



Design and Application of English Learning Platform Based on Data Mining

Yongjuan Chu^(✉)

School of Foreign Languages, Changchun Institute of Technology, Changchun 130012, China
0217053@ccit.edu.cn

Abstract. In order to solve the problem of low overall quality of online English learning platform, a design and application of English learning platform based on data mining was proposed. This paper applies data mining technology to English online learning platform, analyzes learners and learning content through clustering algorithm and association rules, and develops a content organization system based on genetic algorithm, which is applied in this case to provide personalized learning content for learners, hoping that the system can be extended to other online learning platforms in the future. It provides scientific and forward-looking learning guidance for students' learning, overcomes the shortcomings of the previous English learning platform that is only used to view and download information, and is more conducive to the cultivation of students' autonomous learning habits.

Keywords: Data mining · English learning · system design

1 Introduction

The rapid development of the Internet has spurred changes in traditional teaching methods, and the use of online education has become popular. However, the current online learning platform is mainly based on technology such as downloading and uploading materials, lesson plans, and watching video lessons, and there is no communication between teachers and students, teachers can monitor students' progress, efforts, and learning results, and strategies for students research is planned [1]. With the advent of computers, the administration of colleges and universities across the country has mushroomed, and these systems have brought convenience to teachers, students, and educational leaders. As time increases, more information such as teacher and student information, student selection information, student level, etc. will be accumulated in the system. Behind this big data, there is usually some kind of connection that can be useful for decision support and educational analysis. Data mining is a technique used to analyze unique text hidden in multiple files [2].

2 Overview of Data Mining

Data extraction refers to the process of extracting hidden information from multiple files. This raw data can be structured or semi-structured, such as text, images, and pictures. In addition, there are networks without stable data. In addition, there are networks

without stable data. Data mining technology enables data mining in various ways such as statistics, analysis, processing, data recovery, machine learning, experts and known patterns. Currently, data mining technology is widely used and implemented in many fields [3]. A lot of information has been collected as education has developed rapidly. To improve the quality of education, using information technology to find problems in education and find some rules and solutions has become the study of researchers and experts. According to the purpose of this article, the author researches and analyzes the current state of information technology used in educational platforms, mainly checking the information requested from our database only on technology technologies, including cluster analysis, federal policy analysis, and genetic algorithm analysis.

Nowadays, the English learning system of many online education colleges or English learning institutions in society is that students choose their own courses and can learn by paying fees. In the English learning system, most courses are presented by doing various exercises, so that students will accumulate a lot of performance data in the learning process. These data are stored in the database, but they are not effectively used. Teachers or teaching administrators only get a lot of superficial information through simple statistics or sorting, which may be due to the lack of information awareness or information technology, which leads to the knowledge hidden in these massive data can not be fully and effectively used. How to make full use of these learning data and turn the existing data into effective knowledge will not only help network teachers to develop English online courses reasonably, but also help teachers to better understand the learning effect of students and improve the quality of English online learning. These are all issues that we need to consider urgently [4–6].

3 System Design

3.1 System Architecture Design

The system is web-based and uses B/S mode and MVC architecture. The advantage of MVC architecture is that it is “divide and conquer” and can be reused in many places at once. V represents user interaction; M is a model for creating business processes/state operations and business rules. The model receives the requested data from the view and returns the final result. C is a model-matched control that receives user requests and executes user requests [7].

3.2 System Structure Flow

Historical data collection → transformation → integrated data; Display interface (configuration function, display back) ← → Only data module/statistical data. Currently, many English language learning platforms present the learning content through exercises. Since this content is stored in a database and presented to students in the form of a website, it is very important to define the characteristics of the course content. In conjunction with the analysis and design of the first two sections of this chapter, the learning content standards identified in this document are: Number, type, level, breadth, and content, as shown in Table 1. The above indicators are the organization of the learning content, the problem we solve is to find the best for all learners together [8].

Table 1. Indicators of learning content

Order number	Index name	Indicator instructions
1	Number	The number of the learning content and the unique identification of the content
2	Type	The type of learning content
3	Grade	The English level of the learning content
4	Scope	Identify which English module the learning content belongs to
5	Subject	About what the topic of the learning content

3.3 Implementation of System Function Modules

(1) User module

The users of the system are limited to the educational administration personnel, so there is no need to use the role design. This module is only responsible for the management of adding, deleting users and modifying passwords.

(2) Data statistics module

Based on the system operation, the author defined the system architecture, which generally includes four levels: the content library layer, the student information layer, the organizational structure, and the content display layer. Figure 1 is a schematic diagram of the organizational structure [9].

The main realization is to make statistics on the relevant data in the data warehouse and then display it to users with Jfreechart. JFreeChart is an open charting class library on JAVA platform, which can generate pie charts, bar charts, scatter charts, timing charts, Gantt charts and so on. Through the display of pictures, users can obtain information intuitively. The statistical factors considered and realized in the system mainly include students' information, teachers' information, usual English scores, CET-4 and CET-6 test scores and English learning time at CET-4 reference time. The names of the tables are Listening_content, Speaking_content, Reading_content and Writing_content, and the structures of the four tables are the same, as shown in Table 2.

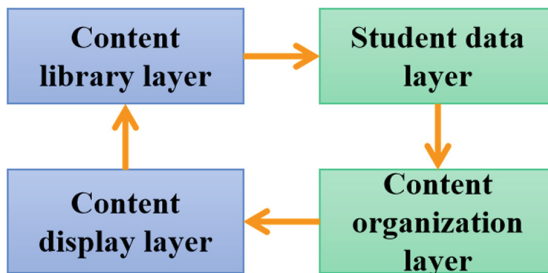


Fig. 1. Architecture of the system

Table 2. Table structure of content library

Field name	Data type	Describe
Id	Automatic number	Content number, primary key
Type	Text	Type of exercises
Rank	Text	The level of the exercises is given
Scope	Text	The exercises belong to the module
Subject	Text	The topic expressed in the exercises

This paper establishes English learning platform based on data mining algorithm. Apply data mining algorithms to English learning platforms, and use group behavior in decision-making. In biology, there is a term called co-evolution, which is the gradual movement of a group to a better state through synergy. As a recommender system, find out the user's favorite option from the available decision options.

The construction of the information resource content of English learning platform is mainly composed of explicit information and implicit information. Explicit knowledge specifically includes software production tools, hardware production tools, scientific research materials, etc. Tacit knowledge is the basis of innovation knowledge innovation, so it is necessary to promote users to make their tacit knowledge explicit through incentive mechanism, which is conducive to the updating and accumulation of platform information resources.

(1) Users of English learning platform

Users are not only the acquirers of resources, but also the components of resources. Cross-border cooperation between users provides human resources for collaborative innovation.

(2) Interactive system of English learning platform

The interactive system is composed of a series of related technologies and software and hardware. It plays a vital role in the improvement and integration of platform information services and the efficiency of information exchange.

After using the platform, the following experiments were carried out to investigate the data mining capabilities of English Innovations. The time interval and training set size of English Innovation data collection are set to 100 and 600 respectively, and the propagation length and intelligent scheduling symbol width of English Innovation data are set to 2100 and 0.15 ms. In order to test the data processing ability of the data mining algorithm, 10 million archive data related to innovation and entrepreneurship data, such as main data type, document data, archive data log data, and user push data, were selected. Platform operation requests select 10 power boost modes, and processing request data selects 10 power boost modes. Figure 2 shows the performance test results of innovation and entrepreneurship information exchange data processing.

Analyzing the data in Fig. 2, it can be seen that as the amount of data processed by batch requests for innovation and entrepreneurship archives in colleges and universities

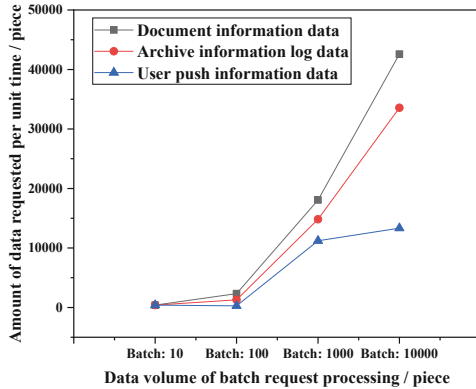


Fig. 2. Data processing performance test results of this platform

gradually increases, the number of requests that the data mining algorithm can support per unit time also increases. Also significantly increased, parallel data connections work better.

For the operation of english learning platform, it is necessary to ensure that the response time of the system is within a certain range, and the access of a large number of users cannot be delayed too much. Otherwise, users may leave the website due to poor user experience, so the system needs to be tested for a corresponding time. This part of the test is mainly conducted in three aspects: browsing preference recommendation response time, similar information recommendation response time and related product recommendation response time. The test results are shown in Table 3.

The sparsity of the user evaluation matrix can be reduced to a certain extent by using the data mining algorithm based on the user characteristic data of this topic. The sparsity of the traditional data mining algorithm is about 86.3%, and the sparsity of the matrix is reduced to 71.2% after the improvement of the algorithm, as shown in Table 4:

Table 3. Response time test

Test type and times		Min (SEC)	Max (SEC)	Average value (SEC)
Similar product recommendation response	1	0.93	4.15	2.55
Time test	2	0.94	4.89	2.92
Related product recommendation response	1	0.82	5.13	2.98
Time test	2	0.91	4.99	2.94
Browsing preference recommendation response	1	1.33	5.68	3.52
Time test	2	1.25	5.12	3.18

Table 4. Comparison of matrix sparsity

Recommended algorithm	Sparsity
Traditional data mining algorithm	86.4%
Data mining algorithm based on user characteristics	71.3%

Table 5. Test comparison of training set and test set

Training set/test set	Traditional algorithm	Optimized algorithm
20%	0.878	0.965
30%	0.865	0.958
40%	0.864	0.950
50%	0.848	0.948
60%	0.837	0.931
70%	0.824	0.912
80%	0.819	0.905

At the same time, different proportions of the training set and the test set are tested [10]. In this test, 20–80% are selected for testing, and the number of similar neighbors is 30. The results are shown in Table 5.

Through the above tests, it can be seen that the currently selected data mining algorithm and its optimization steps are effective and can be improved based on the traditional data mining algorithm to help the recommendation system generate better results.

4 Conclusion

In the process of excavation, it is found that there is not a strong correlation between completing college English learning and passing CET-4; There is a correlation between listening and writing translation in CET-4 and CET-6. If the score of Band 4 is between 425 and 500, Band 6 is not easy to pass; There is a correlation between boys and difficulty in passing Band 4. The main determinant of the passing of CET-4 is the usual grades, and the secondary factor is the situation of teachers. The main determinant of CET-6 passing is CET-4 scores, and the secondary factor is students. It should be noted that the results of the above mining will be different due to the differences of data samples, and the results of changing the settings of support and confidence will also vary. If the data sources are further expanded, more comprehensive conclusions about English teaching will be drawn.

Acknowledgments. Jilin Province Higher Education Teaching Reform Project; Jilin Province Educational Science (Project number: GH20423).

References

1. Luo, X. (2021). Cloud classroom design for english education based on internet of things and data mining. *Hindawi Limited*, 19(4), 205-217.
2. Sun, H., Yao, Z., & Miao, Q. (2021). Design of macroeconomic growth prediction algorithm based on data mining. *Mobile Information Systems*, 2021(7), 1-8.
3. He, Y., Chu, Y., Song, Y., Liu, M., Shi, S., & Chen, X. (2022). Analysis of design strategy of energy efficient buildings based on databases by using data mining and statistical metrics approach. *Energy and Buildings*, 258, 22(7), 262.
4. Jin, X., & Envelope, H. (2022). Research and implementation of smart energy investment and financing system design based on energy mega data mining. *Energy Reports*, 31(14), 288.
5. Zhu, Z., Bi, C., Zheng, R., & Zheng, S. (2021). Design and application of distributed computing environment based on data drive. *Journal of Physics Conference Series*, 41(15), 301.
6. Zhang, C., & Ding, B. (2022). Feature extraction of color symbol elements in interior design based on extension data mining. *Mathematical Problems in Engineering*, 110(32), 452.
7. Wang, P., & Li, J. (2021). Implementation of real-time medical and health data mining system based on machine learning. *Hindawi Limited*, 51(39), 654.
8. Li, H. (2021). The design and development of a ship trajectory data management and analysis system based on ais. *Sensors*, 157(414),745-755.
9. Li, K., & Chen, Y. (2022). Fuzzy clustering-based financial data mining system analysis and design. *International Journal of Foundations of Computer Science*, 33(06n07), 603-624.
10. Xia, H., Wang, P., Wan, T., Zhang, Z. J., Weng, J., & Jasimuddin, S. M. (2022). Peer-to-peer lending platform risk analysis: an early warning model based on multi-dimensional information. *The journal of risk finance*, 212(54), 633.

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.

