



Research and Implementation of Online English Education Learning Performance Prediction Under the Background of Big Data

Han Bao^(✉)

English Department, Heilongjiang International University, Harbin, Heilongjiang, China
Baohan19831204@163.com

Abstract. In the context of big data, mining and analyzing massive data resources in the education field is conducive to making correct decisions for education and teaching based on reference opinions and improving the quality of education. Online English education learning performance prediction is one of the most popular research problems in the field of educational data mining. Based on the background of big data, this article will launch the study of online English education learning performance prediction. This article first introduces the main functions of the online English education learning performance prediction system, then designs the key function modules of the learning performance prediction system, and discusses the learning performance prediction algorithm. This paper conducts experiments on the performance of the designed learning performance prediction system, and it is concluded that the highest accuracy rate of this system for learning performance prediction reaches 87.94%, and the highest accuracy rates of SVM algorithm and decision tree algorithm are 73.41% and 72.13%. This shows that the online English education learning performance prediction system has a high accuracy rate.

Keywords: Education Data Mining · Online English Education · Big Data · Student Performance Prediction

1 Introduction

In the context of big data, the massive data of all walks of life shows an exponential growth trend [1, 2]. Mass data often implies some valuable information, which is a very valuable “soft resource” in all walks of life [3, 4]. The use of machine learning and data mining and other related technologies has promoted the analysis, mining and application of data in the education field, which has become the actual demand and future trend of the development of education informatization [5, 6]. Online English education is a manifestation of education informatization. Through the prediction of students’ academic performance, and according to the predicted results, teachers can teach students in accordance with their aptitude, formulate or improve teaching strategies, optimize resource allocation, and improve education and teaching output [7, 8].

Regarding the study of academic performance prediction, many scholars have carried out multi-angle exploration. For example, Barata G studied the early prediction of student files based on performance and game preference [9]; Deo RC has developed a performance prediction model based on artificial intelligence for undergraduate performance prediction [10]; Bao Y researched student performance prediction based on the similarity of behavior process [11]. It can be seen that the research on student performance prediction has always attracted much attention. Based on the background of big data, the research and design of online English education LP prediction system has practical significance.

This article first briefly introduces the main functions of the online English education LP prediction system, and explains the features extracted from the system's database. Then, this paper designs the key functional modules and algorithms of the system, including user management, student information management, student performance prediction and student performance warning. Finally, this article collects student-related academic performance data sets through the LMS system, and conducts a learning performance prediction test on the system.

2 Design of Online English Education Learning Performance Prediction System Under the Background of Big Data

2.1 Main Functions of Online English Education Learning Performance Prediction System

The network English education performance prediction system mainly realizes the student's academic performance prediction and performance early warning function. The system can predict students' final grades and analyze their knowledge points. The grade prediction is mainly through the data mining of students in online English education, so as to achieve the purpose of predicting academic performance. Knowledge point analysis is the influential features found when doing classification functions. By analyzing these features, we can understand students' current mastery of English knowledge, the correct rate of various types of topics, etc., so that online English educators can have a better grasp of the overall learning situation of students.

Through the machine learning process, the performance prediction module optimizes a variety of methods for the classification target according to the correlation of each feature in the data, and then periodically performs feature extraction on the student data in the database, thereby increasing the training data and improving the model's prediction performance effect in terms of accuracy.

The learner feature extraction that the network English education LP prediction system database needs to pay attention to includes the number of learner submissions, the number of system logins, the number of answer submissions, the average time to do a question, the correct rate, the interval between the average first submission time and the start time of the answer. The explanation of each feature is shown in Table 1.

Table 1. Interpretation of learner characteristics

feature	explain
Number of submissions	Number of attempts by the learner
System login times	Learner attendance
Number of correct questions	Learner's completion
Average question time	Learner's problem-solving efficiency
Correct rate	The learner's correct rate
The average value of the interval between the average first submission time and the start time of the answer	Learner motivation

2.2 Key Functional Modules of the Online English Education Learning Performance Prediction System

- (1) User management: This module includes functions such as adding education users, deleting users, setting and modifying user passwords, setting user permissions, and editing user initial information.
- (2) Student information management: The basic functions of this module include the basic information of students and the addition and modification of students' English learning behaviors. Teacher users can manage student information in the system, and realize functions such as viewing, adding, and deleting basic student information.
- (3) Course score management: This module contains the functions of viewing English course scores and modifying related information. Users can modify the course information that needs to predict student learning scores.
- (4) Student performance prediction: This functional module is mainly to predict the statistical information of students' English performance. This module is the core module in the system, which realizes the function of analyzing and predicting students' answer data. Users can view relevant information about student answer data, and make targeted English learning plans for students. This is to predict the specific questions that the students answer correctly, so the accuracy of the prediction is very high, so that educators can more accurately know the students' English learning situation.
- (5) Academic performance warning: The function of this module includes the addition and modification of warning conditions. The system visually displays the student's status and learning change trend in the most recent period of time by curving the daily data of students, and sets an early warning value. Once a student's performance reaches the bottom, the system automatically sends an early warning message to remind the student. The LP warning function obtains the result through a comprehensive analysis of the student's online English education course performance, and sends a warning to the student by email or text message.

2.3 Online Education English Learning Performance Prediction Algorithm

The online education English LP prediction algorithm mainly analyzes the students' recent English answer data, and then predicts the next LP. Therefore, it is necessary to calculate the student's proficiency in online English education courses, as shown in formula (1):

$$P_j(\theta_i) = P(X_j = 1|\theta_i\alpha_k) = \frac{1}{1 + e^{-(\theta_i - \sum_{k=1}^K q_{jk}\alpha_k)}} \quad (1)$$

Among them, θ_i is the proficiency of student i , K is the total number of problem-solving skills in the transmission model being used; q_{jk} is the item of the transmission model; α_k represents the difficulty of skill k and similar problems, and a higher value means more difficult Skills. In other words, the probability of a special reaction between a special individual and a special topic. Whether an individual student can answer a question correctly depends entirely on the relationship between the student's ability and the difficulty of the question.

3 Online English Education Learning Performance Prediction System Test

3.1 Experimental Environment

Experimental operating hardware: PC main performance CPU-i5, 3.20 GHz, memory 8G;

Experimental system: the system is windows 764 bit;

Experimental tool: Matlab.

3.2 Experimental Sources

The data set of this research was collected by Kalboard360 Learning Management System (LMS). The LMS system provides learners with simultaneous access to educational resources from any device with an Internet connection, and can collect learner-related data with the help of system tracking tools. In this study, a total of 490 students' online English learning data were collected, including 241 men and 249 women.

The target of the learner's academic performance prediction is the learner's academic performance, so this experiment uses the English answer data of the students before the fifth time as the attribute, and the fifth academic performance as the prediction object (the fifth time the student is known in the data set academic performance, listed as unknown data in the experiment). According to the academic performance of the learners in the data set, in this experiment, the predicted performance of the learners is divided into three levels, 0~70 is Low, 71~89 is Middle, 90~100 is High, using L, M, H respectively mark.

Table 2. Confusion matrix of classification results

Confusion matrix		forecast result	
		Positive	Negative
The true situation	Positive	TP	FN
	Negative	FP	TN

3.3 Evaluation Indicators

This article uses accuracy (precision, also referred to as precision), recall (also referred to as recall), F1 value and other three parameters as the evaluation criteria of the experiment. When mentioning these three parameters, it involves the concept of confusion matrix, which is a tabular form of classification results, as shown in Table 2.

TP defines the number of positive samples that are correctly predicted; FN defines the number of negative samples that are incorrectly predicted; FP defines the number of positive samples from incorrect predictions; TN defines the number of negative samples from incorrect predictions quantity.

(1) Accuracy

The accuracy rate represents the ratio of the number of accurate predictions of online English education learners’ academic performance to the number of true examples in the test set. On the basis of the experiment of predicting learner performance in this research, according to the above confusion matrix, we can know the accuracy rate P as shown in formula (2):

$$P = \frac{TP}{TP + FP} \tag{2}$$

(2) Recall rate

The recall rate represents the proportion of the number of samples that are correctly classified and that are positive to the number of samples that are all positive in the real situation, as expressed in formula (3).

$$R = \frac{TP}{TP + FN} \tag{3}$$

The greater the accuracy and recall values, the better the effect.

(3) F1 comprehensive evaluation index

The F1 value is the harmonic average of the accuracy rate and the recall rate, as expressed in formula (4).

$$F1 = \frac{2 \times P \times R}{P + R} \tag{4}$$

In the experiment, the value of F1 comprehensive evaluation index corresponds to the quality of the system.

4 Results of Online English Education Learning Performance Prediction

4.1 Forecast Results

In this experiment, 490 pieces of data are divided into 350 pieces of training set and 140 pieces of test set. Then predict the three levels of learning scores as Low, Middle, and High respectively, and take the average of the prediction results, as shown in Table 3: The three evaluation results of high-level LP prediction are 85.84%, 66.67%, and 75.08%; the three evaluation results of middle-level academic performance prediction are 79.43%, 60.23%, and 68.50% respectively; the three evaluation results of low-level academic performance prediction are 87.69%, 66.67%, and 75.71%.

Table 3. Predicted results of online English education LP

Hierarchical classification	P	R	F1
H	0.8584	0.6667	0.7508
M	0.7943	0.6023	0.6850
L	0.8769	0.6667	0.7571
Avg	0.8547	0.6474	0.7367

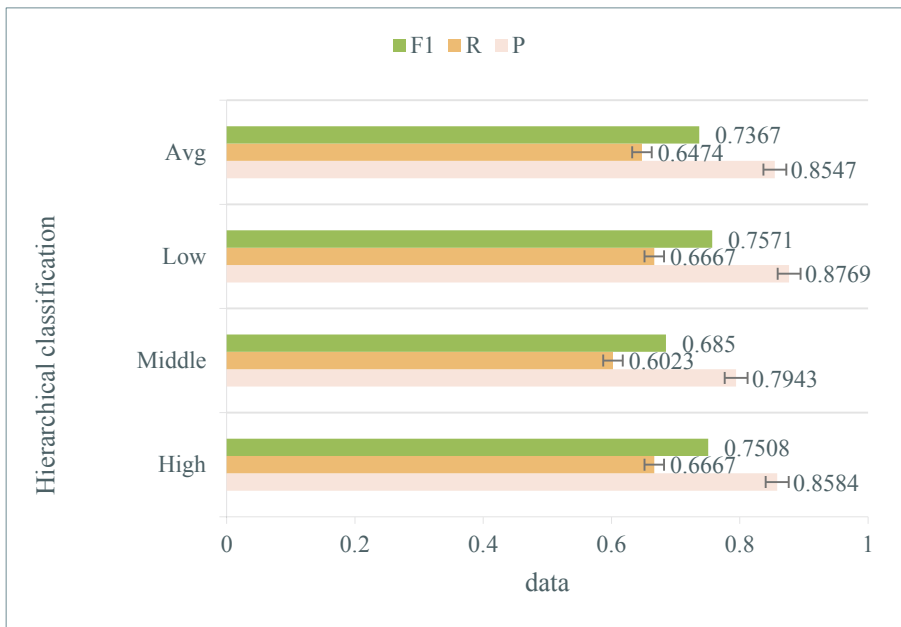


Fig. 1. Predicted results of online English education LP

It can be seen from Fig. 1 that the system has the highest prediction accuracy rate for learners with a low academic performance, which is 87.69%, followed by a high academic performance, which is 85.84%, and the worst prediction effect is Middle, which is 79.43%. The recall rate for the academic performance of Low and High is the same, which is 66.67%, which is greater than the recall rate for the academic performance of Middle. And from the data in the figure, it can be seen that the average accuracy rate in the prediction accuracy rate is 85.47%, and the recall rate is 64.74%.

4.2 Comparison of Prediction Results of Different Algorithms

In this experiment, the same data set is selected, and the performance prediction algorithm of this system is compared with Decision Tree (DT) and Support Vector Machine (SVM), and the results are shown in Table 4.

It can be seen from Fig. 2 that in the three sets of control experiments, the system in this paper has a clear advantage in predicting accuracy, with the highest accuracy rate reaching 87.94%, followed by the support vector machine SVM algorithm, with the highest accuracy rate reaching 73.41%, The next is the decision tree with lower accuracy, the highest accuracy is 72.13%. On the whole, the classification accuracy and recall rate

Table 4. Comparison of prediction results of different algorithms

	P	R	F1
This article	0.8794	0.6667	0.7582
DT	0.7213	0.6474	0.6824
SVM	0.7341	0.6124	0.6678

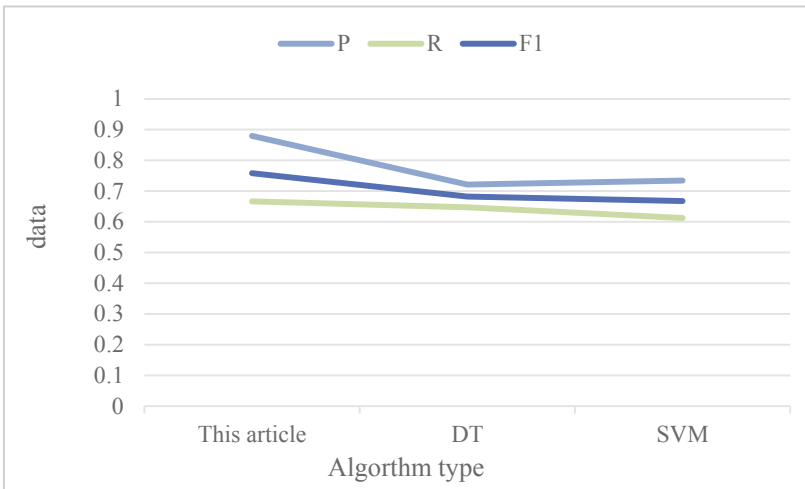


Fig. 2. Comparison of prediction results of different algorithms

of the two traditional machine learning algorithms are lower than the system in this paper. This shows that, in terms of prediction accuracy and time complexity, the online English education LP prediction system is of great research value and significance.

5 Conclusions

The development of big data has promoted the pace of education to digital and intelligent transformation. The use of large-scale digital data in online English education can effectively improve the teaching quality of English online education. The use of relevant data mining techniques in student performance management to extract useful knowledge and rules to continuously improve the level of teaching management is the development trend of school teaching management today. This research aims to establish a network English education learning performance prediction system by analyzing students' English answer data, and the system's learning performance prediction results have a high accuracy rate.

References

1. Li X. Characteristics and rules of college English education based on cognitive process simulation [J]. *Cognitive Systems Research*, 2019, 57(OCT.):11-19.
2. Nie Z. A Study on the Application of Continuing Education Mode in College English Teachers in Internet+ Era [J]. *Boletin Tecnico/technical Bulletin*, 2017, 55(11):504-511.
3. Polyzou A, Karypis G. Feature Extraction for Next-Term Prediction of Poor Student Performance [J]. *IEEE Transactions on Learning Technologies*, 2019, PP (99):1-1.
4. Enhanced Harris Hawks optimization as a feature selection for the prediction of student performance [J]. *Computing*, 2021, 103(7):1417-1438.
5. Saito T, Watanobe Y. Learning Path Recommendation System for Programming Education Based on Neural Networks [J]. *International Journal of Distance Education Technologies*, 2020, 18(1):36-64.
6. Bao Y, Faming L U, Wang Y, et al. Student Performance Prediction Based on Behavior Process Similarity [J]. *Chinese Journal of Electronics: English Edition*, 2020, 29(6):9.
7. Yunanto A A, Herumurti D, Rochimah S, et al. English Education Game using Non-Player Character Based on Natural Language Processing [J]. *Procedia Computer Science*, 2019, 161:502-508.
8. Mahapatra S S, Khan M S. A neural network approach for assessing quality in technical education: an empirical study [J]. *International Journal of Productivity & Quality Management*, 2017, 2(3).
9. Barata G, Gama S, Jorge J, et al. Early Prediction of Student Profiles Based on Performance and Gaming Preferences [J]. *IEEE Transactions on Learning Technologies*, 2017, 9(3):272-284.
10. Deo RC, Yaseen ZM, Al-Ansari N, et al. Modern Artificial Intelligence Model Development for Undergraduate Student Performance Prediction: An Investigation on Engineering Mathematics Courses [J]. *IEEE Access*, 2020, PP (99):1-1.
11. Bao Y, Lu F, Wang Y, et al. Student Performance Prediction Based on Behavior Process Similarity [J]. *Chinese Journal of Electronics*, 2020, 29(6):1110-1118.
12. Yasmin M, Yasmeen A. Viability of outcome-based education in teaching English as second language to chemical engineering learners [J]. *Education for Chemical Engineers*, 2021, 36:100-106.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

