Application of Data Mining in English Online Learning Platform

Ying Sun¹, Yao Jiang²,*

¹Gongqing College of Nanchang University, Jiujiang City, Jiangxi Province, China, 332020
²Gongqing College of Nanchang University, Jiujiang City, Jiangxi Province, China, 332020

*Corresponding author e-mail: sunying@ncu.edu.cn

Abstract. Since the formation of global economic integration, in order to better exchanges and cooperation with countries around the world, master the necessary international common language -- English is very necessary. In the context of today's big era with a variety of English learning methods, it is logical that data mining is applied to the English online platform. As for data mining, it has made remarkable achievements in the commercial field. With the continuous application of data mining technology and the continuous development of online learning system, its application in the field of education is also more and more extensive. It promotes the development of educational decision support model based on evidence and "data". Based on the analysis of the application of data mining technology in the field of education and teaching at home and abroad, this paper summarizes and summarizes the application and development of various data mining technologies in English online learning system. Finally, combined with the actual situation in the field of education in China, this paper puts forward some challenges in the application of data mining technology in this field, and gives the direction and prospect of further development in the future, this paper hopes to improve the learning effect of English online learning platform effectively [1].

Keywords: English Online Learning Platform, Data Mining, Application Analysis

1. Preface

The application and development of data mining integrate the content and technology of many subjects, including database, statistics, machine learning, pattern recognition, artificial intelligence and neural network, etc. It extracts interesting knowledge from large, incomplete, noisy, fuzzy, random data. This knowledge is implicit, unknown to people before it is extracted, but it has its own function. With the development of computer technology and application, data mining has been widely used in banking, telecommunications, insurance, transportation, online retail and other business fields. In these business fields, data mining has made remarkable achievements. For China, the application of data mining in the English online learning platform is indeed an innovation of learning
methods or means. Next, we will analyze the application of data mining in English online learning platform, hoping to improve the learning effect of English online learning platform effectively [2].

2. Cluster analysis application of data mining in platform

Clustering, also known as group analysis, is a statistical analysis method to study the classification of (samples or indicators). Cluster analysis is a process of clustering the data with similar characteristics together by similarity classification of a large number of data. It is based on similarity and there are more similarities between patterns in a cluster than between patterns not in the same cluster. Different from classification, clustering is an unguided learning process, in which the categories of samples are not known in advance, nor the number of categories. In the current English online learning platform, when arranging classes, teachers first divide grades arbitrarily, then the corresponding learning contents are arranged according to the grades of students' actual test results. This method lacks scientific basis and may not be suitable for learners and affect their learning effect. At this point, we can introduce the clustering algorithm in data mining.

Cluster analysis is mainly divided into four parts: data collection, data preprocessing, cluster analysis and result analysis [3].

3. Association analysis application of data mining in platform

Association rules are implication forms such as X-y, in which X and Y are respectively called forerunners and successors of association rules, it reflects the interdependence and relevance of one thing to others. If there is an association between two or more things, then one thing can be inferred from the other things by association rules. Association rule mining techniques are used to mine related rules from sequence items in the user access sequence data base (procedure database). It's one of the most frequently used and widely used data mining technologies, it has been widely used in the decision-making analysis of foreign education and teaching, as shown in Table 1. English learning is a gradual process, and the content of each level has its own sequence. In the learning process, one step at a time must be advanced, otherwise, it will lead to disjointed learning, little effect [4].

| Preamble                 | Succeed                   |
|--------------------------|---------------------------|
| Teaching method          | Student type              |
| Behavioral activity n    | Behavioral activity m     |
| Course type              | Achievements              |
| Teaching effect          | Correlative factors       |
| Course content n         | Course content m student type |
| Book category            | Student type              |

Table 1. Application of association rules in online learning system.

To fully illustrate this kind of application, we will take the learning content of a unit in a certain registration from the database as the association analysis data. Data were collected on the students' basic information, learning purpose, basic English level, listening score, speaking score, reading score, collaboration score, and total score. Among them, the score data is mined as association rules, while the rest data is applied to the analysis of data mining results. Then we preprocessed the data and set thresholds for all kinds of data according to the results of previous cluster analysis to obtain the original data in Table 2.
Table 2. Sampling raw data (5 groups of data).

| Number | Age | Gender | Occupation | Learning purpose | English level | Ability to be improved | Listening Score | Speaking Score | Reading Score | Writing Score | Score |
|--------|-----|--------|------------|------------------|---------------|-----------------------|----------------|----------------|---------------|--------------|-------|
| 1      | 28  | male   | Occupation | Advanced         | Hearing       | 78                    | 81             | 82             | 75           | 79          |
| 2      | 26  | male   | Occupation | Advanced         | Reading       | 72                    | 83             | 75             | 85           | 79          |
| 3      | 29  | female | Occupation | Advanced         | Hearing       | 69                    | 84             | 79             | 70           | 76          |
| 4      | 31  | male   | Occupation | Advanced         | oral English  | 79                    | 77             | 82             | 83           | 80          |
| 5      | 32  | male   | Occupation | Advanced         | Hearing       | 67                    | 82             | 84             | 83           | 79          |

Finally, after association analysis, accurate association rule information geometry can be obtained according to the generated association result network graph, as shown in Table 3.

Table 3. Association rule.

| Rule ID | Association rule                        |
|---------|-----------------------------------------|
| 1       | Reading ←→ total score                 |
| 2       | Reading ←→ spoken language             |
| 3       | Listening ←→ spoken language           |
| 4       | Reading → listening & total score       |
| 5       | Reading → writing                      |
| 6       | Speaking→ listening&total score         |

4. Problems existing in the domestic English online learning platform

Although the application of data mining technology in online learning system has a wide range of prospects and some attempts have been made, it is still a relatively young research field, so far, there are still some challenges. For our country, English teaching and learning in China rise time is shorter, naturally, compared with abroad, the start time of learning English online will also be much shorter, accordingly for this part of the management of our country still has a lot of lack of place, therefore present in front of us online English learning platform overall quality is low [5].

From the perspective of educational and teaching decision-making model, the online learning platform focuses on the theoretical research of learners' learning support, but ignores the importance of practical research. In addition, the data-driven decision support model has not yet formed a habit, and relevant national authorities have not yet formulated the management decisions of online courses according to the specific conditions of online learning platforms. Moreover, relevant courses are not equipped with technical teams specially responsible for development, which leads to the lack of strong
motivation to adopt new technologies in this field, thus hindering the research and development of the application of data mining technology in the learning system.

From the perspective of course selection, we compare overseas and domestic online learning platforms and find that domestic learners do not start from their actual situation when choosing online learning platforms. As a result, in the absence of analysis of their own ability, the arbitrary classification of course levels, lack of scientific theoretical support. It not only consumes a lot of unnecessary time, but also makes scholars feel powerless and give up the idea of learning [6].

5. The development of online English learning platforms in China
There is no doubt that the development of Domestic English teaching platform must be carried out on the basis of solving these problems mentioned above. In view of the above problems of lack of standards and improper selection, the solutions in the author's opinion are given here: Firstly, data standardization should be realized, whose main function is to eliminate dimensional differences between variables, so that data can be comparable. Only by standardizing data and formulating a standardized and unified online course management standard can we better solve various problems encountered in the platform [7].

Secondly, learners' confusion in course selection can be solved by clustering analysis algorithm, that is, learners' actual English mastering ability and level should be evaluated before entering the learning platform. Their different English levels will be established according to their final results, and then English instructors will be arranged to assign different study groups and study rooms according to their English learning levels.

In addition, we can learn from the association rules contained in the data mining and the mining results such as cluster analysis, with the application of modern algorithms, it can help network English teachers to develop tools that can automatically organize learning content according to the actual teaching needs. Moreover, this tool needs to be carried out by teachers in combination with the actual learning situation of learners, so its final appearance in the English learning platform should contain the modification function. This helps to develop different learning contents for different learners at different learning levels and problems existing in learning.

6. Conclusion
With the continuous popularization of computers and the rapid development of multimedia technology, education and teaching mode is no longer just a simple traditional classroom teaching mode, and various online learning systems are more and more popular and valued by everyone. How to apply data mining to this field, mining valuable rules from a large number of data to guide and develop education, promote personalized learning and data-driven learning evaluation, teaching evaluation and educational decision support model. It has gradually become a research hotspot in the field of educational technology and open education. In addition, it is precisely because of the continuous development of The Times that we are more and more closely connected with the world. No matter in politics, economy or culture, we need to communicate. The premise of communication is that everyone uses the same language, which provides conditions for the promotion of English in China. Most of the current English online learning platforms have too single learning content. This paper analyzes the application of cluster analysis and correlation analysis in the data mining of English online learning platforms. It plays a guiding role in improving the rationality of learning content arrangement, diversity of learning content and learners' interest in learning [8].

References
[1] Su Xinning. Data warehouse and Data Mining [M]. Beijing: Tsinghua University Press. 2006.
[2] Wei Shumping. Moodle Platform Data Mining research -- Taking the learning process analysis of an online training course as an example. China Distance Education, 2011(01): 24-30.
[3] Buldu A, Ucgun k. Data mining application on students' data. Procedia Soc Behav Sci, 2010: 5251~5259.
[4] Blei. Dg. A, Jordan.M. Latent Dirichlet allocation. Journal of Machine learning Research[J].2003,3:993-1002. Vapnik V N. The nature of statistical learning theory[M]. New York: Springer, 1995.

[5] Zhang Lulu, Internationalization of Autonomous Learning Platform for College English [J]. Journal of Liaoning Institute of Education Administration, 2017, 34(4):71 73.

[6] Chen Zhongwei, Zhu Jiaquan, Wei Haiqing. Design and Implementation of An Online Independent English Learning platform based on ASP. Net [J]. Computer Optical Disc Software and Applications, 2014(17):257 258.

[7] Ma Chunhua, Lei Nan. Construction of college English Autonomous Learning Platform based on computer Application [J]. Journal of Huaibei Vocational technical College, 2017, 16(1):65-67.

[8] Han Jiawei et al. Data Mining: Concepts and Techniques. Machinery Industry Press, 2012.