journal of Robust and Nonlinear Control*, vol. 30, no. 4, pp. 1351–1372, 2020.
- [64] X. Zhang, F. Ding, and L. Xu, "Recursive parameter estimation methods and convergence analysis for a special class of nonlinear systems," *International Journal of Robust and Nonlinear Control*, vol. 30, no. 4, pp. 1373–1393, 2020.
- [65] F. Ding, X. Zhang, and L. Xu, "The innovation algorithms for multivariable state-space models," *International Journal of Adaptive Control and Signal Processing*, vol. 33, no. 11, pp. 1601–1618, 2019.