JOURNAL OF INFORMATION PROCESSING SYSTEMS JIPS

Evaluation Method of College English Education Effect Based on Improved Decision Tree Algorithm

Fang Dou*

Abstract

With the rapid development of educational informatization, teaching methods become diversified characteristics, but a large number of information data restrict the evaluation on teaching subject and object in terms of the effect of English education. Therefore, this study adopts the concept of incremental learning and eigenvalue interval algorithm to improve the weighted decision tree, and builds an English education effect evaluation model based on association rules. According to the results, the average accuracy of information classification of the improved decision tree algorithm is 96.18%, the classification error rate can be as low as 0.02%, and the anti-fitting performance is good. The classification error rate between the improved decision tree algorithm and the original decision tree does not exceed 1%. The proposed educational evaluation method can effectively provide early warning of academic situation analysis, and improve the teachers' professional skills in an accelerated manner and perfect the education system.

Keywords

1. Introduction

The failure to realize the teaching orientation and the single method of teaching effect evaluation cause the current English education quality to be worrying, further seriously deteriorating the teaching quality and evaluation. In recent years, many scholars have adopted technologies such as cluster analysis and decision tree analysis in teaching system design and learning situation analysis so as to provide guidance and suggestions for teaching decision-making, but ignore the internal relationship of educational effect evaluation [1]. Therefore, in order to reduce the running time of the dataset and avoid the over fitting, this study adopts the concept of incremental learning and the calculation of characteristic interval value to improve the decision tree, and the evaluation model of English education effect is built by combining association rules and provides the possibility of dynamic control and early warning analysis for the control of education quality.

2. Literature Review

According to Li et al. [2] that the predicted flow time series can reflect the human activity flow, and the gradient enhanced decision tree algorithm improved by Kalman filter can shorten the training time of the model. In order to alleviate the mislabeling of a single decision tree algorithm, Hou et al. [3] proposed a new double confidence estimation method for labeled data, and calculated the probability using the weighted algorithm. As the results show, the optimization algorithm has good performance in terms of classification. Di and Xu [4] intervened in the attribute information gain rate of the decision tree branching and pruning strategy. The results show that the execution time of the algorithm in the experimental simulation is effectively improved by 8.75%, and its accuracy exceeds 90% [4]. Wang and Kong [5] improved the decision tree from feature attribute values and information gain to build an air quality prediction model and found that the improved model has higher prediction accuracy and classification performance. Maulana and Defriani [6] applied the logic model tree and decision tree algorithm to the school information data, which strengthens the prediction of students' learning time and provides guiding suggestions for their education and teaching management.

Li [7] established a mathematical model of oral English teaching based on metacognitive means and the characteristics of college English teaching. According to the results, it has high application value. Karjanton and Simon [8] adopted Bruce's taxonomy theory to carry out teaching analysis qualitatively and quantitatively, but it is greatly limited by learning technology tools. Liang [9] developed a teaching effect evaluation model based on artificial intelligence algorithm with the help of support vector mechanism, which has good applications.

The above-mentioned literature shows that the classification decision tree is involved in many fields and behaves as a combination algorithm or model, while the college English education evaluation model is mostly assisted by a single algorithm or theoretical knowledge, and there is a lack of systematic analysis method of evaluation. Therefore, it is innovative and practical to adopt the improved decision tree algorithm and association rules into the education evaluation system.

3. Establishment of English Education Effect Evaluation Model based on Improved Decision Tree

3.1 Improved Decision Tree Based Algorithm ID3

The decision tree is used to realize the classification and model construction of dataset structure based on tree building and pruning, and the induction algorithm ID3 version is commonly used in this algorithm. The mainstream algorithms (C4.5) in traditional decision tree algorithms are over-fitting and their accuracy is easily affected, so the algorithm has poor performance in the operation process. The core of ID3 algorithm is to divide the sample set into subsets by using the information gain rate as the criterion for node attribute selection, so as to reduce the average depth of the decision tree and improve the performance and accuracy of classification.

$$\begin{cases} H(X) = -\sum_{i=1}^{n} pi \, log \, p \, i \\ g(D,A) = H(D) - H(D|A), gR(D,A) = g(D,A)/HA(D) \end{cases} \tag{1}$$

Eq. (1) is the calculation formula of decision tree entropy and mutual information, where $H(D)$ is entropy, entropy indicates the uncertainty of random variables, D is training dataset, $H(D|A)$ is empirical conditional entropy, A is characteristic, $p(i)$ is uncertainty, defines the self-information of information symbols, and $g(D, A)$ means mutual information. $gR(D, A)$ is the information gain ratio. $HA(D)$ is the ratio of entropy of dataset D about the value of feature A, and R is the representation of a certain type of feature set.

The original ID3 decision algorithm cannot determine the relationship between the current value and the optimal attribute value. Therefore, the weighted assignment is investigated, that is, the risk weight is included in the algorithm to mark the importance of attributes, so as to ensure the integrity and usefulness of mining information. The calculation formula of weighted ID3 algorithm is shown in the following Eq. (2).

$$\begin{cases} L\beta(D'|A') = \sum_{j=1}^{n}(P(Vj) + \beta) \sum_{i=1}^{n\Sigma\frac{1}{P(D'ii+Vj)}} P(D'i|Vj)lg \\ g\beta(U,V) = H(D') - H\beta(D'|A') \end{cases} \tag{2}$$

Eq. (2), $L\beta(D'|A')$ is conditional entropy, $g\beta(U,V)$ means corresponding mutual information, β refers to indefinite value, D', A' is the dataset and characteristic information after weighted assignment, and P, V, j represents the probability of any sample, the number of subsets of the sample set and the value of the number of samples, respectively. At the same time, the concept of increment is adopted to design the weighted ID3 algorithm to reduce data overlap, Fig. 1 shows the step flow chart of ID3 incremental learning decision tree algorithm.
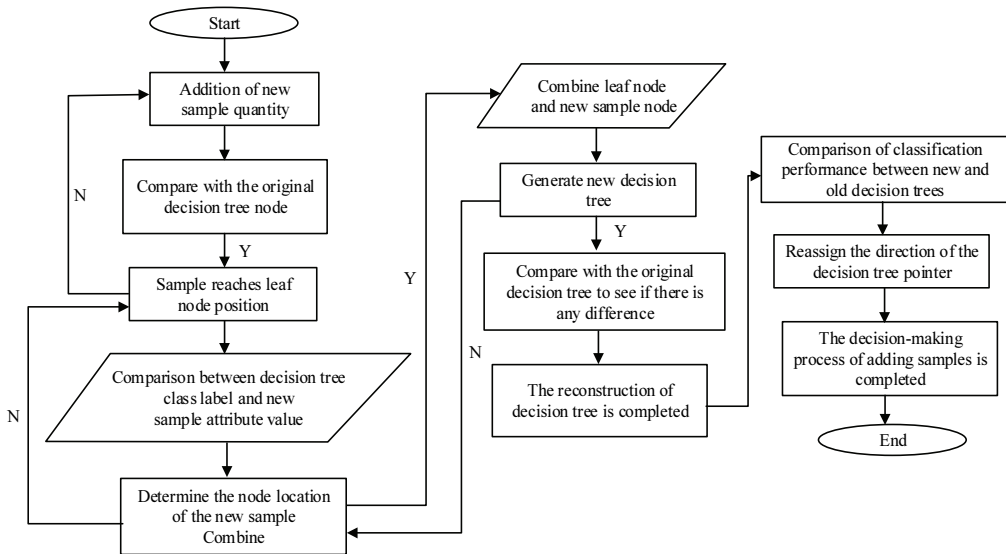

Fig. 1. Step flow chart of ID3 incremental learning decision tree algorithm.

In Fig. 1, the decision tree is reconstructed based on the decision position of the new node to ensure the integrity and optimization of the data. At the same time, the conditional misclassification rate is

adopted based on the weighted idea to realize the continuous correction in the original pruning method. The calculation formula of the corrected conditional misclassification rate is shown in Eq. (3).

$$\begin{cases} r'(t) = P(t/g)\dfrac{(e(t)+0.5)}{n(t)} = \dfrac{(e(t)+0.5)}{n(g)} \\ e'(t) = [e(t)+0.5]P(t/g) \end{cases} \Rightarrow \begin{cases} r'(Tt) = \dfrac{\sum[e(s)+0.5]p(s/gs)}{\sum n(s)p(s/gs)} \\ e'(Tt) = \sum[e(s)+0.5]P(s/gs) \end{cases} \tag{3}$$

In Eq. (3), $r'(t)$ is the conditional misclassification rate of continuous correction, $e'(t)$ is the number of misclassification of node t, g means the source node of the node, $n(g)$ represents the number of instances of the node, and $P(t/g)$ is the conditional probability from the source node to the node. Where, $r'(Tt)$ and $e'(Tt)$ are the corrected formula, Tt is a subtree, $p(s/gs)$ is the source node of node s, and gs means the node condition for splitting into nodes s during classification. When it comes to use the modified decision tree for classification, care should be taken to ensure the scientificity, authority and feasibility of the weight setting when assigning the indicator weights. Then, the judgment matrix is built according to the weight of indicators, the relative importance of indicators is compared, and the corresponding indicator matrix is constructed after adding the weights of indicators screened and checked by experts.

3.2 Establishment of Teaching Effect Evaluation Model based on Improved Decision Tree and Association Rules

When dealing with large datasets, the recursive construction time of the decision tree takes too long time and leads to the over fitting phenomenon of the training set. Therefore, the recursive generation of the decision tree based on the optimal feature and the optimal cut-off point is studied. Eq. (4) is the decision tree algorithm formula for the improved eigenvalue interval partition model.

$$\begin{cases} H_0 = \left\{ X_1, X_2, \ldots, X_{\frac{M}{Q}} \right\} \\ H_{Q-1} = \left\{ X_{\left[(Q-1)\frac{M}{Q}\right]+1}, X_{\left[(Q-1)\frac{M}{Q}\right]+2}, \ldots, X_M \right\} \end{cases} \tag{4}$$

$$\begin{cases} H_0 = \{X | X_{min} \leq X \leq (X_{max} - X_{min})/Q\} \\ H_{Q-1} = \{X | (X_{max} - X_{min})/Q * (Q-1) \leq X \leq X_{max}\} \end{cases}$$

Eq. (4) is the interval division formula for equal precision eigenvalues, where M is the number of samples and Q is the number of given division intervals. The number of samples in a single interval is $\frac{M}{Q}$, and X means a certain characteristic dimension. When the dimensionality fluctuates, equal-precision eigenvalue interval partitioning can hardly handle sample integrity, so variable-precision eigenvalue.

$$\begin{cases} H_0 = \{X | X_{min} \leq X \leq (X_{max} - X_{min})/Q\} \\ H_{Q-1} = \{X | (X_{max} - X_{min})/Q * (Q-1) \leq X \leq X_{max}\} \end{cases} \tag{5}$$

In Eq. (5), X_{max}, X_{min} are the maximum and minimum values of eigenvalues, the division result of variable precision characteristic interval is the average value of sample eigenvalues in the classification node. The division diagram of the two eigenvalue interval algorithms is shown in Fig. 2.
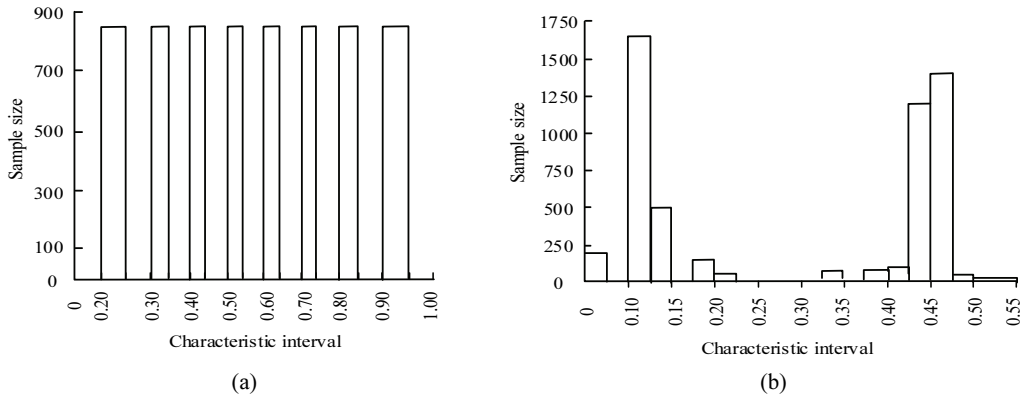
Fig. 2. Schematic diagram of two eigenvalue interval algorithms: (a) equal precision eigenvalues and (b) variable precision eigenvalues.

The effect of English teaching needs to consider several aspects, such as basic knowledge learning and comprehensive quality. Therefore, the research leverages the association rules to comprehensively evaluate the teaching effect, realize the association analysis of system information, so as to provide early warning and prevention for the process of English education and teaching. The Eq. (6) is the calculation formula of association rule.

$$Support(X \Rightarrow Y) = \frac{\|\{T|(X \Rightarrow Y) \subseteq T, T \subseteq D\|}{D}$$
$$\Leftrightarrow \begin{cases} Support(X) = \frac{\|\{d \in D|X \in d\}\|}{\|D\|} \\ Confidenod(X \Rightarrow Y) = Support(X \cup Y)/Support(X) \\ = \|\{T|(X \cup Y) \subseteq T, T \subseteq D\|/\|\{T|(X \subseteq Y), T \subseteq D\| \end{cases} \tag{6}$$

In Eq. (6), X, Y means the transaction, D is the transaction database, and rule $(X \Rightarrow Y)$ is the number of transactions contained in all transactions and the percentage of the number of transactions. Fig. 3 is the structural diagram of the English teaching effect evaluation system based on the improved decision tree and association rules.
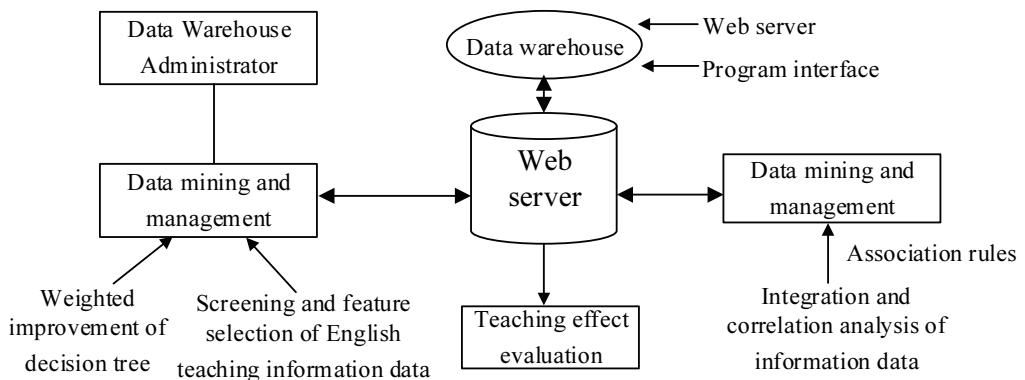
Fig. 3. Structure diagram of English teaching effect evaluation based on improved decision tree and association rules.

In Fig. 3, the feature extraction and correlation analysis of teaching information are realized using weighted improved decision tree and association rules, which provide a basis for decision making in teaching design. The experiment is carried out based on computer hardware equipment. The configuration of computer hardware equipment is shown in Table 1.

Table 1. Configuration of computer hardware equipment and scope of sample dataset

Processor	Intel Core i7-77003, 60 GHz
Memory	32.0 GB
System	Windows version 10.21
Sample size range of experimental dataset	500–15,000 numbers

# 4. Analysis on the Application Effect of English Teaching Effect Model based on Improved Decision Tree

4.1 Effectiveness Test of Improved Decision Tree Algorithm Performance

Then, the establishment method of the English teaching model adopted in this study is compared with the previous research methods. The results are shown in Table 2. The proposed algorithm can enhance the anti-fitting performance, grasp the dynamic change law of both teaching sides and improve the performance of teaching application.

Table 2. Comparison of the advantages and disadvantages of different models in English teaching

Algorithm	Advantages	Shortcomings
Oral English teaching model based on the analysis of metacognitive means of Li [7]	Use metacognitive means to analyze the current teaching situation. The performance of the computer-aided technology discovery model has a good application effect.	There is a lack of support of algorithm technology tools in model construction.
Application model under the Bruce classification theory of Karjanto and Simon [8]	It quantitatively finds the significant statistical relationship between the degree of classroom application effect and students' performance.	Lack of support and verification of technical tools in the application of result exploration.
Support vector machine model under Huang's machine learning [10]	Be able to classify teaching in combination with the distribution characteristics of samples. It has good practical applicability to the evaluation of teaching quality.	It is difficult to select the unique feature vector of kernel data.
Research methods of this paper	Incremental learning algorithm is introduced to improve the decision tree, which improves the accuracy and fitting performance of the algorithm. Association rules can realize the dynamic detection and adjustment of both teaching parties and information data.	Small amount of information sample data.

The decision tree algorithm is introduced into the evaluation and analysis of teaching effectiveness, which helps to integrate and coordinate teaching information and realize the feature extraction of data information. Firstly, the research will improve the decision tree used for performance evaluation, and make statistical analysis on the accuracy of data evaluation with other algorithms. The results are shown in Fig. 4.
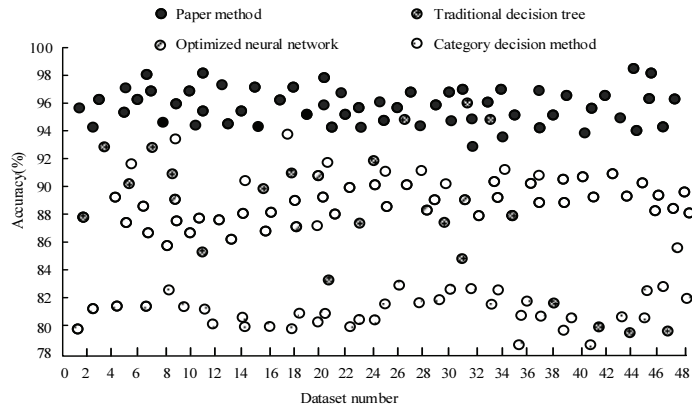
Fig. 4. Statistics of classification and evaluation accuracy of datasets by different classification algorithms.

The results in Fig. 4 show that the evaluation accuracy of the traditional decision tree fluctuates greatly, and there is a lack of data, while the average classification accuracy of the improved decision tree is 96.18%, which is much higher than the information accuracy of neural network algorithm and category decision algorithm (<92%) Then, the classification error rate and fitting performance of different algorithms under the dataset are compared as shown in Fig. 5.

(a)

(b)
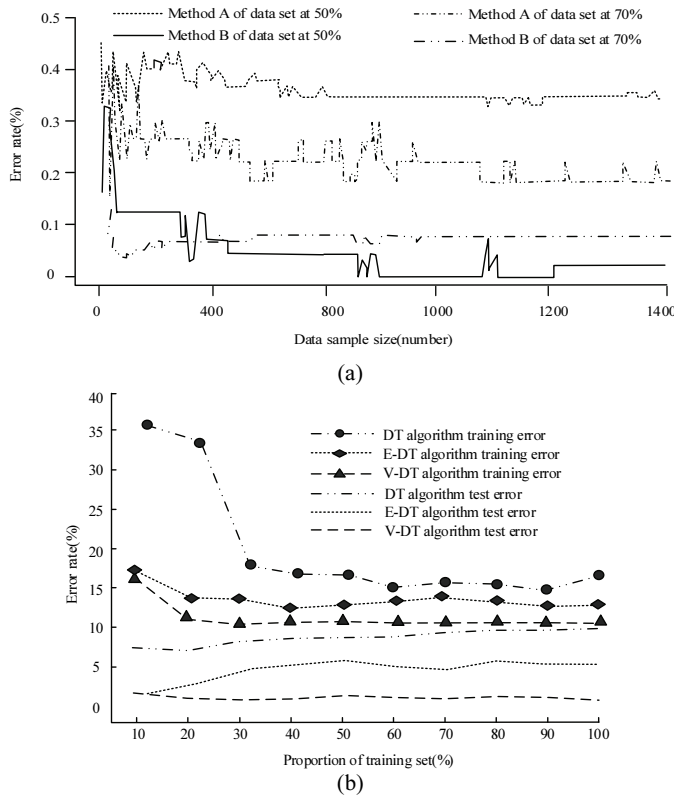
Fig. 5. (a) Statistical results of classification error rate of two methods under different proportion datasets. (b) Comparison results of fitting performance of different algorithms for dataset training and test performance.

Algorithm A and algorithm B in Fig. 5(a) are the traditional decision tree and the improved algorithm, respectively. The classification error rates of algorithm A and algorithm B are 0.25% and 0.05% at the proportion of dataset of 50% and 70%; In Fig. 5(b), "DT" algorithm refers to the decision tree algorithm. "E-dt" and "V-dt" are decision trees with the same accuracy and variable accuracy. The difference of classification error rate between "DT", "E-dt" and "V-dt" is no more than 15%, much lower than the error rate of "DT" (27.5%). The classification of datasets is calculated by different equivalent special interval decision tree algorithms. The results are shown in Fig. 6.

In Fig. 6, the classification error rate between the two eigenvalue intervals and the traditional decision tree under the three classifiers does not exceed 1%, and the classification efficiency is improved. Taking English majors in a university as the research object, this paper explores the correlation between teaching evaluation and influencing factors based on the association rules in the technology of collecting teaching information. The results are shown in Table 3.
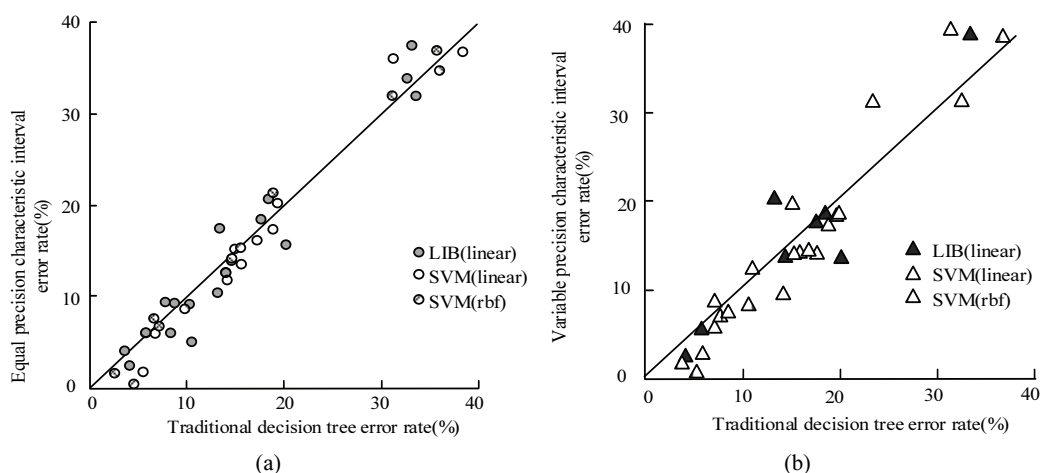
Fig. 6. Statistics of classification error rate of decision trees with different equivalent interval algorithms under three classifiers: (a) equal precision feature and (b) variable precision feature.

Table 3. Statistical results of the correlation between English quality influencing factors and teaching effect evaluation system

Influence factor	Teaching concept	Teaching process	Classroom management	Cultivation of students' skills
Classroom performance	0.25	0.28	0.26	0.31
Basic English ability	0.23	0.21	0.17	0.24
English communicative competence	0.21	0.22	0.20	0.22
Comprehensive quality of students	0.18	0.16	0.15	0.23

In Table 3, the correlation between teachers' teaching philosophy, teaching process, classroom management and the cultivation of students' skills on students' classroom enthusiasm, professional comprehensive ability and quality is greater than 0.1, and the maximum correlation coefficient between students' classroom performance and teaching effect reaches 0.31.

5. Conclusion

Computer technology creates new opportunities and provides technical means for updating the education model and information management of teaching system. Through the research and optimization of the decision tree algorithm, it is found that the evaluation accuracy of the improved decision tree algorithm is much higher than 90%, the anti-fitting performance of the algorithm is good, and the classification error rate does not exceed 1%. In the correlation analysis of the improved decision tree, it is found that the correlation between the two sides of teaching is basically more than 0.1, indicating that the evaluation on English education effect based on the improved decision tree is better and more practical and useful.

References

[1] Y. Wu, "A brief talk on how to use multimedia technology to construct a diversified evaluation system in English teaching," *Journal of Physics: Conference Series*, vol. 1744, no. 4, article no. 042058, 2021. https://doi.org/10.1088/1742-6596/1744/4/042058

[2] L. Li, S. Dai, Z. Cao, J. Hong, S. Jiang, and K. Yang, "Using improved gradient-boosted decision tree algorithm based on Kalman filter (GBDT-KF) in time series prediction," *The Journal of Supercomputing*, vol. 76, no. 9, pp. 6887-6900, 2020.

[3] F. F. Hou, W. T. Lei, H. Li, J. W. Ren, G. Y. Liu, and Q. Y. Gu, "Impr-Co-Forest: the improved co-forest algorithm based on optimized decision tree and dual-confidence estimation method," *Journal of Computers*, vol. 30, no. 6, pp. 110-122, 2019.

[4] J. Di and Y. Xu, "Decision tree improvement algorithm and its application," *International Core Journal of Engineering*, vol. 5, no. 9, pp. 151-158, 2019.

[5] Y. Wang and T. Kong, "Air quality predictive modeling based on an improved decision tree in a weather-smart grid," *IEEE Access*, vol. 7, pp. 172892-172901, 2019.

[6] M. F. Maulana and M. Defriani, "Logistic model tree and decision tree J48 algorithms for predicting the length of study period," *PIKSEL: Penelitian Ilmu Komputer Sistem Embedded and Logic*, vol. 8, no. 1, pp. 39-48, 2020.

[7] X. Li, "Characteristics and rules of college English education based on cognitive process simulation," *Cognitive Systems Research*, vol. 57, pp. 11-19, 2019.

[8] N. Karjanto and L. Simon, "English-medium instruction calculus in Confucian-Heritage Culture: flipping the class or overriding the culture?," *Studies in Educational Evaluation*, vol. 63, pp. 122-135, 2019.

[9] H. Liang, "Role of artificial intelligence algorithm for taekwondo teaching effect evaluation model," *Journal of Intelligent & Fuzzy Systems*, vol. 40, no. 2, pp. 3239-3250, 2021.

[10] W. Huang, "Simulation of English teaching quality evaluation model based on Gaussian process machine learning," *Journal of Intelligent and Fuzzy Systems*, vol. 40, no. 2, pp. 2373-2383, 2021.

Fang Dou https://orcid.org/0000-0002-1044-7338

She graduated from Henan Institute of Education majoring in English education in 2006, and obtained a master's degree in education in 2012. Lecturer and senior lecturer (secondary school series) of Foreign Language Tourism College of Henan Economic and Trade Vocational College, academic and technical leader of Henan Provincial Department of Education, famous teaching teacher of Henan Vocational School, and civilized teacher of Henan Province. In 2018, she participated in the compilation of the basic modules of "English," a new national planning textbook for secondary vocational schools by Chinese Language and Culture Press (Volume 1 and 2), and in 2018, she was the deputy editor of "Student Management" by Beijing Normal University Press. In the past 5 years, she has published seven influential articles in "Henan Education Vocational and Adult Education Edition," "English Campus," and other journals. Participated in six key projects of the Henan Provincial Department of Education, won four first prizes of the Henan Provincial Vocational Education Outstanding Teaching Achievement Award, and two second prizes of the Henan Provincial Department of Education's Humanities and Social Sciences Outstanding Achievement Award. Mentored the first prize of the Henan Provincial Secondary Vocational School Student Quality Ability Competition three times.