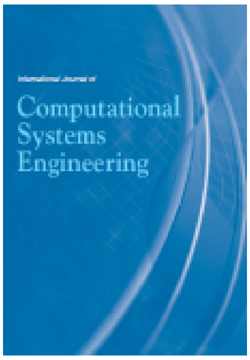
INDERSCIENCE PUBLISHERS
Linking academia, business and industry through research



International Journal of Computational Systems Engineering

ISSN online: 2046-3405 - ISSN print: 2046-3391
https://www.inderscience.com/ijcsyse

Research on the integration of English online teaching resources based on improved association rule algorithm

Hao Liu

DOI: 10.1504/IJCSYSE.2023.10054701

Article History:

Received:	05 August 2022
Last revised:	12 October 2022
Accepted:	14 January 2023
Published online:	19 March 2024

Research on the integration of English online teaching resources based on improved association rule algorithm

Hao Liu

Public Basic Teaching Department,
Henan Vocational University of Science and Technology,
Zhoukou, 466000, China
Email: haoliuhvust@163.com

Abstract: With a series of achievements of intelligent algorithms in various fields, people began to try to apply intelligent algorithms to information-based teaching and learning, and use relevant data analysis techniques to improve existing teaching models. The study integrates RBF neural networks with association rule algorithms, and then constructs an English web-based teaching prediction model. Using a crawler data collection tool, the English test scores of a university were selected to test the model. The results show that the improved model has shorter running time and can be iterated to a stable state faster. The model was used for the prediction of actual grades, and the average accuracy of the model was obtained as 95.7%. Comparing the relative error values of the prediction model with different influencing factors, we found that the average relative error was only 0.041. The improved model can achieve better results when used for English score prediction.

Keywords: prediction models; neural networks; association rules; relative error values.

Reference to this paper should be made as follows: Liu, H. (2024) 'Research on the integration of English online teaching resources based on improved association rule algorithm', *Int. J. Computational Systems Engineering*, Vol. 8, Nos. 1/2, pp.48–55.

Biographical notes: Hao Liu is a University Lecturer. In July 2009, she graduated from Zhoukou Normal University with a Bachelor's in Foreign Language. Currently, she works in Henan Vocational University of Science and Technology. Her research direction is English education and teaching, and she has published one Peking University core paper and two CN papers.

1 Introduction

The rapid updating and iteration of the online education model has generated a large amount of student data, including the length of students' online learning, examination results of various subjects, online interaction times, correct completion rate of exercises after class, etc. However, most of these data are used for simple query and statistics, and have not been further used (Dey and Mukhopadhyay, 2019). How to use data mining technology to reasonably conduct data mining and help schools extract useful data information to improve teaching mode is the focus of data research in universities (Gagnon et al., 2019). Association rule algorithm, as one of the classic algorithms of data mining, is often used in teaching evaluation. It comprehensively evaluates the teaching quality by mining the influencing factors associated with students' grades (Fautley and Daubney, 2019).

The Aprion algorithm among common association rule algorithms has the advantages of simple theoretical derivation and easy implementation, but at the same time, due to too many scans, it is easy to generate a large number of candidate item sets, thus increasing the operation time (Chen et al., 2019). RBF neural network is a feedforward neural network with excellent performance. Traditional BP neural network is prone to fall into local optimisation in the late training period, while RBF will train and adjust the weight of some factors that have an impact on local output (Bin et al., 2019). Therefore, the training time will be shorter, the calculation accuracy will be higher, and the prediction performance and function fitting performance will be better than the traditional BP neural network. In this paper, according to the mining association rule parameters, as the main algorithm to improve the prediction and RBF neural network algorithm data centre selection basis. The previous neural network algorithm usually adopts random initialisation as the data centre selection method. The data centre initialisation method proposed in this paper is probabilistic random initialisation. The accurate prediction of students' English test scores is one of the prerequisites for the integration of online education resources. In this study, data mining is carried out by improving association rules. Aiming at the existing data prediction model, an association rule algorithm design integrating RBF neural network is proposed, and then the model is used in English online teaching prediction.

2 Related work

In the context of big data era, how to find a suitable data mining method in different practical problems, from which the potential value of data can be mined to help management make reasonable decisions and reduce risks has become the research goal of relevant researchers at home and abroad. Tuncalyaman et al. applied algorithms such as fuzzy clustering and regression tree to the analysis of traffic accident severity through data mining. Different influencing factors were analysed and studied (Tuncalyaman et al., 2021). Zhu et al. applied methods such as association rule analysis and cluster analysis to statistically analyse the prescription of conforming Chinese patent medicines, and used data mining techniques to analyse the characteristics of jujube Chinese patent medicines, which provided guidance for modern clinical application and development of jujube (Zhu et al., 2021). For the household electricity load prediction problem, Baker (2021) combined the use of data mining methods such as K-means clustering and K-nearest neighbours (K-NN) classification to achieve experimental research results by building relevant algorithmic models and verified the good performance of the method. Yue et al. (2021) applied support vector machine, logistic regression and information fusion for big data mining so as to conduct an in-depth study of corporate financial risk, and the results showed that the method can classify corporate financial risk well and with high accuracy. Wang et al. (2019) applied association rule mining techniques to the differential evolutionary algorithm parameter adaptive strategy, which uses the evolutionary information generated during the process to enhance exploration by adjusting the control parameters. For the problem of improving the quality and yield of steel, Han et al. (2020) proposed a temporal association rule mining and updating method, which experimentally proved its good performance. Nohuddin et al. (2021) used association rule mining for the study of school children's lifestyle and academic performance to infer the correlation between the two in order to provide an improvement of school children's academic performance provide help. Soufi and Ferdousi (2021) applied data mining techniques to study and analyse the correlation between obesity and breast cancer and experimentally proved the correlation.

With the rapid development of information technology, online teaching mode has gradually become the mainstream. Wang (2022) proposed a new hybrid teaching design paradigm based on the hybrid teaching studied by previous authors, through which the teaching design is used to make reasonable suggestions to teachers to obtain better teaching results. Bose and Gao (2022) explored the reading section in Indian English textbooks to obtain questions about their cultural representations, and the results showed that people's characteristics in terms of race, gender, and ethnicity were related to cultural features in the content of the textbook. Elgllab and Shehata (2019) used a quasi-experiment to explore the methods followed by scholars when searching for information using the minor language and the English language, and the study found that few studies have attempted to explore the use of differences in information search processes when using different languages, and according to the results shown in the data, there was no difference in information search behaviour between scholars from minor languages and English language backgrounds. Karjanto and Simon (2019) investigated the implementation of a flipped classroom for a univariate calculus course in a traditional Confucian cultural context through a one-way ANOVA, and the results showed that the complete significant differences in scores between flipped and single-topic flipping, while the practical significance of the effect size was small.

Through the above-mentioned studies, it is found that data mining techniques have been effectively utilised in many fields, and with the continuous development of the internet era, online teaching will also be commonly popular (Fautley and Daubney, 2019). For the current problem of massive online teaching resources stacking, a suitable algorithm has not been found to mine and collect teaching resources (Chen et al., 2019). Therefore, the study proposes the design of association rule algorithm incorporating RBF neural network, based on which, an English web-based teaching prediction model based on improved association rule algorithm is constructed, and the performance of this model is tested by comparing and analysing with traditional RBF neural network and BP neural network (Bin et al., 2019).

3 Construction of English online teaching prediction model based on improved association rule algorithm

3.1 *Design of association rule algorithm incorporating RBF neural network*

With the rapid development of computer network technology, how to efficiently and safely mine valuable information in big data has become a research goal for more and more scholars (Pollard and Olizko, 2019). Association rule algorithm, as one of the classical algorithms of data mining, is often applied in teaching evaluation, by mining the influencing factors associated with student performance so as to make a comprehensive evaluation of teaching quality. For the existing data prediction model, the study firstly proposed association rule algorithm incorporated with RBF neural network, and then used the model in English online teaching prediction to provide new ideas for English online teaching in the epidemic situation (Lu et al., 2021).

Association rules are rules whose support degree and trust degree respectively satisfy the user's given threshold. The algorithm is generally divided into two steps. High-frequency objects are first discovered from the dataset, and then association rules are discovered from these high-frequency objects. Its high-frequency object mining process is as follows (1).

$$Support(\alpha \Rightarrow \beta) = Freq(\alpha \cap \beta)/N \tag{1}$$

In equation (1), $\{\alpha, \beta\}$ denotes two items in the item group, *Freq* denotes the number of items in the item group, and N is the number of all items in the item group. An item group of k that meets the minimum support is called high-frequency k item group, and according to the algorithm rules, new high-frequency k item groups will be generated from this item group until no longer high-frequency item groups can be found.

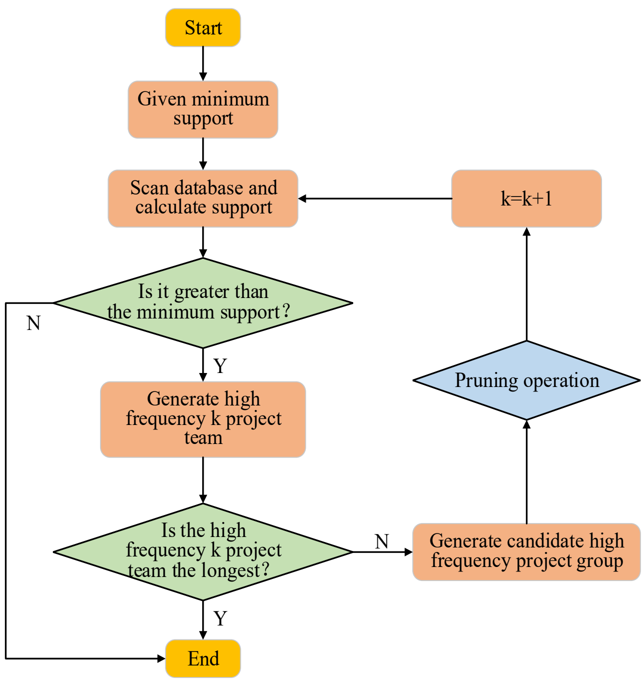
$$Confidence(\alpha \Rightarrow \beta) = Freq(\alpha \cap \beta) / Freq\,(\alpha) \tag{2}$$

Equation (2) is the confidence level calculation formula. The association rule is derived from the HFPG, and a rule is called an association rule if it satisfies the threshold of minimum confidence in the HFPG generation process. The parameter that measures the association rule is the lifting degree, which is calculated as in equation (3).

$$\begin{aligned} &Lift(\alpha \Rightarrow \beta) \\ &= Support(\alpha \Rightarrow \beta) / Support(\alpha) * Support(\beta) \end{aligned} \tag{3}$$

Equation (3) is the lifting degree calculation formula, which is used to determine whether the rule is valuable or not. If the calculation result is greater than 1, it is valid, and vice versa. The flowchart of common association rules is shown in Figure 1.

Figure 1 Flowchart of common association rules (see online version for colours)

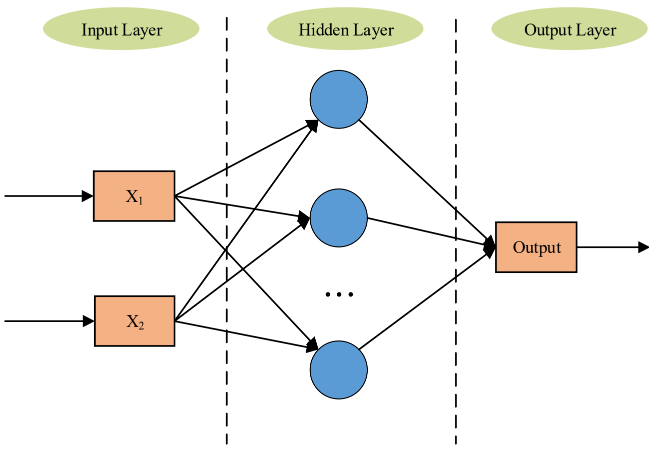


The RBF neural network is improved on the basis of the BP neural network. The BP neural network is prone to fall into the local optimum in the later stage of training, and the weights of some factors that affect the local output of the RBF will be trained and adjusted. Therefore, the training time is shorter and the computational accuracy is higher compared to the traditional BP neural network, and it outperforms the traditional BP neural network in terms of prediction performance and function fit performance. Its calculation formula is shown in equation (4).

$$\varphi = (\|X - X_i\|) \tag{4}$$

In equation (4), X denotes the input of the neural network, X_i denotes the position coordinates of the input, and φ denotes the driving function. The structure of RBF is shown in Figure 2.

Figure 2 Structure of RBF neural network (see online version for colours)

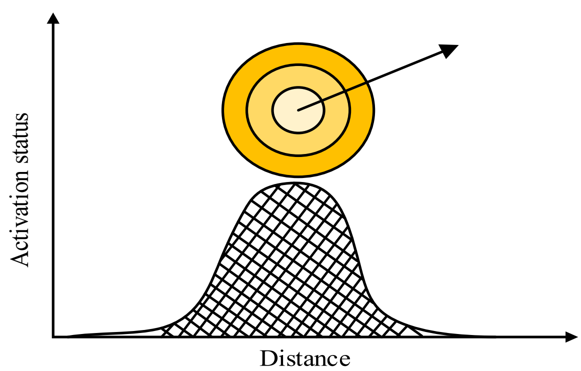


From Figure 2, it can be seen that RBF neural network is a forward three layer neural network, the first layer is the input layer, the second layer is the hidden layer, and the last layer is the output layer. The driving function of RBF neural network is generally a Gaussian function, and its expression is as in equation (5).

$$\varphi = express\left(-d^2 / 2\sigma^2\right) \tag{5}$$

σ in equation (5) represents the extended function of the basis function, and when the σ is small, the driving function is prone to overfitting. The Gaussian function calculates the weights by the distance between the input value and the centre point of the function, and the relationship between the distance of the centre point and the activation state is shown in Figure 3.

Figure 3 Relationship between centroid distance and activation state (see online version for colours)



The relationship between the distance of the input from the centroid and the activation state is shown in Figure 3, where the activation state rises and then falls as the distance gets farther. The gradient descent algorithm is used to learn the RBF, and its error calculation formula is shown in equation (6).

$$E = \frac{1}{2}\sum_{k=1}^{i}(Tk - Zk)^2 \tag{6}$$

In equation (6), i represents the number of neurons in the output layer, Tk represents the desired output of the k neuron, and Zk is the actual output of the k neuron, which is calculated as shown in equation (7).

$$Zk = \sum_{j=1}^{M} W_{kj}^{*}\varphi\left(\left\|X - C_j\right\|\right) \tag{7}$$

Equation (7) is the actual output formula, where M is the number of neurons in the hidden layer, W_{kj} is the weight between the hidden layer and the output layer, the input sample is X, and the neuron data of the j^{th} hidden layer is represented as C_j. To further improve the accuracy of the data prediction, the support and confidence are needed as the basis for selecting the data centre and the data selection is done randomly with the mathematical expression as in equation (8).

$$P_C(X) = Support(X) * Confidence(X) \tag{8}$$

In equation (8), $P_C(X)$ is the probability that the input sample X is selected as a data centre, $Support(X)$ is the support of association rule between the sample and the prediction result, $Confidence(X)$ is the confidence of association rule between the sample and the prediction result, and a higher value of $P_C(X)$ means a smaller gap between the actual value and the prediction result.

$$P_C(X)' = P_C(X) \Big/ \sum_{X=1}^{n} P_C(X) \tag{9}$$

Equation (9) is the normalisation of equation (8), and the probability formula is normalised to avoid the case where the probability sum is greater than 1. The optimal data centre formula is further optimised.

$$P(optimal) = f/(m * n) \tag{10}$$

In equation (10), n represents the number of input items, m represents the number of input data, f represents the number of hidden layer neurons, and $P(optimal)$ represents the probability of selecting the optimal data centre.

$$P(optimal) = \sum_{1}^{n} P_C(X_n) * f/m \tag{11}$$

In equation (11), $P_C(X)$ is the probability that the optimal centre is in the input sample of the n item.

$$\begin{aligned} P(optimal) = {} & P_C(X_1) * f/m + P_C(X_2) * f/m \\ & + \ldots + P_C(X_n) * f/m \end{aligned} \tag{12}$$

The above equation (12) is obtained by expanding equation (11), so the sum of probabilities of a total of n input data is shown in equation (12).

$$P_C(X_1) + P_C(X_2) + \ldots + P_C(X_n) = 1 \tag{13}$$

In equation (13), the probabilities are summed to obtain the sum of probabilities of 1. Finally, the formula for calculating the optimal data centre is simplified as in equation (14).

$$P(optimal) = f/m \tag{14}$$

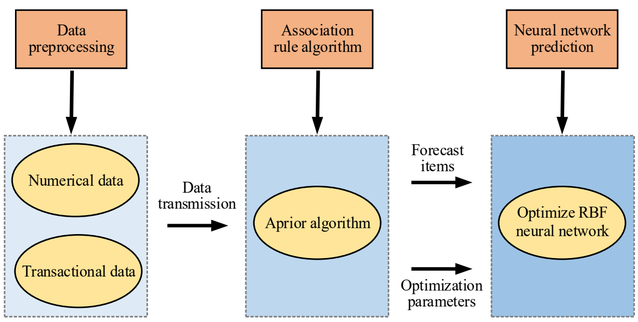
In order to verify the superiority of the algorithm proposed in the study, comparison experiments are setup later and the performance of the network is evaluated by the magnitude of the mean square error, which is calculated as in equation (15). The final obtained algorithmic model of association rules for fused RBF neural networks is shown in Figure 4.

$$MSE = \left((output - target)^2\right)/n \tag{15}$$

3.2 *Construction of English web teaching prediction model based on improved association rule algorithm*

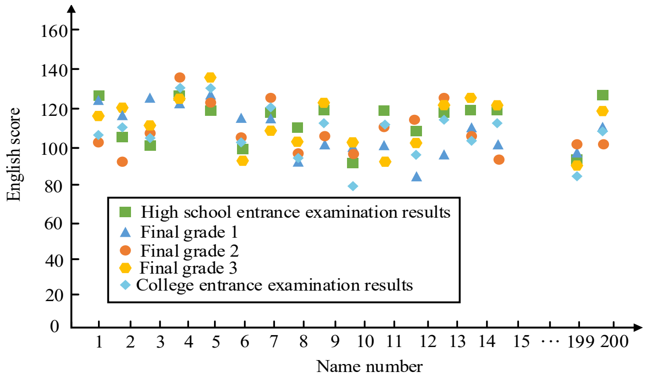
In the whole research of online teaching resources, data prediction is an indispensable part. The rapid update and iteration of the online education model has generated a large amount of student data, including the student's online course study time, test scores of various subjects, the number of online interactions, and the correct rate of after-school exercises. However, most of these data are only used for simple queries and statistics, and are not used further. Predicting and analysing education-related data can effectively improve teaching outcomes and ensure learning quality. This chapter takes an experimental middle school as an example, and uses the model to predict the English grades of its graduating class in the college entrance examination. By predicting the academic grades, the existing network English teaching resources of the school can be improved, and more appropriate teaching objectives and teaching methods can be formulated (Wang and Zhang, 2019).

Figure 4 Algorithmic model of association rules incorporating RBF neural networks (see online version for colours)



The data of this experiment comes from the high school entrance examination scores of 200 students enrolled in an experimental middle school in 2013, the final grades of three grades from senior one to senior three (sequentially numbered final grade 1, final grade 2, final grade 3) and final college entrance examination results. Renumber these score data (for the convenience of later elaboration, the names are numbered sequentially from 1, and the five scores are respectively recorded as A, B, C, D, E) The distribution of all experimental data is shown in Figure 5.

Figure 5 Scatter plot of experimental data distribution (see online version for colours)



Randomly divide the score data of all students in Figure 5 into two groups. One group is the association analysis mining group, which contains the five English scores of 100 students at random to mine the relationship between each data and association rules; The second group is the learning prediction group, which also includes the five English scores of 100 randomly selected students, and then uses the data of another group of students to predict the scores, so as to increase the credibility of the experiment. In this research, the association rule algorithm used is Apriori. The Apriori algorithm is written in R language. In order to simplify the representation of datasets, all data is converted into transaction datasets. If the numerical difference between two scores is small, it indicates that the correlation between the two scores is large. Put some items with differences within a certain range together as a transaction dataset. The sorted transaction dataset is used as the input. After the analysis of the above improved algorithm model, the association rules shown in Table 1 are obtained.

Table 1 Table of association rules among the items

Rule	*Support %*	*Confidence %*	*Lift*
B-D	52.94%	74.15%	2.45
D-B	52.93%	72.64%	2.43
D-E	52.92%	76.32%	2.89
E-D	52.91%	74.21%	2.87
B-C	52.90%	80.18%	2.69
C-B	51.88%	69.34%	2.46
D-C	50.46%	67.82%	2.67
B-E	48.62%	72.46%	1.89
E-B	46.36%	65.32%	1.87
C-E	38.63%	62.69%	1.31
E-C	38.51%	55.64%	1.25

The association rules shown in Table 1 are sorted from high to low according to their support. The support, confidence and promotion of each score range are given. Considering that the prediction model ultimately needs to predict the college entrance examination scores, it is necessary to find out the items related to college entrance examination scores E. It can be seen from the data in Table 1 that since the support of the scores A and E of the high school entrance examination is too low to appear in the above table, it shows that the correlation between their values and the final predicted college entrance examination scores is very small. After determining that B, C, D and E have a high degree of association, the above association rule model of RBF neural network fusion is used for prediction to determine the input of the model, the probability of each sample item being selected as the central data, and the output of the final model. Set the dataset of network training, use the association analysis mining group to select the dataset, take A, B, C, D of 100 students in a group of data as the input items, take E as the output, and calculate the probability of each sample being selected as the central data through equation (14); Then use the learning test group to conduct the final data test. Similarly, take A, B, C, D of 100 students in the two groups of data as the input items, and E as the verification data. After the prediction, calculate the mean square error of the actual value and the predicted value, and default the error target to 0, and the expansion speed of the radial basis function to 1. See formula (15) for the calculation formula of mean square error.

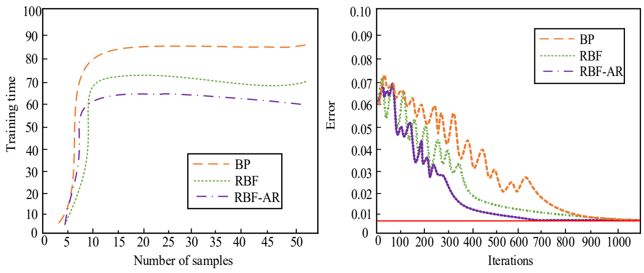
4 Validation of an English online teaching performance prediction model based on improved association rule algorithm

4.1 Comparison of association rule algorithm with fused RBF and traditional algorithm

In order to prove the superiority of the algorithm proposed by the research, the traditional RBF neural network algorithm (referred to as RBF), the BP neural network algorithm (referred to as BP) and the association rule algorithm fused with RBF (referred to as RBF-AR) were used to predict the performance of the model. The three algorithm models are trained with the performance data of the first group, and the respective training time and mean square error values are obtained as shown in Figure 6.

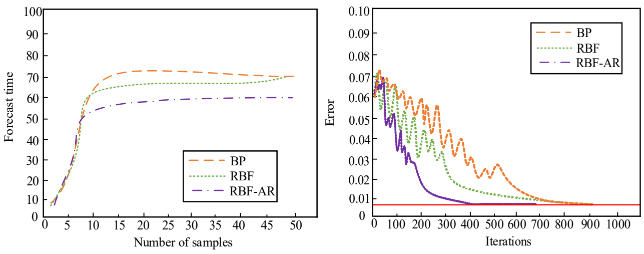
As shown in Figure 6, Figure 6(a) represents the training time of each of the three models in the training group with respect to the number of samples. Comparing the training time of each of the three models, it can be seen that the training time of the RBF-AR algorithm is shorter than the other two algorithms. As the number of samples increases, the RBF-AR algorithm is able to reach a stable training time of 63.5 s as soon as possible; the RBF algorithm has a slightly shorter training time than the BP algorithm and is able to gradually stabilise at 71.8 s; the BP algorithm has the longest training time and is able to reach a stable state at 86.4 s. Figure 6(b) represents the mean square error of each of the three models in the training group with the number of iterations. Comparing the mean square error of each of the three models, it can be seen that the RBF-AR algorithm can reach the steady state at 650 iterations, and the error value stabilises from 0.06 at the beginning to the expected value of 0.01. Both the RBF algorithm and the BP algorithm can reach the steady state after 900 iterations. In summary, it shows that the RBF-AR algorithm obtained from the study is more effective in data training, takes less time, and can reach the steady state faster. To further verify the accuracy of the algorithm model, the model after training is then used for prediction on the second set of data and the prediction time and mean square error under this set of data are obtained.

Figure 6 Comparison of training time and mean square error of training groups, (a) training time comparison chart (b) training mean square error comparison chart (see online version for colours)



(a) (b)

Figure 7 Comparison of prediction time and mean square error for the prediction groups, (a) comparison chart of prediction time (b) comparison chart of prediction mean square error (see online version for colours)



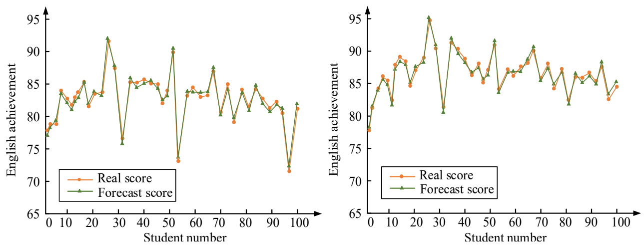
(a) (b)

As shown in Figure 7, Figure 7(a) represents the variation of the prediction time with the number of samples for each of the three models in the prediction group. Comparing the prediction time of each of the three models, it can be seen that the RBF-AR algorithm has a shorter prediction time than the other two algorithms, and also a shorter training time than them. As the number of samples increases, the RBF-AR algorithm is able to reach a stable prediction time at 57.5 s; the RBF algorithm has a slightly shorter prediction time than the BP algorithm and is able to gradually stabilise at 68.2 s; the BP algorithm has the longest training time and is able to reach a stable state at 71.5 s. Figure 7(b) represents the variation of mean square error with the number of iterations for each of the three models in the prediction group. Comparing the mean square error of each of the three models, it can be seen that the RBF-AR algorithm can reach the steady state at 410 iterations and converges significantly faster than the training group, while the RBF algorithm reaches stability at 690 iterations and the BP algorithm still reaches stability after 900 iterations. In summary, it shows that the RBF-AR algorithm obtained from the study can be used in data prediction. Therefore, the model can be used in a series of teaching-related data prediction such as grade prediction and college advancement prediction, and through accurate data prediction, educational resources can be integrated and utilised in a targeted manner.

4.2 *Performance testing of the improved association rule algorithm-based performance prediction model*

The results of the above training group and prediction group data show that the algorithm model designed in this study can be used in actual performance prediction. Randomly select 200 students from another middle school who are preparing to take the college entrance examination, randomly divide the 200 students into two groups, and record the real English scores of these 200 students for data comparison. The performance of the prediction model is analysed and tested, and the accuracy of the English score prediction model is shown in Figure 8.

Figure 8 Comparison of predicted and actual English scores of the two groups of students, (a) the predicted and actual scores of the first group of students (b) the predicted and actual scores of the second group of students (see online version for colours)



(a) (b)

Figure 8 shows the comparison between the predicted and actual scores of the students in the two groups. From Figure 8, it can be seen that in both groups, the results of the students' real scores and the predicted scores of the model are almost the same, which indicates that the accuracy of the prediction model is high. The average accuracy of the prediction model in the two groups is 95.7%, which is much higher than that of the traditional neural network model, obtained by calculation. This result indicates that the prediction system designed in the study has higher accuracy and better performance. Applying the performance prediction system to 160 students who took the college entrance examination, the relative error plot of students' English performance is shown in Figure 9.

Figure 9 shows the relative error data plots of the real English scores and the predicted scores of 160 students, where error 1 (Error1) indicates the relative error value of the prediction model that only considers students' English scores; error 2 (Error2) indicates the relative error of the prediction model that considers students' previous English scores and learning attitudes. It is obvious from Figure 9 that the overall trend of error 1 is above error 2, and only a very small portion of error 1 lies below error 2, indicating that the overall error of error 2 is smaller than that of error 1. The lowest value of error 2 is 0.031 and the highest value is 0.065. In this set of data, the average value of the relative error is 0.041, which is much lower than that of the traditional prediction model. The above results suggest that considering students' learning attitudes in the prediction system will enhance the accuracy of the prediction performance model. To further confirm this conclusion, this English achievement prediction system was used in two groups of students, one group considering only the students' grades and the other group considering previous grades and learning attitudes together. The graphs of the true and predicted grades obtained for the two groups are shown in Figure 10.

Figure 9 Plot of relative error data of English scores of 160 students (see online version for colours)

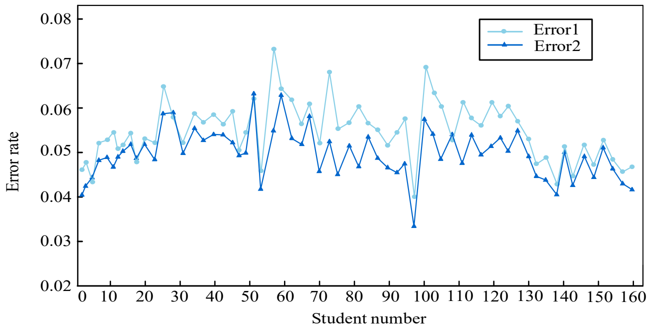
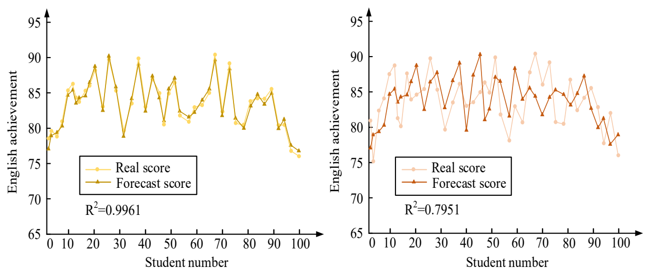


Figure 10 Comparison of the real and predicted scores of the two groups of students, (a) consider past achievements as influencing factors (b) past achievements are not considered as influencing factors (see online version for colours)



(a) (b)

In Figure 10(a) shows the comparison between the true and predicted grades taking into account both the students' past grades and learning attitudes. From the fluctuation of the curves, we can see that the true and predicted grade values are very close in this dataset, and the coefficient of determination in this case is obtained as 0.9961, indicating that the prediction model has a good prediction performance at this time and can accurately predict the students' English. The coefficient of determination is 0.9961. Figure 10(b) shows the comparison between the true and predicted scores considering only the students' scores. In this set of data, the true and predicted scores fluctuate greatly, indicating that the prediction effect of the model is not satisfactory at this time. The value of the coefficient of determination is 0.7951, which is much lower than the prediction model that takes into account both students' past grades and learning attitudes, indicating that the prediction effect of the model is much lower than that of the system model that takes into account both students' past grades and learning attitudes. Based on the results of this comparison, it shows that including students' learning attitudes as a factor affecting students' grades leads to a higher coefficient of determination and improves the performance of the prediction system. In summary, the prediction performance of the model designed in the study is very good in terms of the three dimensions of accuracy, relative error, and coefficient of determination. It is also found that adding students' learning attitude as an influencing factor in the grade prediction model greatly improves the accuracy of the prediction model.

5 Conclusions

The progress of network technology makes the efficient use of network education resources become particularly important. A large number of data generated in online education often lose their potential value because they cannot be effectively used. Through accurate data prediction, we can integrate and utilise educational resources in a targeted way. In order to integrate the existing English online teaching resources, the research attempts to predict the English achievements it produces. Firstly, RBF neural network and association rule algorithm are integrated, and then an English online teaching prediction model based on improved association rule algorithm is constructed. The analysis and detection of the model shows that the training time and prediction time of the model are 63.5 s and 57.5 s respectively, which is shorter than the traditional RBF neural network and BP neural network models; compared with the other two traditional models, this model has fewer iterations and can quickly reach a stable state and complete convergence. The prediction model is further applied to the prediction of actual English scores. The results show that the average accuracy of the prediction model is 95.7%, and its accuracy is higher; by comparing the relative errors of the two factors, it is found that if both English achievement and learning attitude are taken into account, the average relative error of the prediction model is lower, only 0.041; finally, we compare the decision coefficients of the two influencing factors, and find that the decision coefficients of the prediction model that takes into account both students' achievements and learning attitudes are close to 1. From the three dimensions of accuracy, relative error and determination coefficient, it shows that the prediction model has good prediction performance at this time and can accurately predict students' English scores.

In this research, through association rule algorithm, we compared the support between students' scores, determined several factors that have a greater correlation with the final prediction of college entrance examination scores, used the association rule model integrating RBF neural network to predict, and determined the input of the model, the probability of each sample item being selected as the central data, and the output of the final model. Because students' college entrance examination scores are affected by many factors, the past scores alone cannot be completely predicted accurately, and there are many psychological factors. At the same time, due to the different hidden factors of implicit learning among different disciplines, the prediction accuracy of the prediction model constructed by the research for other disciplines is not necessarily high, so it can not be simply applied to the prediction of other disciplines' achievements. In addition, because the integration of English online teaching resources is a big project, it is far from enough to just predict students' English achievements. There is also a need to collect the potential value of other data.

References

Baker, M.A. (2021) 'Household electricity load forecasting toward demand response program using data mining techniques in a traditional power grid', *International Journal of Energy Economics and Policy*, Vol. 11, No. 4, pp.132–148.

Bin, C., Gu, T., Sun, Y. et al. (2019) 'A travel route recommendation system based on smart phones and IoT environment', *Wireless Communications and Mobile Computing*, Vol. 2019, No. 2, pp.1–16.

Bose, P. and Gao, X. (2022) 'Cultural representations in indian english language teaching textbooks', *SAGE Open*, Vol. 12, No. 1, pp.517–542.

Chen, Y.A., Yuan, P.B., Qiu, M.A. et al. (2019) 'An indoor trajectory frequent pattern mining algorithm based on vague grid sequence', *Expert Systems with Applications*, Vol. 118, No. 3, pp.614–624.

Dey, L. and Mukhopadhyay, A. (2019) 'Biclustering-based association rule mining approach for predicting cancer-associated protein interactions', *IET Systems Biology*, Vol. 13, No. 5, pp.234–242.

Elgllab, M. and Shehata, A. (2019) 'Information seeking behavior in Arabic and English: a case study of scholars at Shaqra University', *Information Development*, Vol. 35, No. 3, pp.351–361.

Fautley, M. and Daubney, A. (2019) 'The whole class ensemble tuition programme in English schools – a brief introduction', *British Journal of Music Education*, Vol. 36, No. 3, pp.223–228.

Gagnon, P., Bedard, M. and Desgagne, A. (2019) 'Weak convergence and optimal tuning of the reversible jump algorithm', *Mathematics and Computers in Simulation*, Vol. 161, No. 7, pp.32–51.

Han, Y., Yu, D., Yin, C. et al. (2020) 'Temporal association rule mining and updating and their application to blast furnace in the steel industry', *Computational Intelligence and Neuroscience*, Vol. 2020, No. 3, pp.1–21.

Karjanto, N. and Simon, L. (2019) 'English-medium instruction calculus in Confucian-heritage culture: flipping the class or overriding the culture?', *Studies in Educational Evaluation*, Vol. 63, No. 2, pp.122–135.

Lu, P.H., Keng, J.L., Tsai, F.M. et al. (2021) 'An Apriori algorithm-based association rule analysis to identify acupoint combinations for treating diabetic gastroparesis', *Evidence-based Complementary and Alternative Medicine*, Vol. 2021, No. 17, pp.1–9.

Nohuddin, P.N., Zainol, Z. and Hijazi, M. (2021) 'Study of B40 schoolchildren lifestyles and academic performance using association rule mining', *Annals of Emerging Technologies in Computing*, Vol. 5, No. 5, pp.60–68.

Pollard, D. and Olizko, Y. (2019) 'Current foreign languages teaching issues in higher education art and Esp integration in teaching Ukrainian engineers', *Advanced Education*, Vol. 6, No. 11, pp.68–75.

Soufi, M.D. and Ferdousi, R. (2021) 'Association analysis of obesity/overweight and breast cancer using data mining techniques', *Frontiers in Health Informatics*, Vol. 10, No. 1, pp.60–64.

Tuncalyaman, T., Bilgi, E. and Esen, M.F. (2021) 'Analysis of traffic accidents with fuzzy and crisp data mining techniques to identify factors affecting injury severity', *Journal of Intelligent and Fuzzy Systems*, Vol. 42, No. 1, pp.575–592.

Wang, C., Liu, Y., Zhang, Q. et al. (2019) 'Association rule mining based parameter adaptive strategy for differential evolution algorithms', *Expert Systems with Applications*, Vol. 123, No. 6, pp.54–69.

Wang, N. and Zhang, Y. (2019) 'Application status and promotion strategy of integrated network teaching platform – taking Northwest A&F University as an example', *Asian Agricultural Research*, Vol. 11, No. 4, pp.94–96.

Wang, Y. (2022) 'Implications of blended teaching based on theory of semantic wave for teaching English writing in high school', *Journal of Higher Education Research*, Vol. 3, No. 2, pp.166–168.

Yue, H., Liao, H., Li, D. et al. (2021) 'Enterprise financial risk management using information fusion technology and big data mining', *Wireless Communications and Mobile Computing*, Vol. 2021, No. 1, pp.1–13.

Zhu, Z.W., Zhu, P.S., Miao, Y.Y. et al. (2021) 'Characteristics analysis for Chinese patent medicine containing Jujubea Fructus based on data mining', *China journal of Chinese Materia Medica*, Vol. 46, No. 9, pp.2344–2349.