Research on Data Mining Technology and its application in College English Teaching

Chen Hong

Engineering University of PAP

chenchen4774@sohu.com

Abstract. Under the background of big data era, data mining technology, as a unique data processing technology, is applied and promoted in college English teaching. It can not only guide the English teaching management to develop steadily in the direction of intelligence and digitalization, but also can link the recessive and explicit factors together to improve the comprehensive effect of college students' English learning. According to the accumulated experience of college English teaching in recent years, the application research of data mining technology theory is becoming more and more in-depth, which fundamentally changes the traditional monotonous education management mode and speeds up the pace of college English education reform in the new era. Therefore, on the basis of understanding the current situation of college English education, this paper starts from the perspective of English teaching evaluation, deeply explores the data mining technology and its application direction, takes the college English teaching evaluation system as an example, verifies and analyzes the application value of data mining technology, in order to provide an effective basis for college education innovation in the new era.

Keywords: Data mining technology; Colleges and universities; English teaching; Logistic model; Support vector machine.

1. Introducion

In essence, the purpose of teaching quality monitoring is to provide scientific and accurate basis for practical education activities, and this decision must grasp a large number of reasonable data information and statistical results, so that data can become the basic content of professional education reform. Data mining technology is widely used in college English subjects because of its functions such as data analysis, machine learning and information retrieval. According to the accumulated experience of college English education and language evaluation research in recent years, the most crucial thing to apply data mining technology in college English teaching evaluation system is data cleaning, data induction and data conversion. Among them, data clearing means that teachers should record students' classroom performance, make statistical analysis into a scale and store it in the database. ID3 algorithm is used to conduct in-depth research on students' English performance, and finally define the English learning dimension of each semester. Data induction refers to the establishment of a database of college students' English scores on the network platform. While storing relevant information, it is divided into modules, such as students' names, student numbers, cloze, English error correction, grammar structure, etc., so as to lay the foundation for the analysis and collection of academic scores after forming a database of students' English teaching. Data conversion refers to the conversion and application of evaluation results. [1-3]The purpose is to construct the corresponding English education reform plan according to the evaluation results of students' English learning, so as to improve the management system of English teaching in colleges and universities.

Nowadays, while constructing the network platform for monitoring English teaching, colleges and universities will set up the main modules of class teaching evaluation according to practical education management experience. During practical teaching, English teachers can convert students' specific performances into data and record them in the database through various methods such as classroom performance, English teaching, paper testing and homework completion. Then, by virtue of the data analysis advantages of the network platform, the intelligent analysis of

students' English knowledge level is carried out, so as to clarify the development dimension of English knowledge teaching, comprehensively present each student's English knowledge level and learning difficulties, summarize students' learning problems according to the evaluation data, and finally develop more effective teaching objectives and teaching content. The transformation of classroom teaching mode will inevitably change the traditional educational evaluation system. Therefore, after understanding the current situation and basic requirements of college English teaching in the new era, this paper mainly explores how to use data mining technology to build a new education management platform and teaching evaluation system, and then fully demonstrates the application value of modern data mining technology theory, so as to create favorable conditions for the application of diversified science teaching evaluation.[4-6]

2. Method

2.1 Network Platform

The design idea of the college education management network platform in the new era is to take the education management decision-making department as the control center, to carry out unified and centralized management of all data information, and other departments as workstations, to input, modify, query, statistics, print and other basic operations of the data under the authorization of the competent department. In this system, all the work of the education management department is divided into the grass-roots units, which can ensure that the employees of all departments can process data information scientifically, efficiently and on time. The network architecture diagram is as follows:[7-9]
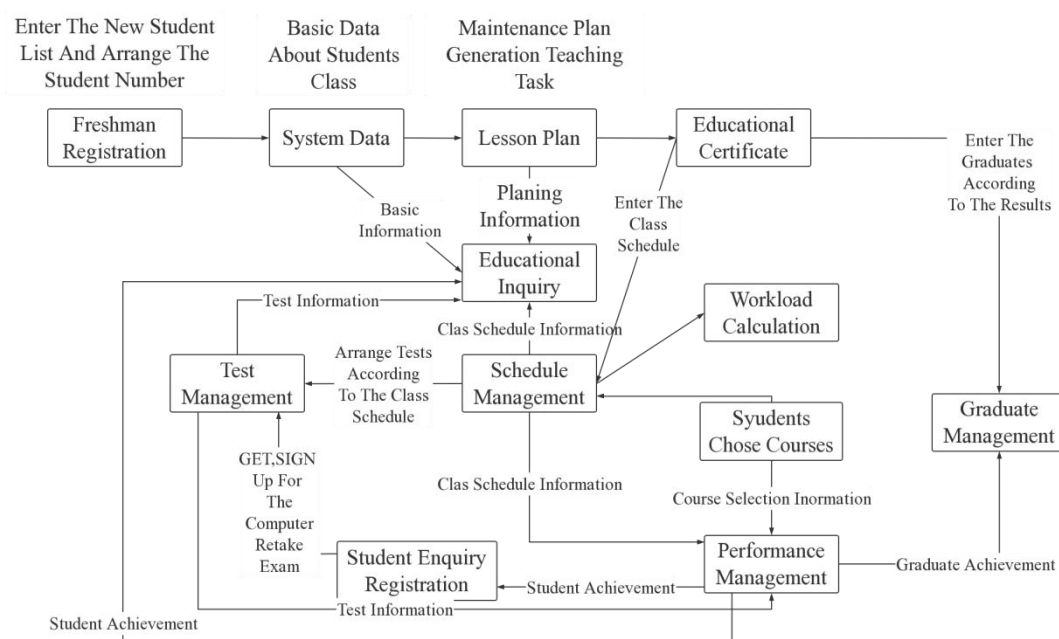

Figure 1. Network diagram of college English teaching management

Based on the above analysis, we can see that the basic functions of the overall system design are mainly reflected in the following points: First, teaching plan. This system can refer to or directly input to generate the teaching plan of each major and each grade, so as to make a statistical analysis of the hours and credits of various courses to accurately assign the basic teaching tasks. After the teaching secretary enters the teachers' opening situation in the new period, the new semester class schedule is generated by the class schedule management system, and the orientation information is managed in a unified way. Second, course scheduling management. This system can automatically generate course scheduling data according to the course opening information and teaching process, and has a very powerful automatic scheduling function. It can automatically arrange different

courses, and flexibly set teaching time, teaching place, course scheduling methods and other constraints, so that the final generated curriculum is reasonable and scientific. It really meets the basic needs of modern college English teaching; Third, educational administration. This system can automatically generate test data according to the information of course selection and opening, and has the basic function of auxiliary arrangement of data such as examination room, examination time, invigilator, etc. It can automatically generate the list of students participating in the English subject examination, convenient for educational management personnel to arrange the information of the examination room and management requirements; Fourth, school management. This system can effectively manage the basic information of students, in the occurrence of abnormal situations the system will automatically manage, all functions can be completed in one interface, can be in accordance with the basic requirements of query download data information, can be in accordance with any project in any range of statistical analysis automatically generate statistical graphics, convenient storage of students' photos and various stages, English learning results; Fifth, textbook management. This system will automatically generate the textbook plan according to the teaching plan and course selection information. After the effective storage of the textbook materials, it will automatically generate the material purchase list according to the inventory situation, learning data, course selection information, and other basic functions, such as fund management, textbook construction, order management, etc. Sixth, score query. The use of network platform to query results has the basic characteristics of speed, security, reliability and so on. After the examination, students can directly query their own results through the school network computer, and then provide effective basis for the following education and learning tasks. Seventh, teaching quality monitoring and information feedback. The evaluation of the teaching quality of the course is mainly made by means of experts' grading, students' grading and the perfection of teaching materials. However, this method has certain limitations due to large quantities of statistical work, and the final result is difficult to fully demonstrate the teaching effect of English courses. Therefore, this research system can use HTML forms to score and manage English courses in networked computers, and quickly summarize and design a series of required weighted statistical results according to CGI functions. This kind of large-scale sampling analysis can guarantee the impartiality of the final data information.[10-11]

2.2 Data mining technology

After defining the college English teaching network platform, in the face of a large amount of data information, to choose the appropriate data mining technology means, the most common mainly reflected in the following two aspects:

On the one hand, binary Logistic model. The Logistic regression model algorithm is mainly used to identify the relevance of feature items of things, synthesize a variety of objective input variables for classification and analysis, and complete classification or prediction according to the characteristics. The application of Logistic model in the evaluation of college English teaching quality needs to combine the objective factors such as teaching environment and teaching tools, get the K, P and I indicators of teaching evaluation and formulate a perfect evaluation system after clarifying the individual characteristics of students. This paper mainly analyzes the binary Logistic model algorithm, and on this basis to improve the optimization. Although the Logistic model has unique advantages such as easy to understand, easy to operate and relatively simple, due to complex nonlinear regression problems among feature items, it cannot quickly adapt to the random changes of feature items, especially when the feature items fluctuate, the classification and recognition effect of the Logistic model will continue to decline. Therefore, some scholars introduced the fluctuation curve of feature items into the Logistic model in their research, and improved the influence of random fluctuations on the accuracy of classification prediction through training, feedback and adaptive measures, so as to improve the accuracy of classification recognition.

On the other hand, support vector machines. This technique algorithm is more effective than the traditional BP neural network, which is based on the statistical learning theory. Among them,

statistical learning theory should make use of the structural risk minimization criterion to minimize the structural risk in minimizing the sample point error, so as to improve the generalization ability of the model, and does not have the limitation condition of the number of data. During the nonlinear classification, high-dimensional space transformation is used to transform the nonlinear into a linear classification problem of high-dimensional space, with the ultimate goal of finding the optimal classification hyperplane under the constraint conditions, as shown in Figure 2 below:
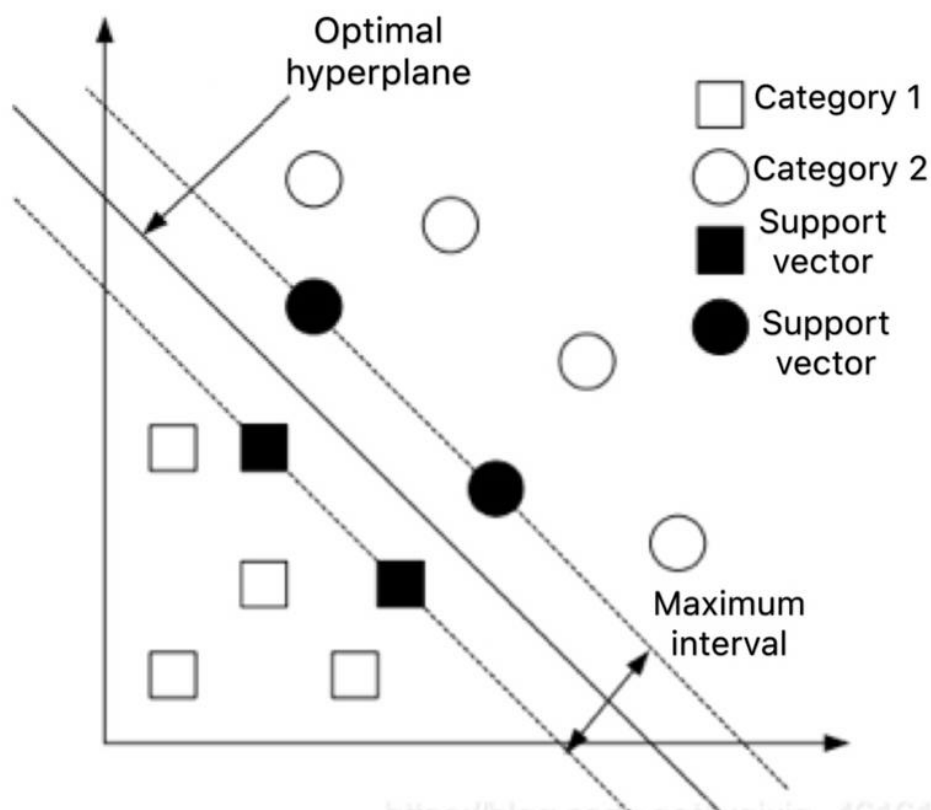

Figure. 2 Optimal classification hyperplane of support vector machine

3. Result analysis

Combined with the above research of college English teaching network platform and data mining technology application content design practice cases, mainly explore the application effect of building a diversified English teaching evaluation system based on data mining technology. In this experiment design, taking the teaching of English major in a university as an example, one class is set as the experimental group, another class is set as the control group, and the application effect of data mining technology and network platform is discussed through comparative analysis.

First of all, English proficiency test results. According to the comparison results shown in Table 1 below, although the two classes use the same English textbooks, the actual designed teaching plans and teaching periods are basically the same, but because of the different evaluation methods used during practical teaching, the final teaching effects are also different. Among them, the experimental group chose the diversified evaluation system with data mining technology as the core, while the control group still continued the traditional education evaluation system, which proves that the reasonable use of data mining technology to build a diversified evaluation system has a positive effect on improving students' English scores.[12-15]

Table 1 Comparison results of English proficiency test results

classes	number of people	average/mean value	The difference between the average	Difference between the mean values of the rear side and the front	standard deviation	P value

			value of experiment al class and control class	side		
experimenta l class	39	63.064	4.295	5.013	5.974	0.010
Control class	36	58.769		0.898	8.211	

Second, learning strategies. During the experiment, the teacher set a number of questions starting from the students' English learning strategies, and asked the students to fill in the questionnaire according to the questions and their own situation, in which "1" to "5" respectively represented whether the actual situation met the requirements, 1 was completely inconsistent, 5 was very consistent. Combined with the survey results shown in Table 2 below, it can be seen that students in the experimental group can accurately understand their learning objectives and complete teaching tasks in an orderly manner in practical teaching, while most of the data of students in the control group do not change significantly.

Table 2 Survey results of learning strategies

classes	project				
	1	2	3	4	5
Front of experimental class (39 people)	1/2.5%	9/23.1%	14/35.8%	14/35.8%	1/2.5%
Behind the experimental class (39 people)	0/0%	8/20.5%	10/25.6%	19/48.7%	2/5.1%
The front side of the control class (36 people)	3/ 8.3%	10/27.8 %	19/52.7%	3/8.3%	1/2.7%
Back of control class (36 people)	2/5.6%	11/30.5 %	18/50%	4/11.1%	1/2.7%

Finally, learn interest. During the experiment, the teacher set a number of questions starting from the students' motivation and interest in learning English, and asked the students to fill in the questionnaire according to the questions and their own situation, in which "1" to "5" respectively represented whether the actual situation met the requirements, 1 was completely inconsistent, 5 was very consistent. Combined with the survey results shown in Table 3 below, it can be seen that students in the experimental group are more and more interested in English learning, while those in the control group are less and less interested in English learning, which proves that the diversified evaluation system with data mining technology as the core can significantly improve students' interest in learning.

Table 3 Statistical results of study interest survey

classes	project				
	1	2	3	4	5
Front of experimental class (39 people)	7/17.9%	17/43.5%	10/25.6%	2/5.1%	3/7.6%
Behind the experimental class (39 people)	5/12.8%	11/28.2%	10/25.6%	10/25.6%	3/7.6%
The front side of the control class (36 people)	9/25%	15/41.6%	7/19.4%	4/11.1%	1/2.7%
Back of control class (36 people)	8/22.2%	14/38.8%	9/25%	3/8.3%	2/5.5%

4. Conclusion

In summary, based on the analysis of the requirements of college English curriculum proposed by the education department, it can be seen that the new college English classroom teaching mode should focus on modern information technology, comprehensively change the traditional single mode of lecturing, and use interpersonal interaction to improve students' comprehensive English ability. Only in this way can we improve the management level of college students' English teaching and truly embody the application value of data mining technology.

References

[1] Shuai Chen,Shixiao Zheng,Dongqi Luo. Research on the application of Data Mining Technology in Teaching quality evaluation of colleges and universities [J]. Chinese Science and Technology Journal Database (Full text) Education Science, 2022(12):4.

[2] Lizhu Cheng. The Application of Data Mining Technology in the Evaluation System of College English Teaching -- A Review of the Research on College English Teaching and Language Evaluation [J]. Science and Technology Management Research, 2022, 42(20):I0013.

[3] Ruibin Xu. Application of data mining technology in teaching quality evaluation of colleges and universities [J]. Educational Research, 2021, 4(5):28-30.

[4] Hongyu Shan. Research on the Application of Data Mining Technology in the Teaching Quality Monitoring and Evaluation System of Colleges and Universities [J]. Office Automation, 2022, 27(8):3.

[5] Hengxu Pan,Weisong Bian,Jie Deng and Wen Xiao. Application research of Data mining technology in Computer network virus prevention [J]. Computer Knowledge and Technology, 2021, 017(010):57-59,62.

[6] Meilian Sun,Cairong Zhan. Research and Application analysis of Data Mining Technology in Sports Training [J]. Contemporary Sports, 2021(35):0129-0129.

[7] Xingjian Wang ,Xin Liu. The Practical Research of Data Mining Method in the Construction of College General Courses -- Taking the course of Computer Network Technology and Application as an example [J]. Progress in Education, 2023, 13(3):8.

[8] Cong Xie. Construction of College English Audio-visual Speaking Teaching Evaluation System Based on Big Data Technology [J]. English Abroad, 2022(16):3.

[9] Jia Lu. Application of Data Mining Technology in the Analysis of Physical Education Results in Colleges and Universities [J]. Science and Technology Information, 2021, 19(27):3.

[10] Baoyi Du. Discussion on the application of data mining classification technology in college teaching [J]. Science Popular: Intelligent Education, 2022(4):2.

[11] Bo Yuan. Application analysis of Data Mining technology in the path analysis of university precision funding [J]. Chinese Science and Technology Journal Database (full-text Edition) Social Sciences, 2021(3):2.

[12] Yawen Zhang. Research on evaluation algorithm of English teaching index based on Data mining technology [J]. Modern Scientific Instruments, 2022, 39(2):207-211.

[13] Xiaomin Li. Thinking on English Teaching in Colleges and Universities Based on Data Mining of English Compound Talent Training Model [J]. Journal of Chongqing Electronic Engineering Vocational College, 2021, 030(002):102-105.

[14] Ruimin Wang,Xingzhi Chen. College English Test Band 4 Data mining and its Enlightenment to Teaching [J]. Modern English, 2023(1):4.

[15] Wen Cai. On the application of Big Data Analysis in English Teaching [J]. Today, 2022(1):2.