
A decision tree-based method for detecting middle school students' behaviour characteristics in online English learning

Feifei Wang and Fengxiang Zhang*

College of Foreign Languages,
Hebei University of Economics and Business,
Shijiazhuang 050061, China
Email: wangffeifei2008@126.com
Email: fengxiang@mls.sinanet.com
*Corresponding author

Abstract: In order to solve the problem of low accuracy and long detection time caused by poor feature extraction effect of online English learning students' behavioural characteristics detection, this paper proposes a method of online English learning students' behaviour characteristics detection based on decision tree. Firstly, the concept and structure of decision tree are analysed, and the classification steps are designed. Secondly, weighted principal component analysis was used to extract the behaviour characteristics of students. Then, the characteristic data is standardised. Finally, the C4.5 decision tree algorithm is used to construct a student behaviour feature detection model to detect students' behaviour characteristics in online English learning. The experimental results show that the feature detection rate of the proposed method is as high as 99.5%, the accuracy is 96.2%, and the detection time is 19.8 s. Therefore, the feature detection effect of the proposed method is good, the accuracy is high, and the detection time is effectively shortened.

Keywords: decision tree algorithm; weighted principal component analysis; WPCA; online English learning; student behaviour characteristics; behaviour characteristic detection.

Reference to this paper should be made as follows: Wang, F. and Zhang, F. (2023) 'A decision tree-based method for detecting middle school students' behaviour characteristics in online English learning', *Int. J. Reasoning-based Intelligent Systems*, Vol. 15, No. 1, pp.54–62.

Biographical notes: Feifei Wang received her Masters in English Language and Literature from the Beijing University of Aeronautics and Astronautics, in 2009. She is currently a Lecturer in the College of Foreign Languages of Hebei University of Economics and Business. Her research interests include the English and American literature, and the college English teaching.

Fengxiang Zhang received her Masters in English Language and Literature from the Hebei Normal University, in 2000. She is currently an Associate Professor in the College of Foreign Languages of Hebei University of Economics and Business. Her research interests include the English and American literature, and the college English teaching.

1 Introduction

With the rapid development of network information technology, digital behaviours based on the internet have gradually replaced the traditional interactive behaviours carried out at the same place at the same time. At the same time, information exchange and sharing through the internet has become an inevitable development trend, and online English teaching is an important part of it (Tang, 2021; Du and Wang, 2021; Herrera Bohórquez et al., 2019). At present, with the popularisation of online open courses and the development of learning technology systems, online English learning behaviours are increasingly occurring in various educational situations, which also brings a large amount of student learning behaviour data (Roychowdhury et al., 2021; Lai et al., 2020). With the widespread use of

online English, the non-intellectual factors of English learners and the influence of online learning behaviours on their learning outcomes need to be deeply analysed and studied in order to guide teachers to formulate corresponding teaching plans. Therefore, it is of great significance to detect the behavioural characteristics of students in online English learning.

At present, scholars in related fields have carried out research on behaviour feature detection. Gong et al. (2020) proposed an analysis method of college students' consumption behaviour based on campus one-card big data. Using big data and data mining technology for the development analysis and research, dig the hidden behind the behaviour of students information, based on the engineering characteristics of the data pre-processing and feature extracting, statistical analysis of the students'

behaviour, circle using clustering algorithm with different behaviour of students, in-depth analysis of the behaviour of different groups of structure and behaviour characteristics, It achieves the extraction and analysis of students' behaviour characteristics, but the method has the problem of low detection accuracy. Luo and Han (2021) studies the behavioural characteristics of students' online learning. This method uses machine learning algorithm to cluster students' online learning behaviours in mixed courses, and extracts typical characteristics of each class of students to realise the study of students' online learning behaviour characteristics. This method is robust to any type of scenes, but it still has the problems of poor feature extraction effect and low accuracy. Bai et al. (2020) proposes a method for detecting students' classroom behaviour based on multi-channel feature fusion called faster R-CNN and transfer learning. Students' characteristics are extracted using pre-trained Inception ResNet-V2 network, and the faster R-CNN target detection framework is used to detect students' behaviours through transfer learning. Combined with the multi-channel feature fusion method, the student behaviour detection model is improved to realise the student behaviour detection, but the detection time of this method is long.

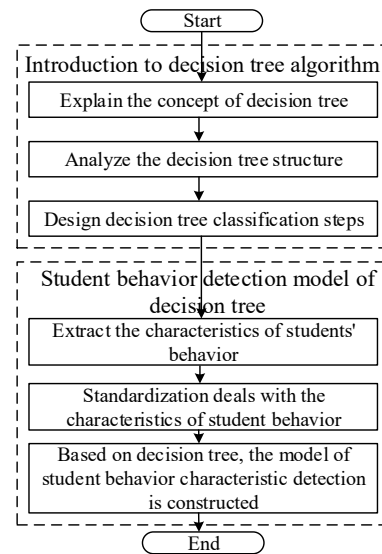
In order to solve the above problems, improve the detection accuracy, reduce the detection time, a decision tree-based method is proposed to detect students' behaviour characteristics in online English learning. Weighted principal component analysis (WPCA) was used to extract the behavioural characteristics of middle school students in online English learning, and the data of behavioural characteristics were standardised. The C4.5 decision tree algorithm was used to construct the behavioural characteristics detection model of middle school students in online English learning, so as to realise the behavioural characteristics detection of middle school students in online English learning. In order to improve the effect and efficiency of online English learning and provide convenient conditions for students.

2 Detection method of students' behaviour characteristics

In order to improve the accuracy and reduce the detection time of students' behaviour characteristics in online English learning, this paper introduces the decision tree algorithm, based on which, combined with the WPCA method, the detection method is optimised. The design flow of detection method is shown in Figure 1.

According to Figure 1, the method in this paper can be divided into two major steps, namely, the introduction of decision tree algorithm and the detection method of students' behaviour characteristics in online English learning by constructing decision tree, in the first part, the basic algorithm of decision tree is introduced. In the second part, decision tree is used to build a detection model to detect students' behaviour characteristics in online English learning.

Figure 1 Design process of student behaviour characteristics detection method based on decision tree



2.1 Decision tree algorithm

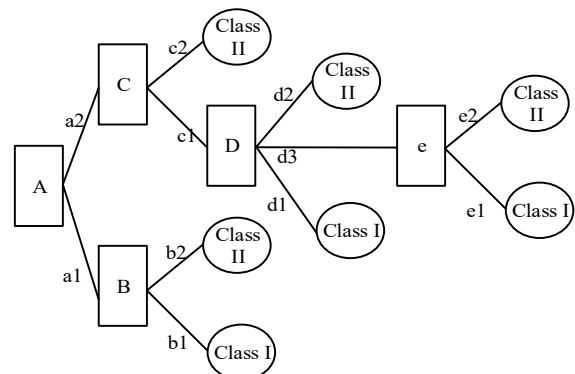
2.1.1 Decision tree concept

Decision tree is a method of 'decomposition and then breakthrough' (Chen, 2020; Zhang et al., 2021; Gan, 2021). The decision tree expresses the relationship of each stage in the decision-making process with an arrow. Each node determines this attribute, and each branch is the result of the decision. Leaf node represents a classification or a result distribution of the final target state of all attributes that can reach the node.

2.1.2 Decision tree structure

Decision tree algorithm can classify and predict data, and has strong robustness and high precision classification performance. The structure of the decision tree is shown in Figure 2.

Figure 2 Decision tree structure diagram



According to Figure 2, start from the root node A to perform the traversal operation, during the search, the inductive algorithm is used to select a branch until the leaf node is selected. Each node on the decision tree is used to represent a type of leaf or a new decision node, and each

branch is the output value of each test. The figure describes that the sample data is divided into two types according to fixed rules, and its attributes are determined in each decision node, and so on to build a decision tree.

2.1.3 Decision tree classification steps

Classification and judgment using decision trees usually include three steps: decision tree generation, pruning and extraction of rules.

- 1 Generation of decision tree: The generation process of decision tree refers to the process of generating a preliminary judgment model by inputting a large number of datasets as training data.
- 2 Pruning of decision tree: In order to obtain the best description effect of decision tree, it is necessary to make the decision tree general and applicable, pruning process decision tree.
- 3 Rules of decision tree: The decision tree represents the final classification results in a tree structure, and can generate rules in the form of if then, which is close to people's cognition and representation of things in the real world (Sun et al., 2021).

Based algorithm is part of the content of the decision tree algorithm, but the application of the part can not only achieve high precision characteristics test, therefore, in this study, based on the application of decision tree algorithm, combined with WPCA, and standardised treatment the student behaviour characteristics, to optimise the online English learning for middle school students based on decision tree behaviour feature detection method.

2.2 Student behaviour characteristics detection method of decision tree

2.2.1 Extracting the behavioural characteristics of students

In order to effectively detect the behaviour characteristics of students in online English learning based on decision tree, WPCA is applied to extract the behaviour characteristics of students in online English learning, which makes preparation for the detection of students' behaviour characteristics in decision tree. WPCA is optimised on the basis of the original PCA, which improves the representativeness of features by giving different weights to data of different attribute types (Xu et al., 2021; Zhang and Tang, 2020; Yuan and Yu, 2021). The specific process is as follows:

- Step 1 Assume that the behavioural characteristic data sample of students in online English learning is c samples with mixed attributes, denoted as H , in which m students are included, and the number of characteristic attributes of each student is c . Through statistical analysis, this sample preliminarily extracted features from online English learning data of middle school students,

including learning duration, learning time, learning content type, etc.

- Step 2 Transpose H to form a matrix of student behaviour characteristics in online English learning, which is described as follows:

$$H = \begin{bmatrix} h_{11} & h_{12} & \dots & h_{1c} \\ h_{21} & h_{22} & \dots & h_{2c} \\ \vdots & \vdots & \vdots & \vdots \\ h_{m1} & h_{m2} & \dots & h_{mc} \end{bmatrix} \quad (1)$$

In formula (1), h_{ij} represents the behavioural characteristic data of students in online English learning.

- Step 3 Normalise the behavioural feature matrix of students in online English learning. The processing formula is as follows:

$$h'_{ij} = \frac{h_{ij} - \min(h_{ij})}{\max(h_{ij}) - \min(h_{ij})} \quad (2)$$

In formula (2), h'_{ij} represents the normalised behaviour characteristic data of online English learning students, and $\min(h_{ij})$, $\max(h_{ij})$ represents the minimum and maximum values of the behaviour characteristic data of original online English learning students respectively.

- Step 4 The mean values of each column in the normalised online English learning behaviour feature matrix are obtained.
- Step 5 Calculate the weights of the data samples of students' behaviour characteristics in online English learning. Calculated as follows:

$$d_i = \frac{1}{\varphi} \times \left(\sum_{j=1}^m r_j \right) \quad (3)$$

In formula (4), d_i represents the sample weight of student behaviour characteristic data in online English learning, r_j represents the normalised value of element h_{ij} of student behaviour characteristic matrix H in online English learning, and φ represents the quantity of student behaviour characteristic data in online English learning.

- Step 6 Multiply the result of step 4 and the result of step 5 to construct a new online English learning middle school student behaviour characteristic data sample to form a new online English learning middle school student behaviour characteristic matrix H' .

$$H' = (u_{ij} \times d_i)^T \quad (4)$$

In formula (5), u_{ij} represents the mean value of each column in the student behaviour characteristic matrix, H' represents the new online English learning student behaviour characteristic matrix.

Step 7 Conduct PCA on the new online English learning middle school students' behavioural characteristic data samples.

Step 8 Calculate the covariance of the new online English learning middle school student behaviour characteristic matrix H' , and form the covariance matrix K .

$$K = (H')_{m \times m} \quad (5)$$

Step 9 According to the cumulative contribution rate β , the first γ new online English learning middle school students' behavioural characteristic data values selected by the cumulative contribution rate are used as principal components, and the first δ principal components are selected as the new online English learning middle school students' behavioural characteristics are expressed as:

$$L_\delta = \beta \times \frac{H'}{K} \quad (6)$$

In formula (6), L_δ represents the first δ student behaviour characteristics, β represents cumulative contribution rate.

Through the above steps, new behaviour characteristics of students in online English learning are selected, and thus, behaviour characteristics of students in online English learning are extracted.

2.2.2 Standardisation deals with the characteristics of student behaviour

The goal of the standardised processing of student behavioural feature data in online English learning is to eliminate the dimensional problem of student behavioural feature data in online English learning from multiple sources, making it uniform (Ustun et al., 2019; Shu et al., 2020). The specific process is as follows:

Step 1 Calculate the arithmetic mean of the behavioural characteristic data of students in online English learning. The calculation formula is as follows:

$$\bar{g} = \frac{g_1 + g_2 + \dots + g_\varphi}{\varphi} \quad (7)$$

In formula (7), \bar{g} represents the arithmetic mean of the student behaviour characteristic data in online English learning, and g_φ represents the multi-source online English learning middle student behaviour characteristic data.

Step 2 Calculate the standard deviation of the behavioural characteristic data of students in online English learning. The calculation formula is as follows:

$$\sigma = \sqrt{\frac{\sum_{i=1}^{\varphi} (g_i - \bar{g})^2}{\varphi - 1}} \quad (8)$$

In formula (8), σ represents the standard deviation of the student behaviour characteristic data in online English learning, and g_i represents the i original multi-source online English learning middle student behaviour characteristic data.

Step 3 Standardise the behavioural characteristics of students in online English learning. The calculation formula is as follows:

$$g'_i = \frac{g_i - \bar{g}}{\sigma} \quad (9)$$

In formula (9), g'_i represents the standardised value of the behavioural characteristic data of online English learning middle school students.

Through the above process, the standardised processing of students' behavioural characteristics in online English learning is completed.

2.2.3 Build the student behaviour characteristic detection model of decision tree

Based on the above analysis, considering that the behaviour characteristics of students in online English learning are complex, the C4.5 decision tree algorithm is used to construct a model to detect students' behaviour characteristics in online English learning. The C4.5 decision tree algorithm maximises the information gain rate, realises the recursive selection of attributes, builds decision tree nodes, and realises the detection of students' behavioural characteristics in online English learning.

The proposed online English learning middle school student behaviour feature dataset Z contains $|Z|$ data, which are divided into x categories and c attributes. $V = \{V_1, V_2, V_3, \dots, V_x\}$ is used to describe the category subset, and $B = \{B_1, B_2, B_3, \dots, B_c\}$ is used to describe the attribute subset. The n attribute B_n contains a total number of different subsets α , then the dataset Z of student behaviour characteristics in online English learning is divided into α disjoint subsets. The subset is described by Z_n^B ($1 \leq n \leq \alpha$), and the total content of samples in subset Z_n^B is described by $|Z_n^B|$. The modelling process of C4.5 decision tree algorithm is shown in Figure 3.

According to Figure 3, the behaviour feature detection model of decision tree is designed in six steps, including selecting training dataset and test dataset, solving the information entropy of each type, solving the information gain of each attribute, solving the information gain rate of each attribute, obtaining the target category of expected behaviour feature detection and decision tree pruning. The specific design content is as follows:

- 1 Randomly divide the standardised online English learning middle school student behaviour characteristic dataset, and select the training dataset and the test dataset.

- 2 Solve the information entropy of each type, determine the sum of the randomness of each type in the sample, and use the formula (10) to describe the information entropy solving process:

$$E(Z) = -\sum_{n=1}^x P(V_n) \log_2 P(V_n) \quad (10)$$

In formula (10), $E(Z)$ represents for information entropy, E represents information entropy and V_n represents the n^{th} category subset.

- 3 Solving the information gain of each attribute: After the online English learning middle school student behaviour characteristic dataset Z is split by attribute B , the information entropy under the new classification condition is described by formula (11) to describe the information entropy solving process:

$$EB(Z) = -\sum_{n=1}^{\alpha} \frac{|Z_n^B|}{|Z|} \times E(Z_n^B) \quad (11)$$

In formula (11), $E(Z_n^B)$ represents the information entropy after splitting the behavioural characteristic dataset Z by attribute B in online English learning. Information gain represents the difference between the original information demand and the new demand obtained after splitting according to attribute B . Formula (12) is used to describe the information gain $G(B_n)$ solution process:

$$G(B_n) = E(Z) - EB(Z) \quad (12)$$

- 4 Solving the information gain rate of each attribute: Use formulas (13) and (14) to describe the information gain rate solution process:

$$S(Z) = -\sum_{n=1}^{\alpha} \frac{|Z_n^B|}{|Z|} \log_2 \frac{|Z_n^B|}{|Z|} \quad (13)$$

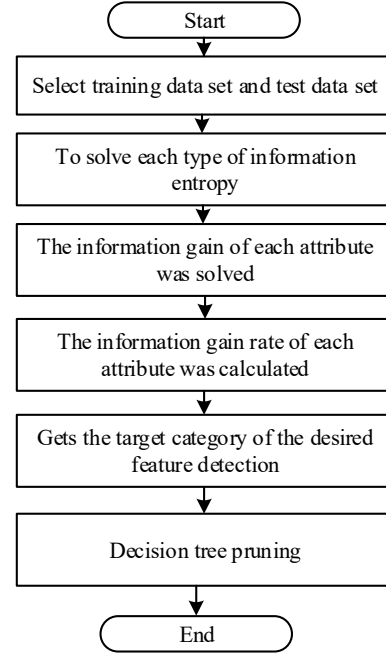
$$G_R(B_n) = \frac{1}{S(Z)} \times G(B_n) \quad (14)$$

- 5 Select the attribute B with the highest information gain rate and use it as the root-leaf point until any end condition is reached. The data of behavioural characteristics of students in online English learning in the leaf point are of the same type, so it is impossible to continue to split, and finally the desired target category of behavioural characteristics detection of students in online English learning is obtained.
- 6 Decision tree pruning: After pruning, the decision tree is pruned, the weighted confidence of each node is solved, and the variance method is used for weighting. The formula is as follows:

$$V_a = \frac{\sum_{i=1}^{\varphi} (g_i - \bar{g})^2}{\varphi} \quad (15)$$

In formula (15), V_a represents the weighted confidence of the node.

Figure 3 Modelling process of C4.5 decision tree algorithm

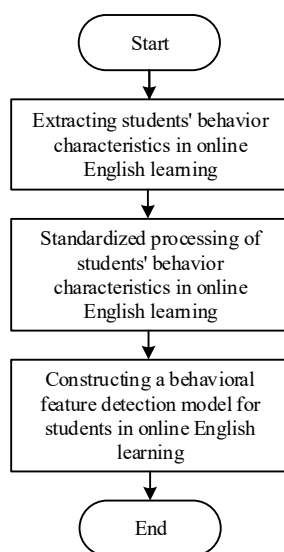


Input data samples of students' behavioural characteristics in online English learning, and set the number of iterations. Through the Boosting algorithm, the weak classification rule G_Z of the decision tree is obtained, and the weight Q_Z of the sample set of student behavioural characteristic data in online English learning is re-allocated, and the weight W_Z corresponding to the weak classification rule is determined before merging. Repeat the above process t times to obtain y sub-classifiers, and use the weighted voting combination algorithm to obtain the final classification rule F_G of the decision tree. Input the acquired data samples of behavioural characteristics of students in online English learning into the classifier, and construct the detection model of students' behavioural characteristics in online English learning according to the classification rules, which is expressed as:

$$D_C = t \times y \times \frac{G_Z \times Q_Z}{F_G \times W_Z} \quad (16)$$

In formula (16), D_C represents the detection results of students' behavioural characteristics in online English learning.

Through the above steps, the construction of the behaviour feature detection model of the students in the online English learning is completed, so as to realise the behaviour feature detection of the students in the online English learning. Its online English learning middle school student behaviour feature detection process is shown in Figure 4.

Figure 4 Flow chart of behavioural feature detection of students in online English learning

According to Figure 4 the designed steps, the online English learning behaviour characteristics of students can be detected based on decision tree. The detection method introduces the decision tree algorithm, which is the core algorithm of the detection method, but only the application of decision tree algorithm, detection can not achieve the desired effect, the improved detection accuracy is low, therefore, compared with the traditional method, the method further optimised the detection method and improved the accuracy of feature extraction by combining the WPCA method.

3 Experimental simulation and analysis

In order to verify the effectiveness of the decision tree-based method for detecting the behavioural characteristics of middle school students in online English learning, a comparative analysis experiment was set up, and the comparison method and the method in this paper were used to analyse the behavioural characteristics of middle school students in online English learning. The comparison methods were Gong et al. (2020) (method of Gong et al. (2020)) and Luo and Han (2021) (method of Luo and Han (2021)). The experimental scheme is firstly to define the research object and parameters, secondly to design the experimental performance index, and finally to analyse the performance of the method according to the index.

3.1 Experimental scheme

In this paper, the comparative method is adopted for experimental analysis to analyse the detection performance of student behaviour characteristics. The experimental process is as follows:

- 1 Setup the experimental environment, explain the research object, and give the specific data information of the research object.

- 2 Set the experimental performance indicators, which are the detection rate, detection accuracy and detection time of student behaviour characteristics, give the calculation method, and explain the relationship between the indicators and detection performance, and use the above three indicators to reflect the detection performance.
- 3 Performance analysis: draw the experimental results into charts and analyse the chart data in depth to reflect the effectiveness of the detection method.

Carry out experiments according to the above experimental plan.

3.2 Setting up the experimental environment

Based on MOOC open dataset, this paper selected 2,500 behavioural characteristics data of online English learning students. The specific experimental parameters are shown in Table 1.

Table 1 Parameter settings

The serial number	The parameter name	The set value
1	Quantity of characteristic data	2,500 pcs
2	The number of students	100 pcs
3	Number of feature categories	10 pcs
4	Number of student characteristics per student	25 pcs
5	Step length	1
6	The number of iterations	20 times

The 100 students are all college students in their freshman, sophomore, junior and senior years. There are 25 students in each grade. The above student behaviour characteristics are collected by the method in this paper, and the set student behaviour type and data length are based on the analysis results of Shu et al. (2020) method's behaviour characteristics. This value affects the detection time of feature detection method, so the value is not set too long. The data types mainly include ten types: learning duration, learning time, learning content, frequency of querying English words, type of querying English words, frequency of querying grammar, type of browsing grammar, correct rate of answering, speed of answering and time of reviewing the content of answering. Each type of data contains 250 data on average.

3.3 Experimental performance indicators

In order to verify the effectiveness of the proposed method, the number of students' behavioural characteristics detected in online English learning was taken as the variable, and the detection rate, detection accuracy and detection time of students' behavioural characteristics were taken as the experimental performance indicators. The detection rate of students' behavioural characteristics is the percentage of the

number of students' behavioural characteristics that can be effectively detected in online English learning compared with the total number of students' behavioural characteristics that need to be detected. The accuracy of students' behavioural characteristics detection is the percentage of the number of students' behavioural characteristics correctly detected in online English learning compared with the total number of students' behavioural characteristics that need to be detected. Student behaviour characteristic detection time is the running time of student behaviour characteristic detection. The higher the detection rate and accuracy of the verification detection method, the better the performance of the verification detection method. The shorter the detection time, the higher the efficiency of the detection method. The specific calculation formula is as follows:

The calculation formula of the detection rate of students' behaviour characteristics is as follows:

$$J_C = \frac{J_G}{J_Z} \times 100\% \quad (17)$$

In formula (17), J_G refers to the number of online English learning students' behavioural characteristics that are effectively detected, and J_Z refers to the total number of online English learning students' behavioural characteristics that need to be detected.

The calculation formula for the accuracy of student behaviour characteristics detection is as follows:

$$A_Q = \frac{J_Q}{J_Z} \times 100\% \quad (18)$$

In formula (18), J_Q refers to the number of correctly detected online English learning students' behavioural characteristics.

The performance index of detection time is calculated automatically by computer software.

3.4 Experimental performance analysis

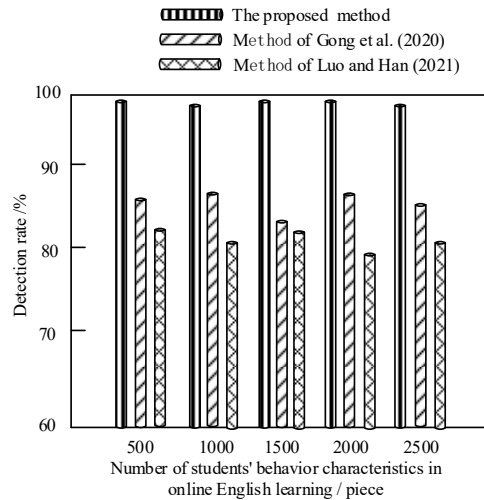
3.4.1 Effect of online English learning students' behavioural characteristics detection

In order to verify the effect of decision tree online English learning, the detection rate was used as an evaluation index. The methods in Gong et al. (2020), Luo and Han (2021) and the proposed methods were compared, and the detection rates of students' behavioural characteristics in online English learning using different methods were obtained, as shown in Figure 5.

According to Figure 5, the detection rate of the proposed method is higher than that of the comparison method when the data amount of different behavioural characteristics is different. Among them, when the number of online English learning students' behavioural characteristics is 2,500, the detection rate of Gong et al. (2020) method students' behavioural characteristics is 85.6% on average. The average detection rate of behavioural characteristics of Luo and Han (2021) method students was 80.9%. The average

detection rate of behavioural characteristics of online English learning students by the proposed method is as high as 99.5%, which is 13.9% and 18.6% higher than that by the comparison method. Therefore, the proposed method has a high detection rate of students' behavioural characteristics in online English learning, indicating that the proposed method has a good feature detection effect.

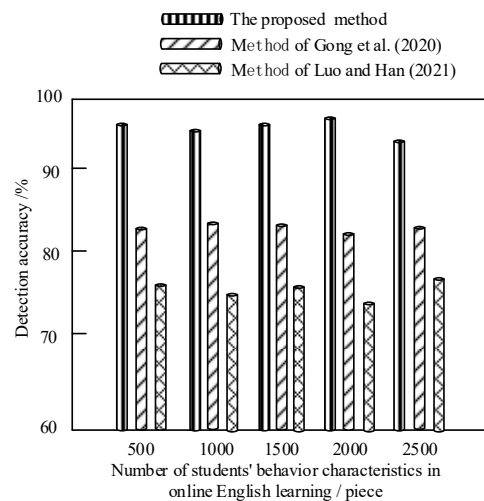
Figure 5 Detection rate of students' behavioural features in online English learning by different methods



3.4.2 Accuracy of students' behavioural characteristics detection in online English learning

The accuracy of the proposed method was further verified, and the detection accuracy was taken as an evaluation index. The methods in Gong et al. (2020), Luo and Han (2021) and the proposed methods were respectively compared, and the detection accuracy of behavioural characteristics of middle school students in online English learning using different methods was shown in Figure 6.

Figure 6 The detection accuracy of students' behavioural characteristics in online English learning with different methods



According to Figure 6, the detection accuracy of the proposed method is higher than that of the comparison method when the amount of behavioural characteristics data is different. When the number of behavioural characteristics of students in online English learning is 2,500, the average detection accuracy of behavioural characteristics of students in Gong et al. (2020) method online English learning is 83.2%. The average accuracy of Luo and Han (2021) method online English learning students' behavioural characteristics detection is 75.6%, and the average accuracy of the proposed method is 96.2%, which is 13% and 20.6% higher than that of the comparison method, respectively. Therefore, the proposed method has a high accuracy in the detection of students' behavioural characteristics in online English learning, indicating that the proposed method has a high accuracy in the detection of students' behavioural characteristics in online English learning.

3.4.3 Detection time of students' behavioural characteristics in online English learning

On this basis, the detection time of students' behaviour characteristics in online English learning of the proposed method is further verified, and the method of Gong et al. (2020), the method of Luo and Han (2021) and the proposed method are used to compare, and the behaviour characteristics of students in online English learning of different methods are obtained. The detection time is shown in Table 2.

Table 2 Detection time of students' behavioural characteristics in online English learning by different methods

<i>Number of students' behaviour characteristics in online English learning/piece</i>	<i>The proposed method/s</i>	<i>The method of Gong et al. (2020)/s</i>	<i>The method of Luo and Han (2021)/s</i>
500	8.5	13.1	15.8
1,000	11.8	17.4	19.3
1,500	14.9	21.2	23.5
2,000	17.3	23.8	25.4
2,500	19.8	26.3	28.9

As can be seen from Table 2, with the increase in the number of students' online English learning behavioural characteristics, the detection time of students' behavioural characteristics in online English learning by different methods increases, but the detection time of the proposed method is lower than that of the literature method. When the number of online English learning students' behavioural characteristics is 2,500, the detection time of Gong et al. (2020) method online English learning students' behavioural characteristics is 26.3 s, and the detection time of Luo and Han (2021) method online English learning students' behavioural characteristics is 28.9 s, while the detection time of the proposed method is only 19.8 s. Compared with the comparison method, 6.5 s and 9.1 s are

reduced. Therefore, the proposed method has a short detection time for students' behavioural characteristics in online English learning.

4 Conclusions

In order to improve the effectiveness of online English learning students behaviour characteristics test, so as to improve the effect of English learning, is proposed based on the decision tree of online English learning for middle school students behavioural feature detection method, using decision tree algorithm, combined with WPCA to detect online English learning for middle school students behavioural characteristics, and verify the feasibility of this method has certain application value.

- 1 Compared with the comparison method, the detection rate of online English learning students' behavioural characteristics is improved by 13.9% and 18.6%, and the average detection rate of online English learning students' behavioural characteristics is as high as 99.5%. Therefore, this method can effectively improve the detection rate of students' behavioural characteristics and is effective.
- 2 In order to further verify the feasibility of the proposed method, the accuracy rate of student behaviour characteristics detection was analysed. The accuracy rate of the method increased by 13% and 20.6%, and the highest value reached 96.2%. Through comparative analysis, the method can effectively improve the accuracy of detection, so as to improve the learning effect of students.

The behaviour feature detection method based on decision tree is also improved in terms of detection time. Compared with the literature method, the detection time of 2,500 students' behaviour feature data is shortened by 6.5 s and 9.1 s, effectively reducing the detection time of students' behaviour feature. Therefore, this method has application value. To meet the needs of online English learning students behaviour characteristics detection.

References

Bai, J., Gao, H.L., Wang, Y.Z. et al. (2020) 'Detection of students' classroom performance based on faster R-CNN and transfer learning with multi-channel feature fusion', *Journal of Guangxi Normal University (Natural Science Edition)*, Vol. 38, No. 5, pp.1-11.

Chen, F. (2020) 'An improved decision tree algorithm using information dispersion index', *Mathematics in Practice and Theory*, Vol. 50, No. 14, pp.76-82.

Du, Y.F. and Wang, J.M. (2021) 'Construction and practice of multidimensional and blended online learning as a regular model: a case study of academic English', *Journal of Sichuan International Studies University*, Vol. 37, No. 1, pp.137-144.

Gan, T. (2021) 'Research on university distance teaching quality evaluation based on decision tree classification algorithm', *Modern Electronics Technique*, Vol. 44, No. 9, pp.171-175.

- Gong, L.G., Gu, K., Ming, X.M. et al. (2020) 'Analysis of college students' consumption behavior based on campus card data', *Journal of Shenzhen University (Science & Engineering)*, Vol. 37, No. z1, pp.150–154.
- Herrera Bohórquez, L.I., Largo Rodríguez, J.D. and Viáfara González, J.J. (2019) 'Online peer-tutoring: a renewed impetus for autonomous English learning', *How*, Vol. 26, No. 2, pp.13–31.
- Lai, S., Sun, B., Wu, F. and Xiao, R. (2020) 'Automatic personality identification using students' online learning behavior', *IEEE Transactions on Learning Technologies*, Vol. 13, No. 1, pp.26–37.
- Luo, Y.Y. and Han, X.B. (2021) 'Research on blended courses classification based on characteristics of students online learning behavior', *China Educational Technology*, Vol. 7, No. 6, pp.23–30, p.48.
- Roychowdhury, S., Alareqi, E. and Li, W. (2021) 'OPAM: online purchasing-behavior analysis using machine learning', *2021 International Joint Conference on Neural Networks, IEEE*, pp.1–8.
- Shu, J.B., Ge, X., Peng, L.Y., Hu, Q.Q. and Liu, S.Y. (2020) 'Analysis of behavioral characteristics based on student's personal big data', *Journal of Huazhong Normal University (Natural Sciences)*, Vol. 54, No. 6, pp.927–934.
- Sun, X.X., Zhong, H. and Chen, H.P. (2021) 'A statistical analysis system of students' test scores based on decision tree classification technology', *Journal of Jilin University*, Vol. 51, No. 5, pp.1866–1872.
- Tang, Q. (2021) 'Hybrid online teaching of vocational english based on POA model – comments on English teaching and internet technology', *China Science Paper*, Vol. 16, No. 12, p.17.
- Ustun, T.S., Hussain, S., Kirchhoff, H., Ghaddar, B., Kai, S. and Lestas, I. (2019) 'Data standardization for smart infrastructure in first-access electricity systems', *Proceedings of the IEEE*, Vol. 107, No. 9, pp.1790–1802.
- Xu, K., Fan, B., Yang, H., Hu, L. and Shen, W. (2021) 'Locally weighted principal component analysis-based multimode modeling for complex distributed parameter systems', *IEEE Transactions on Cybernetics*, Vol. 37, No. 21, pp.1–11.
- Yuan, X.W. and Yu, L.P. (2021) 'An academic evaluation method for assisting experts in empowering and overcoming natural weights: multiple principal component analysis of complete information', *Information Studies: Theory & Application*, Vol. 44, No. 10, pp.57–64.
- Zhang, C.X. and Tang, L.J. (2020) 'Construction of the dimensional evaluation system of middle school students' health behavior in the core area of physical education and health', *Chinese Journal of School Health*, Vol. 41, No. 3, pp.361–364.
- Zhang, X.N., Long, X., Zhang, H.W. et al. (2021) 'Study on influencing factors of health knowledge among pupils in Beijing by logistic regression and decision tree', *Chinese Journal of Health Education*, Vol. 37, No. 10, pp.934–937.