Software Development for Comprehensive Assessment of English Online Teaching Quality in Universities Based on Data Mining

Zhe Zhang(✉)
Department of Applied Foreign Languages, Henan Industry and Trade Vocational College,
Zhengzhou, China
zhangzhe@hngm.edu.cn

Abstract—In the context of big data in education, the education industry is combined with information technology to form an online teaching model. In order to analyze the actual effectiveness of online teaching methods, a comprehensive evaluation model of online teaching quality is developed based on data mining. The model is divided into two parts: an improved K-modes algorithm to evaluate English teachers' "teaching" and a feed-forward neural network to evaluate students' "learning". The improved K-modes algorithm cleansed, analyzed, and mined the teaching data, and improved the calculation of cosine similarity to establish a model for evaluating teachers' teaching status, the neural network model has more excellent index results, where the average error is 0.98, within 1, so the neural network model has a smaller error result. The combined model has a strong feasibility for the comprehensive evaluation of English online teaching quality.

Keywords—K-modes algorithm, feed-forward neural network, web-based teaching, comprehensive quality evaluation

1 Introduction

In recent years, with the outbreak of COVID-19, many schools are facing the suspension of classes. Therefore, network teaching has been rapidly developed and widely popularized. The sudden outbreak of the epidemic has led to a complete suspension of classes, which is not conducive to the development of students themselves, so that network teaching has been focused on. In order to ensure the quality of students' online learning, it is of great significance to evaluate the quality of online teaching. Zou and other researchers use data mining algorithms to establish personalized teaching recommendation models to help learners obtain more suitable learning methods [1]. Wang scholars use data technology to establish a framework for personalized recommendation of online tourism, which promotes the development of online tourism business [2]. Byun and other researchers, based on data mining technology, realized intelligent and personalized recommendation of online exhibitions [3]. The existing research

mainly applies data mining to personalized recommendation model, but neglects the importance of model quality. Therefore, based on the evaluation quality, the research establishes a comprehensive evaluation model of network teaching weight to ensure the quality of network teaching, which is of great significance to promote the teaching evaluation system of colleges and universities and the reform of school education and teaching.

2   Related work

Both K-Modes algorithm and neural network have the characteristics of simple operation and easy understanding. Research teams at home and abroad commonly use K-Modes algorithm and neural network in data mining and prediction to achieve the effects of early warning and evaluation. Kuo and other researchers used the improved K-Modes algorithm to cluster the auto parts suppliers, analyzed the characteristics of the suppliers according to the products provided by the suppliers, and selected the most suitable suppliers from among them. The experimental results showed that the improved K-Modes algorithm based on genetic algorithm has excellent clustering effect [4]. Li and other scholars cluster the classification matrix object data, propose MD fuzzy k-mode algorithm, introduce fuzzy factors, redefine the measure of anisotropy between the classification matrices, provide a heuristic algorithm for the clustering center, and carry out effective verification on five real data sets to clearly classify the customer's consumption habits and preferences [5]. Nguyen's research team proposed a clustering method combining genetic algorithm and K-Modes algorithm to solve the problem that the clustering algorithm terminates the algorithm when obtaining the local optimal solution and considers all attributes equally. They use frequency probability to measure data similarity. Experiments show that the algorithm can remove redundant features, improve clustering performance and shorten calculation time [6]. Ma and other scientific researchers established a BPNN prediction model to determine the heat transfer coefficient within the range of supercritical water pressure. By collecting 14 sets of experimental data, the model was trained. Finally, the experimental prediction results show that the modified model has high prediction accuracy, with the mean error of 0.179, the standard deviation of 4.129, and the root mean square error of 0.179 [7]. Zhang and other scholars proposed an improved RBFNN multi label deterioration prediction model, analyzed the soot and generated a control flow diagram, extracted the nodes and paths to form a multi label data set, and finally used the model for prediction. The experimental results show that the accuracy of the modified model in predicting 2–3 metamorphic relationships is about 80% [8]. Kenanolu and other researchers used the artificial neural network prediction method to analyze the performance and emissions of internal combustion engines. In order to develop the artificial neural network structure, LM learning algorithm was used to adjust the weight in the network structure. The experimental results show that the ANN model can predict the motor torque of the internal combustion engine with 95% accuracy, the motor power with 96.07% accuracy and the NOx emission with 92.35% accuracy [9].

Gamidi research team used ANN model to predict eutectic features of pure component physical properties, using 21 eutectic systems from 8 different model groups as

training data and 9 eutectic systems as prediction data, the experimental results show that the algorithm predicts eutectic features with an accuracy of 98.6% and captures the target physical properties affecting eutectic with high accuracy [10]. Wu scholars Based on online and offline tiered English teaching, the relationship between hybrid teaching and Civics was analyzed through a hybrid teaching method, and improvements to the teaching quality evaluation system were proposed based on the two-part perspective to build a more complete teaching quality evaluation system for online teaching [11]. Yang used a fuzzy clustering correlation mapping model to evaluate the effect of online teaching by investigating 150 The experimental results showed that teachers' knowledge structure and technology integration self-efficacy were positively correlated with students' learning quality and could be used as a relevant basis for online teaching quality evaluation [12]. Based on the construction of online teaching model and the analysis of literature survey, Wang et al. concluded that English online teaching improved students' speaking and reading ability by nearly 30% and students' sense of cooperation among themselves by 80%, which successfully improved students' learning quality. Due to the complexity of English learning, the current quality evaluation of English online teaching is inadequate or inaccurate, and the study aims to the purpose of this study is to ensure the quality of English online teaching [13–15].

3 Improved K-modes algorithm and neural network for English online teaching quality evaluation model

3.1 Evaluation data selection and data structure analysis of the improved K-modes

At present, there is a serious degree of inaccuracy in teaching quality evaluation in universities, which is mainly due to the fact that the data of teaching evaluation at the point of teaching system is too large, teaching managers pay less attention to this data, and the data is relatively difficult to process. By analyzing and processing the data of teaching evaluation, a reasonable and effective evaluation effect can be obtained. The study uses the evaluation data of Z school, based on the online evaluation of teaching, and the evaluation indexes are set up in four items, each of which has five choices. The evaluation indicators are evaluation of quality literacy, teaching attitude, teaching style and interactive links, and the evaluation level is divided into 1–5 levels, with the lowest level indicating failure and the highest level being excellent. After getting the evaluation teaching data, there are inevitably abnormalities in the data, so the data need to be cleaned and processed. The flow chart of abnormal data cleaning is shown in Figure 1 specifically.
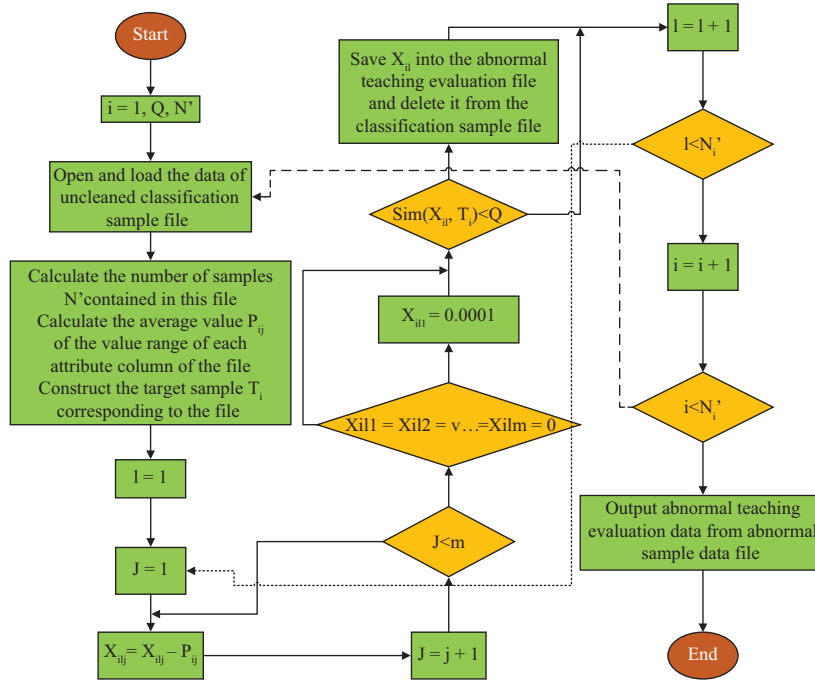
Fig. 1. Abnormal data cleaning flow chart

Data cleaning shall first classify the evaluation teaching data, and the classification can be done in the order of academic year (XN), semester (XQ), course selection number (XKKH), and employee number (ZGH), with a total of 1437 sample data from Z College. The classification was calculated by the improved cosine distance similarity formula, in order to eliminate the results of abnormal data triggered by the inconsistency and the actual situation. There are 5 dimensions of the sample data, and the average value of the sample data dimensions is expressed by P_i, so the value of P_i is 3. In order to avoid the situation that the denominator is 0 in the similarity cosine distance formula, 0.0001 is added to the evaluation value of the first dimension to ensure the accuracy of the similarity calculation as much as possible while solving the situation that the denominator cannot be 0. The improved formula for calculating the similarity and dissimilarity is shown in equation (1).

$$Sim(X,Y) = \sum_{i=1}^{m}((x_i - p_x) \times (y_i - p_y)) / (\sqrt{0.0001^2} \times \sqrt{\sum_{i=1}^{m}(y_i - p_y)^2}) = \frac{X \cdot Y}{\|X\| \times \|Y\|}$$

$$(1)$$

In Equation (1), *X*, *Y* denotes the sample data to be calculated, and *m* denotes the dimensionality of the sample data. The cosine distance similarity comparison requires

a target sample as the target with which another multidimensional data sample is compared. Based on the principle of central distance, the average of each dimension of the sample data is taken as the target sample and is denoted as *T*, whose expression is given in Equation (2).

$$T = (\frac{1}{N}\sum\nolimits_{i=1}^{N}x_{i1} - p_1, \frac{1}{N}\sum\nolimits_{i=1}^{N}x_{i2} - p_2, \cdots, \frac{1}{N}\sum\nolimits_{i=1}^{N}x_{im} - p_m) \tag{2}$$

In Eq. (2), *N* denotes the number of files of sample classification. The anomalous data elimination was performed on the 1437 classified sample data of Z College by equation (1) and equation (2). Figure 2 represents the graph of anomalous data processing results.
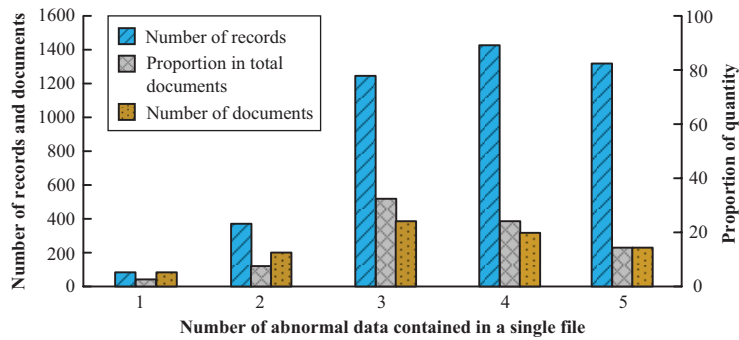

Fig. 2. Abnormal data processing result diagram

After anomalous data elimination, the data also need to be standardized. min-max standardization method is used in the study, which is a method of scaling and mapping the data to the [0,1] interval according to a certain ratio so that all evaluation data are comparable. min-max standardization mapping expression is shown in equation (3).

$$x'_{ij} = \frac{x_{ij} - \min_{1 \le j \le m} x_{ij}}{\max_{1 \le j \le m} x_{ij} - \min_{1 \le j \le m} x_{ij}} \tag{3}$$

In equation (3), $\min x_{ij}$ and $\max x_{ij}$ represent the extreme values of the sample data in the *j* column, respectively, x_{ij} represents the original value of the sample data without normalization, and x'_{ij} represents the normalized value of the sample data. The data in each column are averaged and normalized according to the normalization.

After eliminating and merging the abnormal data as described above, the K-modes algorithm was applied to evaluate the teaching of teachers in Z institutions. The number of clusters K of the algorithm is calculated using frequency similarity, and its calculation formula is shown specifically in Equation (4).

$$AVF(x_i) = \frac{1}{m} \sum\nolimits_{j=1}^{m} f(x_{ij})$$ (4)

In Equation (4), AVF denotes the similarity calculation on the basis of frequency, where x_i is the expression of the sample data and $f(x_{ij})$ denotes the number of occurrences of the sample data in the attribute j. The initial clustering centers randomly selected by the K-modes algorithm may have the problem of unstable clustering results, or even the availability of clustering results being affected by the selected outlier data points. Therefore, to avoid both cases, the sum of squared error (SSE) is used to determine the K initial clustering centers. Equation (5) is the specific expression of SSE.

$$SSE = \sum\nolimits_{l=1}^{k} \sum\nolimits_{x \in L_l} Dist(x, Z_l)^2$$ (5)

In Eq. (5), the number of clustering families is denoted by k, the clustering centers at l are denoted by Z_l, and the similarity between the sample data and the clustering centers is denoted by $Dist(x, Z_l)$. In the traditional K-modes algorithm, the problem of different value sizes and value relationships between the same attribute and different attributes and thus the poor clustering effect is not considered, so the study uses co-occurrence rate as an improvement measure of its algorithm. The co-occurrence rate is the probability of something happening when another thing has already happened. Equation (6) is a distance measure expression based on co-occurrence rate. $d_{ij}(x_{Ai}, y_{Ai})$ in Equation (6) indicates the distance between different values of sample data in one attribute and another attribute.

$$d(x, y) = \sum\nolimits_{i=1}^{m} \sum\nolimits_{j=1,\cdots,m,j \neq i} d_{ij}(x_{Ai}, y_{Ai})$$ (6)

3.2 Model building based on neural network in online learning prediction for college students

The neural network model is built by firstly building a good network structure, determining the network component layers and the distribution of neurons in each layer; secondly, using a better excitation function and making timely adjustments to the input weights of the network; and finally determining the number of times the model is trained to ensure that each weight is adjusted to the best, while preventing the model from over-fitting. The specific flow chart for building the model is shown in Figure 3.
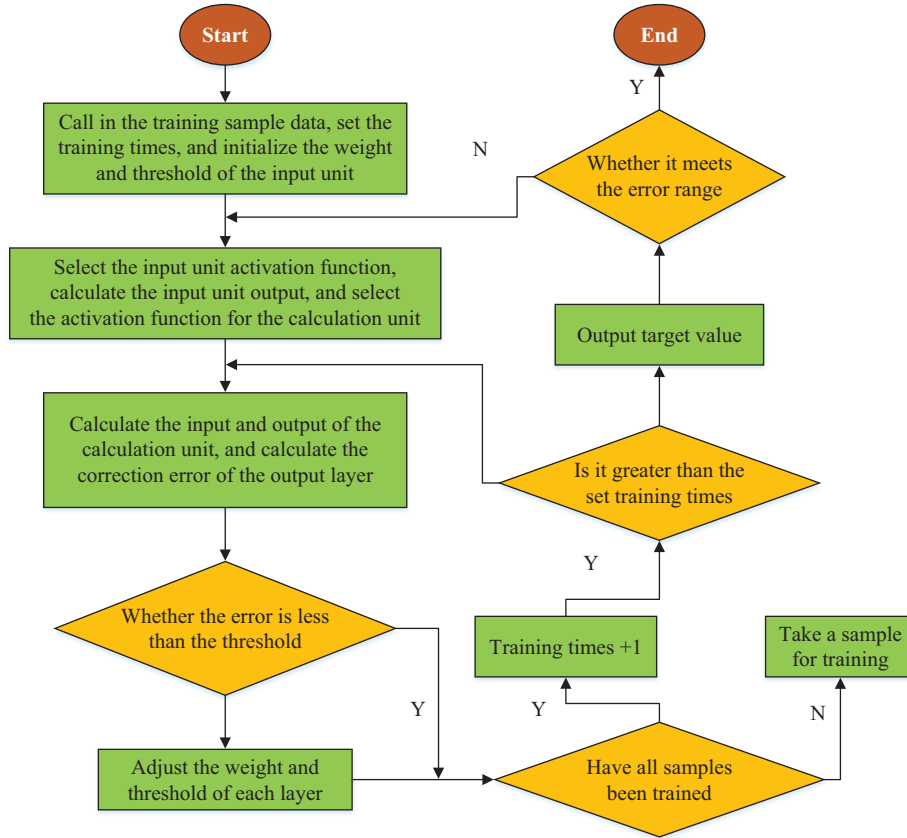
Fig. 3. Specific flow chart of building neural network prediction model

According to the process steps shown in Figure 3, the study is explained with a neural network with the computational unit of *m* layers. The first input of the computational unit is the output value of the input layer, and its expression is given in equation (7).

$$
\begin{cases}
v_k = u_k + b_k \\
u_k = \sum_{i=1}^{n} w_{kj} x_{ij}
\end{cases}
\tag{7}
$$

In equation (7), x_{ij} represents the features of the training samples, where *i* is the number of the training samples and *j* is the number of the sample features. The number of samples is 150 and the feature dimension is 6. w_{kj} denotes the sample feature weights of input units in the input layer, and *k* denotes the number of input units in the input layer. u_k is the expression of each input term and the linear combination of the corresponding weights, b_k denotes the threshold value of the neuron, and $\varphi(v_k)$ denotes the excitation function of each neuron? The second layer input value is the output value of the first layer, $h_k^2 = H_k^1$. The second layer output expression is specified in Equation (8).

$$H_k^2 = \varphi(w_{kj}^2 h_k^2 + b_k^2) \tag{8}$$

Therefore, the input to the *m* level is the output of the previous level of the calculation cell, $h_k^m = H_k^{m-1}$, and its output expression is shown in Equation (9).

$$H = \varphi\left(\sum_{j=1}^{N}(w_j^m h_j^m + b_j^m)\right) \tag{9}$$

The number of layers of the neural network, the number of neurons on each neural network layer, the type of excitation function and the number of training sessions can be adjusted from time to time to obtain the corresponding weights and thresholds for each input term of the model during the learning and training process. The study uses a feed-forward neural network, which can receive several different input units at the same time, and has only one output unit because there is no feedback effect in it. The model can make predictions for student learning effectiveness and academic performance, and occupies an important place in the evaluation of teaching quality.

The model is solved by computer and corrected by multiple solutions. The network is based on a 6-layer architecture with 64 neurons in the first and last layers, 128 neurons in the second and fifth layers, and 512 neurons in the middle third and fourth layers. The excitation functions chosen for the neural network are ReLu function and Sigmoid function. The specific expressions are shown in Equation (10).

$$\begin{cases} \mathrm{Re}\,lu(x) = f(x) = \begin{cases} 0, x \le 0 \\ x, x > 0 \end{cases} \\ Sigmoid(x) = f(x) = \dfrac{1}{1 - e^{-x}} \end{cases} \tag{10}$$

The ReLu function is a unilateral inhibition function, which ensures that the neurons in the model have the nature of sparse activation, which in turn enables the algorithmic model to more efficiently mine data-related features as well as to obtain a better fit to the training data. The Sigmoid function has the property of continuous uninterrupted and derivable in the range of the definition domain, while the S-shaped function is centrosymmetric at $x = 0.5$. Such an excitation function can be used at a smaller granularity in describing the uncertainty of the decision, while the enhancement can improve the convergence speed of the threshold neighborhood and achieve the effect of reducing the number of training times. The experiments were conducted from 5000 training times with each increasing step size of 500 times, and the experimental results showed that the output of the model tends to a stable value after 15000 training times and no more significant changes occur. In the above neural network model, the training was started after determining the model parameters, and the images of its training output results compared with the real results are shown in Figure 4.
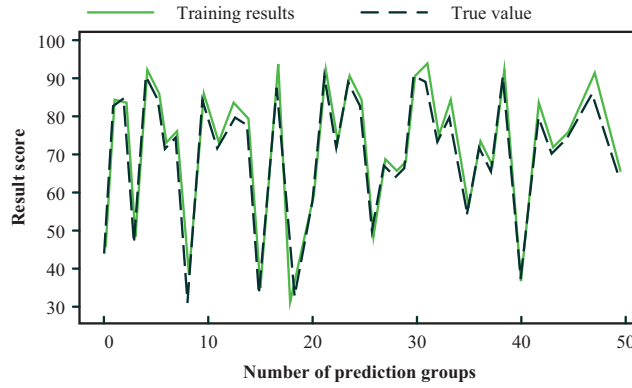
Fig. 4. Comparison image between training output result and real result

The trend of the image shows that the linearity of the training output results overlaps highly with the real results, indicating that the model is feasible. The error of the model output value is judged, and the error of the model output result must be less than the maximum error set by the condition. The maximum error value of the model is shown in Equation (11).

$$\begin{cases} \max(y_i' - y_i) = 5.5 \\ \min(y_i' - y_i) = -5.4 \end{cases} \tag{11}$$

From equation (11), the maximum error of the model is 5.5, while the sum of the errors of the real values of the model and the output results is calculated as shown in equation (12).

$$\begin{cases} \dfrac{\sum_{i=1}^{102}(y_i' - y_i)}{102} \approx 1.26 \\ \dfrac{\sum_{i=1}^{48}(y_i' - y_i)}{48} \approx -1.73 \end{cases} \tag{12}$$

The absolute value of the difference between the actual value and the output value of its corresponding training data is expressed by C_i, and its specific expression is given in Equation (13).

$$C_i = \left| y_i' - y_i \right| \tag{13}$$

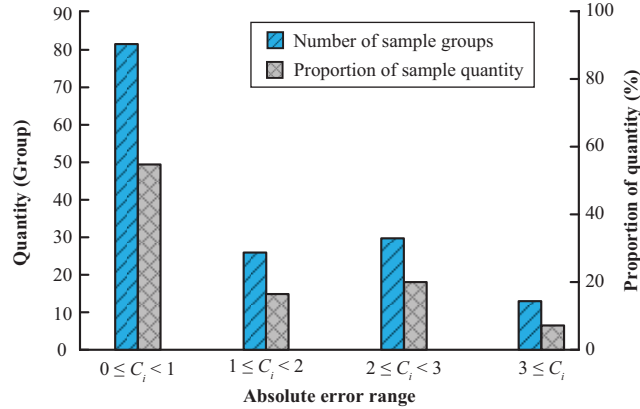The error distribution of the 150 training samples calculated by Equation (13) is shown in Figure 5.

Fig. 5. Error distribution of training samples

From Figure 5, it is known that the absolute error of the model is more in the range of [0,1], so the model has a high learning ability. The mean value of the model is calculated as in equation (14).

$$\begin{cases} \dfrac{\sum_{i=1}^{150} y_i'}{150} = 70.62 \\[4mm] \dfrac{\sum_{i=1}^{150} y_i}{150} = 70.32 \end{cases} \tag{14}$$

The mean value of the actual value in equation (14) is 70.62 and the mean error of the model training output value is 70.32, whose mean error is at 0.3. Equation (15) represents the result of the variance of the actual value and the variance of the training output value.

$$\begin{cases} \dfrac{\sum_{i=1}^{150}(y_i' - \dfrac{\sum_{i=1}^{150} y_i'}{150})^2}{150} = 329.09 \\[6mm] \dfrac{\sum_{i=1}^{150}(y_i - \dfrac{\sum_{i=1}^{150} y_i}{150})^2}{150} = 321.43 \end{cases} \tag{15}$$

The output value variance in Equation (15) is smaller than the actual value variance, indicating that the training output value fluctuates less from the average value, and this situation may occur because the number of data with lower end-of-period scores in the training set is smaller, making machine learning training more difficult and leading to poorer learning results. Therefore, in order to guarantee the machine learning effect, a sufficient number of training samples need to be satisfied.

4 Experimental design and analysis

4.1 Analysis of the results of the improved K-modes algorithm for evaluating the quality of online English teaching

Teaching evaluation is evaluated from both teacher and student perspectives, and the quality of teaching reflects the quality of teaching evaluation of the whole school. The overall teaching evaluation results of students in College Z are stable, although they vary slightly from semester to semester.

Table 1. Data sheet of teaching evaluation of students in a semester of Z College

Evaluation Grade	Excellent	Good	Secondary	Pass	Fail
Basic quality	324	282	171	0	0
Teaching attitude	317	285	171	15	0
Teaching ability	53	366	317	42	0
Extracurricular links	6	77	489	180	27
Total	700	1010	1148	237	27
Proportion	0.2242	0.3235	0.3677	0.0759	0.0086

The presence of three of the four evaluation indicators rated good or above was considered as category 1 teachers; the presence of two to three of the four evaluation indicators rated good and at least one rated moderate was considered as category 2 teachers; and the presence of all four indicators rated moderate or below was considered as category 3 teachers. The clustering results of the K-modes algorithm are shown in Figure 6, which reflects the distribution of the three categories of English teachers by semester.
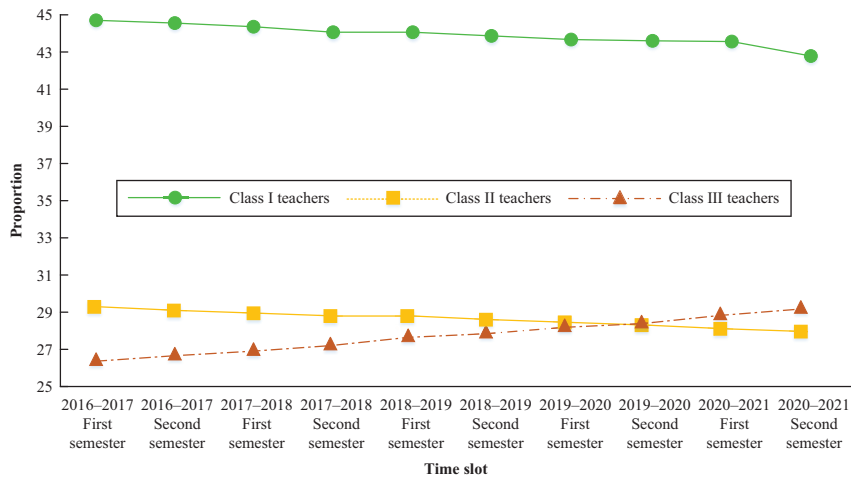

Fig. 6. Distribution of three types of English teachers in one semester

From the clustering results in Figure 6, it can be analyzed that in the two evaluation indexes of basic quality and teaching attitude, the percentage of good and excellent evaluations reached a total of about 77%, and the percentage of those evaluated as moderate, passing and failing was about 23%. In the index of teaching ability, 54% of the students rated good and excellent, while the remaining 46% rated moderate or below, indicating that nearly half of the students were dissatisfied with the teaching ability of the teachers. In terms of the interaction index, less than 10.7% of the students rated excellent and good, while 64% rated moderate, 20% rated pass, and 5% rated fail. The clustering from K-modes algorithm can clearly identify the problematic areas in teaching.

Using the most recent semester of assessment data in institution Z as data, six dimensions of the three K-modes algorithms are compared, which include clustering centroids, number of clustered samples, number of resultant samples, sum of least errors squared, correctness rate, and recall rate. Figure 7 shows the results of the three metrics of correctness, recall, and minimum error squared for the three different clustering algorithms.
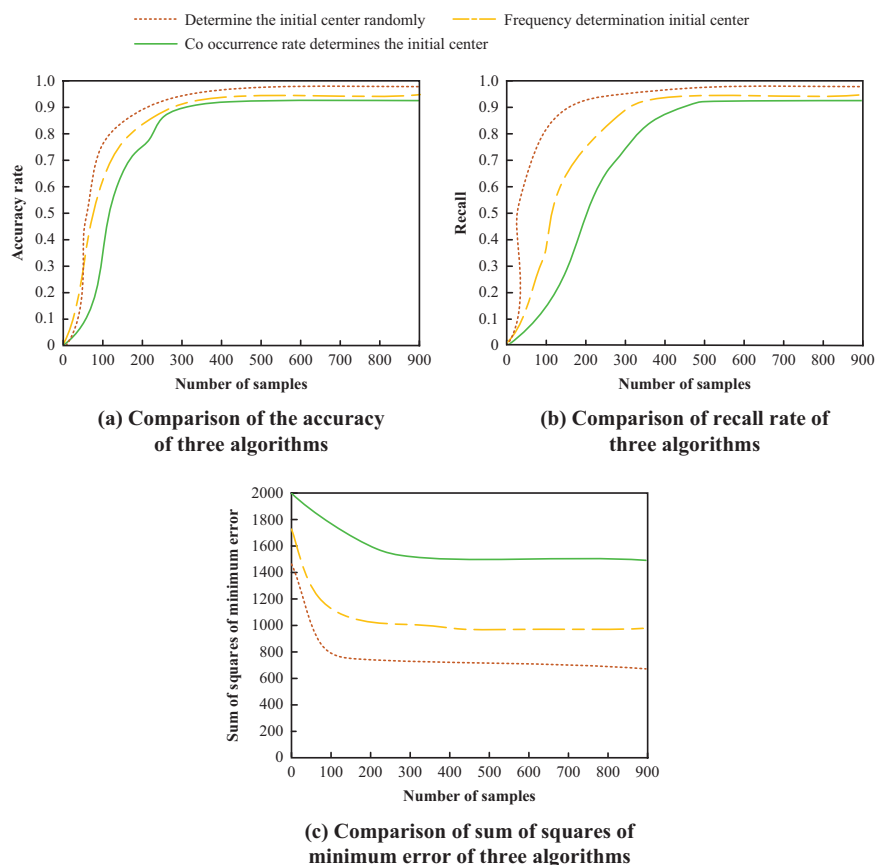

(a) Comparison of the accuracy of three algorithms

(b) Comparison of recall rate of three algorithms

(c) Comparison of sum of squares of minimum error of three algorithms

Fig. 7. Results of three different clustering algorithms: accuracy, calling rate and least square error

Comparing the indicators of the three different algorithms in Figure 7, the minimum error sum of squares of the improved K-modes algorithm is 714 with the same number of samples, and its error is significantly lower than the other two algorithms; the correctness rates of the randomized initial center determination algorithm and the SEE determination initial center AVF calculation distance algorithm are 0.936140 and 0.947674, respectively, while the SEE determination initial center co-occurrence value calculation distance algorithm, i.e., the improved K-modes algorithm, has a correct rate of 0.984583, which is also higher than the previous two algorithms; the improved K-modes algorithm also performs best in the comparison of recall rates, with a recall rate of 0.981612.

The purpose of the clustering algorithm is to cluster the targets as close as possible to the cluster centers, and the distance between the cluster centers is as far as possible. Since the three algorithms have different ways to measure the distance, we cannot simply use the length of the distance to compare the advantages and disadvantages of clustering, so the study uses the average ratio of the distance as the judgment criterion, and the larger the ratio value, the worse the clustering effect, and vice versa, the smaller the ratio value, the better the clustering effect. The comparison table of the average distances obtained by the three algorithms after the ratio value calculation is shown in Table 2.

Table 2. Comparison table of average distance obtained by three algorithms after ratio value calculation

/	Random Determination of Initial Center Algorithm			Frequency Determination Initial Center			Co-Occurrence Rate Determines the Initial Center		
	Z1	Z2	Z3	Z1	Z2	Z3	Z1	Z2	Z3
Z1	0.735	0.847	0.883	0.460	0.753	0.802	0.241	0.708	0.721
Z2	/	0.578	0.821	/	0.379	0.757	/	0.256	0.709
Z3	/	/	0.468	/	/	0.357	/	/	0.226

The mean distance comparison table in Table 3 shows that the minimum ratio values at the same clustering centers are the improved K-modes algorithm, and the minimum ratio values between different clustering centers are also the improved K-modes algorithm. Therefore, the clustering results obtained by this algorithm have stronger correlation, and the algorithm is more accurate as well as effective in solving the student evaluation problem.

4.2 Analysis of neural network models in predicting the results of online learning for college students

Two different sets of data were used for the prediction analysis to reflect the feasibility and effectiveness of neural network models in teaching quality evaluation. One group is the data of students with the same course and the same instructor, and the other group is the data of students with different courses and different instructors. The same

data processing method was used for both groups, and the comparison between the two groups and the real data was obtained, as shown in Figure 8.
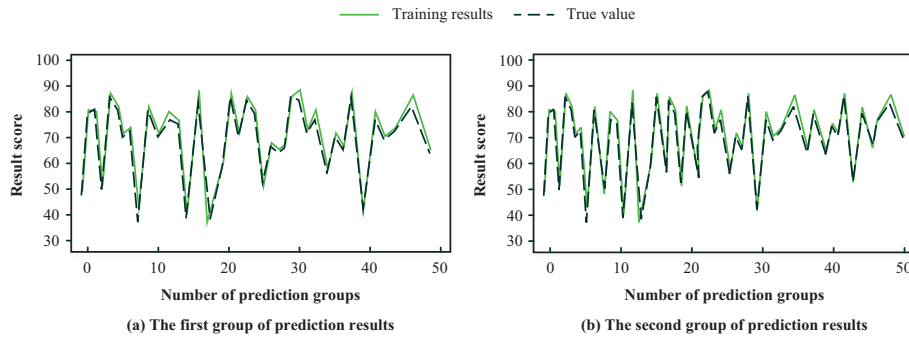

Fig. 8. Prediction effect of two groups of data in neural network model

From Figure 8a and b, it can be seen that the neural network model used for predicting the learning effect of the two groups of courses has excellent results, and the prediction results of the two groups are highly overlapping with the prediction results of the true value, which indicates that the model has high accuracy. The maximum error is obtained by equation (11), and the maximum error of the first group is 6.15 and the maximum negative error is –5.91; the positive and negative errors of the second group are 6.68 and –5.97, respectively. the cumulative positive and negative average errors are obtained by equation (12), and the results of the first group are 1.98 and –2.37, respectively; the results of the second group are the absolute errors of the two groups can be calculated by equation (13), and the distribution of the prediction errors of the two groups is shown in Figure 9.
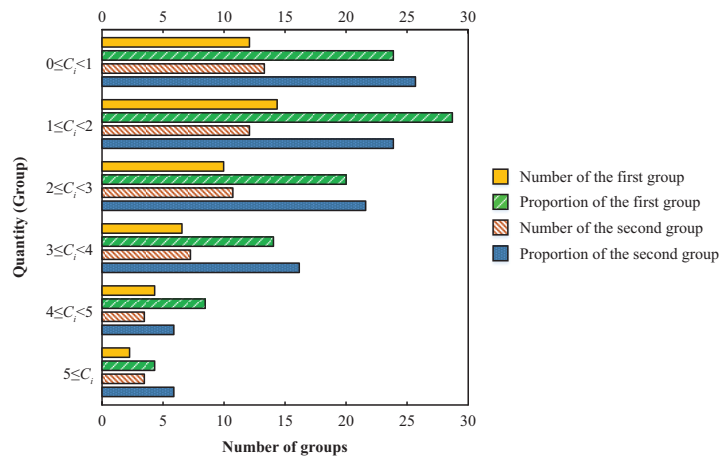

Fig. 9. Distribution of prediction errors of the two groups

In the results of the absolute error distribution, it can be known that the absolute error of the two groups of predictions is heavier in the range [0,2]. The average value of the data can be calculated by equation (14), the actual average value of the first group is 73.26 and the predicted average value is 74.24, the average error value is 0.98; the actual average value of the second group is 74.17 and the predicted average value is 74.88, the average error is in 0.71. After the analysis and comparison of the four indicators of maximum error, average error, absolute error and average value, all indicators There is a small degree of difference between them, but in the mean index, the prediction result of the first group is slightly worse than that of the second group, indicating that the model has stronger adaptability.

To ensure the superiority of feedforward neural networks in predicting student learning outcomes, the study was compared by incorporating a model established by traditional regression analysis. The linear regression model was obtained as in equation (16).

$$y = a_0 + a_1 x_1 + \cdots + a_6 x_6 \tag{16}$$

Where *y* represents the predicted data for final grades, *x* represents the input training data, and *a* represents the linear regression coefficients. The values of the coefficients were obtained by performing linear regression analysis through Excel. The results of the linear regression model for the first set of predictions are shown in Figure 10.
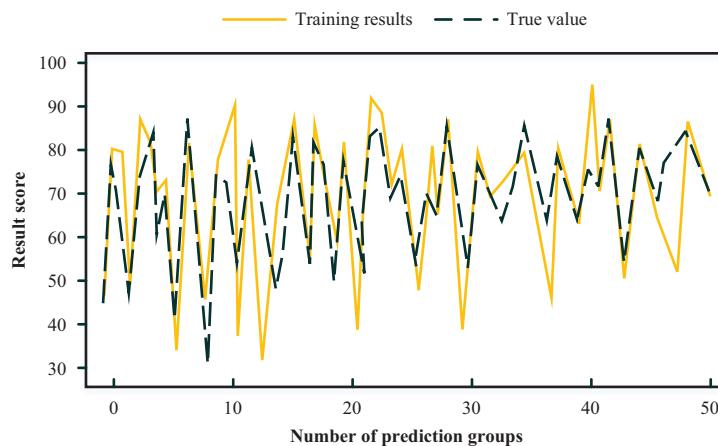

Fig. 10. Data results obtained from the prediction of the first group by linear regression model

The linear trend results in Figure 10 show that there is a large error between the predicted and true values, and the linear trend plot overlap is significantly lower than the accuracy of the neural network model. The study used the same index to test the performance of the regression analysis model, and the positive and negative maximum error values of the regression analysis model were divided into 23.57 as well as −14.30; its positive and negative cumulative error values were 6.95 as well as −3.65, respectively.

the absolute error distribution of linear regression analysis compared with the neural network is shown in Figure 11. The error between the actual mean and the predicted mean of the linear regression analysis model is 3.17.
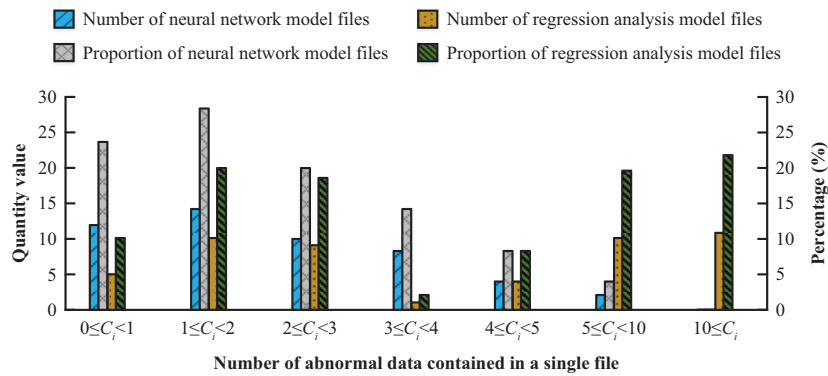

Fig. 11. Comparison of absolute error distribution between linear regression analysis and neural network

After the comparative analysis of the results of the four indexes, the performance of the neural network model was better than the performance of the regression analysis model, thus verifying that the neural network model plays a great role in learning evaluation in English online teaching quality evaluation. Through teaching evaluation data and online learning data, focusing on the evaluation accuracy of the model, it provides valuable reference for the construction of teaching management system and teaching methods in Colleges and universities.

5   Conclusions

In recent years, the rapid development of information technology has brought different degrees of changes to various industries. Meanwhile, meeting the new crown epidemic, the information construction of education industry is becoming more and more rapid. In order to analyze the current situation of online teaching, the study establishes a comprehensive evaluation model of English online teaching quality based on data mining technology [16]. By analyzing the data through experiments, the improved K-modes algorithm can analyze the teachers' teaching methods and clearly find out where the teaching problems lie from each index of teaching evaluation, which provides a strong basis for teaching quality evaluation [17]. The improved K-modes algorithm has the best performance in terms of correctness, recall and minimum error sum of squares compared with the other two algorithms, with a correctness rate of 0.984583, recall rate of 0.981612 and minimum error sum of squares of 714. The feed-forward neural network starts from predicting students' learning and evaluating students' learning. "learning" evaluation, and the prediction accuracy of the feedforward neural network in the two sets of teaching experiments has high overlap with the actual situation, indicating

the feasibility of the model. The neural network model has a maximum error integer value of 6.15 and a maximum error negative value of –5.91; the average errors of positive and negative values are 1.98 and –2.37, respectively; the absolute errors are heavily weighted between [0, 2]; the average error is 0.98, which is within 1. The neural network model has superior results for all four indicators by comparing with the regression analysis model [18].

There are still shortcomings in the research. The teaching evaluation index is not combined with daily teaching, and the behavior data predicted by online learning is not perfect, resulting in the model has not reached the highest evaluation accuracy. Under the current environment, and the government's reform and innovation on education, emphasizing the importance of online teaching and independent learning, which is conducive to the cultivation of talent quality, the following aspects will be further studied [19]. First, in the aspect of teaching operation, the combination of daily teaching and setting up a more detailed evaluation system will form more influential results and provide more targeted innovation basis for teaching reform. Second, collect more comprehensive data, further optimize the model algorithm, and provide more real learning conditions.

6 References

[1] Kim, "A weighted k-modes clustering using new weighting method based on within-cluster and between-cluster impurity measures", *Journal of Intelligent & Fuzzy Systems: Applications In Engineering and Technology,* vol. 32, pp. 979–990, 2017. https://doi.org/10.3233/JIFS-16157

[2] Y. Xiao, C. Huang, J. Huang, et al, "Optimal mathematical programming and variable neighborhood search for k-modes categorical data clustering", *Pattern Recognition*, vol. 90, pp. 183–195, 2019. https://doi.org/10.1016/j.patcog.2019.01.042

[3] A. Aggarwal, V. Vinayak, S. Bag, et al, "Predicting the DNA conductance using a deep feedforward neural network model", *Journal of Chemical Information and Modeling*, vol. 61, pp. 106–114, 2021. https://doi.org/10.1021/acs.jcim.0c01072

[4] R. J. Kuo, Y. Potti, F. E. Zulvia, "Application of metaheuristic based fuzzy k-modes algorithm to supplier clustering", *Computers & Industrial Engineering*, vol. 120, pp. 298–307, 2018. https://doi.org/10.1016/j.cie.2018.04.050

[5] S. Li, M. Zhang, F. Cao, "A MD fuzzy k-modes algorithm for clustering categorical matrix-object data", *Journal of Computer Research and Development*, vol. 56, pp. 1325–1337, 2019.

[6] R. J. Kuo, T. Nguyen, "Genetic intuitionistic weighted fuzzy k-modes algorithm for categorical data", *Neurocomputing*, vol. 330, pp. 116–126, 2019. https://doi.org/10.1016/j.neucom.2018.11.016

[7] D. Ma, T. Zhou, J. Chen, et al, "Supercritical water heat transfer coefficient prediction analysis based on BP neural network". *Nuclear Engineering and Design*, vol. 320, pp. 400–408, 2017. https://doi.org/10.1016/j.nucengdes.2017.06.013

[8] P. Zhang, X. Zhou, P. Pelliccione, et al, "RBF-MLMR: A multi-label metamorphic relation prediction approach using RBF neural network", *IEEE Access*, vol. 5, pp. 21791–21805, 2017. https://doi.org/10.1109/ACCESS.2017.2758790

[9] R. Kenanolu, M. K. Baltacolu, M. H. Demir, et al, "Performance & emission analysis of HHO enriched dual-fueled diesel engine with artificial neural network prediction approaches", *International Journal of Hydrogen Energy*, vol. 45 pp. 26357–26369, 2020. https://doi.org/10.1016/j.ijhydene.2020.02.108

[10] R. K. Gamidi, C. R. Ke, "Analysis and artificial neural network prediction of melting properties and ideal mole fraction solubility of cocrystals", *Crystal Growth and Design*, vol. 20, pp. 5745–5759, 2020. https://doi.org/10.1021/acs.cgd.0c00182

[11] X. Wu, "Research on the reform of ideological and political teaching evaluation method of college English course based on "online and offline" teaching", *Journal of Higher Education Research*, vol. 3 pp. 87–90, 2022. https://doi.org/10.32629/jher.v3i1.641

[12] L. Yang, "Fuzzy cluster correlation mapping for online evaluation of teaching efficacy towards IoT study", *Cognitive Systems Research*, vol. 52, pp. 365–370, 2018. https://doi.org/10.1016/j.cogsys.2018.07.025

[13] L. Wang, Z. Xu, "An English learning system based on mobile edge computing constructs a wireless distance teaching environment", *Mobile Information Systems*, vol. 1, pp. 1–9, 2021. https://doi.org/10.1155/2021/2718859

[14] H. H. Nguyen, "Privacy-preserving mechanisms for k-modes clustering". *Computers & Security*, vol. 78, pp. 60–75, 2018. https://doi.org/10.1016/j.cose.2018.06.003

[15] D. B. M. Maciel, G. J. A. Amaral, R. M. C. R. De Souza, et al. "Multivariate fuzzy k-modes algorithm", *Pattern Analysis & Applications*, vol. 20, pp. 59–71, 2017. https://doi.org/10.1007/s10044-015-0465-3

[16] B. Li, X. Zhuang, "Multiscale computation on feedforward neural network and recurrent neural network". *Frontiers of Structural and Civil Engineering*, vol. 14, no. 6, pp. 1285–1298, 2020. https://doi.org/10.1007/s11709-020-0691-7

[17] S. Masood, M. N. Doja, P. Chandra "Architectural parameter-independent network initialization scheme for sigmoidal feedforward ANNs", *Arabian Journal for Science and Engineering. section* vol. 45, pp. 2901–2913, 2020. https://doi.org/10.1007/s13369-019-04200-2

[18] L. Wang, Z. Chen, G. Yang, "An interval uncertainty analysis method for structural response bounds using feedforward neural network differentiation", *Applied Mathematical Modelling*, vol. 82, pp. 449–468, 2020. https://doi.org/10.1016/j.apm.2020.01.059

[19] S. Cheng, Y. Wu, Y. Li, et al. "TWD-SFNN: Three-way decisions with a single hidden layer feedforward neural network", *Information Sciences*, vol. 579, pp.15–32, 2021. https://doi.org/10.1016/j.ins.2021.07.091

7   Author

Zhe Zhang, Master of Education, is presently working as an Associate Professor and the Deputy Dean of Department of Applied Foreign Languages, Henan Industry and Trade Vocational College. She has published articles in more than 20 journals. Her areas of interest include English teaching, machine learning and vocational education.