Application of Data Mining in English Online Learning Platform

Weiwei Qu*

School of Foreign Languages, Dalian Polytechnic University, Dalian, Liaoning, 116034, China

*Corresponding author e-mail: quweiwei@dlpu.edu.cn

Abstract. English is becoming more and more important in our life and English learning is also conducted anytime and anywhere. With high-tech products more and more popular, learning English through mobile phones and other products is very convenient. There are numerous platforms for English online learning, but they provide a very single learning content. All learners, no matter what their learning purposes are, have the same learning content, thus problems will follow. Based on the analysis of current situation, this paper puts forward solutions, case analysis and conclusion process. The application of data mining technology to English online learning platform provides 80% ideas for the construction of online learning platform. Statistics show that nearly 70 million people study online every year.

Keywords: Data Mining, English Online Learning, Learning Platform

1. Introduction

English is the most widely used language in the world. In our country, English education starts from kindergarten, and English learning has become a compulsory course for students. English is also very important in our work and life. Because of the importance of English, for adults who have joined the work, they can no longer enter the school to learn English, and can only learn English through various platforms, which is convenient for them to learn wherever and whenever they are.

The application of data mining in English online learning platform has attracted the interest of many experts and has been studied by many academic teams. For example, some have found that many employees learn English to meet their personal needs, rather than training for training, but they often see a single, useless learning content on the learning platform, just to complete certain training tasks and random stacking of learning content [1]. Some have found there are fewer and fewer theoretical and practical researches which focus on learners’ learning support service. Many scholars and researchers in China attach importance to the study of learner learning support services and consult the literature. There are countless articles in this respect, but they only stay in the research stage and have little practice. The platform construction of each network college must support the learner's learning support service, but only stay in the slogan stage and the real service is rare. In contrast, the various commercial online learning platforms provide relatively more learning support services because they are for profit. Only by providing more and better services to learners can we...
attract more learners [2]. Many scholars have found that data mining has been effectively applied in
many industries, and its application in education is more and more extensive, but the application in
online learning of English in China is almost zero. Many studies on English online learning do not
mention data mining. As English online learning becomes more and more extensive, a large number of
online learning data are piled up. It is necessary to apply data mining technology to online learning.
Data mining technology can help us to find problems in online learning from the point of view of data,
objectively reflect the problems in online learning platform, and improve the guidance of online
learning quality [3]. Other teams have found that the development of online learning abroad is
relatively mature, the platform of online learning is relatively perfect, the more famous platform is the
online learning platform of British universities, and many domestic scholars and researchers have
studied this more [4]. At the same time, the government, schools, scientific research institutes and
training institutions attach great importance to the construction of online learning platform, fully
analyze the needs of learners, and meet the individualized learning needs of adult learners by
developing rich online learning content. At the same time, the evaluation of online learning is
emphasized to ensure the quality of online learning [5]. Foreign online learning platforms attach great
importance to learners’ learning support services. For example, some online learning platforms have a
module, “Learning support Service”. This module details the content of learning support services
provided by their own platforms. And how different types of learners should learn on their own
platforms [6]. According to the research of foreign scholars, learning support services can be divided
into academic support services and non-academic support services. In terms of academic support,
foreign online learning platforms mainly provide services on the problems encountered by learners in
course learning. These services are very mature abroad and are recognized by many online learners [7].
Although their research results are very rich, there are still some shortcomings.

After more than ten years of development, data mining technology has made a lot of research
results in foreign countries. At the same time, more and more large and medium-sized enterprises have
begun to use this technology to analyze and excavate the current situation of their own companies.
Assist in decision-making on major issues. In China, data mining has gradually changed from simple
research to comprehensive research. In the application stage, the demand for data mining technology
in China is increasing. However, the application of data mining technology in the field of education,
especially in English online learning platform, still has a lot of room for growth. This paper analyzes
and discusses the application of data mining in English online learning platform.

2. Method

2.1. Calculation of Fitness Values
In genetic algorithms, the size of fitness functions is often used to evaluate the advantages and
disadvantages of individuals in a population. The fitness function is obtained by the transformation of
objective function. The larger the value, the better the individual. The fitness function is formula (1),
where the ei is the deviation between the expected value and the actual value distribution of the
attributes such as the range of numbering type and the subject discussed above. wi is the proportion of
each deviation. It can be determined from the formula that when the constraint error of the content
individual to the content organization is small, the larger the fitness value is, which indicates that the
extracted content individual is closer to the content organization [8].

\[ F = \frac{1}{1 + \sum_{i=1}^{5} w_i e_i} \]  

(1)

2.2. Differential Coefficient Analysis
Through the amount of difference in the data sample, the difference reflects the trend of population
separation, That is, the degree of differentiation. This coefficient reflects the different needs of
students in foreign language teaching, Through the coefficient of difference, Weighted by standard
deviation and average, Using the average score as a reference to the difference, (2) is the accounting formula for the coefficient of difference (CV is the coefficient of difference, S standard deviation, V average. Experience shows that, CV values typically range from 5% to 35%, If >35%, It may question whether the average is meaningless; If <5%, The question is whether the value of accounting is wrong, In educational evaluation, Teachers and school administrators also need to analyze differences, To judge the learning differences between students in different subjects and the same subject. Empirical markers of differentiation: if CV<9%, This means little differentiation; If CV>20%, indicating severe differentiation; If 9%<CV<20%, Indicating signs of differentiation) [9]:

\[
CV = \frac{S}{V} \times 100\%
\]  

(2)

2.3. Assessment Scale Method
In the investigation, the five-segment evaluation is called the evaluation scale method. Through the meaning of graph structure, the five-segment evaluation data are analyzed by structural analysis method, and the structural equation is established. The relationship between potential variables such as learning personalization and learning satisfaction is described by structural equation. The mathematical representation of the measurement model is [10]:

\[
X = \lambda_x \xi + \delta
\]

(3)

\[
y = \lambda_y \eta + \epsilon
\]

(4)

According to the task requirements, for mining the strong association rules of each functional module in the test version of the mobile learning platform, it is necessary to use the Apriori algorithm to collect and organize the data before mining. So the minimum support counting formula is (5), and the minimum credibility is 80.

\[
1346 \times 10\% = 135
\]

(5)

3. Experiment
3.1. Source of Experimental Data
The main research object of this paper is adults, who learn English for work needs, or for life needs, etc. Most of the forms of learning are online learning in training institutions or online training platforms within enterprises. The total number of questionnaires issued was 3000, 2600 questionnaires were collected. According to the principle of complete and accurate information, 2531 valid questionnaires were entered into the Excel form. The CronbachAlpha0.938, AKMO value of SPSS22.0, reliability test was 0.955.2531 questionnaires Bartlett qualified have reliable reliability and validity.

3.2. Experimental Design
Collection of data, data preprocessing, analysis selecting, results analysis.

4. Result
4.1. Establishing Learning Content Indicators
At present, the learning content of English online learning platform is single, all learners learn the same content, and do not provide learning content according to their own needs, which leads to the poor learning effect of many learners. Therefore, this paper hopes to provide a set of suitable learning content for each English learner, which can help them master English. At present, the learning content of many English learning platforms is displayed through exercises, which are stored in the database.
and presented to students in the form of websites, so it is very important to define the attributes of learning content. Combined with the analysis and design of the first two sections of this chapter, the learning content index system defined in this paper includes the number, type, grade, scope and theme, as shown in Table 1 below. The above indicators are the key to organizational learning content, and the problem we need to solve is to find the optimal combination for each learner.

**Table 1. Indicators of learning content**

| Order number | Indicator name | Indicator note                                      |
|--------------|----------------|---------------------------------------------------|
| 1            | Number         | Number of learning content, unique identity of content |
| 2            | Style          | Types of learning content                         |
| 3            | Grade          | English Level of Learning Content                 |
| 4            | Scope          | Identify which English module the learning content belongs to |
| 5            | Theme          | Content on what topics                            |

(1) The number is the number of learning content in the database, and the number of each learning content in the content library is unique. (2) For example, the types of exercises in English learning. For example, the case of this paper is EF online learning system. Therefore, for the learning content of this system, the specific types of exercises can be divided into writing questions, oral exercises, matching questions, sorting questions. (3) The application of score index mining in English online learning platform Chapter 1, the application design level of data mining in English online learning platform explains which English level each learning content belongs to. (4) Scope refers to the specific learning content belongs to which English module of listening, speaking, reading, writing, different learning content to cultivate learners' different English ability. (5) Topic refers to the specific learning content is about which topic exercises, such as reading comprehension of climate topics. The content of this paper is defined as the student career scope defined EF online learning.

4.2. Collecting Data

In this paper, the data of students' level test are extracted from EF online learning platform for cluster analysis. The level test is divided into four modules: listening, speaking, reading and writing. At the same time, the data of comparative analysis value are extracted from the basic data as the basis of cluster mining results, including age, sex, occupation, position and learning reasons. Because the research in this paper is mainly online learning, the subjects are adults, the age limit is over 20 years old, occupation, position and learning reasons are based on EF English online learning platform.

**Table 2. Basic information on students**

| Attribute name     | Data type       | Span                        |
|--------------------|-----------------|-----------------------------|
| Age                | Number(10)      | >20                         |
| Sex                | Char(15)        | Male or female              |
| Occupation         | Char(100)       | Accounting/auditing/advertising/marketing/PR/construction, automobile, banking or finance, biotechnology, construction engineering |
| Position           | Char(100)       | Staff, Senior Staff, Manager, Senior Manager, Administration, Senior Administration, others |
| Reason for learning| Char(100)       | Social, professional, academic, travel, other |

4.3. Data Mining

As the first step of the Apriori algorithm, we first perform technical statistics on each item set of previously preprocessed data sources, and the results are shown in figure 1 below.
4.4. Differentiation Coefficient

Because the index distribution cannot reflect the individual's clear cognition of English online learning, it is only in the stages of learning knowledge, learning environment conditions, learners' knowledge, skills and abilities, learners' motivation and so on. Therefore, the difference coefficient analysis of the survey data of the above four standard items is carried out, and the results are shown in Figure 2 below.

**Figure 1.** Statistical counts for all itemsets

**Figure 2.** Differentiation coefficient table
5. Conclusion
Aiming at the problem that the current English online learning system only provides a single learning content, this paper develops a tool to provide individualized learning content for adult learners from the point of view of data mining, and provides guidance and help for teachers’ online teaching. The main research results of this paper are as follows: according to the clustering analysis algorithm, the evaluation results of learners before learning are analyzed, the learners are clustered, their English scores are determined, and the teachers are instructed to arrange the learning content for them. Assign study groups, etc. By clustering students, we can understand the students’ English level more clearly, lay the foundation for the follow-up learning content, and provide more individualized learning content for learners. 2. The analysis of learning content, English learning can be divided into four modules: listening, speaking, reading and writing. According to the analysis of association rules, the relevance of each module of English is analyzed, and the association rules between contents are obtained. Teachers can know which module problems lead to low English proficiency. The results of association rules can also be used as a basis for teachers to provide individualized learning content for learners. There are still some shortcomings in this study, so there are still many problems that need further study.

References
[1] Zhang J, Sui X, He X. Research on the Simulation Application of Data Mining in Urban Spatial Structure [J]. Journal of advanced transportation, 2020, 2020(9):1-9.
[2] Davis D, Chen G, Hauff C, et al. Activating learning at scale: A review of innovations in online learning strategies [J]. Computers & Education, 2018, 125(OCT.): 327-344.
[3] Huminecki U. Models of the Gene Must Inform Data-Mining Strategies in Genomics [J]. Entropy, 2020, 22(9):942-942.
[4] Chen T, Ling Q, Shen Y, et al. Heterogeneous Online Learning for “Thing-Adaptive” Fog Computing in IoT [J]. IEEE Internet of Things Journal, 2019, PP (6):1-1.
[5] A O I, B A A, C Z A, et al. Online Learning Usage Within Yemeni Higher Education: The Role of Compatibility and Task-technology Fit as Mediating Variables in the IS Success Model [J]. Computers & Education, 2019, 136(110):113-129.
[6] B J B A, A K K, C W W B , et al. Decoding and Targeting the Molecular Basis of MACC1-driven Metastatic Spread: Lessons From Big Data Mining and Clinical-experimental Approaches[J]. Seminars in Cancer Biology, 2020, 60(78):365-379.
[7] Zhao Y , Le J , Zhu L F , et al. Study on the Effect of Hypertensive Treatment Based on Drug Factor Analysis Model under the Background of Big Data [J]. Journal of Intelligent and Fuzzy Systems, 2019, 37(11):1-14.
[8] Tomasevic N, Gvozdenovic N, Vranes S. An Overview and Comparison of Supervised Data Mining Techniques for Student Exam Performance Prediction [J]. Computers & education, 2020, 143(Jan.):103676.1-103676.18.
[9] Zhang Y, Zuo L. College English Teaching Status and Individualized Teaching Design in the Context of Mobile Learning [J]. International Journal of Emerging Technologies in Learning (iJET), 2019, 14(12):85-85.
[10] Korf A, Hammann S, Schmid R, et al. Digging deeper - A New Data Mining Workflow for Improved Processing and Interpretation of High Resolution GC-Q-TOF MS Data in Archaeological Research [J]. Scientific Reports, 2020, 10(1):767.