Application Analysis of Data Mining Technology in College English Online Learning Platform

Juan Chen

Sichuan Vocational and Technical College, Suining, Sichuan, China.

How to cite this paper: Juan Chen. (2023). Application Analysis of Data Mining Technology in College English Online Learning Platform. The Educational Review, USA, 7(5), 652-658. DOI: 10.26855/er.2023.05.021

Received: May 10, 2023
Accepted: June 8, 2023
Published: July 6, 2023

Corresponding author: Juan Chen, Sichuan Vocational and Technical College, Suining, Sichuan, China.

Abstract

In order to solve the shortcomings of the existing college English online learning platform research, we use data mining related technologies to analyze students' learning tendency, learning achievements and learning content organization in online learning platform, and optimize teachers' teaching mode and teaching effect analysis. This paper discusses the design principles and data mining technologies of the college English online learning platform, and then briefly discusses the data collection, preprocessing and data sources of the application of data clustering and association rules in the platform. By using data clustering and association rules to analyze learner clustering and learning content association, it is found that different students have high, low and medium levels in English listening, speaking, reading and writing, and some students have English scores above 65. The results of correlation analysis are used to recommend the content of students' English sections and predict their scores.

Keywords

Data Mining Technology, Cluster Analysis, College English, Online Learning Platform

1. Introduction

The learning of college English can enable students to communicate effectively in English in the workplace and daily communication. At the same time, students' comprehensive quality can be improved through the learning of this language, so as to meet the needs of social development and personal development.

Nowadays, more and more scholars have done a lot of research on the online learning platform of college English through various technologies and system tools, and have also made some research achievements through practical research. Nti I K uses big data mining technology in the mobile learning system. Through (EM) algorithm, cluster analysis is conducted on the learning situation of platform students. This technology is used to provide students with personalized learning content recommendations and predict their future performance through their past abilities. The CRISP model is used to build the data mining model of the learning system. It also analyzes the number of English courses and other courses in a week. The analysis shows that 60% of students have reduced the number of English mobile learning and increased the number of other courses. Through the analysis results of this technology, we can create and recommend course content that meets the needs of students (Nti & Quarcoo, 2022). Based on the online learning platform of continuing education and taking College English as an example, Thesini analyzes learners' English learning habits and effects by using data mining technology. The data show that students' test scores and usual scores are positively correlated with the length of online learning, while majors and genders are less correlated. In addition, gender is not related to the comprehensive score. Based on the above analysis, it reflects the shortcomings of the curriculum design, and puts forward some suggestions, which provide a basis for better understanding the online learning effect of adult learners (Thosini & Bamunu, 2019). Moodie I believes that in online intelligent learning, big data is data about students' habits,
cognition and tendency in the learning process, which is diversified and dynamic. Therefore, the author uses big data technology to analyze various evaluation methods and effects of students' learning effects. This research provides a more scientific and effective basis for students' online learning through evaluation and analysis of students' online learning process (Moodie & Meerhoff, 2020). Although the existing research on college English online learning platform is very rich, there are still some limitations in the application of data mining technology.

In order to solve the problem of single teaching content in the current English online learning system, from the perspective of data mining, this paper develops an online learning system that provides personalized learning content for college students, which can provide effective ways for teachers to teach. This paper specifically studies the following contents: providing clustering analysis, clustering learners, determining their English grades, and providing guidance for teachers to design teaching content; Analyze the association rules of the learning content. This paper mainly analyzes the relevance of the main contents of English learning, so as to obtain the association rules between the contents.

2. Design and Analysis of Data Mining Technology in College English Online Learning Platform

2.1 Design Principles of College English Online Learning Platform

The design of an online learning platform that satisfies teachers and students should at least follow the following principles:

(1) Educational principles: the functional design and presentation of online learning platform for college English should be based on online education and students' actual needs, and based on the curriculum principles of various schools and national education goals (Concrete, 2018).

(2) Support multiple learning modes: mainly including allowing students to conduct individual and cooperative learning in the platform, so as to ensure the flexibility and dynamics of the online learning platform (Yw, 2022).

(3) Interactive principle: the principle mainly includes: first, learners can conduct platform learning interaction at any time regardless of place and time; The second is the interaction between different learners and learners and teachers (Rubber, 2022).

(4) The principle of maintainability: when there are changes in the content of curriculum implementation and students have new ideas and needs, they can make timely statements (Concrete, 2018).

2.2 Data Mining Technology

(1) Association rules: analyze the internal relationship of data, find out the relationship between various data items, and give a quantitative result. Use association rules to analyze students' course selection records, and find the interdependence between different courses within each student's course selection records (Vetter, 2022).

(2) Classification algorithm: mining and analyzing the classification rules according to the training data set, then classifying other data of the same type, and providing valuable guidance for the business system (Al-Surmi, 2022). The processing flow of the classification algorithm can be roughly divided into the following steps:

1) Select a complete and symbolic data set, and the characteristics of its original data are consistent with the target data set. Different types of data shall be marked (Hammerschmi, 2021).

2) The training data set is mined and analyzed, and the eigenvalues of each record are calculated to generate classification rules of each type. This rule cannot be achieved overnight. It needs to be tested, adjusted and optimized repeatedly until it meets the requirements (Osadchyi, 2019).

3) Classify the target dataset according to the classification rules, and generate useful guidance rules (Yoshimura, 2018).

(3) Statistical analysis

In the process of data mining, there are certain relationships between superior data, and these relationships can be expressed by functional formulas (Valle, 2018). There are also some connections that cannot be expressed by functional formulas, which are called correlation (Nguyen & Loan, 2018). These correlations can be analyzed by using relevant techniques in statistical analysis, such as regression analysis, principal component analysis, and correlation analysis.

3. Investigation and Analysis of Data Mining Technology in College English Online Learning Platform

3.1 Data Collection and Preprocessing

(1) Cluster analysis data collection and preprocessing
1) Data collection: This paper will collect from the English online teaching system of a university, and conduct clus-
ter analysis on the language competence test data of nearly 200 learners (Preeti Jha, 2021). The ability test consists of
four modules: listening, speaking, reading and writing. The average score of each unit is calculated according to the
percentage system. In addition, this paper is also the data base of clustering mining results from the basic data of nearly
200 learners, covering age, gender, reasons for selection, and majors.

2) Data preprocessing: For college learners learning comprehensive English, the comprehensive English is divided
into six levels, namely novice level, elementary level, intermediate level, intermediate level, advanced level and proficient
level. Therefore, this paper sets the number of clusters as 6, representing the above six English grades (ADHK,
2018).

2) Data collection and preprocessing of rule association data
1) Data collection

<table>
<thead>
<tr>
<th>Student ID</th>
<th>Gender</th>
<th>Major</th>
<th>English level</th>
<th>Ability to improve</th>
<th>Hearing</th>
<th>Spelling</th>
<th>Reading</th>
<th>Writing</th>
<th>Fraction</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Male</td>
<td>Computer</td>
<td>Senior</td>
<td>Writing</td>
<td>78</td>
<td>80</td>
<td>85</td>
<td>69</td>
<td>81</td>
</tr>
<tr>
<td>2</td>
<td>Female</td>
<td>Chinese language</td>
<td>Senior</td>
<td>Hearing</td>
<td>54</td>
<td>85</td>
<td>86</td>
<td>89</td>
<td>82</td>
</tr>
<tr>
<td>3</td>
<td>Male</td>
<td>Computer</td>
<td>Senior</td>
<td>Writing</td>
<td>74</td>
<td>76</td>
<td>72</td>
<td>48</td>
<td>77</td>
</tr>
<tr>
<td>4</td>
<td>Female</td>
<td>Chinese language</td>
<td>Senior</td>
<td>Hearing</td>
<td>61</td>
<td>82</td>
<td>86</td>
<td>77</td>
<td>80</td>
</tr>
<tr>
<td>5</td>
<td>Female</td>
<td>Chinese language</td>
<td>Senior</td>
<td>Reading</td>
<td>78</td>
<td>80</td>
<td>56</td>
<td>79</td>
<td>59</td>
</tr>
</tbody>
</table>

2) Data preprocessing: the results of source data are recorded in percentile system. In Clemen Tine program, in order
to realize that the information required by GRT association rule management is in the form of T/F, all original informa-
tion must be converted into a format. Then, VAB programming in Excel was used to complete the unified management
of information.

3.2 College English Online Learning Data Source

1) Server side data: After learners log in to the platform to select different courses for learning, the server side will
use Web logs to collect information, including learners' online time, online duration and other information.

2) Client data: In addition to data collection based on the server, user data can also be collected based on the client. It
can usually be implemented with the help of script language coding.

3.3 Survey Questionnaire

The experiment object selected 40 students from two classes in School A, among which the average English score of
the first class is 62.14, which is set as the experimental class; the average English score of 2 class is 62.46, which is set
as the control class. The control class tested their scores again after three months of learning under the data mining
technology, while the experimental class was tested after three months of normal teaching. In addition, students from
two classes were surveyed with the questionnaire to count their satisfaction with the online learning evaluation. In fact,
80 questionnaires were distributed and 80 were returned, with a recovery rate of 100%. Excluding some invalid ques-
tionnaires (such as for the evaluation content unrelated to the content of the questionnaire), a total of 78 questionnaires
were recovered, with an effective rate of 97.5%.

4. The Application of Data Mining Technology in College English Online Learning Platform

4.1 Cluster Analysis of Learners on College English Online Learning Platform

The application of learner clustering mainly includes the following steps:

1) Input variables: The first step of K-means clustering is to drag the variables to be clustered into the "Variables"
box. In this paper, we only need to cluster the students' listening, speaking, reading and writing scores.

2) Determine the number of clusters: The number of clusters is set to 6 in this paper, representing English entry level,
elementary level, intermediate level, intermediate level, advanced level and proficient level respectively.

(3) Determine clustering methods: K-means clustering methods mainly include iteration and classification and classification only. Iteration and classification method refers to determining the initial category center point first, and then iteratively classifying according to one mean clustering algorithm. The classification only method refers to the classification only according to the center point of the initial category, without iteration. This article uses the former method.

(4) Clustering result output.

(5) Cluster result analysis: The following provides some practical suggestions for the learning system from the perspective of age and gender through the cluster analysis results.

According to the clustering mining results in Figure 1, we can effectively divide students into different grades, and we can plan courses according to that grade. The proportion of each content in an English course does not need to be evenly distributed. For example, in Figure 1, the student numbered 4 has a cluster category of 5. The student is at an advanced level and should be given a higher level of courses. His listening performance is low and his reading performance is high. At this time, teachers should arrange more listening content for him, and writing courses can be reduced.

4.2 Correlation Analysis of Learning Content in College English Online Learning Platform

The application of learning content association mainly includes the following steps:

(1) Use tools for correlation analysis

Open the Clementine software, select the "Data Source" panel in the node option panel, and you can select the imported original data format according to the data storage type used by the research. Clementine mainly provides the following data formats for import:

1) SPSS is used to import files in format.
2) Excel is used to import files in the format.
3) SAS is used to import files in SAS format.

When mining association rules with, a key step is to establish a data stream. The whole process of association rule
analysis can be realized by establishing a data flow. The specific mining process includes: setting field data type, reading data value, missing value processing, and setting field direction.

4) Insert the GRT node to perform association analysis: set Model Name, use Partition Data, set Minimum Condition Support, set Minimum Rule Confidence, set Maximum Top Items, set Maximum Rules, and display only flag variables with true values. Finally, insert the "Network" node to visualize the relationship between the fields. The Network node displays the strength of the relationship between two or more fields.

(2) Association result analysis

Through the mining results of the above association rules of English online learning scores, this paper applies the results to the course recommendation and score prediction of online learning.

1) Course recommendation for English online learning: After association rules are obtained through association analysis, teachers can arrange courses that meet students' needs according to their learning needs, while courses they have mastered well can be provided less or even not.

2) Prediction of academic performance: the association rules between them obtained in this paper can explain this problem. There is a strong correlation between reading ability and oral ability. If reading ability is strong, oral ability will not be weak, and vice versa. In this way, teachers can predict students' learning at any time in the learning process according to the mining association rules, so as to guide and make decisions on students' learning in the subsequent unit learning.

4.3 Questionnaire results

(1) Algorithm validity

Three months after the experimental design, the test results of the two classes were counted, and the statistical results are shown in Figure 2:

![Figure 2. Test results.](image)

As can be seen from the data in the figure, the performance of the experimental class is not very stable, with an average score of 61.19; the control class has improved with an average score of 75.24. Therefore, the application of data mining to the English online learning platform has a great role in promoting students' academic performance.

(2) Student assessment

For the questionnaire statistics, the proportion of students' evaluation scores on online learning platforms was calculated respectively. The results are shown in Figure 3:
As can be seen from the data in the figure, 64.11% gave more than 8 points, more than half, and only 5 points accounted for 10.25%. It further shows that the application of data mining technology to the English online learning platform is very popular, and the students are also very satisfied with the platform.

5. Conclusion

This paper introduces the clustering analysis and association rules in detail according to the design principles of online learning platform and the main data mining methods. The practical application process of clustering analysis and association rules in college English online learning platform is completely realized. It mainly includes determining the definition of the problem, and adopting data preprocessing techniques such as data collection and preprocessing for data preparation. Through the specific steps of clustering analysis and detailed process analysis of association rule method, learner data clustering analysis and learning content association analysis are carried out. Based on the results of the analysis, specific suggestions for online English learning are provided.

References


Yw A, Tn B, Msr A. Evaluation of continued use on Kahoot! As a gamification-based learning platform from the perspective of Indonesia students [J]. Procedia Computer Science, 2022, 197:545-556.