



The Construction of the Improvement System of Ideological, Political and Educational Effects of College Physical Education Courses Based on Data Discovery

Ning Zou  and Liangyu Zhang  

Changchun University of Finance and Economics, Changchun 130122, Jilin, China
yanbao0101@163.com

Abstract. As humanity enters the information era in its entirety, the area of education likewise enters the information age of education. Several new educational technologies have evolved in this century to accelerate the expansion of education information technology, generating vast volumes of data, which has amassed to build a database. Most data in education can only partially be exploited due to the explosion of data. For this issue, data mining technology is introduced. Course ideological and political education is the main part of the quality of ideological and political work in colleges and universities is the top-level design of college talent training and provides a new perspective for constructing an Ideological and political education effect improvement system in College Physical Education courses. Physical education is the foundation of the general education course module at colleges and universities, and it bears the vital task of shaping a moral individual. The foundation of “Course Ideological and Political” education in sports is educational courses. This work uses data mining techniques to investigate the goals of ideological and political education in colleges and universities and construct a system to enhance the efficacy of ideological and political education in colleges and universities. The paper establishes the improvement system of college sports’ ideological and political education effect, clarifies the ideological and political path of sports courses, and establishes relevant mechanisms by describing the data mining theory, process, and main algorithms, including decision tree algorithm, association rule algorithm, and so on, and based on this algorithm. Finally, it examines how instructors utilize more methods to educate people in Ideological and political education in physical education classes, with 77% using multimedia teaching to form WeChat groups. To assess the ideological and political teaching effect of students’ physical education courses, 41% of students from all categories participated in the analysis and evaluation, demonstrating that the ideological and political education effect of physical education courses is remarkable.

Keywords: Data mining · College physical education courses · Course ideological and political · Sports system

modular design, with intuitive and easy-to-operate interface, good concurrency after testing, and is scientific and practical. The risk assessment in students' entrepreneurship is preprocessed, and the risk assessment model in students' entrepreneurship is established by using support vector machine of big data analysis method. The designing functional modules, the class diagram and timing diagram in UML modeling language are applied and the database structure is designed. In the part of system implementation and test, the realization of functional modules and the results of system test are discussed in turn, and the main contents of realization and the final execution effect are clearly defined. In the future work, The construction of the system will improve the whole environment of college students' entrepreneurship, and provide more support to the whole college students, encourage social participation, and improve the entrepreneurial mechanism of college students.

Acknowledgments. The Research Project of Teaching Reform in Colleges and Universities of Jiangxi Province in 2019. Study on the training mode of innovation practical ability from the perspective in association of create a company Project number:JXJG-19-29-3.

References

1. Guang-xing, Z., MA, et al.: Study the status and path of reform of create a company of civil engineering education system. IOP Conference Series: Earth and Environmental Science, 267. 5, p. 52053 (2019)
2. Bei, H., Liu, H., Zhou, H.: Current situation and educational system construction of create a company of design specialty in colleges and universities. J. Jiamusi Educ. Inst. **012**, p. 237, 239 (2019)
3. Liu, R., Guo, J., Bai, T.: Research on the current situation and countermeasures in students' Create a company Curriculum System—Taking Xi'an University of Engineering as an Example. education forum, 014, pp. 122–124 (2019)
4. Chen, H.: An exploratory research on constructing a model of create a company education for college students based on fuzzy neural network algorithm. Security and Communication Networks, 5, pp. 1–8 (2021)
5. Miloud, T.: Corporate governance and initial public offerings. Int. J. Entrep. Innov. Manag. Entrep. Innov. Manag. **23**(1), 1–22 (2019)
6. Kun, C., Shihua, D.: Research on the design teaching of incorporating into the practice of create a company ——taking the digital media art major of guangdong as the example. Design **000**(005), 32–33 (2018)
7. Li, H.: Mode and practice of create a company education in the major of digital media art design. IPPTA: Quarterly Journal of Indian Pulp and Paper Technical Association, vol. 30, no. 5, pp. 219–226 (2018)
8. Huang, L.: Exploration on the quality evaluation system of create a company education in local universities. The guide of science & education, vol. 000, no. 014, pp. 15–16, 2019
9. Gong, Z., Zheng, H., Wei, C.: Construction of curriculum system of create a company in higher vocational college. Education and teaching forum, vol. 000, no. 035, pp. 262–264 (2018)
10. Cao, J.: The “Create a company Education” curriculum system design on the basis of capability model. Creative Education Studies **06**(3), 227–233 (2018)
11. Liu Yating, H., Yuan, D.M.: Exploration of create a company ecosystem service system. Guide to China's science and technology resources **050**(004), 85–94 (2018)

page. According to the description of the project, the instructor downloads the application form in the attachment, and after reading it, determines whether the project meets the application requirements. If it meets the requirements, select "Approved" in the review status, if it does not meet the requirements, select "Failed", and then fill in the review comments and submit them. The approved items will be handed over to the administrator for approval.

5.4 Project Approval

Project approval is the last pass of whether a project can be established, which is completed by the administrator. The staff first opens the project approval list and sees all the unapproved projects. The staff opens the approval page after clicking the "Approval" link. The examination and approval of the project by the staff is mainly carried out according to the examination opinions of the instructors. Determine whether the project can be established according to the teacher's review opinions and project summary. If the project can be initiated, select "Approved" in the approval status; otherwise, select "Failed".

6 System Test

Data is a systematic process, so white core and black core tests should be used in data and analysis tests to identify key points more accurately., we mainly pay attention to the function and performance of the software, and mainly test the software in black box. The system test method mainly adopts black box test, and the commonly used methods are: Equivalence class partition and boundary value analysis method. According to the non-functional requirements put forward by users, the current mainstream performance testing tool loadrunner is used to test the system performance. This chapter mainly introduces the design method of functional test cases, key test cases and corresponding test results, and makes necessary analysis on the test results.

The functional testing process of this system mainly passes manual testing without using any testing tools. In the testing process, referring to the requirement specification documents, the test cases are designed by the method of equivalence class division and boundary value analysis of black box testing. After testing, most of the functions of the system have reached the needs of users, but a small number of bugs are still found. In addition, the maximum length of the entry of "Introduction to Discipline Competition" is required to be no more than 300 characters, but when the input characters exceed 300 characters, there is no corresponding prompt in the system, which are all problems found during the test. After testing, the bug found was modified and tested again. After the second test, the system function has reached the goal required by the system requirements and basically meets the user application requirements.

7 Conclusion

Through the B/S system architecture, and the use of Web front-end development and Java background development, etc. build a database, design a college student create a company experience system, and run on the Internet to facilitate user operation; The system adopts

model layer, as the main part of the application, is used to represent business data or logic. The view layer adopts JSP technology; The controller uses Servlet to control user interface data display and update model layer data objects according to user input; The comprehensive judgment of obligation, the off-line analysis of data and the actual needs of college students all need to be discussed and judged by multiple indicators [11].

5.2 Realization of Process Management of Create a Company Projects

The project process management part of this module realizes the function of adding project progress, and the user fills in the information related to project progress in the page and submits it; After executing the project progress adding method `addProjectProgress()`, the system first encapsulates the project progress information in temporary variables, and calls the project process management control class `ProcessAction` to verify the project progress information.

The innovation plan management function is realized This data processing process and management process belong to a continuous process, so it is necessary to comprehensively judge and analyze the relationship between mining data, and the management department users can fill The whole information process plays an important role and role.; After submitting the innovation plan information, the system executes the innovation plan adding method `addInnovationPlan()`, verifies and processes the innovation plan information submitted by the user, and finally encapsulates it in the innovation plan entity class `InnovationPlanBeans`. Figure 6 shows the diagram of adding business activities to the innovation plan.

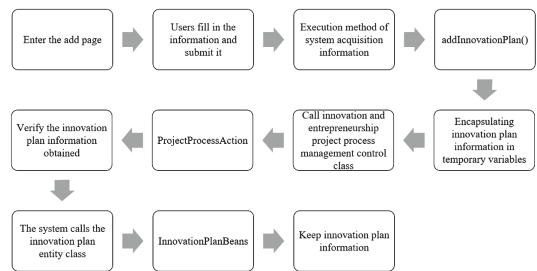


Fig. 6. Add activity diagram to innovation plan

In the innovation plan adding business, the system needs to obtain the innovation plan information submitted by the user and execute the method `addInnovationPlan()`. In the process of packaging and saving, it needs to call the innovation plan entity class and the create a company project process management control class `ProjectProcessAction`.

5.3 Project Review

Project audit is completed by the project instructor. After the students submit the project application, the teacher opens the project review list and sees the projects that have not been reviewed under his guidance. Then click the “Review” link to open the review

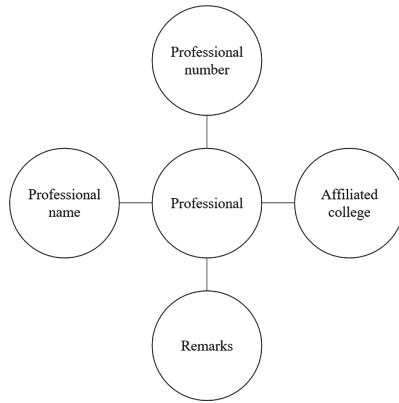


Fig. 4. Attribute map of professional entity

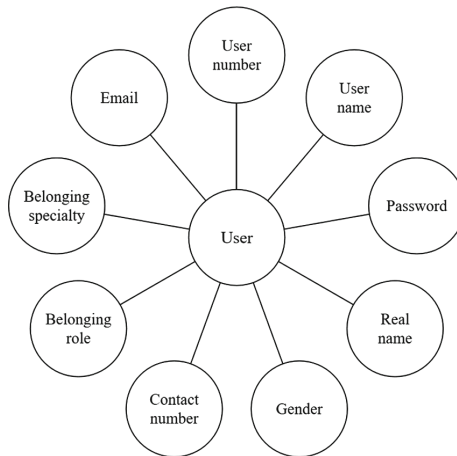


Fig. 5. User entity attribute map

5 System Implementation

In the implementation and testing part in students' create a company management system, the realization of functional modules is the primary task, and the realization results of functional modules are displayed in the form of pages and activity diagrams; Test the system from the aspects of function and performance, and make clear the system execution effect.

5.1 System Architecture

For the language in this matter vector machine, we can adopt the way of network knowledge, build modularization through corresponding websites or network support, and better develop the system. encapsulate them in the code organization structure. The

to integrate local structures, remove redundant attributes, merge different attributes of the same object, and establish the relationship between different entities.

(3) Review global structure

According to the different subjects participating in the review, it can be divided into user review and developer review. College student data, it is judged whether the global structure can reflect the relationship between all entities in the system and whether the attributes of each entity meet the needs of different users in different functional processing.

According to the above design steps, firstly, the E-R diagram is used as an analysis tool to analyze the data entities and their characteristics related to the system functions. College, as one of the basic data of the system, consists of four attributes: college number, college name, college leader and contact telephone number. Among them, the college number is the key attribute used by the system to identify the entity. The corresponding entity attribute map is shown in Fig. 3.

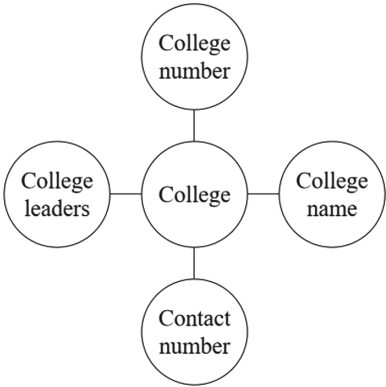


Fig. 3. College entity attribute map

As one of the basic data of the system, the specialty consists of four attributes: specialty number, specialty name, college and remarks. Among them, the specialty number is the key attribute used by the system to identify the entity, and the affiliated college is the foreign key. The corresponding entity attribute diagram is shown in Fig. 4.

As the main body of system operation, user entity contains nine attributes, which are user number, user name, password, real name, gender, contact telephone number, role, major and Email. Among them, the user number is the key attribute used by the system to identify the user entity, and its role and major are foreign keys. The corresponding entity attribute diagram is shown in Fig. 5.

- (5) In the analysis of the overall development environment of college students, we should establish a more accurate model, improve college students, college students' entrepreneurship and the overall analysis results of college students effectively and better, and comprehensively judge the data to complete the continuous analysis of college students' entrepreneurship.

4 System Design

4.1 Design Principle

The system is networking and modular structure. The system adopts B/S structure and adaptive technology, and users can access and operate it anytime and anywhere through a browser such as a computer, a mobile phone or an IPAD. Modular structure is adopted in function realization, which makes the interface clear and easy to operate.

4.2 System Network Architecture

In the system network architecture, it is necessary to clarify the specific deployment environment of the system: students, administrative departments and other users send requests and access the Internet network; On the basis of obtaining the application, the system verifies the security of the access request through firewall filtering; In the process of processing the request, the basic business data is obtained from the database server, and a dynamic JSP page is formed after processing, which is returned to the user through the network, and the user receives the final processing result.

4.3 System Database Design

Database is the foundation of information system, and a database with superior quality. Play an important role. in the stability of information system. Therefore, we must pay attention to the design of database in the process of software design. When designing the database of this system, the design strategy adopted is to design locally and as a whole, and the design steps are divided into the following three steps.

(1) Design local structure

Local structure refers to a specific function, and its associated users hope to browse and process the data structure in this function. This kind of data structure is often displayed according to the user's operation authority. Therefore, when taking the function as the foothold of design, it is necessary to summarize the data of the function operation for all possible users, so as to complete the data design of function points.

(2) Integrate local structures to form the final global structure

If the local structure is adopted as the final result, it will cause serious harm to the following design: first, there is a large amount of data redundancy; second, for the same object, the attributes concerned in different functions are different, resulting in the incompleteness of object description. In view of the above design hazards, it is necessary

In the process of data analysis, it is necessary to collect the relevant data of college students, so as to analyze the attributes of the data, and then form more reasonable indicators and contents to complete the comprehensive judgment of the data, so it is necessary to quantitatively analyze the relevant data of college students.

$$y = w^T \Phi(x) + b \quad (1)$$

Sansheng's entrepreneurship analysis is not only a process, but also a multi-factor analysis result, so it is necessary to comprehensively judge the multi-factor content.

$$\begin{aligned} \min J(w, \xi) &= \frac{1}{2} \|w\|^2 + c \sum_{i=1}^n \xi_i \\ \text{s.t. } y_i (w \cdot \Phi(x_i) + b) &\geq 1 - \xi_i \\ \xi &\geq 0, i = 1, 2, \dots, n \end{aligned} \quad (2)$$

In which: ξ_i represents relaxation factor; c is the punishment degree of error.

Better analysis of college students' entrepreneurial creation can integrate the problems of supporting, symbolizing and encouraging college students' entrepreneurial creation, better analyze and identify, complete the correlation analysis between data, and find out whether there are the main factors of development. And the evaluation function in students' Enterprise leader risk becomes:

$$f(x) = \text{sgn} \left[\alpha_i y_i \exp \left(-\frac{\|x - x_i\|}{2\sigma^2} \right) + b \right] \quad (3)$$

The working steps of establishing college students' Enterprise leader risk assessment model by support vector machine are as follows:

- (1) The indicators of safety create a company and the process of create a company need to involve more problems and the relationship between various problems, which need to be discussed in depth to make up for the corresponding deficiencies and identify and explore the problems existing in college students' create a company.
- (2) To study the existing indicators of single students, find out the better correlation between indicators and put forward assumptions, and make in-depth judgment on the content of assumptions, assumptions and indicators as well as the implementation effect.

$$x_i' = \frac{x_i}{x_{\min}} \quad (4)$$

- (3) Engineering above analysis is a systematic continuous process, so data collection should have continuity, otherwise the data collection will be in an invalid state for better analysis, it is necessary to observe the data continuously.
- (4) Scattering can find the direction in the data, judge the development direction of college students' innovation and entrepreneurship, the future employment direction and the overall results of development, better enhance the employment opportunities of college students, and improve the overall development environment of college students.

(2) Project publisher

Carry out the promotion of the party and employment problems, provide more reference for the employment of other college students, and realize the development of the whole employment, improve the overall employment environment and form more reasonable employment opportunities.

(3) Project reviewer and system administrator

The key issues and key contents in love are analyzed, and relevant policies and financial support are provided for the overall analysis of his life and his life employment, so as to meet the overall development needs of everyone now.

3 Establishment in Students' Enterprise Leader Risk Assessment Model Based on Support Vector Machine

In the whole analysis process, college students play an important role, but there are some shortcomings in college students, so we should make further analysis and research to meet the various needs of boys, and then provide more employment opportunities and support for college students to play their social abilities [7]. With the continuous development of machine learning algorithm, an algorithm We should play our leading role, better analyze and judge the society, form better content and realize the overall analysis of data "over-fitting" results in students.' Enterprise leader risk assessment [8, 9]. Most worried about the actual needs now, the use of analytic hierarchy process, the analysis of college students' ability and analysis of the results of judgment, through the data analysis of the key indicators and key content how to judge, if the existing problems of college students [10].

This problem is mainly to analyze the development of college students, their entrepreneurial ability and entrepreneurial conditions, to judge whether rank fighters can be carried out more accurately, to analyze various kinds of contents, to form a better environment for the overall development of his life, to provide support for the policy proposal and relevance analysis of all parties, and to improve the success rate of college students' entrepreneurship (Fig. 2).

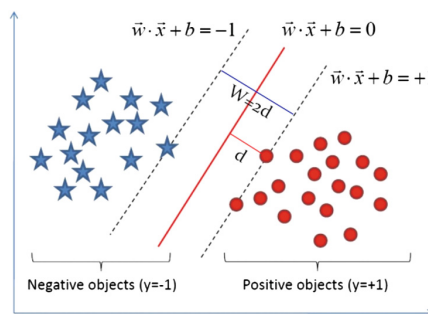


Fig. 2. Schematic diagram of classification plane of support vector machine

management of college students, college students should actively play their student spirit and better enter the student society to complete their historical mission. Third, use their own materials to start businesses and start businesses. Single in the whole construction and ability improvement, can play their own professional advantages, better services for the society, between the community and college students to establish a communication platform, in order to achieve the balance between the public needs, college students can provide more ability for the society, and society to provide more opportunities for college students, in order to create a better social environment, promote the overall development of society.

2
System Functional Requirements

What organizational aspects should be adopted in the analysis of the overall development and ability of Spanish doctors? First, some abilities should be evaluated, and then one point should be added to judge the implementation effect and the role of college students in society, so as to form the relationship between the two. The specific results are shown in the following (Fig. 1).

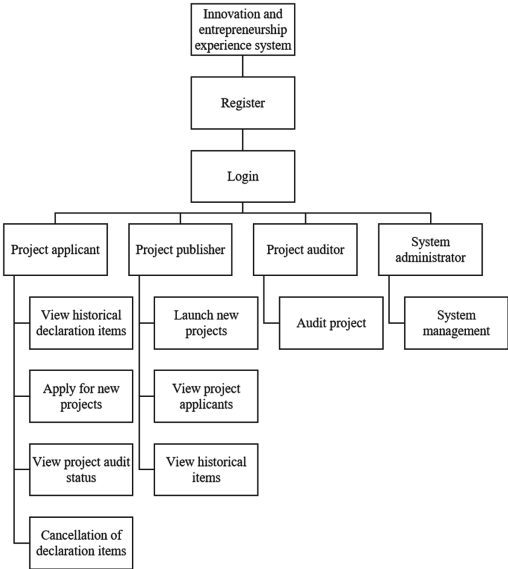


Fig. 1. Function diagram of system role

(1) Project applicant

First of all, we should set up projects for single people’s ability training and entrepreneurship planning, and adopt multi-party expert review to enhance the relationship between cancer in society, so as to provide them with some reasonable entrepreneurship plans.



Design and Implementation of College Students' Innovation and Entrepreneurship Experience System Based on Deep Learning

Yi Yang^(✉)

Jiangxi Institute of Applied Science and Technology, Nanchang, China
531725263@qq.com

Abstract. Cultivating students' Set up a company leader ability is one of the main teaching tasks in universities. On the basis of encouraging teachers and students to participate in Set up a company leader teaching, students' comprehensive quality and practical ability are improved, and more outstanding talents are trained for the society. According to the actual needs in students' credit recognition of create a company, a fully functional experience system in students' create a company based on B/S structure is constructed based on Java language. According to the support vector machine of big data analysis method, the risk assessment model in students' entrepreneurship is established. At last, the functions of the system are realized by ASRNET technology, and the system is tested comprehensively from three aspects of interface, function and performance, and the errors found are corrected, so as to realize the platform and development of the whole system.

Keywords: Support vector machine · The establishment of the company · Project management

1 Introduction

With the rapid development of social environment and social conditions, today's single work is facing a very serious problem, and the social demand for college students is very strong, so it is necessary to cultivate regeneration ability and improve their self-ability to better serve the society [1]. In the current create a company experience system in students, there are some problems such as low coordination degree between departments, irregular process and weak pertinence, etc. Because the school management department lacks effective unified coordination means, it is difficult to grasp the development of create a company management in various departments in a timely and centralized manner, and it is difficult to carry out guidance work for departments; Applying information form to create a company management system in colleges and universities can improve students' comprehensive quality and core competitiveness [2]. College students need more data and ability in the process of establishing, and self-starting and company establishment can provide more employment opportunities for themselves, play the role of students in society and improve the overall skills of college students. For the development and future

17. Wang, G., Department, A.: Application of 3D printing technology in animation contrary teaching design for deaf students in higher vocational colleges. *J. Chang. Univ.* (2016)
18. Yuan, J., Branch, Z.: Application of information technology in the teaching of color class in higher vocational colleges. *J. Zhenjiang Coll.* (2016)
19. Xiao, J.Y.: Application of cloud computing technology in the teaching of computer courses in higher vocational colleges—taking the course of “Cloud Platform Construction and Application” as an example. *J. Hebei Energy Inst. Vocat. Technol.* (2019)
20. Zhao, C.M., University, B.: Discussion on the application of multimedia technology in music teaching in higher vocational colleges. *J. Hubei Corresp. Univ.* (2016)

5 Conclusion

In conclusion, the data mining technology can meet the demand of higher vocational technical college students quality data analysis, break the education work, more intuitive and accurate data analysis results at the same time, eliminate a lot of fuzzy problems existing in the previous work, can give teachers literacy education work to provide a more clear direction, but also to teachers' professional play, More effective cultivation of college students' professional quality.

Acknowledgements. specialized Project in the construction of Double high-level colleges. No.C103. Study on the Measures and Effects of Morality and Technical Education in Vocational Colleges.

References

1. Yadong, L.I.: Application of data mining technology in employment guidance of students in higher vocational colleges. *Theory Pract. Innov. Entrep.* (2019)
2. He, X.: RETRACTED: discussion on the application of computer technology in table tennis teaching in higher vocational colleges. In: *Journal of Physics: Conference Series*, vol. 1992(2):022049 (4pp) (2021)
3. Bao-Quan, M.I.: Application of data mining technology in teaching management in higher vocational colleges. *Softw. Guid.* (2018)
4. Liu, X.: The application of data mining technology in the teaching evaluation in colleges and universities. *J. Comput. Theor. Nanosci.* **14**(1), 7–12, 14 (2017)
5. Zhang, R., Zhang, Y.: Research and application of data mining technology in student information management (2017)
6. Ping, L.: A preliminary investigation and study on professional identity of rehabilitation technology vocational college students **2021**(1), 3
7. Sun, X.: Application of data mining technology in college mental health education (2022)
8. Maerker, M., Guran, S.H.: Application of data-mining technologies to predict Paleolithic site locations in the Zagros Mountains of Iran (2022)
9. Wang, G., Cao, W., Cheng, Y.: Application of data mining technology in college students' employment analysis (2020)
10. Chen, Z., Liu, J.: Research on the application of data analysis in SPOC teaching in higher vocational colleges: a case study of psychological quality training course for college students. *J. Comput. Theor. Nanosci.* (2018)
11. Jiang, P., Tan, Y., Wen-Jun, W.U., et al.: Application of king standard theory in the practical teaching of basic nursing technology in higher vocational colleges. *Smart Healthc.* (2019)
12. Yang, Y., Pei, G.: Application of information technology in the course of food sculpture in higher vocational colleges. *Mod. Food* (2019)
13. Wei-Zhong, L.V.: Application of flipped classroom in the teaching of database application technology in higher vocational colleges. *Comput. Knowl. Technol.* (2018)
14. Shao, M.Y.: The application of information technology in ideological and political education in higher vocational colleges. *J. Hubei Corresp. Univ.* (2017)
15. Yang, Z., Department, H.Q.: Exploration and application of information technology in physical education teaching in higher vocational colleges. *Educ. Teach. Forum* (2018)
16. Wang, Y.T.: Application of modern educational technology in Chinese teaching in higher vocational colleges. *J. Jiangxi Vocat. Tech. Coll. Electr.* (2018)

- (2) Collection and storage of data on major vocational and technical literacy. Develop the adaptation interface of vocational and technical literacy data, connect with various application systems in the school to obtain various heterogeneous vocational and technical literacy data, and use the framework and system of the mainstream of vocational and technical literacy data to uniformly store the vocational and technical literacy data, laying a solid foundation for the mining and analysis of vocational and technical literacy data.

A. Professional and technical literacy data extraction: for various types of professional and technical literacy data that need to be collected by the large professional and technical literacy data analysis platform, the adaptive interface shall be developed respectively [17]. For the existing information system, the corresponding interface module is developed to connect with each information system, and the system that can not realize the data sharing interface of vocational and technical literacy can collect the data of vocational and technical literacy through ETL tools, support various types of vocational and technical literacy databases, such as SQLSERVER, ORACLE, ACCESS and other school system vocational and technical literacy databases, and clean and convert the data of vocational and technical literacy according to the corresponding specifications, So as to realize the unified storage and management of professional and technical literacy data. For other vocational and technical literacy data, it is necessary to implement the relevant interface according to the situation of vocational and technical literacy data, and use the corresponding interface to obtain the relevant vocational and technical literacy data information, so as to complete the extraction of vocational and technical literacy data [18]. B. Pretreatment of vocational and technical literacy data: in order to make the MAPREDUCE algorithm of the big vocational and technical literacy data analysis platform more convenient to process the vocational and technical literacy data, and to make the storage mechanism of the vocational and technical literacy data more scalable and fault-tolerant, it is necessary to combine the data through vocational and technical literacy according to the corresponding relevance, and convert the data into text format, and store it as a file. C. Storage of vocational and technical literacy data [19]: In order to improve the scalability and fault-tolerance of the data storage of vocational and technical literacy, the HDFS file system of Hadoop, the mainstream large vocational and technical literacy data framework, is used to store all kinds of vocational and technical literacy data of the school in a unified textual manner, and the vocational and technical literacy data is stored according to the corresponding rules, so as to realize the daily storage of a complete set of vocational and technical literacy data file sets, Form a data warehouse of professional and technical literacy. (3) The analysis and display of professional and technical literacy data is the core business layer. Through the professional and technical literacy data reporting tool, we can develop diverse and targeted professional and technical literacy data reports according to needs. The professional and technical literacy data analysis system based on Hadoop's MapReduce programming model is used to process the stored professional and technical literacy data, run the algorithm and convert the results, save the results as report files, and form a daily report file set [20]. The report presentation system displays the generated report file in a visual way.

technology and I/O communication interface, and establishes a wireless communication channel between the system and the artificial end device, which can output the data results to the artificial end under the support of the network environment [13]. It is worth mentioning that the default state of the wireless communication channel is closed, only the administrator in the case of their own needs to open the operation, it can carry out data output, on the one hand, to ensure data security, on the other hand, to make the system better cooperate with manual work.

3.2 Application Methods

Higher vocational technical college students through the above design, data analysis, the application of data mining technology in the method can be divided into three steps: first, with the help of the online education platform to generate data source, rely on access to relevant data acquisition program, data acquisition during artificial to literacy listing design work, and through the related tools for data cleaning and transformation; Second, the data in the database is imported into the data analysis module of the system, and the results of machine learning and analysis are output to its own equipment after completion [14]. Third, judging the quality of college students according to the results, and combining subjective experience and systematic suggestions to make decisions. Participation system reality some college students practice project is low, for example, data is characterized by less responsible for the work tasks, completion of tasks, work task quality is low, less so after analysis thinks that the college students' professional attitudes negative literacy problems, and presents "incentive" and "supervision" two policy recommendations, the teacher may choose from the perspective of professional education, And design a detailed education plan, thus targeted education [15].

4 Construction Scheme of Big Data Analysis Platform for Vocational Education

Build a large vocational and technical literacy data analysis platform, extract massive vocational and technical literacy data from other application platforms of the school for management, integration, analysis and utilization, find potential problems and valuable laws from it, and display them in a visual way, which can provide scientific decision-making support for the school management and meet the personalized needs of teachers, instructors and students, So as to improve the quality of information service in our school.

1. Construction objectives

- (1) Realize the sharing and exchange of professional and technical literacy data. Integrate and integrate the vocational and technical literacy data of various application systems of the school, so that various kinds of vocational and technical literacy data with different sources and types can be used with each other, enrich the sources of vocational and technical literacy data, break the information island between systems, and realize the sharing and application of vocational and technical literacy data [16].

Normalization of zero mean. It will map the original data to a distribution with a mean of 0 and a standard deviation of 1. Specifically, assume the mean value of the original features μ . The standard deviation is σ . Then the normalization formula is defined as:

$$z = \frac{x - \mu}{\sigma} \quad (3)$$

The principal component represents the direction of the data that explains the maximum variance. The relationship between variance and information is that the greater the variance carried by a line, the higher the dispersion of data points along the line, the greater the dispersion along the line, and the more information it contains. Calculating the eigenvalues of the covariance matrix is actually calculating the maximum variance, and calculating its corresponding eigenvector is the best projection direction [11]. Calculating the eigenvalues of the covariance matrix needs to be diagonalized. In order to meet the requirements that the covariance between the changed indicators is 0 and the index variance is as large as possible, it is required to solve the maximization problem, which can be expressed as follows:

$$\begin{cases} \max \{ \omega^T \Sigma \omega \} \\ \text{s.t. } \omega^T \omega = 1 \end{cases} \quad (4)$$

$$D(\mathbf{x}) = \omega^T \Sigma \omega \quad (5)$$

Sort the eigenvectors according to the order of the calculated eigenvalues. From high to low, the principal components can be obtained according to the order of importance.

$$\eta = \sqrt{\frac{\sum_{i=1}^d \lambda_i^2}{\sum_{i=1}^n \lambda_i^2}} \quad (6)$$

Finally, knowledge base, which is also a kind of database in essence, is different from the database of data sources. Because the quality data of technical college students in higher vocational colleges is of huge magnitude and various kinds, the knowledge data generated after analysis will increase geometrically, so the knowledge base needs to meet the storage requirements of massive data [12]. In the face of this demand, although ordinary database can through continuous expansion of the operation to achieve the purpose, but this can lead to a huge cost, and with the passage of time, the cost will be rising, so this article chose itself unlimited data storage capacity, and increase operating cost low cloud database as a repository for use (although cloud database itself data capacity is infinite, However, due to data security concerns, the capacity needs to be limited in practical applications. Therefore, capacity expansion is also required.

3.1.3 Output Data Result

Through data analysis, a data model representing college students' literacy will be generated in the knowledge base, and the data model will be analyzed to draw conclusions or decisions. These results must be presented to the human, otherwise they have no meaning. In order to achieve this, this paper chooses the wireless communication network

solve the problem of data duplication, The latter removes all low-quality data that does not conform to the standard format.

3.1.2 Data Analysis

The first is neural network, which has a variety of forms, currently more common are feedforward type, feedback type, convolutional type of neural network, the three are applicable to different conditions, so according to the technical college students in higher vocational education literacy data analysis needs, this paper chooses the feedback type of neural network. Feedback type neural network logic is the feature of again and again, but contact intuitive data, the characteristics that feedback type neural network is suitable for the higher vocational technical college students quality data analysis, namely the analysis work for the most part have mutual influence relationship between factors, such as college students work attitude will affect its working thoughts, and vice versa. Therefore, there is an interactive relationship between the data, and multiple factors should be repeatedly combined in the analysis, so the logic of feedforward neural network moving forward does not meet the requirements [9, 10]. At the same time, the connection between these factors is obvious, indicating that the data are at the same level without convolution, so only the feedback neural network is applicable.

The second is algorithm module, the key point of its design is algorithm selection and model creation. Algorithm selection, because of the higher vocational technical college students core idea is to extract the data characteristics of data analysis, the college students' personal characteristics related to literacy will feedback in the data, thus represents the features of the data related to quality of undergraduates, the feature extraction, can be related to literacy for college students to judge, For example, if college students are very active in learning attitude, the frequency and single time duration of college students' learning behaviors will be at a high level, indicating that the characteristics of data can judge students' personal literacy. According to this train of thought, this paper chose the principal component analysis algorithm, this method is a statistical algorithm, the principle is through the orthogonal transformation to the data set of a set of possible correlation between variables is converted to linear irrelevant, if the transformation results, this group of variables is the main component of a data set that represents the characteristics of this group of data, statistical results show that the algorithm can pass the data, principal component analysis (PCA) was used to extract data from college students' literacy data sets. The algorithm of principal component analysis is shown in Eq. (1).

$$Fp = a1i*ZX1 + a2i*ZX2 + ... + api*ZXP \quad (1)$$

where FP is the principal component of the data group, $a1i, a2i, \dots, api$ is the eigenvector corresponding to the eigenvalue of covariance matrix, $ZX1, ZX2, \dots, ZXP$ is the normalized value of the original variable.

Normalization of linear function. Perform linear transformation on the original data to map the results to the range of [0,1], and realize the proportional scaling of the original data. The normalization formula is as follows:

$$X_{\text{norm}} = \frac{X - X_{\min}}{X_{\max} - X_{\min}} \quad (2)$$

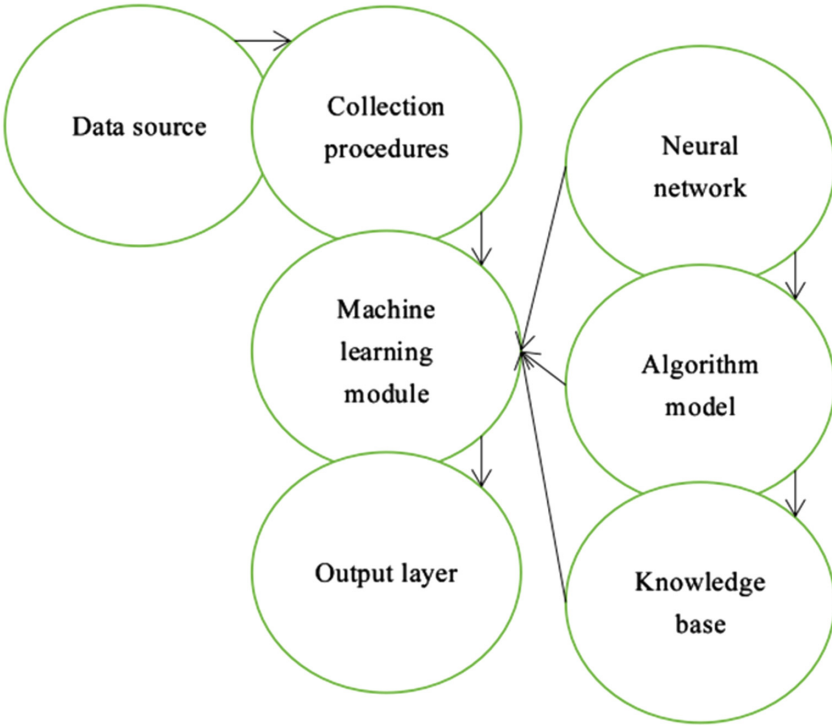


Fig. 2. Overall framework of the system

quality, it also organizes college students to carry out a series of online learning activities, so that the online activities of college students can automatically generate data, and the data will be directly recorded by the online teaching platform to form a data source [5–7]. Therefore, the online teaching platform only needs to be connected with the collection program of the system to directly complete the collection through the program.

Followed by the acquisition program design method of the program is mainly a Java language programming implementation, itself of the structure is simpler, contains the collection list created window, start-stop button two operation project, in which the former is to make artificial created collection list, the list contains all the needed data collected by the project, such as the teacher to the student quality index score results, once the list is created, the representative program will enter the data source according to the items in the list to search for data matching the characteristics of the item and collect it [8]. The latter mainly provides the acquisition intervention conditions for the manual, that is, if the manual wants to start the acquisition immediately, he can click the start-stop button, and if he wants to stop the acquisition immediately, he can click the start-stop button again, so that the loop can control the operation of the program.

In addition, through the data collection of all data will be saved to the database, the realization of data integration, but for the sake of data quality, it is necessary to clean up data in database integration of body, to clean up, generally can choose directly related tools, this paper chose to heavy tools, standard format match removal tool, which can

Table 1. Common data mining tools

Tool name	Characteristics
Rule reasoning	According to the rules of reasoning to get the result, the logic is relatively fixed
Fuzzy sets	Able to deal with fuzzy problems, flexible logic
Case reasoning	Can carry on the reasoning analysis to the specific case, the logic is more fixed

Table 2. Common data set analysis algorithms

Algorithms name	Characteristics
Decision tree algorithm	It can generate and compare relevant decisions to help people make choices
Genetic algorithm	The global optimal solution can be found
Rough set algorithm	Able to deal with problems that cannot be characterized in data sets

3 Data Mining Technology Application System Design and Application Method

3.1 System Design

Focusing on the data analysis needs of technical college students in higher vocational colleges, this paper designed the application system of data mining technology, and the overall framework of the system is shown in Fig. 2.

According to Fig. 2, the system first collects data from the data source, and all the collected data will enter the database and be classified. Second in quality analysis of classification of data in the database requirements by scheduling, form data set, and then driven into the algorithm of neural network module, completed by algorithm, data modeling, and analysis results are given, the results will go into the knowledge base, on behalf of the system have the identification of university students quality data, and carries on the analysis of capacity. Finally, the data model and decision results obtained by the algorithm module will be displayed through the output layer. Combined with the above discussion, this system can be roughly divided into three parts, respectively, data acquisition, data analysis, data result output, each part of the design content is as follows.

3.1.1 Data Collection

First is the formation of the data source method, considering the characteristics of higher vocational technical college students data, artificial cannot directly to collect the data, the natural formation data source under human intervention, so online teaching platform, this paper holds that higher vocational colleges can develop according to teaching principle to transfer work related to the online environment, to cultivate college students' professional

that the model can be purified, which is more conducive to the accuracy of mining results. Data transformation is mainly to some non-quantitative, abstract data into the quantitative data intuitively, facilitate machine understanding, data loading is the key step in the data mining, loading way depends on the selection of data mining tools as a result, the data mining technology includes many mining tools, only to make the right choice, to get accurate results, common tools are shown in Table 1 [4].



Fig. 1. Operation flow of data mining technology

Data set establishment and output is to set the data results obtained after model analysis and establish the corresponding data set. In this way, the data set can be analyzed by special algorithm tools and the decision recommendations can be obtained. Then the decision recommendations can be output to the outside for manual selection. Common algorithms for dataset analysis are shown in Table 2.

retrieval. Build a variety of multimedia network classrooms with complete functions, encourage teachers to use multimedia and computer network teaching in teaching, provide students with a variety of information-based learning environment with pictures, texts and sounds, enhance students' understanding of knowledge and improve their information awareness. Carrying out lectures on literature and information knowledge can not only help college students have a deeper and more professional understanding of literature, intelligence, knowledge and information, but also guide students to acquire and master more extensive information through different ways and methods, so that they can more clearly understand the importance of information resources for personal growth and social development, and thus continuously enhance their strong demand for information, Improve their professional and technical literacy.

At present, many higher vocational colleges have data analysis of technical college students, the reason is that it is only through data analysis, to accurately know the college students' professional development situation, so that the targeted training education, but the data abstraction, not only about the development of the college students' professional quality and large scale, type, manual mode to data analysis, the demands of higher vocational colleges cannot be satisfied, and the development of quality education is forced to stop. The emergence of data mining technology brings a development opportunity, which can transform abstract literacy data into intuitive data, break through the limitations of artificial ability, and quickly process data of huge magnitude and various types, and the artificial only need to make judgments according to the results. Therefore, in order to play the role of data mining technology and break through the current predicament, it is necessary to play a relevant role.

2 Basic Concepts of Data Mining Technology

Data mining technology is a data processing technology, and its operation process is divided into three steps, as shown in Fig. 1.

According to Fig. 1, data initialization can be divided into two steps: One is the data collection, mainly from the data collected in accordance with the requirements of the specific collection of related data, the basic principle is to extract all the characteristics of the data in the data source, if a piece of data characteristics consistent with a particular collection features, will be to collect, and summarized the collection of data to relevant category contains specific collection object. In this way, the data collection is completed and the automatic data classification is realized. Secondly, data integration, as the data collection, each category contains specific collection object is to accumulate data, so want to combine the present demand in data application, confirm need specific collection the data and information required of the category and the category, then these data extraction process, in accordance with the focus on a particular logical model and implements a data integration [1–3]. The integrated data will form a specific data model, which can be used in machine learning, allowing the machine to identify the real things represented by the data and provide support for subsequent work. Data cleaning, transformation and loading are the three important links of data mining, among which data cleaning is also known as “data preprocessing”. The main function of data cleaning is to remove repeated, incomplete, or low-quality data with other problems in the data model, so

The education and cultivation of vocational and technical literacy runs through all stages of life. For college students, strengthening the cultivation of vocational and technical literacy is the most important. Vocational and technical literacy education is to cultivate college students' ability to comprehensively utilize modern information. At present, the vocational and technical literacy education in China's colleges and universities is only limited to the narrow teaching scope of information technology application such as the basic course of computer application and the course of document retrieval, which is far from meeting the needs of college students to adapt to the development of the information age. How to effectively improve college students' professional and technical literacy has become an urgent and important issue.

The standard of vocational and technical literacy is a guide to the cultivation of vocational and technical literacy, which has a macro guiding significance for evaluating the individual's vocational and technical literacy ability and guiding the practice of vocational and technical literacy cultivation in colleges and universities. From the perspective of foreign countries that have developed vocational and technical literacy education well, they have formulated their own standards of vocational and technical literacy according to the connotation of vocational and technical literacy. By comparing the standard system of foreign vocational and technical literacy, Chinese universities should formulate a set of scientific, rigorous and feasible vocational and technical literacy ability standards according to China's national conditions and characteristics of higher education.

Vocational and technical literacy is a comprehensive education. At present, China's colleges and universities have basically set up basic computer courses and document retrieval courses. Computer application is the technical basis for information retrieval. Colleges and universities should not only strengthen the teaching of computer theoretical knowledge, but also strengthen the training of students' computer practical ability. Document retrieval course is a centralized and systematic learning method to improve information retrieval skills. By taking the course of document retrieval, students can not only systematically learn the basic principles of document information retrieval, but also master the retrieval methods of various databases and skillfully use various retrieval tools.

Vocational and technical literacy cannot be cultivated separately from other disciplines. To improve vocational and technical literacy, it is necessary to integrate information technology education into each subject education, cultivate information awareness in regular courses and teaching, and urge students to consciously develop the habit of acquiring and using literature and information resources in the process of professional course learning, improve professional academic performance, master scientific research methods, and improve the level of thesis writing. While learning and mastering professional knowledge, students can also improve their level and ability of document information retrieval.

The library is the document and information center of colleges and universities, with orderly documents and complete retrieval system. In the library, students can not only understand the book classification methods more intuitively, but also learn to use computer retrieval tools through the digital system of the library to improve the level of computer information retrieval and enhance their practical ability of information



Research on the Application of Data Mining Technology in the Quality Data of Technical College Students in Higher Vocational Colleges

Pei Liu^{1,2}(✉) and Xiaolan Guo^{1,2}

¹ Shandong Institute of Commerce and Technology, Jinan 250103, Shandong, China
liupeil1980@163.com

² School of Politics and Law, Baotou Teachers College, Inner Mongolia 014030, China

Abstract. The Modern higher vocational education advocates the cultivation of college students' professional quality, so educational organizations pay great attention to the development of college students' professional quality and hope to analyze it from the perspective of data. In view of this requirement, this paper carries out research based on data mining technology, mainly introduces the basic concept of data mining technology, and then establishes the technology application system, and puts forward the application method of technology in the quality data of technical college students in higher vocational colleges. Through the research, combined with the method in the system, the role of data mining technology can be played to help analyze the data of college students' professional literacy, understand the development of their literacy, so as to adjust the training direction in time, or make targeted decisions.

Keywords: Data mining technology · Vocational technical college students · Professional literacy data

1 Introduction

China's higher vocational colleges are faced with some common problems in talent training, such as focusing on professional construction, offering a large number of professional courses, squeezing professional quality courses, only paying attention to the training of students' skills, while ignoring the training of students' professional quality; Even if there is training of students' professional quality, it does not run through the whole process of education and teaching, and the training content is divorced from the needs of enterprises. In particular, when formulating talent training programs, the lack of social surveys, especially the lack of industry and enterprise requirements for students' professional quality, led to the lack of professional skills in the future work due to the lack of professional quality.

18. Yang J , Liu Y . Application of Data Mining in the Evaluation of Enterprise Lean Management Effect[J]. Scientific programming, 2021(Pt.13):2021
19. Kazanidis, I., Valsamidis, S., Gounopoulos, E., et al.: Proposed S-Algo+data mining algorithm for web platforms course content and usage evaluation[J]. Soft. Comput.Comput. **24**(19), 14861–14883 (2020)
20. Yin, Y.: Research on ideological and political evaluation model of university students based on data mining artificial intelligence technology[J]. Journal of Intelligent and Fuzzy Systems **40**(6), 1–10 (2020)

requirements in the research process, mainly the lack of certain technical support in the practice process, and in the personnel should be limited, so there are certain results deviation.

References

1. Wu, F., Liu, X., Wang, Y., et al.: Research on evaluation model of hospital informatization level based on decision tree algorithm. *Security and Communication Networks* (2022)
2. Peng, X., Shufang, L.I., Zhang, Z., et al.: Fuzzy extension matter-element evaluation model of mine rescue capability. *Safety in Coal Mines* (2018)
3. Sun, H.: Study on application of data mining technology in university computer network educational administration management system. *J. intelligent Fuzzy Systems: Applications in Engineering and Technology*, 2019, 37(3aPt1)
4. Yan, L., Wang, C., Wang, N.: Application of Apriori Association Rule Mining Algorithm in University Management System (2016)
5. Kazanidis, E.: Proposed S-Algo plus data mining algorithm for web platforms course content and usage evaluation[J]. *Soft computing: A fusion of foundations, methodologies and applications*, 2020, 24(19)
6. Ding, L., Zeng, X.: Application of Decision Tree Model Based on C4.5 algorithm in nursing quality management evaluation. *J. Med. Imag. Health Informat.* (2021)
7. Mukherjee, A., Chaki, R., Chaki, N.: Data mining-based hierarchical transaction model for multi-level consistency management in large-scale replicated databases[J]. *Computer Standards & Interfaces* **74**(3), 103485 (2021)
8. Igor, D., Ficko, M., Balic, J.: 2th A model for prediction and evaluation of production processes based on genetic algorithm (2022)
9. Huang, Z., Liang, Y.: Research of data mining and web technology in university discipline construction decision support system based on MVC model. *Library Hi Tech* (2019)
10. Hou, Q., Wang, G., Wang, X., et al.: Research and Application on Spark Clustering Algorithm in Campus Big Data Analysis. 2(1):5 (2020)
11. Xu, L.Q.: The research to the system of teaching quality evaluation based on data mining in School. *Management method and Informatization* (2016)
12. Sun, J., Ling, X.U., Luo, Y.J.: Innovation of E-government Practice Teaching Informatization Evaluation Based on Network Model Application. *Journal of Nanchang Hangkong University(Social Sciences)* (2018)
13. Yang, X., Yijiao, H.U., Yang, D.: Agricultural informatization level evaluation and measurement of Liaoning Province. *Journal of Liaoning Normal University(Natural Science Edition)* (2016)
14. Li, B., Shi, Y. , Wang, Y.: The construction of the evaluation model of teacher education informatization based on linear regression algorithm. *Dynamic Systems and Applications* 29(3) (2020)
15. Wang, S., Zhang, Y., Hu, W.: In-depth Integration Technology of Equipment Management in Construction Enterprises Based on Megadata Mining Algorithm. *Urban Roads Bridges & Flood Control* (2019)
16. Yang, Z.: Application of K-means clustering algorithm in human resource data informatization. In: *CIAT 2020: 2020 International Conference on Cyberspace Innovation of Advanced Technologies* (2020)
17. Zhang, J. , Xuechao, O. , Cao, H. et al.: Research Summary of the Course Selection Algorithm for Teaching Management System in School. *J. Panzhihua University* (2017)

professionals in the field of data analysis in school in China, which restricts the level of management informationization and the development of data mining application in school as shown in information level 4 (Fig. 5).

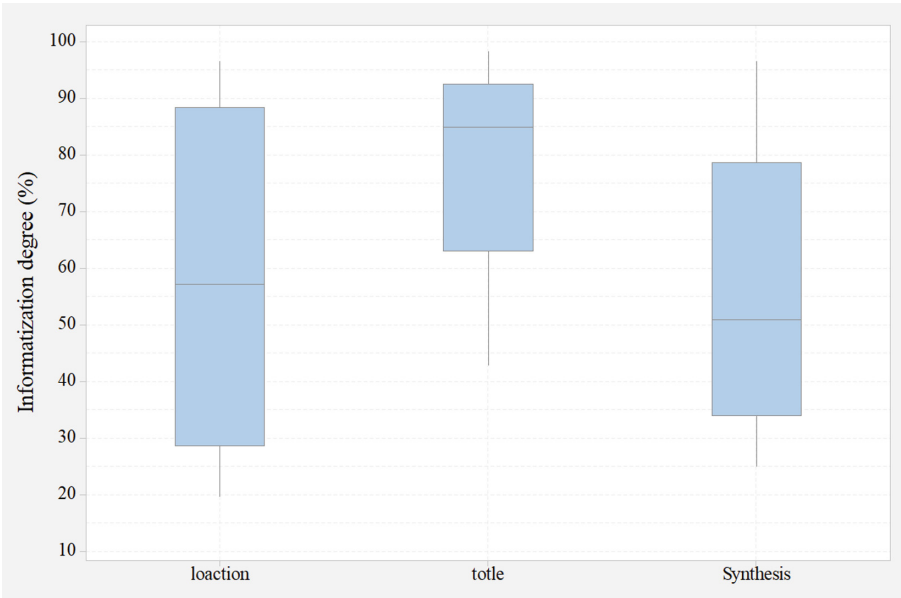


Fig. 5. Comprehensive evaluation of informatization of data mining

Although data mining technology has made some progress, there are still many unsolved data analysis methods and algorithms. How to choose appropriate data analysis methods to improve the accuracy and effectiveness Look at the comprehensive index analysis of data and the overall judgment of data.

5 Conclusion

The management level needs an informationization process and existence. As a support, the content and angle of professional group design in colleges and universities show the results of municipal examinations and future development plans, etc. Therefore, in the process of campus management, data informationization and diversification should be realized as much as possible. At the same time, in the process of data informationization construction, 10 points should be paid attention to the integration of intelligent algorithms and intelligent algorithm platforms, so as to better improve the intelligent effect and in the information processing process. Energy function, and in the process of mental information processing, to remove irrelevant data and irrelevant information, to ensure the overall analysis of data, and in the whole analysis process, the correlation between data and the qualitative cause play an important role, I think that for the main use of intelligent digestion, in-depth consideration of the above issues, there are certain

4.3 Informatization Processing and Accuracy

Through data mining technology, we can provide personalized teaching and learning services for teachers and students, and improve the quality of education and teaching; To provide scientific research management departments with reference for scientific research project declaration and achievement evaluation, and improve the level of scientific research achievements is shown in Fig. 4.

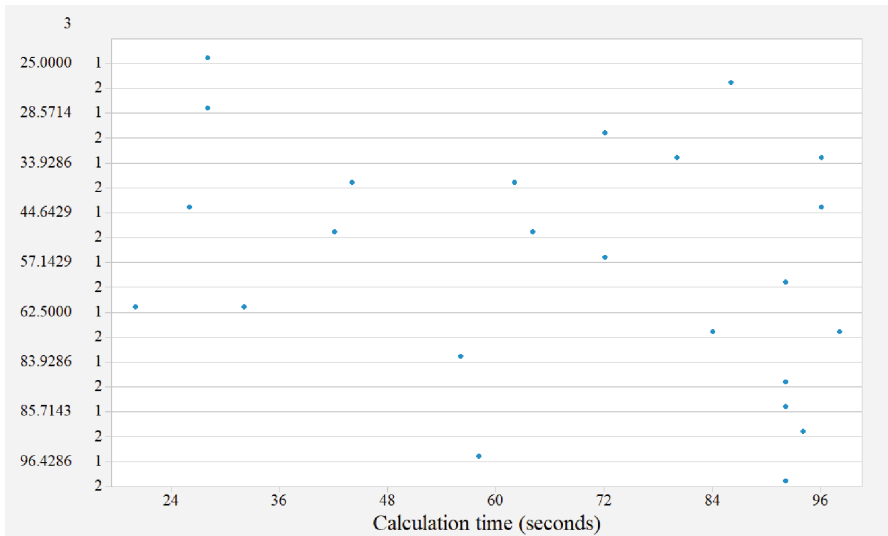


Fig. 4. Accuracy of different calculation times

By judging the accuracy and softness of the data, it can be determined that the data can present normal distribution in the process of processing and meet diversified needs. is shown in Table 3.

Table 3. Comparison of the degree of guidance of different methods

algorithm	Remember to say you robbed	You deal with the effect	error
Judgment of data fusion	47.14	51.43	21.43
The premise is the method	42.86	51.79	25.02
P	40.220	53.211	31.43

The above analysis and judgment found that there are some deficiencies in the research process, mainly because there are large errors in the data processing process, and the error rate is high, which will affect the results of the data and have no obvious influence, which is not conducive to the later competition [18]. There are relatively few

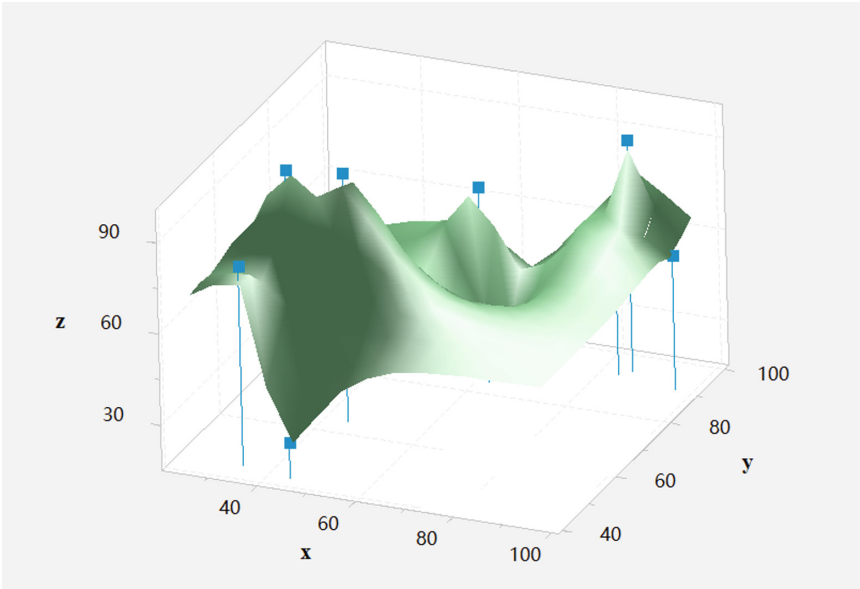


Fig. 3. Processing process of teaching content

4.2 Optimization Ratio of Information

Through data mining technology, we can deeply analyze the use of campus resources, provide scientific decision support for resource allocation, and optimize resource allocation as shown in Table 2.

Table 2. Overall situation of informatization

The critical information recognition rate	The processing standard is certain	Outlier recognition rate
1/2	53.57	78.57
1/3	35.71	87.50
1/4	87.50	80.36
1	66.07	66.07
X^2	83.93	93.57
P = 0.031		

such as missing fields, duplicates, value errors, etc. These “dirty data” need to be processed into regular data. Due to the large amount of data and the variety of data, big data technology needs to be adopted for data cleaning. The system does not require high real-time, so it can process data offline. Data cleaning improves the robustness of the system.

- (3) Comprehensive evaluation: comprehensive evaluation is based on the evaluation indicators and their calculation methods. According to the collected data, the performance of different indicators is evaluated. Finally, according to the weight of indicators, the overall performance is evaluated according to a certain algorithm. According to the diversified analysis of data and the overall framework of data, it is necessary to analyze the comprehensive content and comprehensive indicators of data to complete the comprehensive judgment and improvement of data.
- (4) Visual display: Visual display is to display the evaluation results to users in a friendly way. Some conventional tables can not give people an intuitive feeling. In order to provide a more friendly and readable result presentation interface, this study will select appropriate display methods according to the characteristics of the evaluation results [14]. Common visual display methods include pie chart, bar chart, column chart, line chart, etc.

4 Actual Cases of the Informatization Degree System of School

4.1 Teaching Information

According to the previous data analysis, understand the integrity and standardization of the data to judge, and then deal with the content of the data, data indicators, data structure.

Table 1. Characteristics of different structures

Type	I don't care about you	How much money do you have?	This is really funny
There is enough money to spend now	25.01	4214	34.64
	26.79	43.93	34.29
A flying culture	22.86	49.29	30.36
	29.29	42.14	32.29
The most diverse	27.86	47.86	32.10
	20.36	46.79	32.50

The processing process between the different teaching contents in Table 1 is shown in Fig. 3.

Through data mining technology, we can realize the rapid processing and analysis of a large amount of data, provide timely and accurate decision-making basis for university administrators, and improve management efficiency.

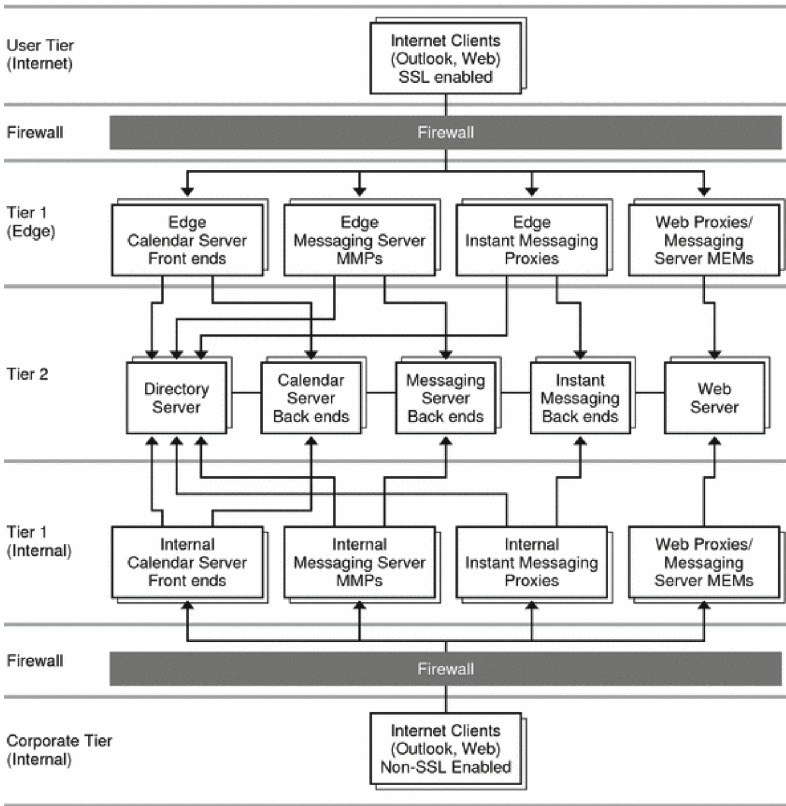


Fig. 2. System architecture diagram

courtesy when carrying out data in various places. The results after data processing should be put in this system and judgment process to complete the record analysis and optimization of data, so as to realize the integrity of data. In addition, in the whole analysis process of data, attention should be paid to a certain weight to complete the plan of data in weight, data volume and data standard, so as to better judge the data comprehensively. Therefore, the method proposed by you can realize the diversified processing of data.

- (1) All these data should be consumed in the processing process, and the integrity, initiation and integration of the data should be carried out to judge whether there are good problems between the data, and the teacher should make comprehensive analysis and comprehensive judgment in the data analysis and data processing, so it is necessary to realize the diversification of the data and the collation of the data in the whole analysis process of the data, um, the comprehensive return database, the Communist Party of China.
- (2) Data cleaning: data cleaning is to process irregular data collected in the evaluation system into regular data. The data collected in the system in this study includes process data and result data. These two kinds of data may include irregular data

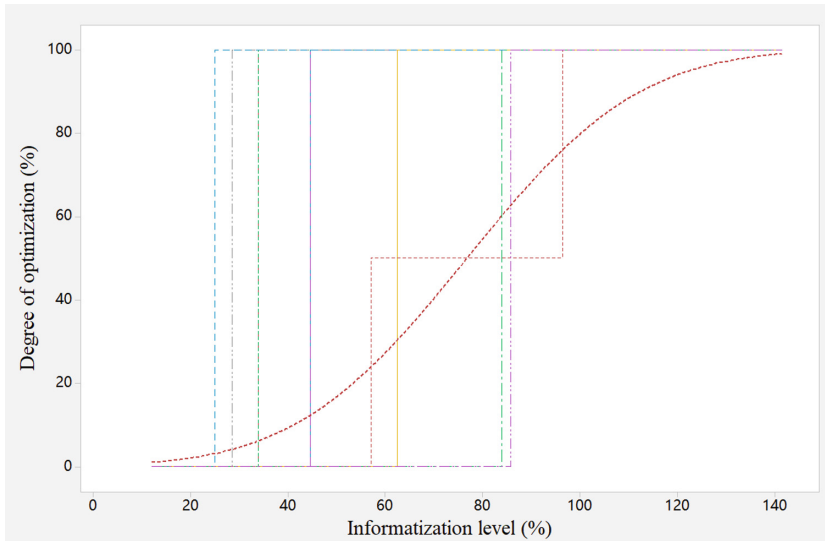


Fig. 1. Informatization of data mining algorithms

3 Development of Informatization Level Evaluation System

With the development of management method, Specific overall processing capacity. by university education management activities is increasing, and the process data contains a lot of valuable information. However, the current evaluation system mainly relies on the result data obtained from the online questionnaire to carry out the evaluation, and it is difficult to collect and use the process data [8]. Therefore, this study combines the big data concept with the university education informatization evaluation, and combines the process data and the result data, and designs and develops a data-driven university education management informatization level evaluation system to achieve accuracy Objective to evaluate the informatization level of university education management objectively. The system can realize the automatic collection of the process data of university education management, can clean the “dirty data” in the process of data collection, ensure the reliability of system operation, can conduct comprehensive calculation and evaluation based on the collected data, and can also display the evaluation results in a more intuitive and visual way.

Through the comprehensive analysis of data resources, the key indicators of fire content and news should be analyzed. In the process of judging the key content and indicators in the object as soon as possible, the overall optimization of data should be realized to complete the comprehensive judgment of data. In the process of overall judgment and comprehensive analysis of data, the comprehensiveness, relativity and standardization of data should also be identified, so as to improve the overall effect of data and realize the comprehensive judgment and comprehensive improvement of data (Fig. 2).

1. The process of data collation and analysis can be judged through the above-mentioned problems. First, the customer should collect data information, and then say

of information and complete the integrity of information. Judge what should be done to the key content and key indicators of data security behavior, so this technology can realize and improve the whole data processing process and processing effect.

More accurate data analysis, and then the efficient management ability in data mining to quantify the quality of data quantification cycle, as shown in Eq. (1).

$$f(x_i) = \sqrt{y_i \cdot \sum x_i} + \xi^2 \quad (1)$$

Upgrading some data requires certain parameters as adjustment, otherwise it is difficult to organize influence, so some parameters should be added.

2.2 Selection of Information Scheme

Hypothesis 2: Through the analysis of students' learning data, provide personalized teaching suggestions for teachers and improve teaching quality; Predict students' academic performance, find students with academic difficulties in advance, and implement targeted teaching intervention is shown in Eq. (2).

$$F(x_i) = \frac{z_i}{f(x_i|y_i)} + w_i \cdot \xi \quad (2)$$

2.3 Processing of Non-digitized Information

The students' behavior data, it provides counselors with early warning of students' psychological status and finds students' mental health problems in time; Through the analysis of graduates' employment data, we can provide employment guidance for students and improve the employment rate of graduates is shown in Fig. 1.

Through the analysis of the use of campus resources, it provides decision support for resource allocation; By analyzing the data of campus safety and energy consumption, we can improve the efficiency of resource utilization and reduce the operating cost.

2.4 Correlation Between Different Teaching Content

There are some differences between the unprovable content and the unprovable method. This content can effectively limit the negative method, while the true method needs some fusion. Content and indicators can be realized. There are certain differences between independent education content and teaching methods in the analysis process. Teaching content and teaching methods can be better integrated, so to improve mathematics knowledge contest, just mobile phone information. On which aspects of mobile phone mining methods and data technology can be more perfect and comprehensive, judge and achieve comprehensive analysis of benefits, complete the synthesis of data, judge and find out the key points and deficiencies, and put forward effective measures for concepts and deficiencies, so as to avoid more complex problems.

means, based on the traditional teaching system, students' learning process is difficult to monitor, let alone "personalized teaching" according to each student's learning ability and progress. With the rapid development of management method, combined with the application of "Internet" and "big data" technology, the problem of teaching process monitoring can be comprehensively solved. The reform of teaching evaluation system can become a necessary means to realize this process. Change the traditional "summative evaluation" into a more scientific "formative evaluation", and continuously observe, record and reflect on the whole process of students' learning, as well as comprehensively evaluate students' performance, achievements, emotions, attitudes. How to effectively use these data to improve the management level of school has become the focus of attention of current college administrators. As a new data processing technology, data mining has been widely used in university management. This paper will discuss the level of university management informationization and the content, situation, current situation, significance and problems of data mining. At the same time, it can also provide employers with a report that can fully understand students. In the process of choosing jobs after graduation, most college students are faced with enterprises' unwillingness to hire new graduates or to spend efforts on the cultivation of new graduates. In the absence of the attention of enterprises, new graduates will also frequently change jobs to seek development, which exacerbates the "discrimination" and distrust of new graduates; The reason for this phenomenon is that under the traditional teaching mode, the evaluation of students' learning effect is mostly based on the results of the "final examination" [2]. This evaluation model with only transcripts, for employers, lacks image recognition and can not have a comprehensive understanding of the ability of students just out of school, thus forming an embarrassing situation that is similar to a vicious circle. Based on the above phenomena and problems, at the beginning of the establishment of teaching resources and training tasks, based on the attributes of business talents and the knowledge attributes of these resources, we carried out the matching and design of relevant ability attributes, so as to finally form a "student portrait" of "full knowledge structure and multiple ability dimensions" through students' learning and training process, results detection and big data analysis, Show students' abilities to enterprises in an "image", so that enterprises can get satisfactory talents and students can get a stage of development. The informationization level of university management refers to the degree of informationization construction in education, teaching, scientific research and management, including the comprehensive level of hardware facilities, software application, information resources and information services. Is to take you as an important information processing process and information content has an important role and significance., which mainly includes the following steps: data preprocessing, data mining algorithm selection, model construction, result analysis and evaluation.

2 Related Concepts

2.1 Mathematical Description of Data Mining Algorithms

The relationship between data and the sporadic of data in the whole process of data mining has important guiding significance for data mining and data processing. Moreover, data mining can improve the level of data management, realize the comprehensive guidance



Evaluation Model of University Management Informatization Level Under the Genetic Algorithm

Suiying Liu^{1,2(✉)} and Yuling Dan^{1,2}

¹ Foreign Languages School, Sichuan University Jinjiang College, Meishan 620800, Sichuan, China

d175966@163.com

² Institute for Rural Development, Sichuan Academy of Social Sciences, Chengdu 610000, Sichuan, China

Abstract. Information management in school is a new mode of its information level development, which is of great significance to the overall management and business level comparison of school. Financial garbage algorithm can improve the information management level of school, optimize data, realize data comprehensiveness, judge and achieve the overall culture of data, and complete the comprehensive evaluation of information. Forget this sentence, it can realize the whole information, how to realize the correlation between data and enhance the research of data.

Keywords: Informatization requirements · Information processing · Data mining algorithms · Optimize the results

1 Introduction

There is still a certain gap between the level of management informationization and the application of data mining in Teaching in school.. On the one hand, some school have insufficient investment in informatization construction, and the application level90 20 is low; On the other hand, the application of data mining technology in some universities is not mature enough, and there is a lack of professional data analysis talents. These problems restrict the development of management informationization level and data mining application in school. With the development of the information age, the era of educational information 2.0 is also a new chapter of educational reform in the next decade. Educational informatization role in optimizing the educational structure, rationally allocating educational resources, promoting educational balance, providing high-quality education, cultivating innovative talents, and even comprehensively realizing educational equity and building a learning society [1]. It is an inevitable choice for China's educational development to lead and support educational modernization with educational informatization. In the era of relatively lack of management method

and improve learning outcomes. Finally, data mining can also be applied to predict students' academic performance and progress in experiments. By analyzing students' data and behavioral patterns during the experimental process, predictive models can be established to help teachers predict students' learning performance, design appropriate teaching strategies and measures in advance, and thereby improve students' learning outcomes and grades. By using the formation of a series of strong association rules data for comparative analysis, so as to effectively analyze the data of physical education practice teaching, provide technical support for classification management, and provide reasonable, effective and reliable data for future physical education practice teaching directional data and suggestions.

References

1. Park, H.S., Park, J.J.: Formation process and content of pre-service physical education teacher's pedagogical content knowledge in teaching practice phase. *J. Korean Soc. Study Phys. Educ.* **23**(1), 37–53 (2018)
2. Batterham, M., Neale, E., Martin, A., et al.: Data mining: potential applications in research on nutrition and health. *Nutr. Diet.* **74**(1), 3–10 (2017)
3. Sung-Geon, P., Kwang-Won, et al.: The related word analysis on korean society of leisure and recreation based on text mining and network analysis: focused on KAHPERD. *Korean J. Phys. Educ.* **56**(3), 389–402 (2017)
4. Sinelniko, V.O.: The influence of content knowledge on pedagogical content knowledge: an evidence-based practice for physical education. *J. Teach. Phys. Educ.* **37**(2), 1–35 (2018)
5. Parmezan Bonidia, R., Duilio Brancher, J., Marques, B.R.: Data mining in sports: a systematic review. *IEEE Lat. Am. Trans.* **16**(1), 232–239 (2018)
6. Yuan, C., Yang, Y., Liu, Y.: Sports decision-making model based on data mining and neural network. *Neural Comput. Appl.* (2), 1–14 (2020)
7. Lee, Y.S., Wang, J.R., Zhan, J.W., et al.: Data mining analysis of overall team information based on internet of things. *IEEE Access* (99), 1–1 (2020)
8. Link, D.: Sports analytics. *German J. Exercise Sport Res.* **48**(1), 13–25 (2018)
9. Kostakis, O., Tatti, N., Gionis, A.: Discovering recurring activity in temporal networks. *Data Min. Knowl. Disc.* **31**(1–4), 1–32 (2017)
10. Porter, A.L.: Data analytics for better informed technology and engineering management. *IEEE Eng. Manag. Rev.* **47**(3), 29–32 (2019)
11. Thabtah, F., Zhang, L., Abdelhamid, N.: NBA game result prediction using feature analysis and machine learning. *Ann. Data Sci.* **6**(1), 103–116 (2019)

Table 2. Network education forms that teachers think are more suitable for College Physical Education

Form	Number of people	Percentage
Network education	242	34%
After class guidance	152	21%
Blended learning	325	45%

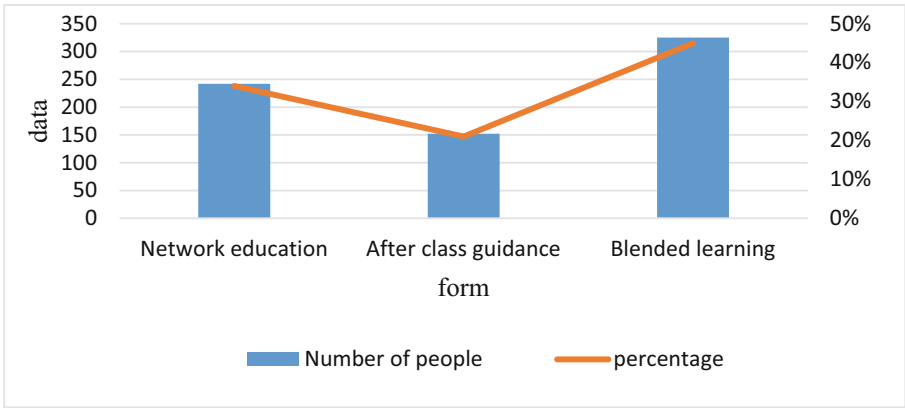


Fig. 2. Network education forms that teachers think are more suitable for College Physical Education

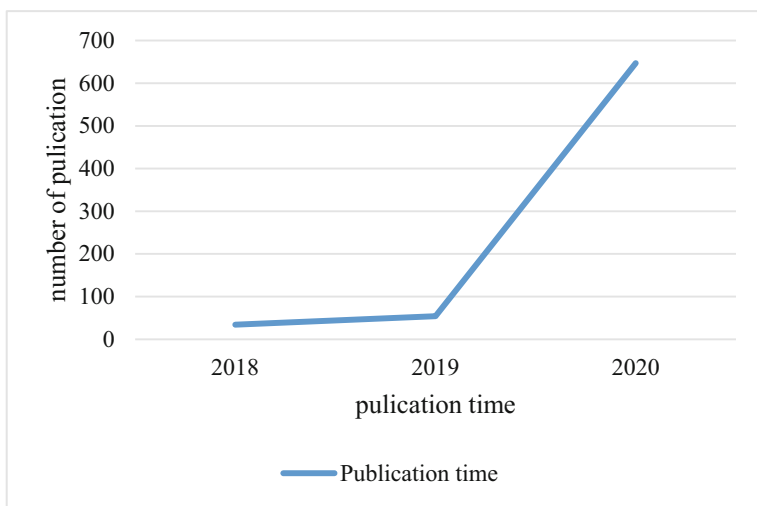
From the above, the teachers think that the maximum number of PE Teaching in the form of mixed learning is 325, and the number of teachers who carry out PE in the form of after-school tutoring is at least 152.

5 Conclusion

The application of data mining in physical education experimental teaching guidance can provide teachers and students with more comprehensive and accurate learning support and personalized guidance. By analyzing the large amount of data generated by students in experiments, such as action data, time data, and energy data, data mining can reveal students' learning patterns, difficulties, and potential. Data mining technology can help teachers understand students' learning and performance in experiments, such as accuracy, speed, and strength of actions. By analyzing these data, teachers can identify students' strengths and weaknesses in different skills and movements, and provide targeted feedback and guidance to help students improve and improve. In addition, data mining can also discover students' learning patterns and preferences, helping teachers differentiate teaching based on individual learning characteristics of students. By analyzing students' learning trajectories and performance data, teachers can provide personalized learning support and guidance to help them better explore their potential

Table 1. P.E. Textbook Publication

Publication time	2018	2019	2020
Number of publications	34	54	647

**Fig. 1.** P.E. Textbook Publication

It can be seen from the above that in terms of the number of copies of university sports textbooks, 34 copies will be published in 2018, 54 copies will be published in 2019 and 647 copies will be published in 2020. The results are shown in Fig. 1.

It can be seen from the above that the number of copies of college PE textbooks is increasing with the growth of time. In 2020, the number of copies of college PE textbooks is up to 647.

4.2 College Physical Education Teachers' Views on College Physical Education Learning

On the role of college sports blended learning, mainly through the form of more suitable for college sports network education, college sports blended learning has led to the reform of PE teaching and sports learning methods, the development of college sports blended learning is shown in Table 2.

It can be seen from the above that the number of teachers who think that PE teaching is carried out in the form of network learning is 242, accounting for 34%; the number of teachers who think that PE teaching is carried out in the form of after class counseling is 152, accounting for 21%; the number of teachers who think that PE teaching is carried out in the form of blended learning is 325, accounting for 45%. The results are shown in Fig. 2.

Promotion degree means that when project set A exists, project set B also exists, but it also controls the popularity of project set B, as shown in Eq. (3).

$$\text{Lift}(A) = \frac{\text{support}(A, B)}{\text{support}(A) \times \text{support}(B)}. \quad (3)$$

3 Experience

3.1 Extraction of Experimental Objects

College Students' physique data aspects: physique measurement data, physique and sports performance. In terms of, due to different grades, different schools offer a wide range of courses. Therefore, this paper is mainly, management and analysis of the other two data. Through DM to classify and associate the students' physique data, timely grasp the, so as to help schools effectively improve the students' physique and health level. In the development of information mining system, the most important work is the design of base table in database. The base table can be simply said to be all the database objects in the database. In the base table, data is organized in the form of a spreadsheet. When designing a database, the responsibility of the system designer is to decide what kind of base table is needed, and the name, type, length, primary key, foreign key and index of the specific field of each base table.

3.2 Experimental Analysis

First of all, we should establish a sports data platform to provide information learning for scientific research. Secondly, popular sports tools and cute screen can eventually develop the integration of DM technology and motion simulation system. Through the system, scientifically formulate scientific lineup, counterpart, and for managers to plans and on-site for team members. The construction of the first simulated examination system needs to be solved through "finding problems, analyzing problems, solving problems, and building a system". Therefore, in the process of constructing the content system of PE teaching theory, we should first find out the system itself, that is, the system itself and the specific PE teaching process, then classify and comprehensively the problems, so as to clarify the problems existing in the content system theory. According to the construction principle of content system, the content system of PE teaching theory is constructed.

4 Discussion

4.1 Positioning of Sports Theory

Sports theory is the operation of university sports value, the excavation of sports cultural content and the orientation of sports learning content. The names of PE teaching materials in Colleges and universities were investigated, as shown in Table 1.

people's pursuit of their own physical beauty, because many factors of human beauty are closely related to sports. People understand information, learn knowledge, enrich life and fill psychological space through TV and Internet. Sports is not only a fashion, but also has penetrated into every family and become an indispensable part of people's daily life.

(2) Physical education practice teaching

In the process of teaching, there are still some deficiencies in the selection of teaching content, teaching organization form, teaching methods and so on. The evaluation in PE curriculum: Zhang Qingxin pointed out the following problems in the process of putting forward the problems and skills in the selection of sports and health knowledge content: (1) lack of guiding and practical consciousness for students to carry out lifelong sports. The teaching content of self exercise is not correct; (2) the bad life style and mental health teaching in the selection of teaching content are not paid attention to; (3) the knowledge of health education is not well integrated into PE, but is bound by the latter. In the new curriculum standard, the goal of health education should be realized in PE. In the process of implementation, in order to achieve these goals, PE teaching is constantly changing. However, in terms of teaching management, most schools adopt a modular approach to health knowledge teaching in theory class, while only a few schools combine health knowledge teaching with PE practice teaching.

(3) College Physical Education

From the particularity of sports practice and sports teaching, the development of sports informatization in Colleges and universities is inseparable from sports classroom practice teaching. To develop in the field of network teaching, we must actively explore and try in blended learning. The research and practice of blended learning should be different from the development of other disciplines. As a kind of education, PE as a kind of education, in addition to sports practice, there are other educational elements. Blended learning can have a variety of structures, and PE can develop blended learning in an all-round way.

2.3 Support

Suppose that a and B are two itemsets satisfying $A \in I, B \in I, A \cap B = \emptyset$. $A - > b$ means that a and B appear at the same time, and $\text{support}(a - > b)$ means the percentage of times that item set a and B appear at the same time in all transactions in transaction set D , as shown in formula (1).

$$\text{support}(A \rightarrow B) = \text{support}(A \cup B) = \frac{\text{count}(A \cup B)}{|D|}. \quad (1)$$

Decision tree algorithm can get less unstable decision tree by segmentation technology. Set s with s sample data and have different value category attribute s , as shown in formula (2).

$$C_i(i = 1, 2, \dots, m) \quad (2)$$

2 Method

2.1 Data Mining

(1) The concept of DM technology

DM technology refers to data mining technology, which is a method and tool for mining valuable, implicit, and unknown knowledge and patterns from a large amount of data. Data mining is the discovery of potential patterns, correlations, trends, and anomalies from large-scale datasets through the application of technologies and algorithms such as statistics, machine learning, and pattern recognition [5]. DM technology involves multiple steps, including data preprocessing, feature selection, data model selection and construction, model evaluation and optimization, etc. It can be applied in multiple fields, such as business, healthcare, finance, education, etc. [6]. In these fields, it can help enterprises and organizations discover market trends, predict demand, improve decision-making efficiency, and improve service quality. Through DM technology, useful information and knowledge can be extracted from massive data, helping people make more accurate predictions and decisions [7]. It can help discover patterns and trends hidden behind data, conduct intelligent data analysis and prediction, and effectively support business development and problem solving. Data mining technology not only has theoretical research significance, but more importantly, it has broad application value in practical applications, providing data-driven insights and decision support for various industries [8].

(2) Application of DM in sports

The results of DM are applied to the daily sports training and competition, and the previous steps are modified according to the implementation situation until the reasonable results are obtained [9]. At the same time, the idea of this model is applied to other projects to realize the reuse of knowledge [10]. Before the emergence of DM, data was only used as the most basic query. DM technology is to obtain deeper information from data [11]. With the massive growth of data, the value of DM has been greatly improved. DM technology can classify and predict students' behavior. For example, the teaching quality evaluation, poor students' judgment and college students' mental health can help efficient managers make more effective decisions.

2.2 Physical Education Practice Teaching

(1) Physical exercise

If the basic skills are practiced purposefully and systematically, the movements will tend to be automatic and reach the skill level. These skills must be based on solid theories. Motor skill is a series of motor actions that students can complete through learning. Through repeated practice, it forms an automatic, perfect and creative action system. Therefore, it has a certain fixed form and variable form, different sports need different motor skills. PE is reflected in people's physical and mental adjustment, which can meet

With the continuous, many experts have studied sports. For example, association rules in curriculum construction, analyzed the relationship between gender and physical education, and discussed the influence of gender factors on PE teaching practice. It is suggested that the major of physical education should focus on pedagogy and train qualified reflective PE teachers. We strongly recommend that more human resources be invested in teaching practice throughout the teacher program. From the teaching practice, extract the elements different from the program, use the idea of DM to provide decision-making basis for PE teaching, trying to reinterpret. The practice of PE reflects strongly the influence of PE on the theory of physical culture while investigating the practice process of extracurricular physical culture. In of school PE management, this paper puts forward the school, and constructs. Using BD mining the allocation of school SM resources, using the information of school SM resources, using the school sports md. In this paper, fuse school physics md, and a physics md is. The school physics management are analyzed by using. The classification and recognition of neural network model improves. The scientific DM of sports performance management is introduced into the new framework. Taking association rule mining as an example, this detail the method of scientific management of sports performance by using DM technology, and briefly analyzes the mining results [1]. Some experts have studied the feasibility of applying DM tools and technologies to the field of sports, especially in the mining of sports competition itself. This paper introduces several typical applications of BD mining visualization. Based on DM technology classification learning decision tree, this the quality of life of PE teachers and their views on the education system. This of DM at home and abroad, in sports competition technology and tactics analysis, national fitness monitoring and sports training. This paper introduces DM technology into scientific management of sports achievements, and framework of scientific management of sports achievements. Taking association rule mining as an example, this the method of scientific management of sports performance by using DM technology, and briefly analyzes the mining results [2]. Some experts have also studied the sports. Through the DM and sorting of sports research literature, using the research method of sports statistics, the sports research literature is analyzed from the aspects of SP, competition, training, application, teaching, review, teaching and so on. The results show that the current on simple theoretical analysis, while the establishment and application of sports field database is relatively simple, which leads to the one sidedness of the research. In the aspect of data, the computer as the database platform, adopts the three-tier architecture to realize each functional module of sports performance management, and uses the analyze the correlation of the input sports performance. This principle of DM, and applies critical to predict the performance of the system [3]. Although there are many achievements in PE research, there are still some deficiencies in the DM in PE practice teaching guidance [4].

In the application of DM in guiding PE practice teaching, this paper studies DM and PE practice teaching, and finds the support. The that DM to the development of PE practice teaching.



Application of Data Mining in Physical Education Experiment Teaching Guidance

Lv Hao^{1,2}, Wang Lan^{3(✉)}, Yi Shao⁴, Xinxin Guan⁴, and Feng Gai⁴

¹ Physical Culture Institute, Yili Normal University, Yining 835000, Xinjiang, China

² Key Laboratory of College Student Physique Monitoring Center, Yili Normal University, Yining, China

³ School of Cyber Security and Information Technology, Yili Normal University, Yining 835000, Xinjiang, China

LH46490325@126.com

⁴ Yantai Vocational College, Shandong 264670, China

Abstract. The application of data mining in physical education experimental teaching guidance can situation and performance in experimental courses, and provide personalized teaching guidance and feedback. By analyzing students' experimental data and behaviors, data mining can discover their learning patterns, difficulties, and potential, and provide data-driven teaching strategies and resources for teachers. Through data mining technology, teachers can analyze students' data indicators in experiments, such as action data, time data, energy data, etc., in order to identify students' strengths and weaknesses in different actions or techniques, and help them provide targeted feedback and guidance. Teachers can use data mining techniques to find students' learning patterns, such as their mastery speed and process of different experimental skills, so that they can conduct differentiated teaching based on students' learning patterns. In summary, the data mining in physical education experimental teaching guidance can provide teachers with a comprehensive understanding and students' learning situation, and help teachers provide personalized teaching guidance and feedback. This will promote the improvement of students' experimental abilities and the optimization of learning outcomes.

Keywords: Web technology · Data mining · Physical education · Teaching practice

1 Introduction

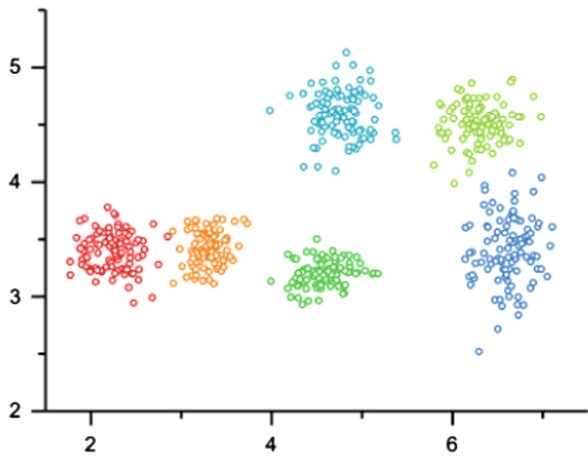
The progress of computer and NT not only for university, but also puts forward higher requirements for students' physique. In sports, there are many kinds of data at different stages. The sports industry is using these for improving performance, avoiding injury, organizing, researchers, coaches, players, game planning, decision-making and even forecasting purposes.

learning analysis technology promotes accurate and comprehensive portraits of learners in the digital age, and promotes the deepening of personalized support services. When dealing with diversity constraints, we first try to give a complete description of the problem in high-dimensional space. Then, the diversity is characterized by classifying the participants and controlling.

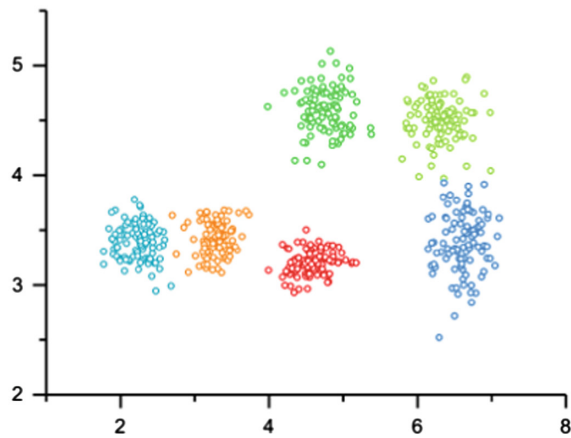
This paper proposes data mining education data mining based on multiple constraint models is the core of digital learning and educational big data analysis, and learning data elements provide strong support for the analysis of multiple constraint models. Data mine the teaching model from multiple levels and angles, to judge the data categories that meet the query conditions, and to query data according to different data categories, so as to realize the database optimization query.

References

1. Qing, C., Cheng, Z.: Data mining model of mobile terminal user preference under interval constraints. *Comput. Simul.* **37**(11), 4 (2020)
2. Tao, Z., Si, Z.: Learner model design and computational research in educational big data mining. *Electron. Educ. Res.* **41**(9), 7 (2020)
3. Peng, G.: Chinese language and literature education and humanistic quality education of college students. *New West: Late Ten Days Theory* (3), 2 (2016)
4. Qianqian, H.: Analysis of the current problems in Chinese language and literature education. *Commodities Q.* **000**(038), 280 (2017)
5. Chun'e, S.: Innovative research on chinese language and literature education. *Arts Lovers: Educ. Teach. Edn.* (2), 1 (2018)
6. Pengfei, X., Qinhua, Z., Yaohua, C., et al.: Research on learner modeling in educational data mining. *China Distance Educ.* (6), 7 (2018)
7. Yumei, J., Hang, C.: The value positioning, realistic limits and effective strategies of educational big data mining. *Educ. Theory Practice* **41**(19), 6 (2021)



a. Set 1 original clustering partition



b. SGA clustering results

Fig. 6. Experimental scatter analysis results

5 Conclusions

That is, the “three-dimensional” curriculum goal must be qualitatively improved, the curriculum structure must match the professional quality of Chinese teachers, the curriculum content should be close to students’ life, the curriculum implementation method should highlight students’ subjectivity, and the curriculum evaluation should focus on students’ development. The universities contains a lot of useful information for students, teachers and school management departments. As explore the law of education from four aspects: student characteristic analysis, talent factor analysis, student employment prediction and student source quality analysis. The development of educational big data and

4 Experimental Results and Analysis

The experiment of this study aims to explore the effectiveness of data mining in Chinese education based on multiple constraint models, and draw corresponding conclusions from the experimental results. The data set used in the experiment is a group of students' Chinese learning data, including students' gender, age, learning time, academic performance and other factors.

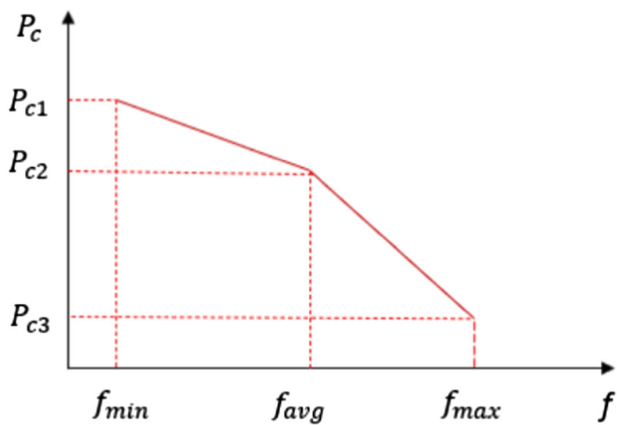
Before conducting the experiment, data preprocessing and feature extraction were first carried out. After cleaning, deduplicating, and filling in missing values on the data, multiple features such as student gender, age, study time, homework completion, and exam scores were selected as input to the model.

Then, Chinese education data mining can be carried out based on a multi constraint model. The advantage of multivariate constraint models is that they can accurately control the correlation between data and avoid interference factors in data analysis. When implementing it, we used the Python programming language and the Scikit learn machine learning library.

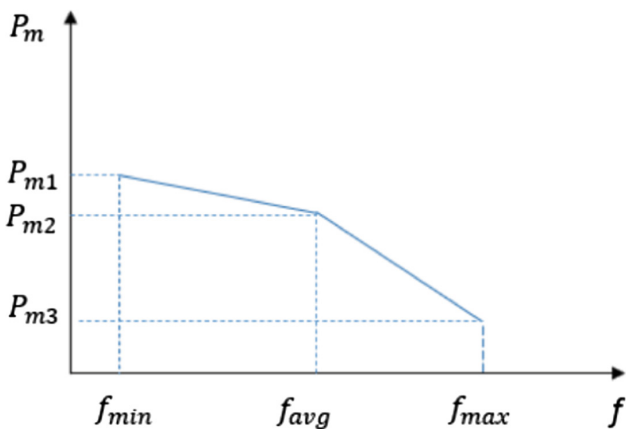
The experimental results indicate that the multi constraint model is very accurate in analyzing Chinese language education data. After using PCA algorithm for principal component analysis, we found that students' age, learning time, and homework completion were most significantly correlated with Chinese language grades, with learning time having the most significant impact on grades. By ranking the importance of different features, we can obtain an overall ranking of data feature importance, which can be used to propose corresponding educational strategies based on data mining results. The experimental scatter analysis results are shown in Fig. 6.

For example, according to the conclusion that students' learning time and homework completion have the greatest impact on their grades, strengthening homework guidance and organizing appropriate learning plans can be helpful in improving students' Chinese language grades. In addition, other features can also be sorted by importance through data mining, which can be used to guide the practical operation of Chinese language education.

In summary, the data mining of Chinese education based on multiple constraint models has achieved satisfactory experimental results. This model can accurately grasp the correlation between data, avoid interference factors in data analysis, and provide beneficial data analysis and decision support for the field of Chinese language education. In practical applications, it is necessary to further enrich the dataset and improve the algorithm model to improve the accuracy and value of data mining.



(a) The trend of PC changes



(b) The changing trend of pm

Fig. 5. Adaptive curve

itemsets of frequent itemsets. When the support threshold is small, a large number of “long” association rules will appear, so the cost of the algorithm is very high, and it is easy to have the problem of combinatorial explosion.

3.2 Diversity as Upper and Lower Bound Constraint Analysis

As an optimization problem, we must first have quantitative optimization objectives. For each transaction, the subset function is used to find out all subsets of candidates in the transaction, and each such candidate is semantically calculated. Semantic computing combines semantic content and semantic links to establish attributes that can describe the function of knowledge elements, and produce knowledge elements with the smallest granularity. The knowledge element extraction is shown in Fig. 4.

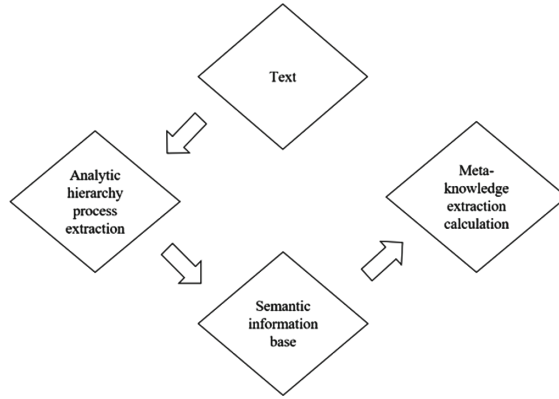


Fig. 4. Knowledge element extraction process

First, generate sets at the same time. The solvable set of linear gauge problems of unequal bundle is $R(A) = \{x: Ax \leq b, x \geq 0\}$. If a is a unitary matrix, then for any integer vector b , all vertices of $R(A)$ are integers. That is, in the constraints of different levels of particles, high-level coarse particles are transformed into low-level fine particles, and low-level fine particles form a richer understanding of the nature of things from different perspectives. The number of constraints is equal to the length of the maximum frequent itemset, so it is necessary to scan the database several times.

$$\max \sum_{i,j} w_{ij} \cdot x_{ij} \quad (3)$$

w_{ij} --The ij first preference scoring matrix;

x_{ij} --Final match.

The types of knowledge points are presented by semantic annotation. The inheritance, generalization and combination of knowledge points are completed by semantic nodes. The adaptive curve is shown in Fig. 5.

For the case where both local and global constraints appear in the extracted sub matrix, we can multiply the row corresponding to the local constraint by—1 and add it to the global constraint (the elementary transformation does not change the determinant). In this way, we transform the sub matrix into a form with only local constraints without changing the determinant of the sub matrix. Local constraints need to generate candidate

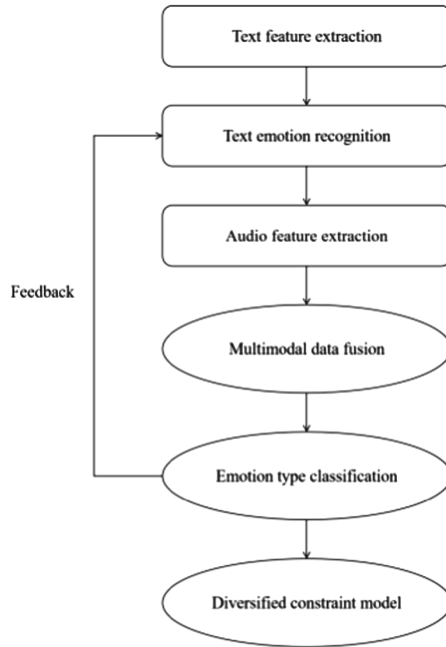


Fig. 3. Multiple constraint model

By processing the data in the database, the data variables can be obtained, and more than 90% of the database information characteristics can be obtained. The change of data is based on the collaborative change of sub model, showing the learning level of a certain stage and the overall degree of understanding things. According to the empirical data:

$$e(y,z)=\left[\sum_{j=1}^pb_{ij}|y_j-z_j|^s\right]^{\frac{1}{2}} \tag{2}$$

b_{ij} --The j and j data attributes in the database

Classify provide accurate data basis for data mining. If there is no way to collect the preference list of numerical values, we can artificially convert the order in the preference list into numerical values. Finding all frequent itemsets quickly and efficiently is the central problem of classification processing and the standard of data mining. Just define a descending function for sorting. The ranking result, as the initial value of the next level model, causes the internal structure of the next level model to adjust and update the collaborative relationship of sub-models. The dynamic structure and feedback mechanism of each level model constitute the dynamic system and iterative cycle of various constraint models, and the model transformation between levels is realized by the learning activity flow composed of a series of learning events.

data collection based on multimodal perception is realized, and the potential characteristics of learners such as “knowledge, cognition, emotion and interaction” are mined.

It is necessary to construct the initial feasible solution to meet all the hard constraints, which can usually be given by random sampling or experience.

Data mining includes algorithm design and analysis model establishment. Encode a specific problem into a string in binary form to generate the original feasible solution. Learner feature analysis model can be designed from six dimensions: basic information, learning style, emotional state, interest preference, social network and cognitive level. By integrating scene awareness modeling method and frequent sequence mining algorithm, learner feature values of each dimension in different learning scenarios can be calculated.

The post-processing stage includes result interpretation, output, evaluation, analysis and use, and explains and evaluates the results. The constraints are easy to describe, and can usually approach or even reach the global optimal solution in specific calculation examples. Visual graphics and colors are used to represent learners’ cognitive process of knowledge, and adaptive knowledge navigation is used to allow learners to view and reflect on their mastery of knowledge, discover myth concepts and promote the improvement of metacognitive ability.

3 Establishment of Diversity Constraint Model and Analysis of Diversity as Upper and Lower Bound Constraints

3.1 Establishment and Analysis of Multiple Constraint Model

In the diversity of the same data attribute is fully considered. Therefore, unnecessary cumbersome data features can be reduced. The establishment of multiple constraints model collects the activity information of different learning service systems across platforms through unified description specifications and operation mechanism, so as to reflect learners’ social status and information sharing. The multi constraint model is shown in Fig. 3.

The following formula can be used to calculate the change parameters of different data attribute characteristics:

$$e(y, z) = \left[\sum_{j=1}^p |y_j - z_j|^s \right]^{\frac{1}{s}} \quad (1)$$

y --The amount of data in the database;

z --Number of attributes;

s --Difference of attribute characteristics.

The analytical mechanism of big data and learning analysis provides technical support for objectively understanding learners’ individual characteristics and establishing perceptual digital portraits of learners, and realizes the calculation and analysis of learners’ multi-dimensional characteristic values in the fields of cognitive level, knowledge structure, behavior and emotion integration.

The methods and technologies of data mining mainly include statistical analysis method, association rule method, decision tree method and so on. To some extent, there is a problem of three-dimensional goal imbalance, but does not pay enough attention to the goal of ability, process and method, emotional attitude and values.

Statistical analysis method is a method to describe and infer the overall information and knowledge by analyzing the sample data in the population. These information and knowledge suggest the internal laws in the population. The statistical analysis method can be completely described as high-dimensional matching.

It is a simple but practical analysis rule. Assuming that the capacity lower bound constraint is not considered, we use Da algorithm to obtain a stable solution, but it is not satisfied with the capacity lower bound constraint [6]. Association rules are widely used in transaction data analysis to guide sales, catalog design and other market decision-making through the analysis results.

Decision tree method is derived from machine learning. It is an inductive learning method based on examples. Characterize the process of learners acquiring knowledge, standardize the data exchange of learners' information in different systems, standardize the learners' information with static features with semantics and syntax, and focus on the presentation of learners' internal relations.

2.2 Data Mining Process

Every semester, the school has to conduct classroom teaching evaluation survey, which has accumulated a lot of data. The knowledge level and mastery level of learners are described through the universal characteristics of knowledge extraction, but the solution of problems in the field of learning practice can not accurately describe the change of knowledge structure and specific cognitive state. We will introduce data mining based on diverse constraint models [7]. The basic model of association rule mining is shown in Fig. 2.

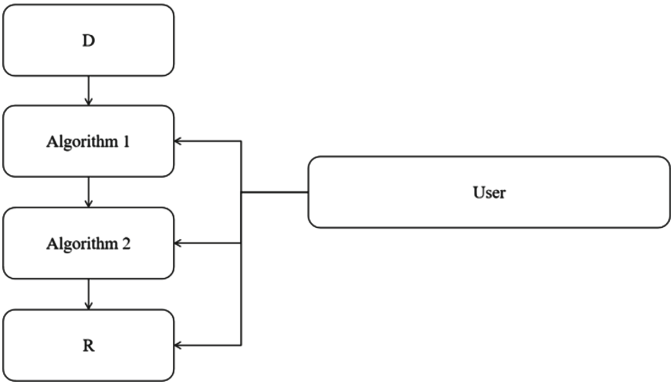


Fig. 2. Basic model of data mining

Preparation stage this stage includes problem definition, object understanding, data collection and other preparations, collects the whole space-time and multi-dimensional

nature of the problem, and the researcher's goals. Here are several common data mining methods:

1. Classification and regression analysis

Classification and regression analysis is a supervised learning method. They are based on labeled datasets and predict the labels or numerical results of new unlabeled data through classifiers or regression models [4].

2. Cluster analysis

Cluster analysis is a unsupervised learning method. It divides data into different sets or groups by searching for similarities.

3. Correlation analysis

Association analysis is a method of mining frequent patterns in data. It can reveal the connections between different elements in the data, such as shopping basket analysis, which can be used to find frequently purchased items together.

4. Abnormal detection

Anomaly detection is a method of determining whether data is abnormal or does not conform to expected patterns. It can detect data points that are different from other data and find data that is slightly different from the general public. Application scenarios such as fraud detection, determining whether equipment needs maintenance, etc.

5. Time series analysis

Time series analysis is a method of processing time series data. It can identify and predict trends and periodicity in time series, and is applied in fields such as stock market and weather forecasting.

The above are some common data mining methods, but there are actually many other methods, such as decision trees, support vector machines, etc. When selecting data mining methods [5], it is necessary to consider the specific scenario and problem nature, and comprehensively select the most suitable method.

Figure 1 shows the relationship between data mining and other fields.

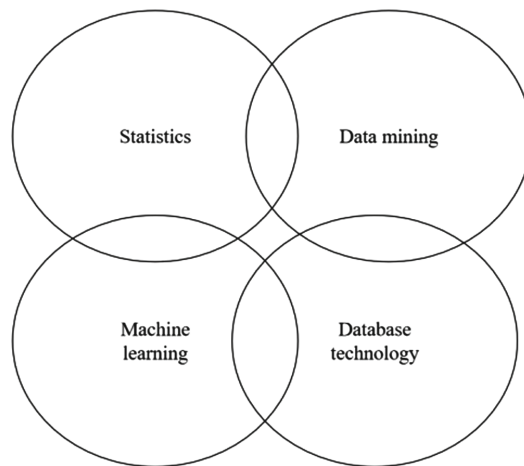


Fig. 1. Data mining is a collection of many disciplines

correlation between data. Therefore, if successfully applied to data mining in Chinese language education, it will greatly improve the efficiency of education and teaching.

Therefore, data mining methods have become a hot topic in the field of information management and have attracted the attention of many scholars. Data mining originated in the mid-1990s. This is a young and active research field. It is a product of the combination of multiple disciplines and technologies. It has broad application value [1]. This also brings great convenience. The most commonly used is data mining based on multi constraint models. Establish a digital portrait of learners, explain the correlation logic between models, and provide intelligent teaching decision support for personalized learning and learners' comprehensive development.

At present, although it is necessary to have a full understanding of its target audience, namely students. At the same time, combining measures and strategies of characteristic education, form a management system with its own characteristics. The problem of course arrangement is not only about simply allocating courses within a reasonable time frame, but also considering the various needs of teachers and students in choosing courses [2]. In theory, it can enrich curriculum theory, deepen people's understanding of the relationship between curriculum reform, and clarify the goals and theoretical basis of curriculum reform. If we want to reflect the different requirements of teachers and students in the course arrangement results, then we need to set some constraints in the course arrangement process to make the course arrangement results meet our expectations.

This study aims to mine Chinese language education data through a multi constraint model, and adopt corresponding educational strategies based on the mining results to improve the quality of education and promote students' learning outcomes. This study will be based on a certain amount of Chinese language education data, including educational goals, learning materials, educational methods, etc., and use multiple constraint models to model and analyze these data.

Through research, we will discover which factors have the most significant impact on improving students' Chinese literacy, and propose corresponding educational strategies accordingly [3]. The research significance of this study is to provide a new approach for the field of Chinese language education, while also providing reference for other educational fields.

Through the experiments and explorations of this study, we are expected to achieve some remarkable results that may be applied in practical education. This will play an important role in the development of future Chinese education and the improvement of students' Chinese literacy. Therefore, this study has important theoretical and practical significance.

2 Data Mining Method Based on Diverse Constraint Model

2.1 Main Methods of Data Mining

Data mining is a process that utilizes computers to automatically analyze large-scale datasets, using various complex algorithms and techniques to extract valuable information from data. The choice of data mining methods depends on the type of data, the



Data Mining of Chinese Language and Literature Education Based on Diversified Constraint Model

Xiaolan Wei^{1,2}✉

¹ Nanchang Institute of Science and Technology, Nanchang 330108, Jiangxi, China
565607996@qq.com

² School of Economics and Management, Guilin University of Electronic Technology,
Guilin 541000, China

Abstract. China culture. When teaching Chinese language and literature, teachers should vigorously and guide the teaching, so as to effectively enhance students' learning enthusiasm. All colleges and universities have established their own teaching management information systems, which have greatly improved the level of teaching and management, and accumulated a large amount of teaching and management data. But at present, most of these information systems are on-line transaction processing, lacking the ability of comprehensive analysis and decision-making. Data mining technology is produced to solve this problem. If the function of classifying and deep-level analysis of the collected data cannot be provided, the managers cannot understand the relevant information in time in the face of complex situations, and the leaders do not have practical data support in decision-making. As the core component in the field of digital education practice, diversity constraint model understands the structural form of Chinese language and literature education data mining in the field of learning practice with internal psychological mechanism and external behavior structure.

Keywords: Chinese language and literature education · Data mining · Multiple constraint model

1 Introduction

With the advent of the information age, data mining has become an important tool for research in various fields as an effective means. In the field of education, Chinese language education has always been a highly concerned field, and educators have been trying to explore the best educational methods through continuous practice to improve students' Chinese literacy.

However, the research on data mining in Chinese language education is relatively limited compared to other fields. This study will explore data mining in Chinese education based on a multi constraint model. The multivariate constraint model not only captures the complex relationships between data well, but also accurately controls the

4 Conclusions

The wide application of big data, Internet and other information technologies in music communication, on the one hand, enables us to enjoy music works more conveniently and quickly, on the other hand, it realizes sharing, from music communication methods such as cloud music, online concert and show live broadcast. The rapid development of new technology and its increasingly wide application provide unparalleled imagination for music communication. It brings us not only convenience and quickness, but also “sharing”. This paper combines data mining technology with music theoretical knowledge, takes various elements constituting a complex cultural system as the analysis object, analyzes the regional diversity and Preference Diversity of music, in order to study the cultural structure of China’s online music society, and puts forward a research on the data construction of music culture communication based on data mining technology.

References

1. Ziyue, Z.: Research on music style classification method based on data mining technology. Nanchang University **23**(1), 4 (2013)
2. Linlin, Q.: Research on the application of data mining technology in music pop trend prediction. Southwest University **25**(2), 8 (2017)
3. Ziyue, Z.: Research on music style classification method based on data mining technology. Nanchang University **43**(1), 5 (2014)
4. Ziyue, Z.: Research on music style classification method based on data mining technology. Nanchang University **35**(2), 4 (2016)
5. Tingting, Z.: Application of data mining technology based on association rules in music classification. Modern Electron. Technol. **46**(7), 6 (2020)
6. Xiao, C., Shuo, C., Min, C.: Research on music curriculum model under cloud computing and data mining technology—comment on algorithmic composition theory and practice. Chinese Sci. Papers **14**(7), 1 (2019)
7. Yan, D.: Research on music curriculum model under cloud computing and data mining technology. Autom. Instrum. (9), 3 (2018)
8. Ziru, Z., Muxuan, M., Huan, X., et al.: Research and analysis of Chinese folk song industry based on data mining. Inf. Technol. Inf. (10), 4 (2019)
9. Xingyu, Y.: Research on music playing data mining based on FP growth algorithm and ARIMA model. South China Agricultural University **33**(5), 6 (2017)
10. Yan, L.: Research on fine marketing of music brand value-added services based on data mining. Pr World (22), 2 (2022)

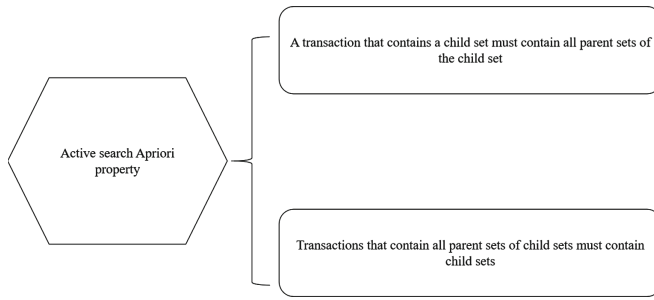


Fig. 4. Properties of active search Apriori

Cloud computing is a kind of pay by usage pattern, this pattern provides available, convenient, on-demand network access, enter the configurable computing resources Shared pool (resources including network, servers, storage, applications and services), these resources can be quick to provide, just in the management of the very few and or little interaction with service providers. At the same time, users can also easily and quickly retrieve music works that meet their requirements through cloud computing, thus effectively saving users' time and energy spent on music retrieval. Kullback-leibler divergence, or relative entropy, used in statistics to calculate the match between two probability distributions, and in machine learning to measure the approximation between two functions, is calculated by the following formula:

$$KL(p\|q) = \sum p(x) \log \frac{p(x)}{q(x)} \quad (3)$$

The active search Apriori algorithm only needs to scan the transaction database once, and when calculating the support of frequent itemsets, it uses the active search strategy to replace the passive scan, which effectively reduces the time complexity. Therefore, we should fully grasp the advantages brought by new media, constantly refine and spread our music culture, and actively push our excellent music culture to the world. In the previous research, the knowledge ontology in the music field has been constructed, in which four sub-ontologies are defined to describe the superior conceptual structure, namely, music events, music works, music figures, music places and music technologies. For each track, we add a unit value to the elements of the track vector corresponding to tags a, b, and c if the playlist that contains tag information (such as tags a, b, and c) contains the track. Overall, the future trend is that cloud computing, as the bottom layer of computing resources, supports the upper layer of big data processing; and the development trend of big data is real-time interactive query efficiency and analysis capabilities. The time complexity of this algorithm and the number of times of scanning the transaction database are lower than those of the algorithm based on transaction compression technology; compared with the matrix-based algorithm, the time complexity is the same, but the space complexity is lower than that of the matrix-based algorithm.

of cloud music providers to create their own private cloud, then upload all their collected music works to their own private cloud, and then use relevant music playback software to play music on computers and share the music works uploaded to the cloud on mobile phones and other devices, so that users don't need to spend time and effort to store and copy the collected music works, and can give full play to the characteristics of network interaction to the greatest extent. When applied to music style differentiation, each itemset needs to have a decision attribute D . When connecting to generate candidate $K + 1$ itemsets, two frequent K itemsets that can be connected must meet the following requirements:

$$(l_a, l_h) \in \{(l_1, l_2) | \text{diff}(l_1, l_2) = 1, d_1 = d_2\} \quad (1)$$

It is worth noting that the distribution of regional diversity of individual users in Tibet and Xinjiang shows a certain degree of non-monotony in regions with low regional diversity. In the music industry, the data into the hands of the many music company executive, they through a variety of data collection and analysis of comprehensive, not only can clearly understand the grades in each song on flow media service web site, the popularity of the audience, and, more importantly, we collected data through to the large audience feedback information, statistics, analysis and mining, I can clearly grasp the situation of the main audience and the music types they like, and even the listening habits of the audience can be well informed. Considering the correlation between these diversities and the economic development level of each region, it can be found that users in regions with higher economic development level are more likely to prefer minority music with low regional diversity, revealing the relationship between economic development level and the personalized music preference of users.

The MID music file format was put forward in 1980s, which is different from WAV in that MID music file is not a sound waveform signal but a series of instruction signals, such as notes, time values and playing skills, etc. These instructions can directly control MID equipment to perform music.

3.2 Classification Rule Extraction

The extraction of classification rules is to complete the data preprocessing first, and then use the association rules in data mining to classify music styles. In the era of big data, digital processing of music content alone is far from meeting the needs of effective music dissemination, and a more convenient way of management and calculation is needed, that is, dataization. For the active search type Apriori algorithm, its advantages are that it is simple, reliable, and easy to implement. It is a very mature mining algorithm for association rules. In the environment of large transaction databases, the active lookup Apriori algorithm has absolute advantages compared with the matrix-based Apriori algorithm, and the specific advantages are shown in Fig. 4.

In the process of extracting music classification rules using active search Apriori algorithm, it is necessary to set a confidence threshold, extract rules from a frequent itemset, and calculate the confidence of the rules:

$$\text{Confidence} = \frac{\text{support}}{|\text{matched}|} \quad (2)$$

only promote the circulation of music culture products, but also promote the free flow of music culture market. The music culture system has opened up development space for the dissemination of music culture while realizing the optimal allocation of music culture resources.

At this stage, many technologies can carry out audio recognition, but one of the most widely used music retrieval technologies based on fingerprint can effectively extract the fingerprint in the original waveform music, then analyze the extracted fingerprint information and automatically match the music works with the highest similarity. Apriori property is that all non empty subsets of frequent itemsets must be frequent itemsets; The supersets of all non frequent itemsets must be non frequent itemsets. In modern music culture, various social thoughts appear and there are many ideological schools. The advantage of the algorithm is that the algorithm is simple, stable and easy to implement. The disadvantage is that it scans the transaction database too many times and costs a lot.

3 Introduction and Preprocessing of MIDI Music Files

3.1 Introduction to MIDI Files

In this era of Digital music, music file formats are very diverse, and we are familiar with MP3, WAV—like, MID is also a common music file format, its full name is Musical Instrument Digital Interface (Musical Instrument Digital Interface). Music culture, from tradition to modernity, from classics to network, presents the important characteristics of rich variety and long history. It is extremely difficult to spread all music culture. MIDI music files are used as the source data for music style classification analysis, and the reasons are shown in Fig. 3:

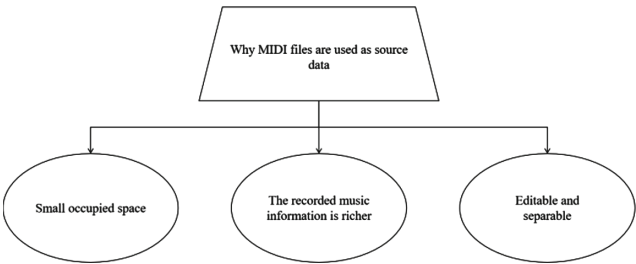


Fig. 3. Why MIDI is the source data

Therefore, when disseminating music culture, it is necessary to take the essence, remove the dross, scientifically sort out, select and refine, spread the ideological value and widely influential music culture in Chinese music culture, and fundamentally occupy the new media The development space, actively create music culture, let people understand music, love music culture. When the total amount of data reaches a certain level, other characteristics of things hidden behind the data will be revealed, and more connections between things can be found. Cloud storage and cloud computing are simple manifestations of cloud computing technology. Users can use the cloud storage service

new technologies and models for storage, calculation and transmission, which brings us many unexpected applications. When the disseminator spreads the basic information of music culture to the network, the receiver will respond in the first time. Besides, it can also forward the basic information of music culture to others, thus fundamentally realizing the basic characteristics of two-way interaction and multiple interactions in the development process of music culture.

2.2 Association Rules Mining Algorithms

Artificial intelligence, statistics and database technology have laid a solid foundation for the birth of data mining technology. Artificial intelligence and statistics are both disciplines that study how to find knowledge from existing data and predict through existing knowledge. The research on the feature mining of cultural products can give full play to the advantages of big data and intelligent analysis means. The process of classification method is to classify the first layer, including melody analysis and harmony analysis. After obtaining the corresponding decision classification, take the results of the first layer classification as the conditions of the second layer classification. Music style classification is a bottom-up process, as shown in Fig. 2.

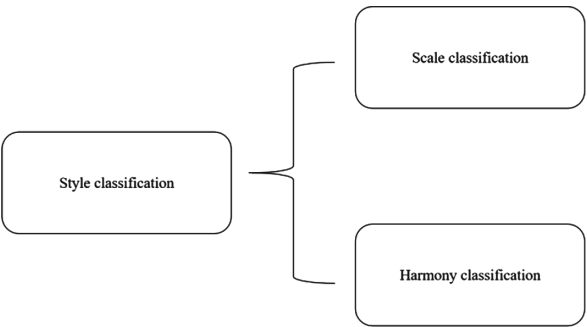


Fig. 2. Music style classification process

Classical algorithm is a classic algorithm for mining frequent itemsets from database. This algorithm connects and prunes the itemsets in database according to the nature. Finally, according to the preset threshold, frequent itemsets and strong association rules can be obtained. Traditional statistics usually adopt the method of sampling statistics. In the era of big data, we analyze the relationship between things by analyzing all existing data related to topics. Now, according to the understanding of the music culture industry, we can know that there is a lack of a unified and standardized cultural system in the process of communication and development of Chinese music culture, which fundamentally restricts the communication of Chinese music culture. Traditional text-based music search has its inherent limitations. Searchers must know the related text content of music works, including the title and introduction of works, while audio-based search and discovery will largely make up for the defects of text search. As an important part of the development of music culture, the music culture market can not

2 Outline of Basic Theory

2.1 Overview of Data Mining

In the past century, the Internet has become an important channel for people to store, obtain and transmit information. However, with the development of Internet technology and the improvement of storage media, information explosion is an important problem that can not be ignored. Music platform operators to accurately compared to users recommend music works, you need to pass the identity of the user ID and mobile client download information to the user, often listen to the music, and comment on information behavior related to the real-time collection, on the basis of the filter, filter, after analyzing and mining the data information that exist in the contact and characteristics, In this way, users' behaviors and habits can be fully grasped, and corresponding tracks can be pushed accurately for different end users. Data mining technology is more dependent on the powerful computing power of computers, and the whole process of knowledge discovery is roughly shown in Fig. 1.

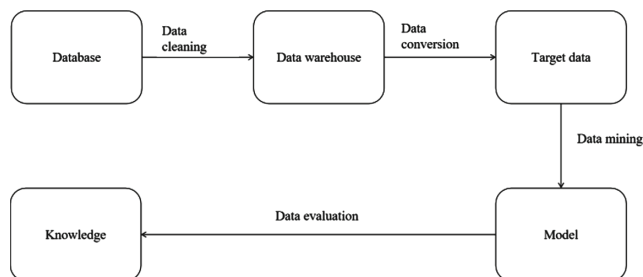


Fig. 1. Process of knowledge discovery

Data on the Internet continues to grow exponentially. New media has various forms of expression and can provide equal opportunities for people to accept music in the development process. To be precise, big data means to collect as many different variables as possible from all aspects to describe various characteristics of people, things, things and phenomena. Each participant can participate in it through different platforms and communicate with each other about music culture, becoming the subject of the new media era. For music perception, we have acquired rich basic cognition in personality influence, social identity, cultural background and cross-cultural influence, and established a relatively complete theoretical framework. In the process of music transmission, music must be connected with human behavior, and the amount of human behavior information data is very large. In the face of the ever-increasing need to record these variables, the number of retrieved data is also increasing rapidly, so the database capacity is getting larger and larger, and finally it is close to infinity. Therefore, a large amount of information will be submerged in the ocean of data. When the traditional media spread music culture, the disseminator is the main center, forming a point-to-point communication mode. In the new media era, the communication of music culture has realized the communication between point and point, point and surface. Processing massive data requires

used as a theme to lead the trend of the new era [3]. With the advent of the new media era, the media dominated by digital technology and network technology have played an important role in the inheritance and dissemination of music culture, but the ways and methods of communication have also changed [4]. The inheritance and development of music culture takes place in the process of communication, and communication, as the basis for the survival of music culture, is inseparable from the media at every stage of development [5].

As a music industry that is highly dependent on communication technology, it is one of the first fields to be deeply affected by the communication technology revolution in the era of big data. In addition, music has played an important role in the evolution of human group culture, including the preservation of human knowledge, the promotion of ritual activities and collective behavior, and the promotion of human group cohesion. In particular, digital music, which has gradually developed in recent years, is likely to innovate and change the way and channels of popular music transmission [6]. First of all, the development of digital music brings possibilities for the innovation of communication methods and channels. Digital music stored in digital format is mainly spread through the Internet, and its biggest feature is that no matter how many times the same piece of music is played or downloaded, the quality of the piece will not change [7]. From the perspective of the public, people can no longer enjoy music only through radio and television stations and buying records. They can obtain high-quality music with only a multimedia computer.

The emergence of the new media era has broken the development of traditional music media in China, which is reflected in the time, scope and space of communication. In addition, in the use of new media, media users can send or receive audio, video and text images related to music culture at any time and any place. Music is not only a work of art, but also a cultural consumer goods widely disseminated and consumed by audiences at different levels [8]. This feature of digital music gets rid of the medium carrying music works, and is the necessary basis for it to adapt to the high-speed development of the Internet [9]. In the era of big data, human beings have recorded massive amounts of electronic data through a large number of sensor devices deployed in their daily lives. We use social perception as a research tool to analyze the big data of human social behaviors and interactions, study social culture and its evolution rules, and try to discover human behavior patterns and build explanatory models for them. This feature enables digital music to adapt to the development requirements of the Internet, and completely gets rid of the traditional media that carries music works. In recent years, various music styles emerge one after another, fully demonstrating the personality of music creators [10]. The diversity of music makes music style an important feature of Internet music retrieval, and most people are used to choosing the music they might like according to their favorite style. The openness of new media makes the world a whole, and music cultures from different countries and regions can be displayed on the new media. In the new era when big data and cloud computing become mainstream information processing technologies, the dissemination of digital music needs a new communication model.



Data Construction of Music Culture Communication Based on Data Mining Technology

Sufang Wang(✉) and Guixiang Lou

Dongying Vocational Institute, Dongying 257091, Shandong, China
wangsuofangyin@163.com

Abstract. In recent years, my country's big data and information technologies such as the Internet have made great progress, and have been widely used in various fields. The application in the field of music communication has greatly impacted the traditional music communication methods and channels. Internet technology has promoted the development of the music industry, and various emerging music styles have emerged one after another, and users' appreciation preferences also vary from person to person. But at the same time, whether it is a folk song or other excellent pieces of traditional music, after being discovered, with the intervention of the media, it will bring a certain amount of hype and commercial purposes, and they will be deleted and modified, and in the end only unrecognizable "boutiques" will be left. The rapid development of society is inseparable from the joint action of many factors. Culture is one of the important factors driving the evolution of social self-organization system. To develop music culture, we should first preserve the original flavor. Only correct preservation can provide a good foundation for all kinds of development in the future, and only better maintaining the tradition can better nationalize. Based on data mining technology, this paper expounds the basic characteristics of music culture in the new media era, and studies the basic strategies of music culture communication and protection in the new media era, in order to promote China's music culture towards sustainable development.

Keywords: Data mining technology · Communication of music culture · Data-based construction

1 Introduction

At present, China has entered the era of big data information, and various new Internet technologies such as big data and cloud computing have revolutionized many aspects of our lives [1]. The inheritance and development of music culture takes place in the process of communication, and communication, as the basis for the survival of music culture, is inseparable from the media at every stage of development [2]. As a classic art form, music has had a profound influence on people's lives. Music can be used as a way of entertainment to relax. It can also be used as a medium to spread culture; Can be

5 Conclusions

Under the traditional education idea of physical education teaching method is easy to kill the students' individual character, also has restricted the students' all-round development, under the new curriculum standard of physical education teaching method, should be people-oriented, attention to student's individual differences, to conform to the education of students individual differences, timely feedback of students' classroom learning, also want to in many ways to inspect students' sports ability, reasonable use a variety of ways of evaluation, comprehensive evaluation of students, let the students to be able to own a objective cognition [10]. Teachers should actively apply the curriculum concept of "caring for students' personal differences" in the new curriculum standards into physical education teaching, take "health first" as the guiding ideology of physical education teaching, innovate physical education classroom teaching methods based on students' individual differences, and fully implement them to lay a foundation for the healthy growth of students in the future.

References

1. Huang, S.-C., Wu, C.-F., Chiou, C.-C., et al.: Intelligent fintech data mining by advanced deep learning approaches. *Comput. Econ.* (21), 213–215 (2021)
2. Virupaksha, S., Dondeti, V.: Anonymized noise addition in subspaces for privacy preserved data mining in high dimensional continuous data. *Peer-to-Peer Netw. Appl.* (32), 215–218 (2021)
3. Agany Diing, D.M., Pietri Jose, E., Gnimpieba Etienne, Z.: Assessment of vector-host-pathogen relationships using data mining and machine learning. *Comput. Struct. Biotechnol. J.* (2), 210–212 (2020)
4. Thomas, M.L., Martin, A.S., Eyler, L., et al.: Individual differences in level of wisdom are associated with brain activation during a moral decision making task. *Brain Behav.* **6**, 9–11 (2019)
5. Jonathan, M.P. Wilbiks, A.B.: Individual differences in multiple object tracking, attentional cueing, and age account for variability in the capacity of audiovisual integration. *Attention Percept. Psychophys.* (11), 82–84 (2020)
6. Dapeng, Y., EungSoo, O., Yingchun, W.: Hybrid physical education teaching and curriculum design based on a voice interactive artificial intelligence educational robot. *Sustainability* (19), 12–15 (2020)
7. Wenwu, H., Liaokun, Y.: Impact of big data technology on the diversity of physical education teaching methods. *J. Phys. Conf. Ser.* (4), 174–176 (2021)
8. Jingwen, Z., Yoo Jung, O., Patrick, L., et al.: Artificial intelligence chatbot behavior change model for designing artificial intelligence chatbots to promote physical activity and a healthy diet: viewpoint. *J. Med. Internet Res.* (9), 22–25 (2020)
9. Cong, D., Sang-Bing, T.: Step-counting function of adolescent physical training app based on artificial intelligence. *Math. Problems Eng.* (7), 4–7 (2020)
10. Liu, Y.R., Vijayalakshmi, S.: An artificial intelligence and machine vision based evaluation of physical education teaching. *J. Intell. Fuzzy Syst.* (2), 40–43 (2021)

experiment and calculation, students in a class of a certain region were selected as data samples to calculate the learning efficiency of students under different physical education teaching methods. According to the relevant calculated data, Fig. 1 can be drawn as follows.

At the same time, the corresponding formula was used to calculate the learning quality of students in different physical education teaching methods, and the data were compared and analyzed, as shown in Fig. 2.

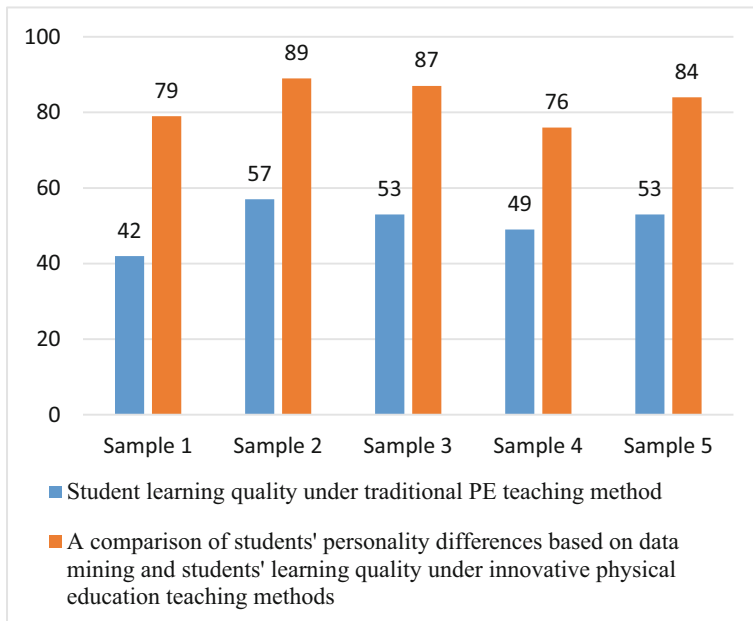


Fig. 2. Comparison of students' personality differences between traditional physical education teaching methods and data mining and comparison of students' learning quality under innovative physical education teaching methods

The comparison of students' personality differences based on data mining and its innovative physical education teaching methods are based on the differences of students' learning ability, knowledge, hobbies, personality, intelligence, structure, personality and thoughts in teaching, changing the traditional physical education under the guidance of the concept of exam-oriented education. For the unity of the traditional style of education content, education target, education method, to change the teaching evaluation, according to the students' learning basis, cognitive ability of personalized, differentiated education, vary from person to person to choose the education goal, teaching method and content, such as respect for each student's development rights, and promote students' academic progress and individual development, Provide sufficient possibility for students' personality growth, so that students can really learn and gain something in the physical education class.

Table 1. Comparison of students’ personality differences between traditional physical education teaching methods and students’ personality differences based on data mining in a certain school class and different classroom participation of innovative physical education teaching methods

The teaching method	The class 1	The class 2	The class 3	The class 4	The class 5
Traditional PE teaching methods	65.8%	58.9%	62.9%	69.1%	64.6%
A comparison of students’ personality differences based on data mining and its innovative physical education teaching methods	86.3%	79.8%	82.4%	89.6%	83.8%

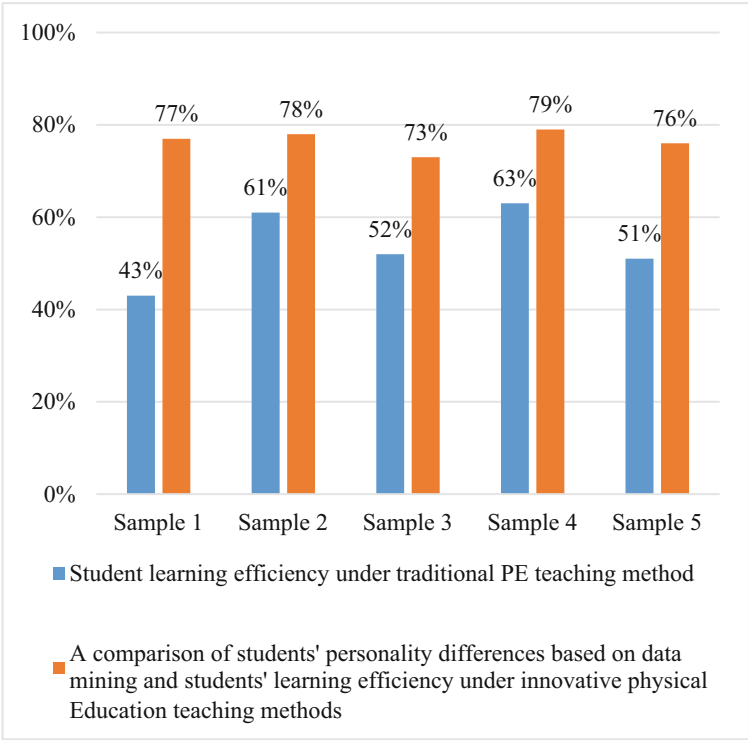


Fig. 1. Comparison of students’ personality differences between traditional physical education teaching methods and data mining and comparison of students’ learning efficiency under innovative physical education teaching methods

teaching method and students' personality differences based on data mining and its application in the innovation of physical education teaching method is as follows:

$$P = (Z_1 - Z_2)2^n \cdot x^k x^{nk} \quad (2)$$

4 Comparison of Students' Personality Differences Based on Data Mining and Data Analysis Under the Application of Physical Education Teaching Method Innovation

4.1 Comparison of Students' Personality Differences Based on Data Mining and Advantages of Innovative Physical Education Teaching Methods

Physical education teaching methods based on the individual differences of students, change the original teacher's single setting of teaching objectives and methods, according to the actual situation of students, detailed analysis of educational objectives, so that each student can learn physical knowledge and skills according to their own goals, so as to maximize the realization of the purpose of physical education classroom. The physical education teaching method based on students' personality differences pays attention to the needs of individual students, so that the education received by different students can be more effective on the students. Teachers changed their teaching methods on the basis of flexible and varied teaching methods to better take into account individual differences and encourage students to learn more in the physical education class and improve their personal development.

In the process of carrying out physical education, teachers should try their best to ensure that students take an active part in teaching. Under the traditional physical education model, organizational elements have always been the main index to measure the effect of physical education, and educational activities emphasize the unity of students. In fact, from the concept of health education, every student's physical quality, sports performance, sports needs are very different. In the process of carrying out physical education teaching, teachers should follow the characteristics of students' individual differences to ensure that students' health is not affected, and on the basis of students' different interests and hobbies, carry out physical education activities purposefully.

By investigating some school class of traditional sports teaching method compared with students individual differences based on data mining and its different classroom participation of innovative teaching methods, we can see students for traditional sports teaching methods and students individual differences based on data mining to compare different attitude and innovative teaching methods, further reaction of classroom teaching, The details are shown in Table 1.

4.2 Analysis of Calculation Results

Based on the meta-ancestor theory, we calculated the differences of students' personalities between traditional PE teaching methods and data mining, and the efficiency and quality of students' learning under innovative PE teaching methods. According to the

teaching method to improve students' enthusiasm and the PE difficulty teaching method to set up difficulties, etc.

3 Study on the Comparison of Students' Personality Differences Based on Data Mining and Its Application in the Innovation of Physical Education Teaching Methods

3.1 Innovative PE Teaching Methods Based on the Comparison of Student Personality Differences Based on Data Mining

In physical education, there are many factors that should be considered in the determination of educational goals. However, in actual physical education, many teachers ignore the differences of students' personal level when setting educational goals, which is easy to occur the phenomenon of "equalizing the water in a bowl" without considering the actual situation of students. At the same time, the traditional physical education teaching method is very single, with the "collective teaching method" as the main teaching method. This teaching method seems to face all students, but actually ignores the differences of students. Educational evaluation is an indispensable link of education and an important activity to judge the value of educational process and results. However, in the evaluation of students in physical education, the individual differences of students are ignored, which does not reflect the real significance of educational evaluation.

"Physical Education and Health New Curriculum Standards" pointed out that in the teaching process should pay attention to individual differences, teachers are required to pay attention to the similarities and differences between students, not only pay attention to the teaching content, but also pay attention to the actual situation and students combined. Teachers must allow students of different levels and talents to see their starting points and conditions for developing their potential, so that students can achieve important learning through hard work, improve their self-confidence, stimulate their enthusiasm, allow students to devote their whole body to physical education and lay a good foundation for their future.

3.2 Data Calculation Formula

Based on the meta-ancestor theory, excluding all the influencing factors, the research focus is placed on the influence of physical education teaching methods on students' learning quality and learning efficiency, and then the collected data samples are classified, summarized and integrated. The calculation formula of learning efficiency of the traditional PE teaching method and the comparison of students' personality differences based on data mining and its application in the innovation of PE teaching method is as follows:

$$H = (P_1 - P_2) \div P_g + v_2^2 - v_1^2 \quad (1)$$

On this basis, the specific calculation formula of learning quality of physical education teaching method under the comparison between traditional physical education

2 Students' Personality Differences and Physical Education Teaching Methods

2.1 Personality Differences Among Students

Personality difference refers to individual differences in physical characteristics, psychological structure, personality characteristics and other aspects influenced by biological heredity and social environment [6]. The scholar Gardner put forward the theory of multiple intelligences, dividing human intelligence into linguistic intelligence, visual-spatial intelligence, introspective intelligence, interpersonal intelligence and so on, and demonstrated the individual differences between people from the aspect of intelligence. The famous psychologist Erich Fromm once said, "All people have their own unique personalities". This is determined by the characteristics of an individual's living environment. Students' learning activities and a person's growth process have unique personalities, and students' personality differences are necessary factors in educational and teaching activities. Only with the promotion of educational reform can educators and others better realize the value and significance of education [7].

The differences in education received by students, the neglect of the collection and research of learning situation before teaching, the outdated teaching philosophy of teachers, and the unscientific evaluation methods of teachers are the main reasons for the individual differences among students, which are obtained from the routine education and teaching and the observation of students' learning practice and situation.

2.2 Physical Education Teaching Methods

With the acceleration of quality-oriented education, the reform of physical education mode in middle schools has become an important link in the improvement of education level, and also an important factor affecting students' all-round development of morality, intelligence, physical education, beauty and labor. Not only that, high school physical education teachers from the perspective of as much as possible, a comprehensive understanding of the new curriculum goal put forward specific requirements in the reform, and understand the dominant position of student education, insist on many aspects to improve education pattern, the education content innovation, constantly improve the students' learning desire, a more scientific way, improve the comprehensive quality of students, to carry out the quality education concept, further complete the teaching goal.

According to the famous theory of "Optimizing the Educational Process" established by Pedagogy Babanski, the physical education discipline refers to the educational content and teaching methods of physical education. The best teaching plan should be formulated for students according to their individual differences and then flexibly applied to improve the teaching efficiency [8]. Physical education teaching methods should highlight the key points and difficulties, consider the students' own acceptance ability, take care of the general students, pay attention to the feedback collection of students' learning in teaching, so that students can carry out more targeted sports theory learning and sports practice activities [9]. There are many PE teaching methods, including the PE stratified teaching method based on the concept of people-oriented, the PE pair teaching method based on students' personality differences, the PE competition

teaching methods of the subject, in order to achieve the training goal of comprehensive development of students [1]. However, teachers will still encounter many difficulties in PE teaching under the guidance of the new curriculum standard, which requires teachers to adhere to the people-oriented, consider the actual situation of students, and constantly reform teaching methods. In physical education teaching, the teacher must admit the individual differences between students is objective existence, professor in the same sports knowledge, accepting and understanding degree of students is not the same in different, so in the actual teaching process, teachers if not adopt the unified teaching way, we should combine the actual situation of students, attention to each student's individual differences, To ensure that every student's learning needs can be met, only in this way can all students be effectively promoted on the basis of the original sports [2].

Since the mid-1980s, with the progress and pace of domestic physical education reform, the innovation and reform of physical education teaching modes and methods in primary and secondary schools have gradually developed into the research focus of experts and scholars in the field of physical education in primary and secondary schools [3]. They contain special ideas and philosophies of physical education teaching, as well as targeted physical education teaching objectives, in the unique physical education classroom environment, to realize their own physical education teaching functions, constitute a more effective physical education classroom teaching framework. It can not be ignored that different teaching concepts and implementation methods of PE classroom will form more diversified PE classroom teaching modes. At present, China's research on physical education teaching methods has been deeply explored, but due to the late start of research and short development process, the basic theory of school physical education is relatively weak [4]. In recent years, the research has made some progress and achieved good results. However, the research of physical education is still facing the dilemma that there is no breakthrough. Scientific research on foreign education model has experienced a long time. Due to the cultural background of each country and other complexity, the understanding of overseas study has a variety of definitions [5].

This paper studies the comparison of students' personality differences based on data mining and its application in the innovation of physical education teaching methods. Based on the meta-ancestor as the basis of the calculation model, the calculation algorithm is constructed according to its theoretical knowledge. Calculate the teaching efficiency and teaching quality of the traditional physical education teaching method and the comparison of students' personality differences based on data mining and the application of the physical education teaching method in the innovation of physical education teaching method. The paper discusses the significance of the physical education teaching method for traditional physical education based on the comparison of students' personality differences based on data mining and its application in the innovation of physical education teaching method.



Comparison of Students' Personality Differences Based on Data Mining and Its Application in the Innovation of Physical Education Teaching Methods

Jianxin Zhang^{1,2}, Xiaoyu Shi^{1,2(✉)}, Shi Minfa³, and Ruyong Zhang³

¹ Physical Culture Institute, Yili Normal University, Yining 835000, Xinjiang, China
shixiaoyu1221@126.com

² Key Laboratory of College Student Physique Monitoring Center, Yili Normal University,
Yining, China

³ Shandong Institute of Commerce and Technology, Jinan 250103, Shandong, China

Abstract. The emergence of data mining technology promotes the analysis and comparison of individual personality differences. The deepening of the new curriculum reform promotes the reform of physical education teaching methods. The traditional physical education teaching methods can no longer meet the requirements of the society for the cultivation of people with all-round development of morality, intelligence, physique, beauty and labor. The difference of individuation is an objective phenomenon that cannot be ignored in the process of educational practice, which is also the basic premise of our educational practice. Comparison of students individual differences based on data mining and its use in the sports teaching method innovation can fully take into account the differences in each student's body and mind and personality, and fully consider the school and the parents, by actively participating in various sports classroom teaching, so as to respect the rights of every student, care and to develop each student's personality as the teaching goal, So that every student in its physical and mental and social development direction, growth rate and the ultimate degree of development can be achieved in the same. Based organization based on the data as the main algorithm, based on data mining to the student individuality difference comparing the sports classroom and traditional classroom teaching quality of practical teaching efficiency and the actual calculation, by analyzing the calculated respectively by 18.7 and 23.1% higher than that of the traditional sports teaching means in our country.

Keywords: Data mining · Personality Differences · Physical Education · Teaching Methods

1 Introduction

With the deepening of the reform of the new curriculum standard, quality-oriented education has been paid more and more attention, and the all-round development of students has been paid more and more attention. The society and teachers continue to reform the

8. Shaari, M., Rahman, M.S.A.: Utilisation of data mining technology within the accounting information system in the public sector: a country study—Malaysia (2022)
9. Chen, C., Xu, W.: Innovation and application of college students' education and management based on big data. In: ICBDE '20: 2020 The 3rd International Conference on Big Data and Education (2020)
10. Ghatge, M., Bhoite, S.D.: Improving College Students Performance Using A Data Warehouse (2016)
11. Chen, L., Wang, L., Zhou, Y.: Research on data mining combination model analysis and performance prediction based on students' behavior characteristics. *Math. Problems Eng.* (2022)
12. Huang, Z., Liang, Y.: Research of data mining and web technology in university discipline construction decision support system based on MVC model. *Lib. Hi Tech* (2019)
13. Yang, Z.H., Pan, et al.: Research on electronic commerce information management system based on data mining. *Agro Food Ind Hi Tech* **28**(3)(–), 2627–2631 (2017)
14. Luan, J.: The development of library information management system based on data mining technology. *Autom. Instrum.* (2016)
15. Han, S.H.: Discussion on computer data mining technology and its applications in archive information management system. *Electron. Sci. Technol.* (2016)
16. Wei, R., Ting, M.: An optimized design and development of student management information system in college and university based on MVC. *Revista de la Facultad de Ingenieria* **32**(16), 330–336 (2017)
17. Zhong, J.: Research on the data mining technology in college students' attendance system based on the big data architecture (2020)
18. Liu, J., Shi, G., Zhou, J., et al.: Prediction of college students' psychological crisis based on data mining. *Mob. Inf. Syst.* **2021**(23), 1–7 (2021)
19. Xie, Y.C., Liu, W.L., Yao, Y., et al.: Design and implementation of college students' comprehensive evaluation management system based on B/S mode. *Comput. Inf. Technol.* (2019)
20. Sheng, X., University, H.: Research of the innovation of college students' funding management model based on deep learning. *J. Chongqing Univ. Technol. (Natural Science)* (2019)

or in person, so as to give solutions or responses as soon as possible, and then feed back to the Internet; Even if the students may not have reasonable opinions, they should also give timely feedback, correct unnecessary students' misconceptions and misunderstandings, and create a harmonious learning and living environment for students [20]. On the one hand, it enhances the students' sense of participation, provides a channel for the student work system and students to communicate with each other, enhances the transparency of the work, and makes the work better accept the suggestions and supervision of teachers and students; On the other hand, by connecting the current news trends at home and abroad, the current main work of the school and other information, we can effectively reflect the requirements of the "Three Represents", occupy the ideological work position of students on the Internet, and make the campus network become another important form of students' ideological and political work.

5 Conclusion

This paper introduces the design and implementation of university management information system in detail. The system designed has practicability: it almost includes all aspects of traditional management information and modern information resource development in colleges and universities, and the system design fully considers practicability. At the same time, due to the modularization of the system, the future system expansion is superimposition rather than replacement. Therefore, with the change, adjustment and expansion of the functions of the management department, any part of the value-added platform can be expanded relatively independently. Progressiveness. The functional design incorporates the latest concepts and research results of higher education research, modern resource management and development, information management systems and other multidisciplinary fields. The system design adopts advanced and mature computer technology, thus ensuring the progressiveness nature of the system and having a longer service life.

References

1. Yin, X.H.: Construction of student information management system based on data mining and clustering algorithm. *Complexity* **2021**(2), 1–11 (2021)
2. Yang, H., Zhang, W.: Data mining in college student education management information system. *Int. J. Embedded Syst.* **15**(3), 279–287 (2022)
3. Liu, X.: The development of sunshine sports information management system based on data mining technology. *Revista de la Facultad de Ingenieria* **32**(13), 471–476 (2017)
4. Yang, L., et al.: The Application of data mining techniques in college students information system (2018)
5. Ren, G., Wang, H.: Design and implementation of university management information system based on data mining. *Comput. Meas. Control* (2016)
6. Pan, F., Huang, F.: Development of library information management system based on data mining technology. *Autom. Instrum.* (2019)
7. Zu, E., Shu, M.H., Huang, J.C., et al.: Management problems of modern logistics information system based on data mining. *Hindawi Limited* (2021)

As the student work in colleges and universities is a systematic project with complicated process and content, this paper establishes ten subsystems according to the main content of student work and the actual situation of the student work in our school. According to the basic situation of students, the relevant field names are set according to the requirements and characteristics of each system; At the same time, the student number and name are run through each subsystem, and the field name of “student number” is used to associate different subsystems, so that relevant data can be easily queried when a record is queried [18]. For example, when querying the information of a graduate, you can click the student number to find out the rewards and punishments, military training, etc.

(2) Function analysis of student information management module

The database of student information management module adopts C/S architecture. The backstage is supported by department datasets and global data warehouses. Users use OLAP tools or data mining tools on the front end for data analysis. Its architecture is shown in Fig. 3.

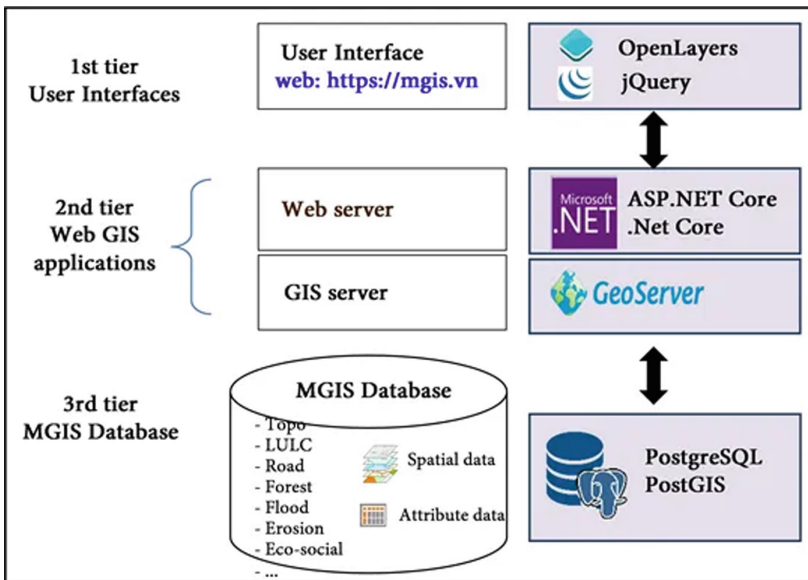


Fig. 3. Database architecture diagram

Because the network provides a direct and fast information feedback environment, and this system provides powerful retrieval and query functions, it can timely collect students’ current hot spots and focus issues, understand students’ requirements and suggestions, and systematically summarize them [19]. For the hot spots and focus issues collected and sorted out, the good side is to feed them back to the systematic forum for publicity and praise; For students’ reasonable opinions on the work of the school, they will be sent to the relevant departments through the network (e-mail), telephone, written

In data mining, an important task is to find frequent itemsets in large databases. When a large number of patterns exist in databases, finding frequent itemsets becomes a costly task. The algorithm is an association rule mining method that does not generate candidate item sets. As an algorithm for mining frequent item patterns, a given database needs to be scanned many times, so I/O reading and writing has become a big bottleneck. It is to solve this problem that the data structure of a tree is introduced by using the divide-and-conquer strategy. No matter how much data is, only two data sets need to be scanned, thus improving the efficiency of the algorithm.

4 Function Realization of Data Mining Information Management System

In the process of data mining and data information control, the teaching management data analysis system needs to be analyzed from the perspective of data information processing, information dimension, granularity, etc. In the process of establishing information indicators, it needs to be analyzed from the perspective of data analysis, information processing, etc., so as to improve the information processing effect of the teaching management data analysis system [15]. The specific indicators are as follows:

Indicator: It is mainly to adopt and analyze different subjects and different information in the teaching management data analysis system, so as to achieve comprehensive improvement of data evaluation and data information analysis effect.

Dimension: It can be processed with different uploaded evaluation information of different users in the teaching management data analysis system. On the premise of analyzing the data information transmission path and data information processing, the comprehensive improvement of data information processing effect can be achieved.

Granularity: refers to the detailed procedure for describing various teaching management-related information in each indicator dimension of the teaching management data analysis system.

In the process of processing and channel construction of mining data information in the teaching management data analysis system, we need to improve it from the perspectives of data information processing, logical model building and data mining [16]. Moreover, after data mining, we need to control the data transmission channel, and combine the conceptual model application of data mining and the perspective of data analysis to analyze the node relationship of data information. From the perspective of information channel design, on the basis of realizing data management and information channel construction, take the snowflake model as the basis of data constraint relationship, and analyze data information transmission and information dimension structure according to data logic relationship, so as to improve the data mining function of the teaching management data analysis system. In the process of designing the information channel, we need to control it from the perspective of information dimension, so as to achieve the comprehensive improvement of data dimension and information transmission control effect [17].

Take the student information management module as an example to introduce its contents and functions.

(1) Content of student information management module

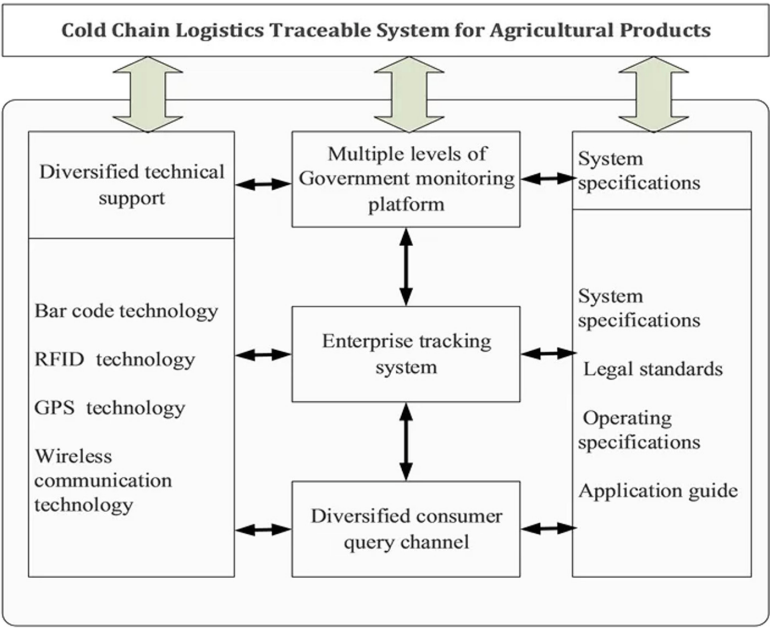


Fig. 2. College student management information system

system should take management needs as the core, and take the relatively perfect and prominent management functions as the main goal [14]. Therefore, the main functional modules of the system should include student information management, teacher information management, teaching information management, book information management, teaching material information management, employment management, training management, system management, etc., which can realize the comprehensive, scientific and standardized management of university resources. Table 1 below shows the management information values of college students.

Table 1. Management information of college students

Scale type	Academic and scientific activities	Construction of campus culture	Social practice	Construction of student work team
Social scales	0.68	2.41	1.11	0.61
Information scale	0.79	1.38	0.66	0.85
Personality scale	1.25	6.2	1.21	0.56

It can be seen from this that when there are different requirements for data clustering, one of them is data clustering.

high similarity, and the objects between different clusters have low similarity. In the process of analyzing the practical application of clustering method, the core of network data processing and control is to conduct quantitative analysis of data information change, data collection and other aspects from the perspective of clustering analysis, and then realize the processing and application of data information. At present, combined with the clustering analysis of the database and based on the database technology, the data information is analyzed to improve the data information processing and control effect.

2.3 Design Objectives of Data Mining Information Management System

The university information management system stores a wealth of teaching, scientific research, management and other aspects of information, and the scope is very broad. It should be said that data mining on this basis will have many directions to choose from. However, at the same time, the university information management system is also a very complex large system. If the problem is too big or too vague, it will have a negative impact on the practical significance of mining. Therefore, for this topic, finding a clear goal, that is, the key link to be solved is the primary task [12].

There are many mining algorithms available in the university information management system. This paper mainly uses association rules to mine the existing data according to the characteristics of the data in each module of the system. This paper refers to data mining theory and other mining tools to enable the system to achieve the following functions:

- 1) Ease of operation. That is, the simplicity of the mining process and the readability of the mining structure.
- 2) Efficient data processing capability, simple and effective data processing.
- 3) Provides a simple path to integrate with other tools.

3 Design of University Student Management Information System Based on Data Mining Technology

The research of university management information system mainly includes two levels: management level and technology level. On the management level, the system should meet the application requirements of university resource management; On the technical level, the system shall adopt advanced and applicable software development mode and software development process. Corresponding to these two levels, this paper discusses the university resource management system from the perspectives of system function, development mode and system implementation [13]. The structure of college student management information system is shown in Fig. 2.

The level of information collection and processing is directly related to the overall management level of the unit to a certain extent. Due to the complexity of information management in colleges and universities, it is required to have a reasonable division of labor, clear responsibilities and sound management; The constant expansion of information requires that information management must be accurate and efficient; The openness of management requires multiple staff to contact on the open network with different permissions, interact in real time, and complete their own tasks or common purposes. This

The purpose of developing this system is to apply modern information processing technology to school work management, improve work efficiency, and liberate managers from complicated manual operations. Furthermore, in the process of building a high-level university, the construction of information management system is also one of the important links [7]. Therefore, the development of university information management system is an important part of university information construction. At present, there are the following difficulties in collecting and analyzing various existing information in colleges and universities:

- (1) Decentralized storage location: all kinds of information in colleges and universities are stored in various departments of colleges and universities, and data consistency is difficult to ensure.
- (2) Poor information integrity: departments work independently and lack unified management and coordination on information collection and update, which makes it difficult to ensure the integrity of information.
- (3) Poor information accuracy: the information collection point is not unique, the departments lack cooperation, and the data accuracy is difficult to ensure.

The university information management system is a powerful assistant to assist the school leaders and staff in various departments to carry out various work. It is a university management information system that takes the basic information of the school's software and hardware resources as the main body and takes high timeliness and high-quality service as the criterion [8]. It not only conforms to the actual situation of the school, but also has advanced management ideas. This system has noticed the openness and practicability of the system, which is convenient for operation.

2.2 Data Mining Methods

In the process of research and analysis of data mining analysis methods, the commonly used methods are mainly classification, regression analysis, clustering, association rules, deviation analysis, web page mining, etc. [9]. In the process of practical application, data information needs to be mined from different perspectives.

Data classification and mining control need to be controlled from the perspective of data mapping, data prediction and analysis. In the process of extracting and managing important data information, important data can be obtained by mining and extracting multi-data information. In the process of data classification, CLS algorithm, C45 algorithm and other means can be used to control the important data information [10]. On this premise, control the classified data to improve the effectiveness of data mining analysis.

Formally, if there are m unmarked samples in the sample set $D = \{x_\gamma, x_\gamma, x_\alpha\}$, and each sample $x_i = (x_{i1}; x_{i2}; x_{in})$ is an n -dimensional feature vector, then sample D is divided into k disjoint clusters $\{C_l \mid l = 1, 2, \dots, k\}$, $C_l, C_i = \emptyset$ and $D = \prod_{l=1}^k C_l$, by the clustering algorithm,. Correspondingly, $\lambda \in \{1, 2, \dots, k\}$ to represent the cluster mark of x , that is, $x_j \in C_{\lambda_j}$. In this way, the result of clustering can be expressed as $\lambda = (\lambda_1; \lambda_2; \dots; \lambda_m)$. According to different needs,

In the process of analyzing the application of clustering algorithm, based on the principle of maximizing intra-class similarity and minimizing inter-class similarity, data mining and control are realized through clustering [11]. The objects within a cluster have

Assumption 2: the program selection function q_i is and the weight coefficient is $F(x_i)$, then the choice of information management intervention options for college students is shown in Eq. (2).

$$F(x_i) = z_i \cdot \sqrt{f(x_i|y_i)} \sqrt{\prod q_i \cdot \xi} + \quad (2)$$

In terms of the external environment of university student management, we should pay attention to providing students with good external environment and high-quality logistics services. Colleges and universities attach importance to the combination of professional training of student administrators and guiding services, and pay attention to the treatment of students in a democratic way. Colleges and universities can better serve students by setting up a variety of functional departments. Colleges and universities pay attention to the growth of students in a relaxed and harmonious atmosphere, so as to fully tap their own potential, form good habits, and let students realize that their rights are respected in the process of management, So as to consciously abide by the school system, which is conducive to achieving the training objectives of colleges and universities; Colleges and universities pay attention to providing services for students through a perfect socialization mechanism, and advocate a simple management concept. Social institutions are responsible for the management of students in colleges and universities [6]. As shown in Fig. 1 below, there are bright clusters between blue and red in the student information management set diagram, and each cluster contains 5000 data. If there are outliers in the data, the blue cluster may contain more outliers by visual inspection, because the data points are relatively scattered (Fig. 1).

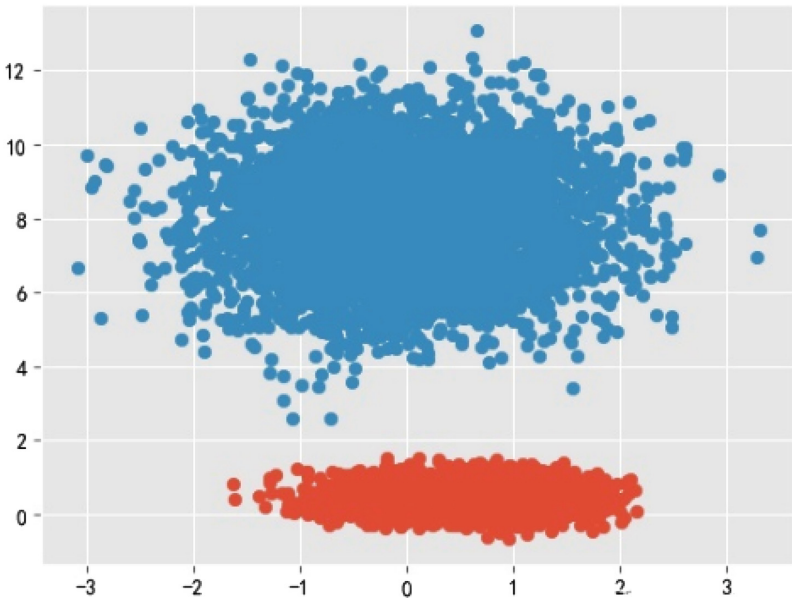


Fig. 1. College student management information system (Color figure online)

attributes, nor can they obtain important knowledge. This also leads to the “survival in name only” of the management information system [3].

A set of perfect university management information system developed by using data mining technology is conducive to introducing modern management theories and methods into the actual management of the school, promoting the transformation of school management from empirical management to scientific management, making full use of the human, material, financial and other resources of the school, and more effectively realizing the talent training, scientific research [4]. The function of serving the society and its short-term and long-term development goals.

2 Related Work

2.1 Purpose of Developing Data Mining University Management Information System

At present, the management level of some school information in China is still on the basis of paper media. Such a mechanism can no longer adapt to the development of the times, because it wastes a lot of manpower and material resources. In the information age, this traditional management method will inevitably be replaced by computer-based information management.

The information management system is an indispensable part of an educational unit, and its content is crucial for the decision-makers and administrators of the school. Therefore, the information management system should be able to provide users with sufficient information and quick query means. In recent years, with the continuous expansion of enrollment in colleges and universities, the number of students in school has risen sharply, which has greatly increased the workload of management staff engaged in various aspects of the school [5]. The traditional manual management mode can no longer meet the needs of the current work. This management mode has many shortcomings, such as low efficiency and poor confidentiality. In addition, a large number of files and data will be generated over a long period of time, which brings many difficulties to search, update and maintain.

College students' information management is to use the information scale to comprehensively evaluate management problems, and find management problems in the management evaluation according to multi-dimensional indicators, integrate the evaluation results, and finally determine the causes of management problems. Combining the information scale and using information mining and intelligent algorithms to comprehensively evaluate the results of college students' information management can improve the evaluation level of college students' information management.

Hypothesis 1: The content of the scale is, the result set is y_i , the information standard is $set \sum x_i$, and the judgment function of college students' information management is $f(x_i)$ shown in Eq. (1).

$$f(x_i) = \sum \underline{x_i | y_i} \cdot \int \xi \quad (1)$$

Data mining technology has been applied in many fields, such as commerce, finance, insurance and telecommunications, and has achieved satisfactory results. In the field of university management, with the deepening of the reform of the education system, the competition faced by the university is also increasingly fierce. This requires the school administrators to constantly update the education concept and talent training mode, and constantly make new scientific decisions according to the development of the school. Applying data mining technology to the college enrollment management information system can, on the one hand, coordinate the work of all departments well, facilitate the timely transmission and processing of information, and realize the sharing of information resources among all departments, thus simplifying the registration procedures for new students, saving the school's manpower and material resources, and improving the efficiency of enrollment work. On the other hand, by analyzing the enrollment data over the years and mining effective information, It also helps to provide decision-making reference for candidates' voluntary filling, college enrollment publicity, specialty setting and other aspects, scientifically and effectively assist college administrators to make decisions, and play an important role for colleges and universities to take the initiative in the fierce competition in the future.

With the development of modern management theory and decision-making science and their application in the management of colleges and universities, China's colleges and universities will further change from the past experience based management to scientific management based on modern management theory and methods. In the past, the school education management only emphasized the particularity of the education field, ignored the commonness of education management and general management to a certain extent, overemphasized the particularity of school education, and emphasized the management mode based on experience. For example, colleges and universities generally have not established management information systems (MIS), which is an important modern management tool. In sharp contrast, many colleges and universities themselves are developing various management information systems for governments and enterprises. Many colleges and universities do not realize the importance of management information system in college management. Even though some colleges and universities have designed and used corresponding college management information systems, most of these systems are connected to large databases, in which a large amount of original information is stored, including the staff information of all teachers and students and a large amount of data in the teaching process. These data are seemingly unrelated [2]. In the face of the "mountains" of data sets, traditional data analysis methods are still difficult to cope with, whether in the sense of time or space. Administrators are unable to understand and effectively use these data, resulting in an increasingly serious "data disaster", which forces school leaders to adopt two countermeasures that may lead to "decision-making disaster": one is "poor in coping", the other is "ignoring", In fact, no matter what kind of countermeasure is taken, it is out of a "helpless". Traditional data analysis methods (such as statistics) can only obtain the surface information of these data, but can not obtain the internal relationship and implicit information of data



Design of College Students Management Information System Based on Data Mining Technology

Yan Zhao^(✉) and Jian Li

Zaozhuang Vocational College of Science and Technology, No. 888, Xueyuan Road, Longquan Street, Tengzhou City, Zaozhuang, Shandong Province, China
celia1888@126.com

Abstract. The design of student information management system is a process of defining the overall architecture of information infrastructure. It describes how an organization will use and manage its information resources (e.g., data, applications, databases) to support the decision-making process, improve service to users, and reduce costs to achieve specific goals. Essentially, it is a blueprint that defines what “infrastructure” will be built at the beginning of the project. A good example of this is Microsoft Office or Adobe Acrobat Reader, which are designed by one person or team, but millions of people use them every day since they came out. The data mining technology is applied to the teaching information management module to analyze students’ test scores, and determine the correlation between students’ usual performance and final exam scores. The paper elaborates the design scheme, functional characteristics and key technologies of the system, which provides a set of effective methods for scientific and standardized management of colleges and universities.

Keywords: Data mining technology · college student · management information systems

1 Introduction

Colleges and universities have always attached great importance to the first link of various management work for students at each level, as well as the important link of determining whether the school’s teaching, teaching aids and logistics management mechanism can operate normally. Before the registration of new students, the school leaders should convene a coordination meeting of enrollment work involving all teaching departments and departments of academic affairs, enrollment, student management, logistics and other departments to make detailed deployment and arrangement. The check-in method of each school is also basically the same: relevant departments set up check-in points in the campus, and freshmen go to each department to handle relevant procedures according to the procedures [1].

6. Hua, S.: Research on the construction of higher vocational english ubiquitous learning resource platform based on big data. *Modern Educ. Sci.* (7), 75–81 (2016)
7. Jian, S.: Research on the categories and characteristics of higher vocational english learning needs based on data mining technology. **2019**(8), 52–58 (2021)
8. Hua, G., Meili, L.: Research on e-commerce english teaching strategies in higher vocational education under the background of big data. *Friends Humanities* (017), 110–110 (2018)
9. Han, W., Haiyun, H.: Application of promoting data mining technology in English A-level score prediction of higher vocational students—Taking Sanya aviation tourism vocational college as an example. *Liberal Arts Enthusiasts (Education and Teaching)* (05), 39–40 (2019)
10. Fen, S.: Construction of multiple monitoring model for Higher vocational english autonomous learning based on network environment. *New Campus (Early Ten Days)* **000**(5), 155 (2018)
11. Rongzhen, S.: Construction of multiple monitoring model for Higher vocational english autonomous learning based on network environment. *Changjiang Ser.* (027), 103 (2016)
12. Jinhong, Z.: Research on the construction of english autonomous learning model for higher vocational students in the “micro era”. *J. Lanzhou Inst. Educ.* **032**(008), 133–134137 (2016)
13. Yang, F.: Analysis of students’ autonomous learning achievements based on educational data mining—taking liaoning ecological engineering vocational college as an example. *Inf. Comput.* (016), 219–220225 (2019)
14. Luxi, C.: Research on autonomous learning of English viewing, listening and speaking in higher vocational colleges based on smart phones. **2017**(4), 83–84 (2021)
15. Jian, S.: Research on the categories and characteristics of higher vocational english learning needs based on data mining technology. *Mech. Vocat. Educ.* (008), 52–58 (2019)

5.4 Communication Learning

The learning exchange space for students includes communication tools, online learning forums, and Q & A areas for teachers and experts. Communication tools include synchronous communication tools and asynchronous communication tools. There are many kinds of synchronous communication tools, including video conference system, chat software, various chat rooms, etc. Asynchronous communication tools include e-mail, electronic bulletin board, etc. Online learning forum can provide students with open classified discussion area services. Students can express their views on a problem, ask others for help, ask some questions or exchange learning experiences with each other. In addition, it also provides entertainment, emotional communication and other sections for students. In students' autonomous learning, some unanswerable questions will inevitably arise. At this time, students can enter the question answering area of experts and teachers for help. Experts and teachers help students complete their learning by providing students with ideas to think and solve problems. Students' learning exchange platform provides a convenient and effective platform for students' communication. At the same time, it also provides conditions for students to form a learning community. Through the communication and communication on the communication space platform, students can find other learners with the same goals and similar backgrounds and form a learning community. With the participation of teachers and students and the multi-level communication between online and offline, learners can not only get the guidance and help of teachers on the Internet, but also argue with other students.

6 Conclusions

This paper puts forward an open platform to help summarize and make use of the existing network and information technology, trying to provide the most open service for all people who are interested in sharing learning information resources. The purpose is to provide an interactive platform for learners and researchers, improve the efficiency of obtaining learning information, realize knowledge discovery and reduce the regional differences in the distribution of educational resources.

References

1. Haiyun, H., Han, W., Chao, L.: Analysis of the relationship between higher vocational english A-level examination and employability based on data mining. *Higher Vocat. Educ. J. Tianjin Vocat. Univ.* **028**(004), 86–90 (2019)
2. Xiaohua, C.: Construction of multiple monitoring model of higher vocational english autonomous learning based on network environment. **2016**(1), 119–120 (2019)
3. Huanxiang, H.: Construction of vocational english autonomous learning micro platform based on vocational ability in the information age. *Chinese Foreign Entrepreneurs* (029), 192 (2016)
4. Jie, Z.: Research on the construction and practice of higher vocational english autonomous learning model based on network platform. *J. Hubei Correspondence Univ.* **29**(9), 3 (2016)
5. Weiguo, Z.: Application and practice of higher vocational teacher management system based on data mining. *Comput. Age* (307)(01), 86–89 (2018)

and practice, and highlight the key points and difficulties. For Higher Vocational English Autonomous Learning in the multi-modal environment of information system, teachers' supervision function is far weaker than ordinary classroom teaching. Therefore, in order to give full play to the advantages of autonomous learning, an important content is to stimulate students' motivation and curiosity. This content is one of the keys to the success of autonomous learning in the multimodal environment of information system.

5.2 Teacher's Key Explanation

Under the correct guidance of teachers, students have mastered most of the knowledge content through autonomous learning. For individual knowledge content, students can not fully digest and absorb it. At this time, teachers should focus on explaining this part of the content through the network platform. Teachers are not an independent element of the network learning environment, but they are indispensable participants in the process of students' Autonomous Learning under the network environment. They are a member of the learning community in the learning environment. Teachers play a constructive role in the Network-based Autonomous Learning environment. The Network-based Autonomous Learning environment needs teachers to build. The early construction of the Network-based Autonomous learning platform includes the supplement of various resources such as network courses, the construction of learning tools, and the provision of relevant learning situations, learning strategies and supports [15]. The Network-based Autonomous Learning environment requires teachers to give full play to their guiding role. The network is mainly characterized by a large amount of information. In autonomous learning, students will be interested in swimming in the rich information, and the situation of being lost in the network also occurs from time to time. Teachers' timely guidance can help students get rid of confusion and return to the learning process in time.

5.3 Interactive Learning

Learning resources, learning strategies, learning tools, learning support and communication space constitute the intelligent learning system of College Students' autonomous learning environment based on network. Intelligent learning system provides students with rich learning resources, diverse learning strategies and tools, flexible learning support and communication space. Students choose according to their own needs and get strong support from the system. The interaction between students and intelligent learning system is mutual. The system provides a number of support for students' learning, such as learning information resources. Students can choose and use according to their own needs. At the same time, students can also enrich various elements in the learning environment according to their personal experience and needs, and carry out new integration to promote their own learning. Due to the development of the network, the original data transmission and message segmentation methods have been changed, so that various media can spread in the network, text interaction becomes multimedia interaction, and single interaction becomes two-way interaction, so as to realize high-level interactive learning.

(1) Content delivery module. (2) Student behavior feature library. Store student input and learning behavior data marked with time stamps in the learning system. The database includes two parts of data. One is the real-time related data directly obtained through online learning system and educational administration system. The other part is the data through instructor feedback, personal filling in and questionnaire survey. (3) Preprocessing module. Be responsible for cleaning the massive data of students, removing the data with excessive deviation according to the filtering conditions, and improving the accuracy of the system processing results. The running results in the recommendation module are displayed to the teacher in a visual way. (4) Recommended modules. According to the data provided by the preprocessing module, the appropriate learning content is recommended to learners based on the collaborative filtering algorithm. (5) Diagnostic feedback module. Teachers can make human intervention according to the operation results of the recommendation module to correct the recommendation results. (6) Student information system. It stores students' basic personal information data.

From Fig. 4, we can see the operation process of the whole system. Preprocess and extract the students' dynamic original data in the learning information system and the students' behavior feature database, and submit it to the recommendation module after cleaning. The recommendation module uses the preset recommendation algorithm to analyze the preliminarily recommended learning content, and submit it to the content delivery module and visualization module at the same time. Through the analysis of the content, teachers decide whether to use the diagnostic feedback module to intervene the data [13]. The content delivery module delivers the learning content after the intervention to the students, receives the feedback data from the students, and dynamically submits it to the student database to form an adaptive intelligent learning assistant system.

5 Overall Design of English Autonomous Learning Information System Based on Data Mining

5.1 Multimodal Environment Model of English Autonomous Learning in Higher Vocational Colleges Based on Big Data

Combining the basic characteristics and learning characteristics of higher vocational students with the elements of learning environment, and establishing a multimodal environment model suitable for higher vocational students to carry out English Autonomous Learning Based on information system, the middle school students' information database in the autonomous learning environment based on information system is to meet the students' timely and objective understanding of their own situation in the process of autonomous learning. According to their own learning situation to control the learning progress, formulate the future learning plan and grasp the personal learning progress and quality, and establish a database containing students' basic information, learning history and learning evaluation information. Higher vocational English Autonomous Learning in the multimodal environment of information system is not completely separated from teachers [14]. In this teaching process, students' learning objectives and standards should be formulated by teachers. The learning objectives and tasks shall meet the requirements of the curriculum outline, fully reflect the combination of theory

U represents the set of all students and C represents the set of all items. The utility function s is defined to calculate the recommendation degree $s: U \times I \rightarrow R$, the research problem of recommendation algorithm is to find the most interesting item $I' \in I$ for each student $u \in U$ by calculating the recommendation degree.

In the recommendation system, both students and teachers can be represented by a set of different attributes or features. In the process of application, the recommended algorithm must extrapolate s to the whole $U \times C$ in the space of C . According to different extrapolation methods, many recommendation algorithms are derived, such as collaborative filtering, content-based recommendation, association rule recommendation and so on. As a traditional recommendation method, collaborative filtering is one of the more advanced algorithms. As shown in Fig. 3.

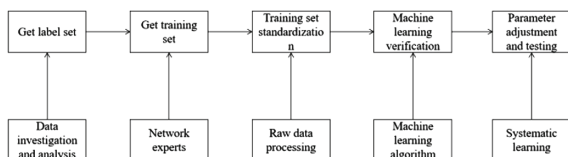


Fig. 3. Autonomous learning label standardization system

Collaborative filtering algorithm is based on the attributes of teachers, calculates the information attribute distance between different teachers, and makes relevant recommendation by finding the nearest teacher. At present, there are two main types of collaborative filtering algorithms: memory based collaborative filtering algorithm and model-based collaborative filtering algorithm. Memory based collaborative filtering uses the whole student teacher score data set for calculation, and each student is an integral part of the score prediction process. Memory based collaborative filtering selects some neighbor students with similar interests for the target students, and predicts the score of the target students to the teachers according to the score of the neighbor students.

The system includes; Six modules: as shown in Fig. 4.

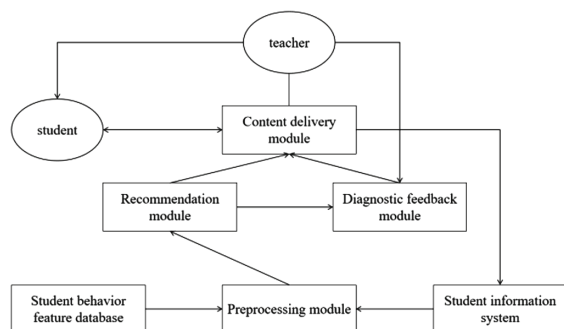


Fig. 4. Composition of adaptive learning system based on big data

the quality of service, and it is difficult to design the algorithm. Using machine learning method for optimization can improve the accuracy of analysis and reduce the process of algorithm design.

3 Overall Framework Design of English Autonomous Learning Information System

As the core function module of the system, whether the data analysis module is designed reasonably directly affects the operation performance of the whole system. Therefore, in order to ensure the reliable, stable and safe operation of the system, software developers should pay attention to the design and implementation of the module [12]. This module mainly refers to that users should complete the analysis and processing of the original data from three aspects: data audit, data cleaning and data integration according to the actual situation of data calculation before calculating the data, so as to improve the accuracy of data calculation and ensure that the data form can meet the requirements of the algorithm, so as to reduce the complexity and complexity of data calculation, It lays a solid foundation for further improving the efficiency and effect of data calculation. The data preprocessing method is shown in Table 1.

Table 1. Data preprocessing methods

Existing problems	Pretreatment method
Inconsistent data	Data audit
Invalid, missing, duplicate data	Data cleaning
Data format/Size mismatch	Data transformation
Data processing/Consolidation processing	Data Integration
Data missing label information	Data Annotations
Unordered data	Data sorting

By completing this step, the scientific design and implementation of the data analysis module can be realized, which creates good conditions for further improving the efficiency and effect of data analysis.

4 Recommendation Algorithm Based on Data Mining

In the teaching system, the most important intervention for students is to make students understand and familiar with the course or teacher through recommendation, or make the system have the characteristics of some social networks through similar student introduction, so as to improve students' satisfaction. Its basic working method is recommendation. For any recommendation system, the recommendation algorithm is shown in formula (1).

$$\forall u \in U, i_u = \arg \max_{i \in I} s(u, i) \tag{1}$$

e-learning system based on data mining technology is to design personalized learning schemes of different levels and categories for the learning of the same subject knowledge.

Traditional teaching is carried out by entering the school. Its characteristic is that it not only aims at knowledge learning in class, but also attaches importance to moral, intellectual and physical education [10]. Due to the mandatory curriculum, it can ensure the learning quality of students and cultivate students' interest in autonomous learning. As shown in Fig. 1.

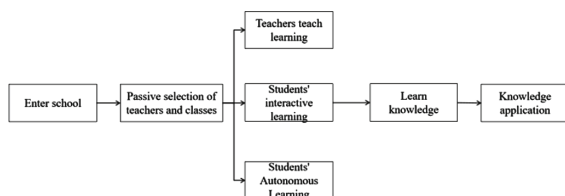


Fig. 1. School learning process

Figure 1 school learning process is different from traditional education. Autonomous learning information system is an educational method in which teachers upload courses in the form of digitization for students to learn and explore independently under the condition of relatively popular network. Most of the autonomous learning information systems do not provide direct training direction, but an educational method dominated by students' autonomous interest. When students learn about a subject through other channels and feel interesting, but traditional teaching can not meet students' desire for autonomous learning, the advantages of autonomous learning information system appear. Autonomous learning information system has autonomy, instability and strong directionality. Therefore, how to make the autonomous learning information system more interesting, in line with and can strengthen students' interest in learning is the main problem, as shown in Fig. 2.

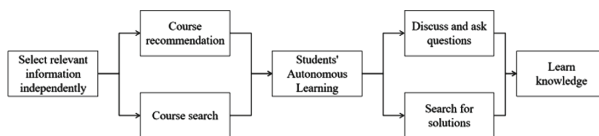


Fig. 2. Self regulated learning process of systematic education

For autonomous learning information system, curriculum providers (teachers) and curriculum learners (students) themselves cannot intervene. In order to make it easier for students to learn the knowledge they want to learn, the system should have a full understanding of the three objects of teachers, students and courses, and then carry out services that meet the needs of users [11]. The traditional algorithm extracts specific data according to research and other methods, and uses a fixed algorithm to analyze users and projects. However, this method may not be able to reasonably process some data under specific circumstances, resulting in excessive proportion of some data, affecting

On this basis, using educational data mining and English learning analysis to build relevant models in the field of education and provide effective decision support for education and teaching will become the trend of development in the future. Especially for higher vocational colleges, this application is particularly important [2]. Higher vocational colleges belong to the junior college level. Compared with undergraduate students, there are common phenomena that students' English learning motivation is insufficient, their English learning enthusiasm is not high and their English learning efficiency is not high [3]. How to effectively guide and intervene higher vocational students' English learning, improve students' English learning enthusiasm, clarify the objectives, methods and plans of English learning, and further provide all-round services for higher vocational education decision-making, teachers' teaching intervention and higher vocational students' Personalized English learning [4]. Through the storage and integration of massive resources of higher vocational education, an adaptive English learning system for Higher Vocational College Students Based on big data is established. The system compatible considers the unique semantic relevance of students' English learning behavior in the field of higher vocational education. Through the analysis of students' cleaning data, the recommendation module based on collaborative filtering algorithm is introduced to realize the decision-making of analyzed data on Higher Vocational Education [5].

2 Introduction of Data Mining Technology

2.1 Concept of Data Mining Technology

Data mining technology is actually a process of extracting and mining data with surface difficult to find but useful information and knowledge from massive, incomplete, unclear, random and noisy databases [6]. At present, most of the use of data mining technology in China is in the commercial field. Members of this research group hope to use this technology to accurately reflect the key factors affecting teaching quality, so as to further provide a theoretical basis for the reform of relevant links and build a scientific and multi-dimensional teaching quality management system in higher vocational colleges. Data mining algorithms contain a large number of models and algorithms. These algorithms do not have any conflict problems in the process of use, and can solve various problems. Therefore, it is necessary for software developers to strengthen their understanding and understanding of different types of algorithms [7].

2.2 Design Concept of Autonomous Learning Information System Based on Data Mining

This paper puts forward an open, dynamic and developing design concept. Because the current information technology and educational technology and concept in China are not mature enough, the learning system has great deficiencies in the analysis and treatment of individual differences of learners and bad alienation in each learning stage [8]. Therefore. The network learning system based on data mining should establish a correct idea at the beginning of design, that is to use data mining technology to realize the analysis and processing of individual differentiation [9]. The key to the design of



Construction of Higher Vocational English Autonomous Learning Information System Based on Data Mining

Jian Xi^(✉) and Zhu Hao

Zibo Normal College, Zibo 255130, Shandong, China

xijian@sohu.com

Abstract. With the continuous in-depth development of school digitization, informatization and smart campus construction, the application of campus information system has been integrated into all aspects of students' learning and life. There are a large number of student behavior data in each information system. Analyzing and studying these behavior data is the focus of current research. The data warehouse technology is used to integrate the student behavior data existing in various information systems into the data warehouse, and establish the behavior analysis model of higher vocational students through data mining analysis, which can provide help and suggestions for students to modify learning methods, teachers to adjust teaching methods, and management departments to accurately manage students. This paper studies the adaptive learning system of Higher Vocational Colleges Based on big data analysis, puts forward the semantic association model of students' behavior, studies the learning recommendation algorithm based on data mining, and finally constructs the architecture of adaptive learning system. The research can effectively guide and intervene higher vocational students' English learning.

Keywords: Big data · information mining · autonomous learning · education system

1 Introduction

The comments and articles of College Students' autonomous learning of English are studied. Most of the comments on the construction of autonomous learning mode and the analysis of autonomous learning effect, but few of them explain the reasons why the implementation efficiency of autonomous learning is not improved [1]. Based on this situation, this paper expounds the reasons for the low efficiency of English Autonomous Learning in higher vocational colleges, that is, the students' learning process is not subject to too high constraints. Therefore, we should establish an appropriate learning monitoring system to monitor and restrict the students' English autonomous learning process in Higher Vocational Colleges.

5. Yi-Liang, X.U.: On the application of ideological counsel in emotional education on college students. *J. Shanxi Youth Vocat. College* (2017)
6. Shi, X.: Emotional data mining and machine learning in college students' psychological cognitive education. (4) (2019)
7. Li, Y., Zhang, Y.: The innovation research of college students' academic early-warning mechanism under the background of big data. In: *International Conference on Management Science and Engineering Management*. Springer, Cham (2018)
8. Ranellucci, J., Poitras, E., Lajoie, S., et al.: Understanding emotional expressions in social media through data mining. *Emotions Technol. Soc. Media* 85–103 (2016)
9. Zheng, C., Zhou, W.: Research on information construction and management of education management based on data mining. *J. Phys.: Conf. Ser.* **1881**(4), 042073 (6pp) (2021)
10. Wang, J., Li, L.: Research on the college graduate employment education based on data mining technology. *Anthropologist* **23**(1–2), 231–235 (2016)
11. Hui, F.: Research on the construction of emergency network public opinion emotional dictionary based on emotional feature extraction algorithm (2022)
12. Yin, Y.: Research on the cultivation of talents quality under the condition of innovation education based on data mining. *J. Comput. Theor. Nanosci.* (2016)
13. Fang, L.: The prevention and analysis of college students' psychological crisis based on data mining technology (2021)
14. Zhuang, M.J.: Research on the innovation of college students' psychological health education from the perspective of big data. *J. Heilongjiang Vocat. Inst. Ecol. Eng.* (2019)
15. Wu, Y., Ji, Q., Zhao, A., et al.: The Construction of Mental Health Prediction Model Based on Data Mining Technology. Springer, Singapore (2022)
16. Gkintoni, E., Halkiopoulou, C., Maria Georgiadi, et al.: Association of life orientation index and parameters of emotional intelligence in university students: a data mining approach. In: *27th European conference of psychiatry*. Warsaw, Poland, 6–9 April approved (2019)
17. Xi-Jun, H.E., Zhu, X.Y.: An empirical study on deep mining of college students' teaching evaluation data. *Heilongjiang Res. Higher Educ.* (2019)
18. Fang, G., Lin, W., Dong, L., et al.: The structure mining of college students' emotional semantic network based on micro-blog community. *Chinese J. Ergon.* (2017)
19. Hou, Y.: Students' emotional analysis on ideological and political teaching classes based on artificial intelligence and data mining. *J. Intell. Fuzzy Syst.: Appl. Eng. Technol.* **40**(2) (2021)
20. Zhang, S.: Application of data mining technology in the analysis of e-commerce emotional law. *J. Phys. Conf. Ser.* **1852**(2), 022044 (2021)

a relaxed attitude. On the basis of full conversation, conduct comprehensive analysis and research, so as to achieve targeted, practical guidance and help students. Student education managers must be mindful of students, sincerely make unremitting efforts for their interests, be able to understand students' difficulties, and do not seek complete blame and boast about them. Fault finding. We should be able to make good use of the two different educational means of criticism and praise to enable students to grow and improve. Students who occasionally make small mistakes can be handled with a generous attitude on the premise of not violating the seriousness of rules and regulations, so that they can understand their mistakes through reflection and avoid repeating them in the future. Once they find the bright spots on students, they should be affirmed and praised [20]. And encourage them to perform better in their future work. When students encounter difficulties in life or emotional fluctuations, they should be able to find them in time and help them tide over the difficulties. In particular, they should be given strong enough spiritual support and encouragement, and continue to give them correct guidance, so that they can find the right way to solve problems, so as to improve students' adaptability. The ability to study, live and work in school.

5 Conclusion

To sum up, emotional education is a contemporary interdisciplinary education mode, and an important means for college education managers to give full play to their personality charm and creatively use the art of management in their work. Emotional education can be said to be an important supplement to college system management education and an important guarantee for the smooth development of system management. College education management. Only by working hard to create a harmonious, harmonious and pleasant working atmosphere, let students study and live in a relatively loose environment, and properly organize some lively and rich amateur activities. Through the development of various recreational and sports activities and competitions, everyone has the opportunity to display their talents and enhance individual self-confidence and collective return. Emotional education is an important means that can not be ignored in the education and management of college students. Only by making rational use of institutional education and emotional education can the education and management of college students be carried out smoothly.

References

1. Mei, C.: An optimized research mode of college students psychological health and education based on data mining technology. *Revista de la Facultad de Ingenieria* **32**(9), 93–99 (2017)
2. Li, H., University, CN.: research on the theory and practice of college students' moral education under the background of internet era. *J Shanxi Inst Econ Manag* (2018)
3. Sang, H.: Analysis and research of psychological education based on data mining technology. *Secur. Commun. Netw.* **2021**(7), 1–8 (2021)
4. Liu, H.: Research on the innovation of college students' educational management under the new media technology (2019)

University Formal management system and informal system, etc. both formal system and informal system are good experience summary and accumulation in all aspects of colleges and universities, and become rules that people consciously abide by [15]. The institutionalized principle is an inevitable requirement for the order and efficiency of student management. Only through institutionalized management can we continuously promote the order, effectiveness and science of student education management Sex.

4 Implementation and Application of Emotional Education in College Student Education Management

- (1) To go deep into students' hearts and accurately understand students' emotional needs, we must fully love and trust students

To fully understand and master the ideological and emotional dynamics of students, we must first fully respect, love and trust students, and treat them sincerely with a frank attitude, which is the basic criterion that students' education managers should follow. According to Maslow's theory of psychological needs hierarchy, we know that love, respect and trust are what students want to be satisfied in their communication with teachers. Students hope to thrive in an environment surrounded by love, respect and trust [16]. Students' education and management workers need to understand and study students' individual moral quality, academic performance, interests and hobbies, personality characteristics, family background and social growth environment. They should be able to go deep into students' groups and be good at exchanging ideas with them, Communicate feelings and care about their daily life, learning status and mental status. If student workers can treat students with heartfelt love and trust and have friendly and deep conversations with students from the bottom of their hearts, students will regard you as a close friend and open their hearts to you unreservedly, so that you can find the best education method [17]. As an education management worker The author should be aware of "teaching without class" We should be good at discovering the advantages and strengths of each student, give praise and affirmation when we see their bright spots, encourage them to cultivate self-confidence, and guide students to move forward on the right track of life development. On the premise of fully loving and trusting students, the management mode of emotional education will achieve good results Effective.

- (2) To go deep into the hearts of students and accurately understand their emotional needs, we must treat students sincerely. As a student education manager in Colleges and universities, we must pay sincerely, make sincere dedication, and think for students wholeheartedly. The starting point and foothold of our work should be on the sincere care and help to students [18]. We must strive to guide them to grow up and become talents on the road of life, and we will win students. Students' education managers must go deep into students' lives, not only understand their learning and living conditions in time, but also fully master their daily life habits and strive to become their life mentor and close friend. Students' education managers should be good at heart to heart talks with students, and properly choose the way/skill and attitude of heart to heart talks, so as to create a kind of integration [19]. The atmosphere of conversation enables teachers and students to open their hearts and talk freely with

activities to promote the all-round development of students by guiding, standardizing and serving students. Student management can be divided into broad sense and narrow sense. The broad sense of student management is what we usually call student work, including ideological and political education, student affairs management, student work technology and methods, student work assessment and evaluation, student growth and development guidance and so on [12]. In a narrow sense, student management refers to the content of student affairs management. It focuses more on some daily affairs than thematic and developmental education, that is, class construction and management, freshman enrollment management, safety guidance and management, student rewards and punishments, loan assistance, dormitory management, employment guidance, career development services, etc. College student education management is an important part of college management. The level and quality of college student education management are directly related to the quality of talent training, the reform and development of colleges and universities, and the establishment of social image of colleges and universities [13]. Therefore, it is the mission and obligation of colleges and universities entrusted by the times to do a good job in the education and management of college students.

According to China's regulations on the administration of students in Colleges and universities, the educational management of college students refers to "Colleges and universities should focus on cultivating talents, follow the national educational policies and educational laws, and constantly improve the quality of education; run the university according to law, strictly manage, improve and perfect the management system, and standardize the management behavior; combine management with strengthening education, constantly improve the management level, and strive to cultivate qualified socialist builders and reliable successors."

3.3 Principles of College Student Education Management

The coordinated operation and effective management of college student management should follow the following principles:

- (1) The principle of practicality. The principle of practicality is to make the education and management of college students proceed from reality and integrate theory with practice. Therefore, the key to the education and management of college students is to proceed not only from the actual situation of the school, but also from the actual situation of students [14]. By proceeding from the reality of the school and students, decompose the functions of student education and management, and establish and improve the education and management of college students. The organization of education management, determine the objectives of student management, and study the ideal mode, system and means of college student education management.
- (2) Institutionalization principle. The institutionalization principle requires university student workers to formulate and apply various systems for student education and management according to the relevant national laws and regulations, combined with the training objectives and school running style of colleges and universities, so as to make the education and management of college students have rules to follow. The student education and management system of colleges and universities mainly includes the laws and regulations formulated by the state and formulated by the

3 College Student Education Management

3.1 Visualization of Emotional Analysis Results

After the emotional analysis results are obtained, the analysis results should be displayed by visual means such as charts, so as to provide decision support services for users or enterprises.

Considering the quantity of online course text data and the accuracy of the experiment, the method of emotion dictionary is used for emotion analysis, which mainly depends on the number of positive and negative emotion words in the sentence to calculate its overall emotional tendency. Considering the influence of degree adverbs and negative words in sentences, the basic calculation formula is as follows:

$$E_s = \sum_{i=1}^n Q_{w_{ai}} * Q_{w_{bi}} * W_i \quad (6)$$

In formula (6), E_s It represents the final emotional polarity of the sentence, n represents the number of emotional words in the sentence, W_i represents the extreme strength of emotional words, w_{ai} represents the weight of degree adverbs, w_{bi} represents the weight of negative words.

Generally, the data encountered in the actual experiment process will have the difference between structured data and unstructured data. Structured data such as students' education background, height, etc., unstructured data such as students' text profiles, published discussion content, etc. Structured data can be used after vectorization, while unstructured text data often needs to be analyzed and processed by vector space model to obtain feature representation before it can be used [11]. When the data to be processed is text data, the feature extraction of text resources, The commonly used TF-IDF (term frequency-inverse document frequency) algorithm is implemented by extracting the n keywords with the highest weight and calculating their weights. The weight calculation formula is as follows:

$$Q(w, t) = \text{tf}(w, t) * \log(n_w + \beta) \quad (7)$$

where $Q(w, t)$ represents the weight of word w in text t , $\text{tf}(w, t)$ represents the word frequency of word w in text t , n represents the number of texts in all training sets, and n_w represents the number of texts in which word w appears in the training set, β As an empirical value, the general values are 0.01, 0.1 and 1

According to the constructed content feature vector, learn the user's preference features to construct the user feature vector. The commonly used learning algorithms include DT, KNN, NB, Rocchio algorithm, etc.

According to the content feature vector and user feature vector obtained from the analysis, the matching degree is calculated using cosine similarity, Pearson coefficient and other methods, and the top N items with the highest matching degree with users are recommended to users.

3.2 Definition of College Student Education Management

College student education management is a kind of student education management. Student management, also known as student affairs management, refers to the organizational

Step 1: understand the knowledge of application field. Before starting knowledge discovery, the first and most important thing is to understand your data and business problems.

Step 2: data integration and data cleaning. Data integration combines the data of multi file or multi database operating environment related to the research problem. Data cleaning solves the semantic fuzziness in the data, corrects inconsistencies, and deals with omissions, noise and dirty data in the data.

Step 3: data reduction and preprocessing. Data reduction will get the compressed representation of the data set [10]. Mining on the reduced data set will be more effective and produce the same or almost the same analysis results, so as to improve the quality of data mining. Preprocessing is to organize and arrange the data orderly according to a specific algorithm.

Step 4: select data mining function. Select the corresponding mining functions according to the needs of mining tasks, such as classification, clustering or association rule mining.

Step 5: select the appropriate data mining algorithm for data mining.

Step 6: model evaluation. Evaluate the mined patterns, visualize, transform and express knowledge.

Step 7: expression of knowledge.

Figure 2 shows the basic process of data mining.

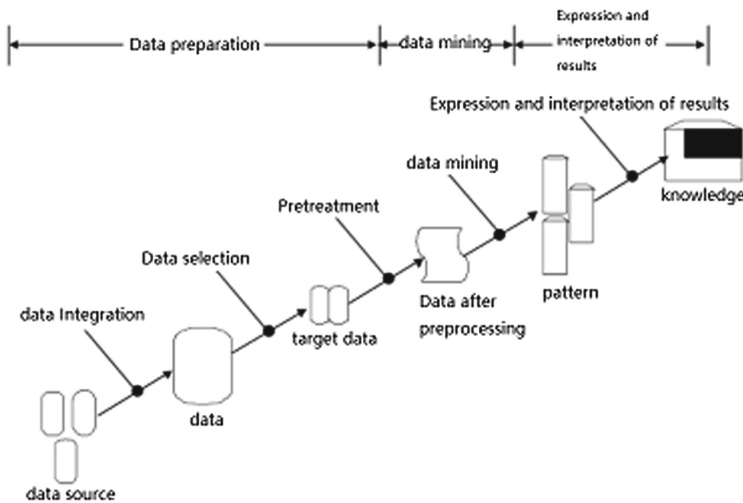


Fig. 2. Data mining process

2.2 Data Mining Process

Nominal attribute: its value is the name of some symbols or things. Each value represents a certain category, code or state, so the nominal attribute is also regarded as classified. These values need not have a meaningful order. In computer science, these values are also regarded as enumerations. It is meaningless to find the mean or median in the nominal attribute, while the mode is meaningful. Binary attribute: It is a nominal attribute with only two categories or states: 0 or 1, where 0 usually indicates that the attribute does not appear, and 1 indicates that it appears. Binary attribute is also called Boolean attribute. If the two states correspond to true and false. Ordinal attribute: there is a meaningful order or rank evaluation between its possible values, but the difference between successive values is unknown [8]. The central trend of the ordinal attribute can be expressed by its mode and median (the median of the ordered sequence), but the mean value cannot be defined. Numeric attribute: The numeric attribute is quantitative, that is, it is a measurable quantity, expressed as an integer or a real value. The numeric attribute can be interval scaled or ratio scaled. (1) Interval scale attribute: measure in equal unit scale. The values of interval attributes are ordered and can be positive, 0 or negative. Therefore, in addition to the rank evaluation of the value, this attribute allows us to compare and quantitatively evaluate the difference between the values. (2) Ratio scale attribute: numeric attribute with fixed zero point. In other words, if the measurement is scaled by ratio, we can say that one value is a multiple (or ratio) of another [9]. In addition, these values are ordered, so we can also calculate the difference between values, as well as the mean, median and mode. The definition is always smelly and long, not as simple as the summary of symbols and tables, as shown in Fig. 1.

- The type of an attribute depends on which of the following properties it possesses:
 - Distinctness: $= \neq$
 - Order: $< >$
 - Addition: $+ -$
 - Multiplication: $* /$
 - Nominal attribute: distinctness
 - Ordinal attribute: distinctness & order
 - Interval attribute: distinctness, order & addition
 - Ratio attribute: all 4 properties

Fig. 1. Quantitative value attribute

Data mining is a multi-stage data processing process, which mainly includes the following steps:

For document m , the number of topics is K , and the probability distribution of the n th word $w_{m,n}$ in document m is expressed as follows:

$$\mathbb{P}(w_{m,n} | \mathbf{m}) = \sum_{k=1}^K \mathbb{P}(w_{m,n} | z_{m,n} = k) \mathbb{P}(z_{m,n} = k | \mathbf{m}) \quad (1)$$

The process of generating a document according to the LDA model. For document m , the probability model of generating the document is expressed as follows:

$$\mathbf{P}(t_m, z, w | \alpha, \beta) = \mathbf{P}(t_m | \alpha) \prod_{n=1}^N \mathbf{P}(z_{m,n} | t_m) \mathbf{P}(w_{m,n} | z_{m,n}, \beta) \quad (2)$$

The formula can be obtained by integrating the formula twice:

$$\int_{m=1}^M \mathbf{P}(t_m, z, w | \alpha, \beta) d_{t_m} = \int_{m=1}^M \mathbf{P}(t_m | \alpha) \prod_{n=1}^N \mathbf{P}(z_{m,n} | t_m) \mathbf{P}(w_{m,n}, \beta) d_{t_m} \quad (3)$$

The LDA model formula of the whole document set is obtained:

$$\mathbf{P}(\vec{z}, \vec{w} | \alpha, \beta, q) = \int_{m=1}^M \mathbf{P}(t_m | \alpha) \prod_{n=1}^N \mathbf{P}(z_{m,n} | t_m) \mathbf{P}(w_{m,n} | z_{m,n}, \beta) d_{t_m} \quad (4)$$

The LDA model should achieve the purpose of extracting document topics through continuous iteration to convergence. Gibbs sampling has greater usability because of its simple process of training the LDA model and its ability to converge to the global optimum.

$$\mathbf{P}(z_i = k | z_{-i}, w) - \mathbf{P}(z_i = k, w_i = t | z_{-i}, w_{-i}, w_{-i}) \quad (5)$$

Emotional analysis, also known as opinion mining, for its definition, some researchers believe that text emotional analysis is the process of analyzing, processing, summarizing and reasoning subjective texts with emotional colors⁴. Some researchers also believe that text emotion analysis is the process of mining and analyzing the subjective and objective, viewpoint, emotion and polarity of the text through computing technology to classify and judge the emotional tendency of the text [6].

Large scale data set is the research object of data mining, which is vividly described as “the source of knowledge”. It can be structured, such as data in relational database, or semi-structured, such as text, graphics, image data, and even heterogeneous data distributed on the network. Data mining technology is application-oriented. It makes micro or macro statistics, analysis, synthesis and reasoning on specific data to guide the solution of practical problems, attempts to find the relationship between events, and even uses the existing data to predict future activities. In this way, it improves people’s application of data from low-level end query operation to providing decision support for business decision-makers at all levels [7]. It should be pointed out that the knowledge mentioned here is relative. It should be oriented to specific fields and have practical application value under specific preconditions and constraints [2]. At the same time, it should be easy to be understood by users, and can even be expressed and described in natural language.

The teaching process is a bilateral activity process between teachers' teaching and students' learning. In this process, teachers are guides and students are subjects. How to give full play to the role of the guide and the main body, make teachers happy to teach and students happy to learn, so as to obtain the effect of happy teaching, in which emotional education plays a link role [4]. The infection and incentive function of emotional education in teaching management cannot be erased. Perhaps when it comes to education, people first think of the tangible results of "preaching, teaching, and dispelling doubts", while ignoring the intangible teaching management process. However, teaching management is the top priority of all school work. Only by doing a good job in teaching management can we ensure the scientific and orderly progress of education. So, how to do a good job of teaching management under the people-oriented education concept? Emotional education is obviously an effective catalyst.

In the process of teaching management, if divided, there are both pure teaching and pure management. The former can be attributed to the teachers and the latter to the head teacher. However, no matter what kind of emotion, it is inseparable from the bond that maintains the relationship between teachers and students. Of course, perhaps pure teaching is only for the so-called "preaching, teaching, and dispelling doubts", which seems to be realized independently without emotion. However, the management of the head teacher cannot be separated from emotional education in any case. Because the work of the head teacher is the work of love and emotion. If it is separated from emotional education, the work of the head teacher will look pale. Then, let alone the teaching management staff who focus on both teaching and management, emotion is undoubtedly the link in their work.

Emotional education, as a unique educational concept, its development and context trend are also very different from other educational concepts, and it is generally believed that the corresponding is cognitive education. Emotional education is to comprehensively form an internalized self-ability through the learning of various knowledge, so as to achieve the purpose of cultivating college students to form correct value feeling and emotional experience [5]. Educators need to strive to make college students have emotional identity in the process of education and teaching and management, draw nutrition from it, and finally turn it into their own ideals and beliefs, so as to promote their healthy growth and all-round development. Therefore, based on data mining technology, this paper analyzes and studies the emotional education on the educational management of college students.

2 Data Mining Technology

2.1 Overview of Data Mining

Data mining (DM) is a decision support process. It is mainly based on artificial intelligence, machine learning and statistical technology. It highly automatically analyzes the original data of enterprises, makes inductive reasoning, excavates potential patterns, and helps decision makers make correct decisions. It is generally believed that data mining is a link of knowledge discovery KDD in database. It is the most important step to automatically and efficiently extract useful patterns from data by using specific data mining algorithms in KDD.

excellent college students and qualified talents of socialist modernization is an important topic in the current management research of colleges and universities.

The so-called emotion, in short, is a kind of psychological reflection produced by people in the process of cognition of objective things. Emotional education is to educate and infect students with emotion on the premise of fully exploiting this psychological reflection. It is the so-called “appeal to reason, appeal to emotion” and “appeal to people with emotion, convince people with reason” [2]. The process of emotional education is the blending and perception of mind and soul, the spiritual dialogue between teachers and students, and the common reflection and exchange of objective things between teachers and students.

If management is a science, then teaching management can be said to be an application of this science and the practice of management science; If education is an art, then emotion will make this art more personified. Perhaps, in many cases, we will think that emotion is the means for literati to play by emotion, the only way to carry out ideological and political education, and the research category of pedagogy and psychology. However, to some extent, the use of emotional education in teaching management is of great benefit, and it is worth our in-depth discussion.

Bai Juyi once said, “No one can move the heart without first feeling”. It can be seen how important emotion is in the communication between people. If education is an art, I think it is an art that makes people feel happy physically and mentally. Otherwise, there will be communication barriers between educators and educatees. So how to eliminate this obstacle? The answer is to be “emotional”. The answer is simple, but the implementation is not simple.

The people-oriented quality education focuses on the common development of students, teachers and schools, which requires us to look at quality education from the perspective of development. In the process of education, we should respect personalized development and strive to provide students with time and space for independent learning and innovation. This process needs to be realized by emotional education. Get closer to students, more in-depth and more understanding, and impress students with the truth.

In the past, teaching focused on examination. Teaching management often directs the teaching work of the school around this focus. The teacher takes the examination goal as the teaching goal, and the students mechanically accept the book knowledge. There is a lack of communication between the teacher and the students, let alone the blending of mind and soul. The emotional closure between teachers and students directly restricts the development of emotional education [3]. This has led to the educational phenomenon of “high scores and low abilities” and “high IQ and low quality”. Compared with exam-oriented education, the former is a new teaching mode. It aims at improving the overall quality of people and looks at the future education trend from the perspective of development. From the actual needs, comprehensively improve the quality of students. For students, it is particularly important to cultivate their self-esteem, self-improvement, enterprising and other good psychological qualities in terms of cognition, emotion and interest, and emotion is the most active factor in psychological quality. Therefore, teachers should update their ideas, especially teaching managers, and should step out of the past exam-oriented education model and use emotional education for teaching management.



Analysis and Research of Emotional Education on College Students' Education Management Under Data Mining Technology

Xiangzhou Liu^(✉) and Zhang Chen

Yunnan College of Business Management, Yunnan 650106, China
liuxiangzh365@aliyun.com

Abstract. In the traditional college education management mode, more attention is paid to and reliance on system management. On the one hand, system management promotes the standardization and institutionalization of student education management, and is a solid fortress of education management; On the other hand, simple system management is easy to make management work in trouble, so we should pay more attention to the artistry in the process of management work. Under the background of the increasing pursuit of humanistic care in contemporary education, the intervention of emotional education can be said to be an essential lubricant in the educational management of college students. Moving with emotion and convincing with reason, based on data mining technology, through emotional communication and exchange, open the psychological gate between teachers and students, and use the means of emotional education while implementing system management, which will achieve fruitful results.

Keyword: Emotional education · Data mining · Student education management · Emotional needs

1 Introduction

At present, the situation faced by colleges and universities is undergoing extensive and profound changes. Contemporary college students who are receiving higher education are the pillars of the country in the future. Their quality will be directly related to the development of China's socialist modernization and the great rejuvenation of the Chinese nation. China's Higher Education Law stipulates that the task of China's higher education is to cultivate senior professionals with practical ability and innovative spirit, develop socialist science, technology and culture, and promote the development of socialist modernization. College student education management is an important part of realizing the goal of talent training in Colleges and universities, and plays an irreplaceable role in the reform and development of higher education [1]. College student education management reflects the school spirit and appearance of a school, and is an important symbol of the level of school management. How to make the education and management of college students scientific, institutionalized and legalized, and then cultivate a large number of

management, and ensuring the promotion of research and practical implementation of digital university English teaching.

References

1. President Office of Hebei GEO University, Shijiazhuang, China Hebei Geo University, Shijiazhuang, China. Study on Construction of Open Network Teaching in University. Balkan Region Conference on Engineering and Business Education, vol. 2, no. 1, pp. 183–188, 2017
2. Rickyelton.: The digitization of construction. *Concret. Constr.* **64**(4), 64 (2019)
3. Odenthal, J., Jacobs, G., Kramer, A., et al.: FutureSite—test center for the development of tomorrow's construction site. *ATZheavy Duty Worldwide* **12**(4), 58–61 (2019)
4. Chaudhry, A.S.: Re-conceptualization of knowledge organization: imperatives of networked resources and digitization. *Int. J. Knowl. Content Dev. Technol.* **6**(2), 93–108 (2016)
5. Müller, L., Tipold, A., Ehlers, J.P., et al.: Digitization in teaching?—An analysis of requirements for implementing lecture recording in veterinary education. *Tierärztliche Praxis Ausgabe K: Kleintiere/Heimtiere* **47**(3), 164–174 (2019)
6. Mason, A., Richardson, S., Bugajewska-Waller, A., et al.: Digitization of toxin development. *Toxicon* **190**, S48–S49 (2021)
7. Davies, I.W.: The digitization of organic synthesis. *Nature* **570**(7760), 175–181 (2019)
8. Noh, Y.: A study to evaluate the digitization level of Korean libraries (part II). *Library Hi tech* **34**(2), 359–403 (2016)
9. Shi, Y.: Research on the application of university digital campus platform based on data mining. *J. Huainan Normal Univ.* **19**(1), 3 (2017)
10. Zhang, F., Zhou, B., Xu, F., et al.: Research on digital resource sharing portal based on big data. *China Educ. Inf.* **7**, 3 (2016)
11. Du, H.: Research on the cultivation of data mining ability of digital accounting talents. *Bus. News* **225**(35), 190–192 (2020)
12. Zhang, Z., Wu, F.: Research on precision teaching mode based on big data and data mining technology. *Inf. Technol. Inf.* **10**, 3 (2018)
13. Wang, G.: Recommendation analysis of high-quality education resources for college students based on big data mining. **9**(8), 183–184 (2019)

$$1 = (P_y - Y_i)(Y_{i+1} - P_y) \leq 0 \quad (3)$$

Digital English teaching resources can enable students to select learning materials in the resource library according to their own needs and carry out autonomous learning after class when completing teachers' classroom teaching tasks. Students can make their own learning time and learning plan, and gradually develop the autonomous learning ability of "learning to learn".

4 Conclusions

Resource pool and network-assisted teaching mode integrate the advantages of modern media information, timeliness and interactivity into teacher teaching, students' writing learning inside and outside class, homework evaluation and other teaching links, and gradually show its unique advantages, which are well received by students. The arrival of the digital era has promoted the process of digital education, and the construction and application of digital educational resources have become a new trend in the today. In the field of college English teaching, the construction and of digital resources based on data mining technology incorporate learning process visualization, optimization of learning plans, and other content into teaching strategies to improve course participation and academic quality.

The difficulty in constructing digital educational resources lies in how to collect, store, and process data. Large data can provide a more realistic and accurate data foundation. How to obtain useful information about personal learning habits, academic achievements, learning goals, and personalized plans from the data is an important academic issue faced by college English teaching. Through data mining, hidden and useful information and patterns can be extracted from a large amount of data, and digital educational resources can be constructed and optimized to improve teaching quality and effectiveness.

In the construction of digital educational resources, it is necessary to accurately collect, classify, and organize data based on students' personalized needs. By analyzing students' basic data, preference data, and research data, we provide more personalized digital education resources, including courseware, reading materials, specialized exercises, and discussion questions, to meet the learning needs of different students.

In the application of digital educational resources, data mining can help teachers continuously discover and improve teaching methods and educational resources, providing more choices and opportunities for students' growth. Through the analysis of learners' data, we can better understand students' learning process and needs, and recommend practical and effective educational resources to students, such as improving oral ability, writing ability and Critical thinking, so as to promote students to form effective learning strategies and habits.

Of course, in the teaching process, any application of educational resources and technologies must comply with educational regulations and ethical standards, ensure students' rights and privacy, and continuously enhance scientific and authoritative aspects. Therefore, the construction and application of digital educational resources require the joint efforts and pursuit of the entire society, strengthening relevant supervision and

2. Databases can provide students with richer learning resources: In English teaching, databases can of learning resources, such as teaching courseware, recordings, and videos. At the teachers' teaching content and students' learning content can be mutually supplemented and optimized, and on this basis, comprehensive English learning knowledge can be provided to students. By analyzing students' learning data and user preferences, teachers can continuously improve teaching resources and provide better support for students' English learning.
3. The database can provide data support for college English teaching: through the analysis of English learning business data, students' learning situation can be found, such as students' favorite Syntax error, needs for entrance examinations, students' interests and preferences, etc. The analysis of these data can help teachers and students better understand their learning situation and effectiveness, and then develop targeted teaching plans and plans English teaching.
4. The database can facilitate online interaction and easy communication: In college English teaching, the database provides a more convenient online communication and discussion environment, making communication between students and teachers easier, more natural, and more efficient. Students can ask questions and view answers at any time through the digital database system. Teachers can manage paper information such as papers through the database, provide targeted learning suggestions, and enhance students' awareness of self-directed learning.
5. Databases can provide data protection for college English teaching: In the field of English teaching, data protection is very important, not only responsible for the, but also to ensure and risk control of the teaching system. Databases can ensure the security of data and supporting electronic devices by encrypting, backing up, and fault-tolerant processing of data, and prevent various unexpected situations to ensure the stability of data protection in college English teaching.

In summary, databases play a crucial role in the field of college English teaching. Through the systematic management and analysis of student learning data, students can receive better learning support, and English teachers can timely understand students' learning situation, adjust and optimize teaching plans accordingly, improve the quality and effectiveness of college English teaching, and contribute to learning

3.2 Application of English Teaching Digital Resource Bank in Teaching

Third, explain the reading methods, writing methods, listening skills and translation skills; Fourth, timely check the completion of students' homework and make an objective and fair evaluation. The judgment of the particularity to be judged means that if the point is on the edge of the polygon, the direction of the vector $\overrightarrow{P_i P}, \overrightarrow{P P_{i+1}}$ is the same, and the judgment point P is between the line segments $P_i P_{i+1}$, then the following three discriminant functions are established at the same time:

$$j = (X_i + 1 + X_i)(P_y - Y_i) - (Y_{i+1} - Y_i)(P_x - X_i) = 0 \quad (1)$$

$$k = (P_x - X_i)(X_{i+1} - P_x) \leq 0 \quad (2)$$

Digital resources include classroom teaching courseware, teaching background and other auxiliary materials, audio, video and picture materials, teachers' teaching plans, teaching plans and massive resources on the Internet. Digital resources provide students with learning contents and tips in time through the network, which has reformed the tradition that students' English learning depends entirely on classroom, changed Students can make learning goals and plans according to their own learning abilities and needs. The richness of digital resources meets, and the interest of digital resources stimulates students' enthusiasm and motivation for learning. The reasonable application of digital resources embodies the educational concept.

3 Application of Digital College English Resource Database

3.1 Concept and Characteristics of Database

Data warehouse should have the features of quick retrieval, permanent consistency of data, easy-to-use browsing tools, complete and reliable data, etc. The data warehouse framework is shown in Fig. 3.

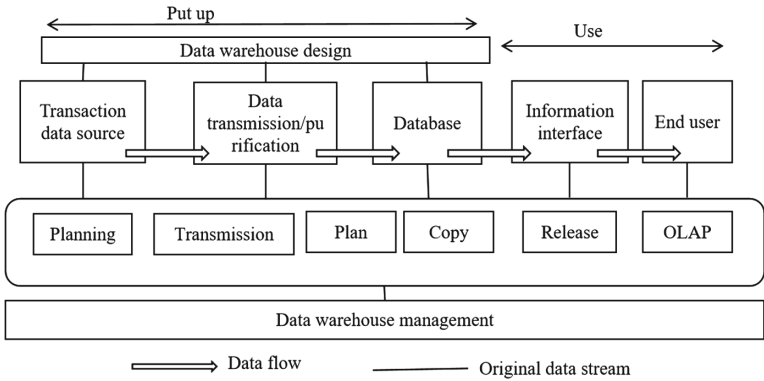


Fig. 3. Data warehouse framework

In today's digital age, databases are role in college English teaching. A database is a system that can store and manage data. Through a database system, universities can collect and analyze students' learning situations, provide targeted teaching plans based on their different needs, and also adjust and optimize teaching content in real-time to improve teaching effectiveness. The following is the role of a database in college English:

1. Teachers can obtain students' learning situation through the database and analyze them in a timely manner. In college English, teachers can use databases to understand students' learning processes and outcomes, thereby helping teachers better formulate teaching plans and plans. For example, by analyzing the evaluation results of students and data on various exam scores, teachers can identify students' weak links and provide corresponding teaching suggestions.

The update of English teaching methods is also one of the challenges that need to be faced in college English teaching. Although many universities have already applied various teaching methods in English teaching, such as online teaching, interactive teaching, task based teaching, etc., many universities still adopt traditional teaching models and methods, lacking a diversified English teaching system. This situation directly affects the effectiveness of English teaching and fails to meet the learning needs of students in different aspects. The diversity of student groups and needs in English teaching has also become one of the challenges faced by college English teaching. With the increasing emphasis of the country on the diversified development of higher education, the diversity and personalized needs of student groups are also increasing. Therefore, c adjusted and optimized from various aspects such, and teaching resources to different students. In short, the challenges faced by college English teaching are multifaceted. Only by carefully considering and striving to improve these challenges can we promote the continuous college English education and provide better English learning environments and conditions for students.

2.2 Importance of Digital Resource Construction

Which is the premise of “realizing new teaching and learning methods and changing traditional teaching structure”. It is the focus and direction of our research to dig deep into the function of resource database and make it serve for teaching to the maximum extent. The digital English teaching platform is shown in Fig. 2.

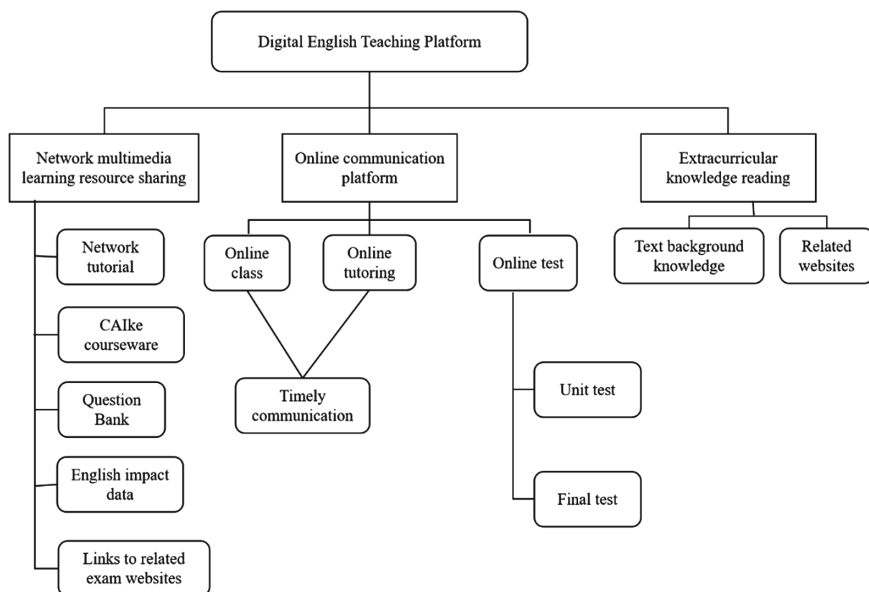


Fig. 2. Main disciplines involved in educational data mining and learning analysis

2 Construction of Digital College English Teaching Resources

2.1 Challenges Facing College English Teaching

The is centered on teaching materials. Students can only obtain knowledge by listening to classroom lectures and reading books, journals, newspapers and other paper media [8–11]. However, multimedia teaching means, English learning software and campus network teaching website have not been fully developed and utilized, resulting in the deviation of teaching resources from network media and the lack of novelty and richness [12]. Paying attention to textbook reading and written expression ability, emphasizing grammar teaching, ignoring students’ listening and speaking training and daily communicative English, “Campus English corner” is difficult to establish, or the effect is poor after establishment, so that many students’ oral and listening levels can not meet the requirements of the outline [13]. Even if there is training in listening and speaking, it often takes meeting the requirements of the examination outline as the means and aiming at passing the examination. The result is low score and low ability. For example, many college students who have passed CET-4 and CET-6 are difficult to communicate effectively with international friends in oral English, is still dominated by teachers. The “duck feeding” passive teaching mode of teachers’ speaking, students’ listening and recording does not establish the concept, which leads to the fact that classroom teaching is full of teachers, lack of interaction and inactive classroom atmosphere, which affects students’ enthusiasm and initiative in learning English. The main disciplines involved in shown in Fig. 1.

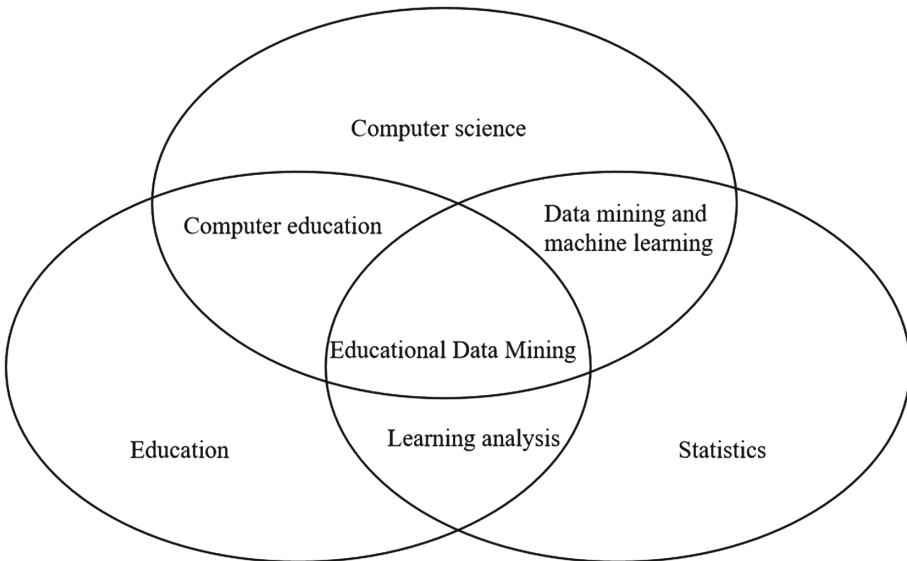


Fig. 1. Main disciplines involved in educational data mining and learning analysis

1 Introduction

In the universities has continued to expand, and universities with more than 10,000 people are not a few. On the other hand, the digital construction of colleges and universities has also made certain progress. The campus network has been initially established. If these data are not effectively developed and utilized, they may become a burden or even rubbish. The digital resource library for college English teaching is the basis and guarantee for effectively applying modern information technology in teaching, promoting teaching reform, and improving teaching quality [1], but for the digital campus, there is a lack of systematic research, which is worthy of our further discussion and research, so that data can truly become a school's resources, and use it to provide services for the school's own business decision-making and strategic development.

One of the difficulties in the development of resource database is how to make the framework design and implementation function of resource database more suitable for teachers and students and better serve teaching [2].

Although the construction of subject teaching resource database is only a small step in the modernization of education, it has already made us see the prospect of education development, spied on the way of education and learning in the information age, and inspired us to continuously study in this field, so that every new thing can play its due role [1]. The structure of this article is arranged as follows. Firstly, the construction and application of digital education resources will be sorted and analyzed. Then, the current application status and prospects of data mining technology in the field of education will be briefly described. Next, taking college English teaching as an example, we will explore the construction and application of digital educational resources based on data mining in college English teaching [3, 4]. Finally, the research achievements and contributions of this article will be summarized, and future research directions and prospects in related fields will be prospected. In the construction and digital educational resources have become an indispensable part and development of higher education. In addition to the development of digital technology, the widespread application of digital educational resources has further promoted the construction of university courses, improvement of teaching methods, and continuous improvement of students' learning experience [5, 6]. The famous President of Tsinghua University, Qiu Yong, also stated at the 2020 National Education Conference that digital educational resources will bring more profound changes to the education field. However, the development of digital education also faces certain challenges. In the process of constructing digital education resources, complex technical support and high costs are common challenges in the construction of digital education. In terms of application, the pertinence, practicality, and effectiveness of digital education resources also need to be further explored and improved. In addition, the static way of digital education resource construction can no longer meet the demand [7]. It also needs to provide more dynamic and intelligent support for the education field through Big data mining and other technical means. In order to promote the construction and application of digital education resources, data mining technology has been applied in the field of digital education. Data mining technology establishing mathematical models, analyzing and mining large amounts of data, and discovering hidden patterns and associations within the data.



Construction and Application of Digital College English Teaching Resources on the Basis of Data Mining

Xiaomin Yang^(✉)

Guilin University School of Language and Literature, Guangxi 541002, China
25871095@qq.com

Abstract. The arrival of the digital era has brought many challenges and changes to traditional teaching models, teaching methods, and teaching resources in higher education. Among them, the construction and application of digital educational resources has also become an important trend in the future development of higher education. This article mainly introduces the construction and application of digital college English teaching resources based on data mining. In terms of the construction of teaching resources, this paper uses the method based on data mining technology to analyze and process a large number of student learning data and teacher teaching data within a certain period of time, mine the patterns and rules of student learning, and establish knowledge expressions, Knowledge graph and curriculum knowledge trees related to the curriculum; In terms of application, online teaching platforms are used to provide students with rich and diverse digital teaching resources, allowing them to achieve targeted learning through various social learning methods such as independent learning, autonomous dialogue, and group discussions. Practice has proven that digital college English teaching resources based on data mining can better serve the teaching practice of higher education. On the one hand, this teaching resource makes students' learning more targeted and personalized, which is not only conducive to interaction between teachers and students, but also helps to improve students' English learning efficiency; On the other hand, this teaching resource can provide data support and reference for teachers' teaching practice, so as to better grasp students' learning situation and teaching progress, and achieve optimization and improvement of the teaching process. In summary, the digital college English teaching resources proposed in this article based on data mining are a feasible and effective teaching model teaching. Based on this model, we can continuously optimize teaching content, methods, and management, further improve the and quality of English teaching, and lay a solid students' English and career development.

Keywords: Data mining · Digitization · teaching resource

so as to save the time of looking for resources, improve the reading experience of college students, and cultivate the character of College Students' loving knowledge, Cultivate the habit of active reading and achieve the behavioral goal of lifelong love for learning.

6 Conclusions

Reading is a necessary way for college students to grow up and become talents, and it is also the cornerstone of knowledge society. The library should find out the reasons from its own perspective, increase the publicity of the library, even if sending push information to every classmate, try its best to increase the awareness of the library in readers' minds, actively improve the service quality of the library, provide more personalized services and make readers more willing to use the library. The successful application of data mining technology in the commercial field has brought inspiration to researchers in the field of university library. This paper studies the related knowledge of data mining technology and recommendation technology, and by combining three common technologies of data mining technology with university library services, especially reader services, constructs a collection recommendation system of university library based on data mining, which mainly mines association rules of circulation data of existing systems. Regular user data can lay a good foundation for the accurate recommendation service of the library, and help the library accurately retrieve the target resources of readers in a large number of literature data, so as to achieve the service objectives of reducing resource acquisition time, simplifying resource retrieval methods, optimizing resource recommendation efficiency and improving reader satisfaction.

References

1. Huang, W.: Research on accurate recommendation service mode based on reader reading data mining. *Lib. J. Henan* **39**(6), 3 (2019)
2. Zhu, Y.: Research on personalized recommendation model of university library based on data mining. *Times Financ.* **26**, 2 (2017)
3. Wwang, Z., Hu, D.: Research on virtual college student reader model and its reading recommendation algorithm. *Inf. Theory Practice* **41**(6), 7 (2018)
4. Zhang, R., Zhang, Y.: Research on intelligent recommendation technology of electronic library based on data mining algorithm. *J. Southwest Normal Univ.: Nat. Sci. Edn.* **42**(7), 5 (2017)
5. Ping, L., Wang, Y.: Research on book recommendation system of university library based on data mining. *Comput. Prod. Circ.* **11**, 1 (2019)
6. Liu, X., Cai, N.: Research on the application of data mining technology in book recommendation service. *Archit. Eng. Technol. Des.* **000**(030), 1650 (2017)
7. Wang, Q.: Research on personalized fast recommendation algorithm of library based on data mining technology. *Modern Electron. Technol.* **42**(5), 4 (2019)
8. Sun, W., Zhang, N., Li, Y., et al.: An empirical study on the reading needs of library readers based on improved K-Means algorithm. *New Century Lib.* **5**, 7 (2020)
9. Zhu, D.: Research on personalized reading promotion for college students based on data mining technology-taking Anhui Normal University as an example. *J. Tongren Univ.* **19**(7), 7 (2017)
10. Wang, H.: Research on the application of book recommendation algorithm based on data mining technology. *Modern Inf. Technol.* **3**(23), 3 (2019)

where, $R_{u,c}$ represent the user U 's rating on item C , $R_{u,c}$ and $R_{v,c}$ represent the average rating of user U and user V respectively.

5 Design of Collection Recommendation System Model in University Library

5.1 Decision Tree Algorithm Modeling

Using $C_{5.0}$ decision tree algorithm establishes the model, takes the reader's "major", "gender" and "grade" as the input variables, and "book category" as the prediction variables to establish the decision tree model. There is an obvious characteristic of College Readers' reading, that is, reading only when it is useful. What books are useful to readers will be read by readers. For example, some of the list belong to exercise guidance reference books. Readers read such books simply to assist learning, which has obvious pertinence and purpose [9]. Accurate recommendation calculation method is the technical guarantee to improve the efficiency of Library recommendation service. The library can choose the appropriate recommendation calculation method according to the actual reading situation of readers. The decision tree will recursively branch all borrowing data based on the input variable values. The first branch, also known as "root", contains all borrowing data records. Then, the root will be sorted according to the importance of input fields such as major, grade and gender, and then divided into several subsets or sub-branches. Each sub-branch can be further divided into secondary sub-branches, which can be further divided, and so on. Branches that are no longer divided are the lowest branches of the tree. Such a branch is called a blade. Finally, an analysis node is added to this model, which can determine whether the accuracy of the model can meet the special requirements.

5.2 Building Models Using Clementine Data Mining

Using the export node in Clementine data mining software, you can create a new field of "Borrowed Quantity of Books by Categories" according to the field of "Balance Quantity of Books by Categories" in the borrowed data, and make the borrowed quantity of books by categories less than 3 false and greater than 3 true, and establish a data mining model. The construction of library's accurate recommendation service mode based on reader data mining technology needs to take the dynamic data generated by readers' reading as the research ontology. The library resource recommendation system not only emphasizes the modeling of literature resources, but also attaches importance to the modeling of user data. The library should also make full use of big data mining algorithms to improve the matching degree between the target documents of readers' retrieval needs and the documents recommended by the library system, make full use of big data mining technology to analyze readers' reading rules and explore the internal relationship between data [10]. In the next step, we will mine and analyze the use of various electronic resources, provide personalized browsing and navigation of the network and personalized push services of network information resources, avoid readers' blind browsing on the Internet, and realize the filtering of network information resources,

where p is the point in the space representing the object, and m is the central point of cluster c .

Hierarchical method decomposes a given data set object hierarchically until certain conditions are met. According to the way of decomposition and formation of hierarchy, hierarchy method can be divided into two types: condensation method and splitting method [8]. Density-based method the density-based method overcomes the disadvantage that the distance-based algorithm can only find the circle-like clustering algorithm. As long as the density of nearby areas (the number of data points or objects) exceeds a certain threshold, clustering will continue.

Nearest neighbor collaborative filtering recommendation technology is the most successful recommendation technology at present. Its basic idea is to generate recommendations to target users based on the score data of nearest neighbors with similar scores, that is, generate a recommendation list for target users according to the views of other users. The analysis of library circulation data based on association rules is actually the association mining of the bibliographic information borrowed by readers, and then find out the effective borrowing rules, and then recommend the books that may be interested to readers according to the readers' borrowing interest. Whether the nearest neighbor query of the target user is accurate or not is directly related to the recommendation quality of the whole recommendation system. In order to accurately query the target user, it is necessary to accurately calculate the similarity between different items, so how to calculate the similarity between items becomes the key to improve the recommendation accuracy.

4 Cosine Similarity

The user rating is regarded as a vector in the N -dimensional item space. If the user does not rate the item, the user rating for the item is set to 0, and the similarity between users is measured by the cosine angle between vectors. Let the scores of user U and user V in the N -dimensional project space be expressed as vectors U and V respectively, and the similarity calculation method between user U and user V is shown in formula 2.

$$\text{sim}(u, v) = \cos(u, v) \frac{u \cdot v}{|u| \times |v|} \quad (2)$$

In the formula, the numerator is the inner product of two user scoring vectors, and the denominator is the product of two user vector modules. The smaller the included angle is, the higher the similarity is.

(1) Correlation similarity

Suppose that the item set scored by user u and user V is represented by I_{uv} , then the similarity $\text{sim}(U, V)$ between user u and user V is measured by Pearson correlation coefficient, as shown in formula (3).

$$\text{sim}(u, v) = \frac{\sum_{C \in I_{uv}} R_{U,C} \cdot R_{V,C}}{\sqrt{\sum_{C \in I_{uv}} R_{U,C}^2} \sqrt{\sum_{C \in I_{uv}} R_{V,C}^2}} \quad (3)$$

the analysis of reader reading data mining technology and library accurate recommendation service, this paper puts forward a specific service mode implementation scheme from four aspects: front-end application, algorithm implementation, role modeling and data collection (see Fig. 3) [5].

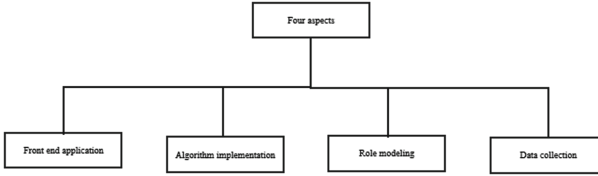


Fig. 3. Four aspects

Find out the hidden association network in the database through association analysis: on the one hand, use association rules to find the bibliography with strong association; On the other hand, the use of classification technology to determine the formation of multiple reader groups with different styles and interests, and recommend the bibliographies that have been consulted more frequently in the reader group, so as to establish a library personalized recommendation model. At the level of data collection, the library should fully consider the situation of readers' historical browsing records, readers' behavior and professional reserves, so as to recommend information resources with higher value for readers and improve readers' experience satisfaction. Classify data types and establish data sources [6]. At the same time, according to the characteristics of library data, through association analysis, find out the hidden association network in the database and establish the model.

3.2 Cluster Analysis Algorithm Selection

The formal description of cluster analysis is: the given data set is $V = \{V_1, V_2, \dots, V_n\}$, where v_i is called data point, and the data set is divided into K groups according to the similarity between data points, namely: $\{C_1, C_2, \dots, C_n\}$, $C_i = \{C_{i1}, C_{i2}, \dots, C_{in}\}$ ($i = 1, 2, \dots, k$). The algorithm of library circulation data mining chooses K-Means algorithm in clustering method, hierarchical method [7].

K-means algorithm is an algorithm based on centroid. It divides an object into clusters (parameters), so that the similarity within clusters is high and the similarity between clusters is the lowest. It mainly refers to dividing n objects into k categories according to their attributes, $K < n$, then locating the central position of K categories, calculating the distance from n objects to K central points, reclassifying n objects into the nearest category, recalculating the central position of each category, and pushing them in turn until the end. Similarity calculation is based on the average of all objects in a family (regarded as the center of gravity of a cluster or cluster). Generally, the criterion function is the sum of squares of distances, as shown in formula (1).

$$E = \sum_{i=1}^k \sum_{p \in c_i} |p - m_i|^2 \quad (1)$$

of data mining to mine the knowledge hidden in the book circulation data, and master the characteristics, behavior and trend of readers' reading, so as to improve the service quality of the library [4]. Pre-process the data and determine the type of data mining to be carried out. The basic steps of converting data into an analysis model CRISP-DM model include: business understanding-data understanding-data preparation-model building, model evaluation and model implementation. Figure 2 describes these steps and their relationship. In data mining, cluster analysis, association analysis and decision tree are very important methods, and scholars generally only use one of them to study. In industry, media and research, data mining usually refers to the whole process of knowledge discovery, that is, data mining in the broad sense above, so data mining can be defined as: data mining is the process of mining interesting patterns and knowledge from a large amount of data.

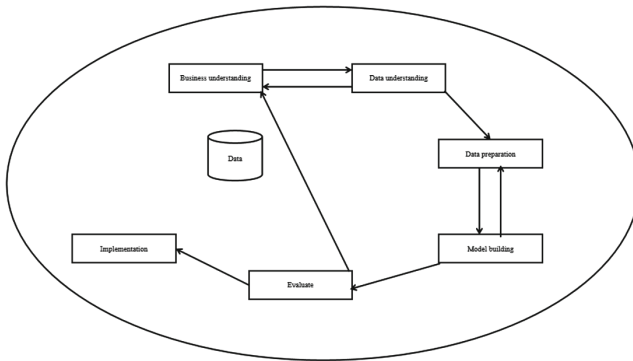


Fig. 2. CRISP-DM data mining process

3 Algorithm Selection of Library Circulation Data Mining

3.1 College Students' Reading Data Collection

In the face of a large amount of library lending information and students' personal information, data suitable for data mining applications are screened out. Because of the differences in readers' reading purposes, ways and habits, the reading behavior varies from person to person, but there are common manifestations. According to their own knowledge discovery, knowledge construction, readers gradually establish their own reading behavior. The purpose of constructing accurate recommendation service mode based on reader reading data mining in library is to provide personalized and distinctive push experience for readers. Data mining and association analysis based on the characteristics of reader reading data can highlight this individuality and characteristics. Through the research of classification technology and mining algorithm in data mining, this paper classifies the data type characteristics such as borrowing type, author type and college major in the existing lending records of libraries, and establishes data sources. Through

2 Data Mining Related Theory and Technology

2.1 Data Mining Concepts

With the rapid development of information technology and Internet technology, data shows an explosive growth trend. As the construction and exchange center of document resources, library pays more and more attention to the construction and improvement of database. Data mining and knowledge discovery are rising with the development of information technology such as database, data warehouse, machine learning and so on. Using the algorithms of association rules, clustering and decision tree in data mining technology, this paper models different situations respectively, and excavates the information and knowledge about students' borrowing habits, so as to guide the reading promotion of college students. As a new technology, data mining has been applied in some fields and made some achievements. The most typical case is the case of beer and diapers. The data mining technology such as association analysis is applied to the mining database to obtain the readers' borrowing habits, interest patterns and reading trends, so as to establish the readers' borrowing behavior model. Data mining can find the potential laws hidden behind the data through certain algorithms. Therefore, data mining provides a new way to solve this problem. It uses data mining technology to analyze the circulation data of the library, and then master the reading characteristics and reading trend of readers [3]. The data mining system consists of various databases (see Fig. 1), pre-processing module, mining operation module, pattern evaluation module and knowledge output module, which constitute the architecture of the data mining system. Data mining, as the core technology of knowledge discovery, is a process of extracting or mining information or knowledge from large-scale data sets, which is hidden and unknown to people in advance and valuable for decision-making. Simply put, data mining is a process of "mining" or extracting knowledge from specific data sets such as databases, data warehouses or other information bases.

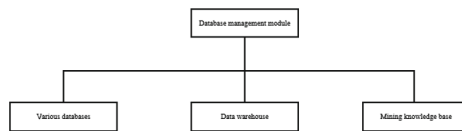


Fig. 1. Database management module

2.2 Research Status of Data Mining

Compared with foreign countries, domestic research on data mining started a little late and immature, and is currently in the stage of development. Latest development: in the research of classification technology, we try to establish its set theory system to realize massive data processing; Constructing fuzzy system identification method and fuzzy system knowledge model; Construct intelligent expert system. Data mining system is an application software system integrating information retrieval, management, expert system, analysis and evaluation, data warehouse and so on. Scholars use the key technology

the application of data mining technology in accurate recommendation service, which requires the library to analyze and model from massive data resources, screen out the target resources that match readers' needs to the maximum extent, and recommend them to readers. In recent 10 years, China has been advocating nationwide reading activities. As an important force in the industry, university library is also an academic institution for teaching and research services in colleges and universities, and it is the second classroom for college students. Reading promotion has become a routine activity and duty requirement of university library. However, there are some urgent problems to be solved in the construction of library digital service, such as how to improve the management level and service efficiency of the library, how to promote college students to make more efficient use of library resources and so on. University libraries have accumulated a large amount of data in the process of automation and informatization. Because the resources are updated every day, it is difficult for readers to obtain the latest resource information at the first time in most cases. Moreover, because the readers of the library have different backgrounds such as their age, major and identity, the demand for books is also different [1]. Therefore, in order to solve these problems, we need to carry out personalized recommendation through data mining, so that it can make a scientific, objective and reliable judgment on the needs of users.

As a relatively novel technology, data mining technology has been widely used in commercial fields such as banking, telecommunications and other industries, which provides a new perspective for the development of university libraries. University library, as the biggest reading source for readers, is also an important position of university education, and it is more important to understand the change of readers' reading tendency, which is of great theoretical and practical significance for library collection construction, readers' reading guidance, readers' education, etc., and also of great significance for giving full play to the educational function of university library [2]. In a broad sense, digital library is a computer processable and orderly information collection, which can also be regarded as the storage of digital information. It uses digital technology to organize and manage information resources. Users can query and retrieve information efficiently and conveniently through the network to obtain information services, and its information storage and user access are not limited by time and region. As the main reading group of society and colleges, College Readers' reading behavior has become an important research object. In practical work, this kind of accurate recommendation is the concrete expression of the library's active service concept. Accurate recommendation can provide readers with highly targeted and personalized information services, which fully reflects the library's reader centered service idea. In order to improve the sustainability and effectiveness of reading promotion in university libraries, university libraries need to make extensive analysis of readers' borrowing, resource browsing and downloading, and use data mining technology to find out the relevance, understand the knowledge needs of college students at different stages of school, and actively carry out reading promotion.



College Student Reader Model and Its Reading Recommendation Algorithm Based on Data Mining

Xin Peng^(✉) and Lixin Nie

Xi'an Eurasia University, Xi'an 710000, Shaanxi, China
pengxin@eurasia.edu

Abstract. This paper expounds the necessity of applying data mining technology to the recommendation service of university library, and introduces the present situation of applying data mining technology to the literature resource recommendation service, literature resource retrieval service and literature resource management service of library. The informatization construction of colleges and universities is one of the important fields of social informatization construction in China, and it is an important measure to comprehensively improve teaching quality and scientific research ability. The reading behavior of university readers is biased, utilitarian, the reading content tends to popular culture, the classical reading consciousness is relatively weak, and the reading style is digital. University library is an indispensable part of cultivating high-quality talents, and its informatization construction degree affects the cultivation level of college students' overall quality to a certain extent. As a service institution of higher education, university library should fully understand the reading trend of readers and grasp their reading interests, which directly affects the service quality of university library. The existing university library management system has accumulated a large amount of reader information and borrowing history information, which provides a data basis for recommendation service. This paper analyzes the three influencing factors of the construction of precision recommendation service model based on College Students' reading data mining technology, namely database construction, precision recommendation calculation method and reader oriented operation process.

Keywords: Data mining · College students · Readers · Model · Reading recommendation · Algorithm research

1 Introduction

As an important supporting body of scientific research services, the university library provides a platform and necessary conditions for the quality education of university readers, and plays a leading role in reading for the whole people with the advantages of collection, personnel, technology and equipment. With the development of readers' reading diversification and individuation, the library pays more and more attention to

3. Rongfan, M.A., Yang, J.: Research on the identification mechanism of university young teachers' political institutional based on binary logistic regression analysis. *Revista De Cercetare Si Interventie Sociala* **65**, 248–259 (2019)
4. Jacoby, J.W., Lesaux, N.K.: Language and literacy instruction in preschool classes that serve Latino dual language learners. *Early Childhood Res. Q.* **40**, 77–86 (2017)
5. Goulding, A., Dickie, J., Shuker, M.J.: Observing preschool storytime practices in Aotearoa New Zealand's urban public libraries. *Libr. Inf. Sci. Res.* **39**(3), 199–212 (2017)
6. Barnes, N., Brighton, C.M., Fives, H., et al.: Literacy teachers' beliefs about data use at the bookends of elementary school. *Elem. Sch. J.* **119**(3), 511–533 (2019)
7. Pinto, M., Fernandez-Pascual, R., Puertas, S.: Undergraduates' information literacy competency: a pilot study of assessment tools based on a latent trait model. *Libr. Inf. Sci. Res.* **38**(2), 180–189 (2016)
8. Zhang, H., Rao, H., Feng, J.: Product innovation based on online review data mining: a case study of Huawei phones. *Electron. Commer. Res.* **18**(1), 3–22 (2017). <https://doi.org/10.1007/s10660-017-9279-2>
9. Brown, C., Zhang, D., Xu, N., et al.: Exploring the impact of social relationships on teachers' use of research: a regression analysis of 389 teachers in England. *Int. J. Educ. Res.* **89**, 36–46 (2018)
10. Medaille, A., Beisler, M., Tokarz, R., et al.: Honors students and thesis research: a study of information literacy practices and self-efficacy at the end of students' undergraduate careers. *Coll. Res. Libr.* **82**(1), 92 (2021)
11. Tan, X., Yu, F.: Research and application of virtual user context information security strategy based on group intelligent computing. *Cogn. Syst. Res.* **52**, 629–639 (2018)
12. Herakleioti, E., Pantidos, P.: The Contribution of the human body in young children's explanations about shadow formation. *Res. Sci. Educ.* **46**(1), 21–42 (2016)

structural similarity of two nodes is 1, it means that the two nodes are equivalent. The structural similarity matrix can be calculated by the following matrix.

Get:

$$S = D^{-1/2}MD^{-1/2} \quad (4)$$

D is a diagonal matrix of diagonal elements $D_{ii} = M_{ii}$. It can be seen that the S matrix is symmetric, and the diagonal elements are all 1 s.

In this algorithm, we first extract the feature information of the network to be detected, and use the obtained feature matrix Φ_l as the input data of the clustering algorithm. Each row of the feature matrix corresponds to a node in the network. As a result, the community detection problem is transformed into a clustering problem in the scale space. Finally, the obtained feature matrix is clustered by combining the quantum clustering algorithm that introduces the adjacency information, which achieves good results and reduces the time complexity. Spend.

4 Conclusions

The promotion of educational informatization makes preschool teachers' information literacy more and more valued. Teachers should actively establish information awareness and apply information technology to teaching practice through training and self-study, so as to promote their information literacy to be continuously improved. Information literacy education is a kind of information enlightenment and ability training, which can prepare children for learning and lifelong development. Therefore, we should vigorously cultivate preschool teachers' information literacy in many ways, improve the overall information knowledge level of preschool education teachers, further improve children's learning environment and jointly promote the development of early information education for children in China. With the progress of overall social cognition, people are more and more aware of the great value of preschool education. In order to promote the progress of preschool education in China, we need to pay attention to the cultivation of preschool teachers' professional quality. In addition to the traditional professional quality, it also includes modern information quality, which requires relevant colleges and universities, kindergartens and preschool teachers to pay more attention, gradually improve the information quality of preschool teachers from various aspects such as training mechanism, awareness concept and form innovation, and improve the level of preschool teachers' query, analysis, integration and application of teaching information.

References

1. Michael, P.J., et al.: The influence of children's temperament characteristics on teachers' decision strategies. *Am. Educ. Res. J.* **19**(2), 165–181 (2016)
2. Zhao, R., Zhen, L.I., Huang, Y.: Correlation research on the practice of school administrative ethics and teachers' job morale and job involvement. *Revista De Cercetare Si Interventie Sociala* **72**, 44–55 (2021)

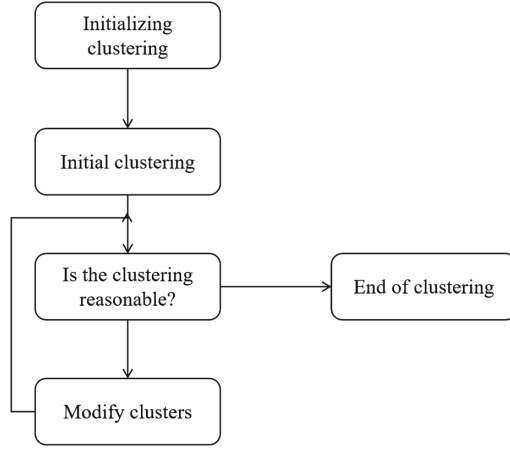


Fig. 3. Dynamic clustering process

We define a set i containing node i and points directly connected to node $\Gamma(i)$ through a common edge, then the tightness of node i and j can be expressed by the following formula:

$$M_{ij} = |\Gamma(i) \cap \Gamma(j)| = \sum_{u \in \Gamma(i) \cap \Gamma(j)} \omega_{iu} \cdot \omega_{ju} \quad (1)$$

However, the formula can't express whether two points in a set are connected with each other, so this index can be modified by a constraint, that is, when nodes i and j are not connected, let $M_{ij} = 0$. The matrix expression of post- M_{ij} constraint is as follows:

$$M = (A^2 + 2A + I) * (A + I) \quad (2)$$

In formula (1), A represents the adjacency matrix of the network, I is the identity matrix of the same order as the adjacency matrix, in which the operator $*$ represents the matrix multiplication, that is, the number multiplication at the corresponding position of the matrix of the same order. This modified matrix M is called the common adjacency matrix.

Next, the concept of structural similarity is introduced, and the formula of structural similarity is as follows:

$$S_{ij} = \frac{M_{ij}}{\sqrt{|\Gamma(i)| \cdot |\Gamma(j)|}} = \frac{M_{ij}}{\sqrt{\sum_{u \in \Gamma(i)} \omega_{iu}^2} \cdot \sqrt{\sum_{u \in \Gamma(j)} \omega_{ju}^2}} \quad (3)$$

The purpose of introducing this concept is to better define and measure the similarity between any two nodes. The return value of institutional similarity is between 0 and 1. The more common neighbors between a pair of nodes, the higher the structural similarity between the two nodes. If the two nodes are borderless, the similarity is zero. If the

role in life and teaching. Using information technology with computers and networks as the core to collect, process, generate and exchange information is a skill that teachers should have in the information age. For preschool teachers who do not major in computer science, the main purpose of learning computer, multimedia technology and network technology is to improve their work and improve teaching effects. At present, most kindergarten teachers have been able to use basic computer projection and other multimedia equipment, and are proficient in using basic application software such as word, Power Point, and windows drawing tools for lesson preparation and teaching, Collect and obtain required teaching information resources such as pictures, sounds, animations, etc. However, information technology should not only become a teaching tool for preschool teachers to demonstrate, but also a tool used to create teaching situations, organize children's thinking, and promote children's cognitive development. The essence of the Internet is sharing and communication, and many front-line teachers only search for information from the Internet, and do not publish their educational and teaching achievements on the Internet to share with everyone, and promote their own growth through mutual exchanges. Due to the influence of traditional concepts, information technology has entered the kindergarten poorly, and each kindergarten has little investment in hardware and software for modern educational technology, and only has a dozen computers for teaching, educational work management, and a small amount of multimedia, teaching software and other information resources. All these objectively restrict the cultivation and improvement of preschool teachers' information literacy.

3.2 Using Fuzzy Clustering Algorithm to Analyze the Application of Data Mining in Preschool Education

Clustering is the process of dividing physical or abstract objects or data according to the similarity between objects, and dividing similar objects or data into multiple classes or clusters. In short, it is the ancient idea of "birds of a feather flock together". Clustering is different from the classification in the field of machine learning. The learning process of clustering is unsupervised, and training samples and prior knowledge required by the classification process are not needed. Only according to the similarity between objects or data, the similarity between objects in each category is the largest, while the similarity between categories is very small, so as to discover the potential structure and potential effective information of data. Clustering analysis can be divided into hard partition and soft partition in terms of data partition. The dynamic clustering process is shown in Fig. 3.

Cluster analysis relies on the similarity between data as the basis for classifying each category, so how to measure the similarity between data is an important issue. At present, the commonly used similarity measurement methods mainly include distance-based measurement methods, similarity coefficient and correlation coefficient-based methods, information entropy-based methods, and other measurement methods for different types of data. Different from the previous quantum clustering-based community detection method, we introduce the concept of structural similarity in the data preprocessing stage. Next, we discuss the structural similarity matrix in detail.

In the process of using information technology, preschool teachers should first put morality first, strictly abide by the corresponding moral standards in many links such as information acquisition and processing, and have certain ability to distinguish the authenticity of various information. In actual operation, we should fully respect factual property rights, respect personal privacy, abide by network laws and regulations, actively resist bad information, and refrain from spreading false and harmful information, so as to fully assume social responsibilities and obligations as a citizen. Teachers' specialization means that teachers' profession has its own unique professional requirements and conditions, which is also a developing concept and a deepening process. The improvement of teachers' professional level always needs to acquire new knowledge, update knowledge and apply knowledge. According to Vygotsky's "zone of proximal development" theory, we advocate preschool teachers to find the zone of proximal development for children, provide effective "support" for children's learning, and arouse children's enthusiasm for active exploration and active learning. As the main undertaker of preschool education, we also need to find the "nearest development zone" for preschool teachers to promote their professional growth. It is in this sense that the information literacy of preschool teachers is of great significance to improve the professional level of teachers.

3 Research on the Application of Data Mining in Preschool Education

3.1 Application Status of Data Mining Technology in Education

In the field of educational management, data mining technology is slowly being popularized, mainly applied in library management, educational administration management, teaching evaluation, data analysis and other affairs, but the research is not mature on the whole. At present, the major normal universities in China are also aware of the importance of data mining, and have set up "data mining" and other related courses in the curriculum of education. With the continuous development and improvement of association analysis, clustering, concept description, deviation detection and other technologies, data mining will play an increasingly important role in the field of education. At the same time, with people's increasing attention to this technology, it is believed that its application range in the field of education will be wider and wider, so as to quickly promote the reform and development of education. In the field of preschool education, due to the lack of managers in technology, the relevant applications of data mining are basically blank. More often, they use Excel charts or SPSS to summarize and analyze the survey results and draw conclusions. With the wide application of information technology, experts in preschool education also began to pay attention to the cultivation of information literacy and investigate and study the information literacy of kindergarten teachers. However, they failed to make more meaningful discovery and reasoning about the obtained data, the internal relationship and mutual restriction between problems.

Preschool teachers in the information society have better information awareness, and can correctly and quickly recognize information, use information, and create information. The teaching information is novel and the teaching vision is broad, which is conducive to cultivating children's divergent thinking and developing children's creativity. Most preschool teachers agree that information technology plays an important

2.2 Information Literacy that Kindergarten Teachers Should Have

Because children's attention concentration time is short, they mainly focus on graphics and image thinking, and early childhood is the most active period of their creative thinking. Therefore, in early childhood activities, making full use of information technology can easily, quickly and accurately provide a large number of teaching materials and develop a suitable teaching environment, which requires preschool teachers to have certain information literacy. The information literacy required mainly includes: ① optimize teaching and make the form of education more flexible. In teaching, teachers are the provider of children's learning resources, the organizer of learning activities, and the most powerful helper in the process of children's learning and growth. For example, teachers with information literacy can understand children's interests and needs from children's life activities according to children's physical and mental characteristics, and take it as the basis for presupposing teaching themes; Then collect relevant pictures and small toys around the theme, use the network to consult relevant text materials and give guidance in all aspects; You can also use computers and projectors to play animation to help children learn language; Use computer drawing tools to help children learn painting; Through a variety of simple computer game software to help children learn numbers, and so on. Children experience and find problems in computer activities, and communicate with peers and teachers. In this way, learning and playing in the information-based teaching environment brings endless happiness to children's learning, so that their initiative, practical ability and creative imagination can be exercised and developed. At the same time, preschool teachers with information literacy can communicate with children's parents and peers in the form of network to explore educational experience, so as to enrich and flexible educational forms, so as to promote the all-round development of children. ② Cultivate children's information literacy and stimulate children's interest in learning. Like many toys and tools, computers and some simple game software can start from children's curiosity and interest, provide children with a variety of interesting opportunities, and timely seize useful educational opportunities to support children's learning and thinking. Only when preschool teachers have information literacy can they teach children to use computers and children's computer game software as in other activity classes, so that children can have more opportunities to practice problem-solving and make decisions. The overall structure of the integrated educational data system is shown in Fig. 2.

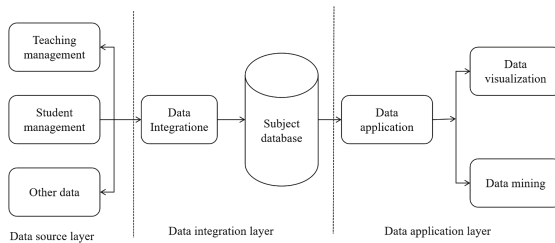


Fig. 2. The overall structure of the integrated educational data system

2 Information Literacy - The Professional Literacy that Preschool Teachers Should Possess

2.1 Connotation of Preschool Teachers' Information Literacy

In the field of preschool education, preschool teachers as the main body are a special young group. They generally have low educational background, limited research level and lower learning ability than primary and secondary school teachers, but they are a vibrant and positive group. Due to the special educational environment and teaching objects, they need more efficient teaching means and teaching modes to enrich the classroom, attract children's attention and improve the teaching effect, and information technology is the best tool. Therefore, how to strengthen preschool teachers' information technology ability and improve preschool teachers' information literacy is a very important task. The frame design of the preschool education public welfare evaluation index system is shown in Fig. 1.

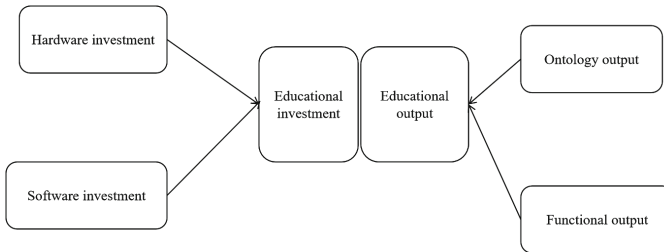


Fig. 1. Framework design of preschool education public welfare evaluation index system

Generally speaking, information literacy refers to the ability to master information-related theories, knowledge and methods, and effectively use information equipment and resources to obtain information, process information and create new information. Then, the information literacy that preschool teachers should possess in the information age includes: being aware of the important role of information in the information age, having a positive internal demand for information, being able to feel their own demand for information all the time, knowing what kind of information is easy for young children to accept and understand, promoting various activities and teaching, being able to integrate effective information into the teaching of the courses they teach, being good at quickly connecting information phenomena with children's life and environment, and being good at finding out the key to solving problems from information; Master the basic knowledge of information, information sources, transmission and acquisition; Can use network media, various multimedia and various softwares to develop and make preschool teaching courseware and preschool activity environment platform; Teachers, as transmitters of information, should filter out the mixed information before transmitting it to children to ensure their physical and mental health.

requirements for preschool education teachers [2]. The informatization of education has gradually become an important platform for educational modernization, and preschool education is the cornerstone of the whole education building. Its informatization plays a strong role in promoting the informatization of the whole education. Educational informatization is a systematic project. With a good software and hardware environment, teachers have high information literacy, which is of great significance for the promotion of educational informatization [3]. The popular understanding of information literacy is to search, process and process information, and then combine their own knowledge structure to internalize information and put forward their own views [4]. This kind of quality and ability is particularly important in the network era, and it is also an important quality for preschool teachers to develop professionally [5]. Kindergarten teachers' information literacy should be "know how to use computer and network technology to obtain relevant preschool education and teaching information, creatively develop and enrich kindergarten education resources, and reasonably and flexibly use a variety of information to solve practical problems in kindergarten education and teaching" [6]. The integration of Internet and education has become the development trend of the education industry. As an important participant in enlightenment education, preschool teachers must have basic information literacy to keep pace with the times and make the process of early childhood education more dynamic [7].

In recent years, the informatization level of early childhood education has been continuously improved, and the society has put forward higher requirements for the information literacy level of early childhood teachers [8]. Information literacy is an internalized cultivation. There are many indicators and factors involved in the evaluation. Its evaluation is often highly subjective. How to objectively and scientifically evaluate the information literacy level of preschool teachers is worthy of our careful study [9]. Data mining technology has been successfully applied in some fields, and its application in the field of high-level education management has been paid more and more attention. Not only normal universities and comprehensive universities, but also many large companies and enterprises have also participated in this business for the benefit of the future [10]. In the field of preschool education management, data mining is still a new topic. This paper will use data mining methods to analyze the information literacy of kindergarten teachers in the process of using information technology, including: ① personal information awareness and attitude, including teachers' professional ability, information environment construction, teachers' information awareness, and their initiative to solve work and life problems by means of information. ② Personal information knowledge and skill level, including teachers' use of multimedia software, multimedia courseware making ability, network skills, knowledge skills and courseware making ability. ③ Comprehensive factors and learning objectives of personal information literacy, including information teaching awareness, information ethics and security, information technology training requirements and information technology difficulties [11]. Through the research and analysis of the above three aspects, we hope to find out the subjective and objective factors affecting the improvement of teachers' information literacy, so as to guide the formulation of teachers' information technology training programs, and contribute to the formulation of pre service teaching plans for preschool education teachers and the adjustment of training objectives [12].



Application of Data Mining of Preschool Teachers' Information Literacy Based on Fuzzy Clustering

Lihong Shi¹(✉), Cheng ZhenHua², and Jie Zhang²

¹ Yunnan University of Business Management, Anning 650300, Yunnan, China
1454621886@qq.com

² Shandong Institute of Commerce and Technology, Jinan 250103, Shandong, China

Abstract. With the support of national policies and the active promotion of all sectors of society, China's information construction has also entered a new stage. While making progress in information construction, people need to correctly deal with the shortcomings in the development of information construction, actively learn advanced construction experience, and make the information construction of preschool education move forward stably. Multimedia courseware, modern information technology, computer, information technology education, information literacy and information construction have become the hot words in the information research of preschool education in China. In order to implement the planning requirements of the Ministry of education on strengthening the information literacy of preschool education teachers, using modern new media means to build a digital platform to improve the quality of children, promote the harmonious development of children's body and mind, and enhance the professional ethics and professional level of preschool educators, which is an important means to lay an important foundation for the steady and all-round development of preschool education in China. Preschool education is the most basic part of basic education. Improving Preschool Teachers' information literacy is particularly important to carry out early childhood information technology education and promote the development of preschool education. In view of the current situation of preschool teachers' information literacy in China, this paper puts forward to strengthen the construction of hardware and software, create an information technology application atmosphere, improve the management system and provide guarantee for information technology training.

Keyword: Fuzzy clustering algorithm · Preschool education · Information literacy

1 Introduction

With the rapid development of computer technology, network technology and communication technology, information technology has been widely used in various fields of Education [1]. The rapid development of information technology puts forward higher

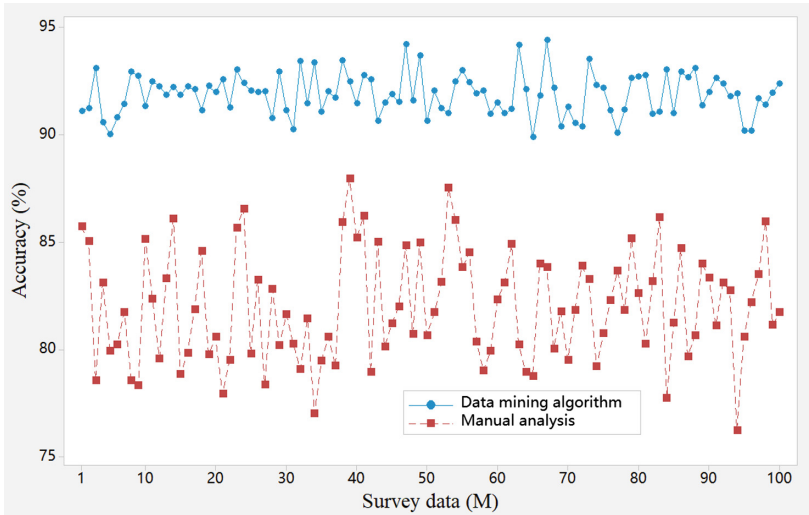


Fig. 4. Research accuracy of different methods

4 Conclusion

Under the condition of philosophical teaching, this proposes a algorithm for the physical education and the theory of big data to optimize the physical education. At the physical education standards of are the effect collection of physical education of college students. The research physical and synthesize college students' physical education Evaluation. However, in big data mining, too is paid to evaluation and analysis, in the lack of index correlation analysis.

References

1. Cai, B.: Cultivation of innovative ability of college physical education students based on the Internet of Things multimedia environment. *Int. Trans. Electr. Energy Syst.* (2022)
2. Ding, X.F., Tang, N.: Investigation on depression of college students majoring in physical education and nonphysical education: a study based on the age region and gender of 374 students. *Evid-Based Complement. Altern. Med.* (2022)
3. Du, X.F., Liu, Z.X.: Influence of physical education on the mental health of college students. *Revista Brasileira De Medicina Do Esporte* **28**, 311–314 (2022)
4. Li, F.Z., Liu, J.X., Qiu, F.S., Liu, H.T., Xin, S.F., Yang, Q.: Changes in mental health levels among Chinese physical education college students from 1995 to 2019. *Front. Psychol.* **13** (2022)
5. Liu, X., Dong, M.Y., Li, Y.C., Gao, S.S., Ye, M.H.: Analysis and research on influencing factors of college students' satisfaction with physical education. *J. Environ. Public Health* (2022)
6. Wang, P.P.: Construction and effect analysis of college students' physical education teaching mode based on data mining algorithm. *J. Sens.* (2022)
7. Wu, C., Liu, G.: Analysis of physical education based on deep learning on college students' mental health and social adaptability. *Front. Psychol.* **13** (2022)

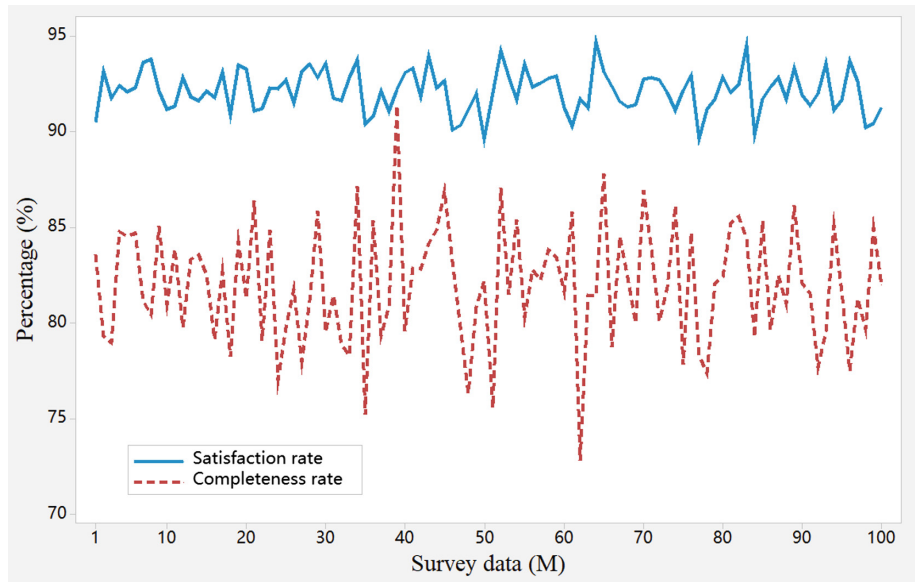


Fig. 3. The change process of physical education indicators of college students

the data mining, while the evaluation of the manual analysis method is relatively stable Uneven. The average results of the are shown in Table 3.

Table 3. Comparison of evaluation accuracy of different methods

Algorithm	Precision	Magnitude of change	Error
Data mining algorithms	92.78	3.51	2.78
Manual analysis	83.51	7.63	3.51
P	7.231	7.633	4.612

From Table 3, it can be seen that the analysis of teaching effectiveness for college students based on data mining algorithms can provide in-depth insights into the teaching process and student performance. The comprehensive results of data mining algorithms have higher accuracy than manual analysis methods. By analyzing the teaching process, the effectiveness of different teaching methods and strategies can be evaluated, and suggestions for improving teaching can be provided. The accuracy of data mining algorithms is greater than 90%, and there is no significant change in accuracy. In order to further verify the superiority of the data mining algorithm, the data in Table 3 was analyzed, and the results are shown in Fig. 4.

It Fig. 4 that the change range of the is small, while the change range of the manual analysis method is small, which further proves the advantages of the algorithm.

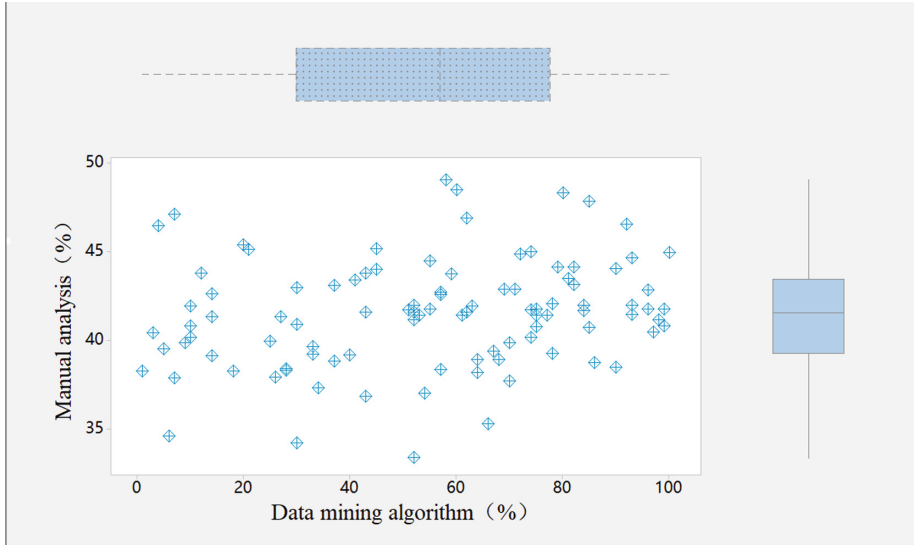


Fig. 2. The processing process of physical education for college students

Table 2. Overall effect of physical education indicators for college students

Time period	Satisfaction rate	Complete rate
1–4 weeks	82.47	86.60
5–8 weeks	83.51	87.63
8–12 weeks	86.60	89.69
Mean	87.63	90.72
χ^2	89.69	91.75
P = 0. 721		

college students’ were selected, and the results of college students’ teaching were shown in Table 2.

The results in Table 2 are shown in Fig. 3.

The data in Fig. 3 shows that the process of change in satisfaction and completeness rates is relatively stable, and the overall result is better. Therefore, the overall effect of the data in Table 2 is better, indicating that the research method in this paper is more effective.

3.3 The Accuracy and Stability of College Students’ Sports Evaluation

The results are compared with manual analysis to verify the data mining algorithm’s accuracy. The data mining the manual analysis method, indicating that the evaluation of

physical education of college students, and the index weight of the results the accuracy of the path big data mining [4]. If the data students are distribution, the evaluation resul, reducing the accuracy of the evaluation. In order to improve the level of evaluation, it to select the indicators of college students', and the specific index is shown in Fig. 1.

The physical education indicators of college students, which the objective facts [5]. The evaluation index has no directionality, indicating that the students has study of 8–12 weeks. The evaluation index meets the usual requirements, the evaluation index, irrelevant indicators, and the default index so that the full result is strong.

3 Practical Cases of Physical Education Teaching for College Students

3.1 Introduction to Physical Education for College Students

In order to facilitate the evaluation, the physical education of college students in civil aviation transportation under complex circumstances, with 12 paths and the evaluation results of specific physical education teaching are shown in Table 1.

Table 1. Relevant parameters of physical education for college students

Teaching content	Period	Learning effects	Appraise	Teaching format
Specialized courses	1–8 weeks	90.72	87.63	Online, Offline
	8–12 weeks	90.72	86.60	Online, Offline
Elective course	1–8 weeks	82.47	88.66	Online, Offline
	8–12 weeks	87.63	89.69	Online, Offline
Instruction classes	1–8 weeks	82.47	89.69	Online, Offline
	8–12 weeks	86.60	85.57	Online, Offline

The processing process of physical education effect of college students in Table 1 is shown in Fig. 2.

It from Table 1 that with the manual analysis method, the of the data mining algorithm are closer. In terms of the rationality and adjustment degree of evaluation index selection, it is analysis method. From the result change in Fig. 3, it can stability and faster judgment speed [6]. Therefore, the teaching optimization speed, intelligence level and of data mining.

3.2 Satisfaction Rate of Physical Education Teaching for College Students

The evaluation of the on college students, including satisfaction and implementation rates. After the pre-selection of the algorithm, the preliminary physical education results of college students are obtained, and the feasibility education teaching data [7]. In the effect of students' physical education teaching more accurately, different periods of

2 Related Concepts

2.1 Mathematical Description of Big Data Mining

Big data mining is to use the learning results, and according to the indicators in the learning results, find the outliers in students', integrate the corresponding results, and finally judge the of feasibility. The big data theory to quantify the physical, which can improve the accuracy of evaluation [3].

Hypothesis 1: The for college students is, the learning set is, the learning d_i index is, and the set_i learning y_i judgment is $F(d_i \geq 0)$ as shown in Eq. (1).

$$F(d_i) = \sum x_i \cap y_i \cdot \xi \quad (1)$$

2.2 Selection of Evaluation Indicators

Hypothesis 2: The indicator selection function is, and the teaching coefficient is $z(d_i)$, then the indicator selection w_i is shown in Eq. (2).

$$z(d_i) = \cup z_i \cdot F(d_i, y_i) \rightarrow k \cdot w_i \cdot \xi \quad (2)$$

2.3 Processing of Physical Education Data for Students

Before big data mining, the teaching data should be analyzed discretely, and the results should be mapped into a two-dimensional plane to eliminate abnormal results. First, the

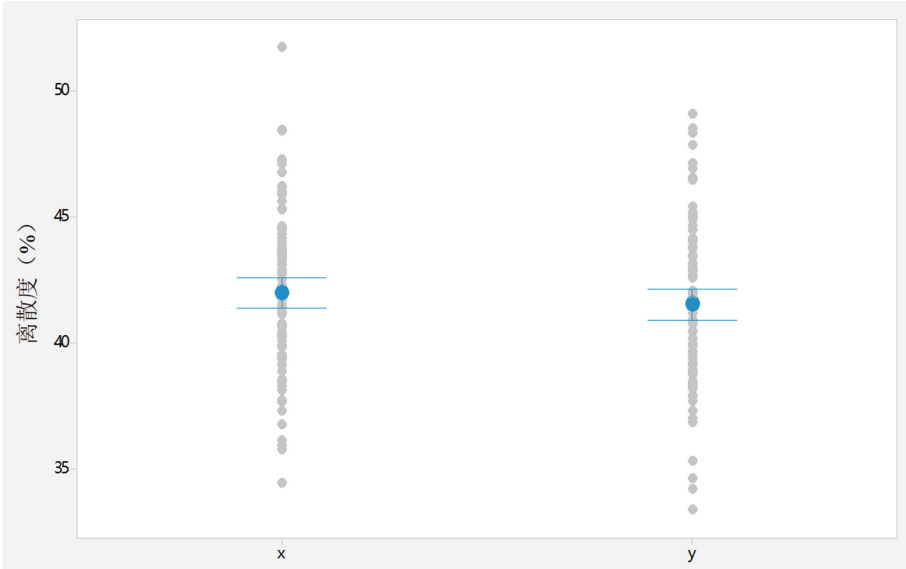


Fig. 1. Selection results of physical education indicators for college students



Analysis on the Teaching Effect of College Students Based on Data Mining Algorithm

Haifeng Xu^(✉), Yang Tong, and Liao Jun

Modern College of Northwestern University, Xi'an 710130, China
guanjun029@163.com

Abstract. The role of learning effect physical, but there the learning effect is not satisfactory. Manual analysis cannot solve the problem of poor learning effect in physical education of college students, and there are fewer learning indicators. Therefore, this to construct an evaluation model for the effect students' physical education. Firstly, the big data theory to evaluate the effect of college, and the evaluation collection is divided to the teaching norms the evaluation. Then, the big data theory is used to evaluate the physical education teaching of college students, form an evaluation set, and comprehensively evaluate the results. MATLAB that under teaching requirements, the d evaluation accuracy and evaluation time are better than the manual analysis method.

Keywords: big data · theory · university students · Sports · Learning · Effect

1 Introduction

The learning effect is one of the important evaluation contents of college students', which is of great significance for optimizing college students' teaching. However, in the actual evaluation process, there is a results, which has a certain impact on the physical education. Some scholars believe that learning intelligent algorithms into college students' physical education can effectively analyze the evaluation results and for evaluation. On this basis, this paper proposes the physical education of and verify the model's effectiveness. Firstly, we will collect a large amount of student learning data, including student grades, exam answers, and homework completion [1]. These data will be used to construct a comprehensive student learning dataset. Next, we will apply data mining algorithms such as association rule mining, classification algorithms, and clustering algorithms to discover hidden patterns and patterns in student data. By analyzing the correlation between students' learning behavior and academic performance, we can gain a deeper understanding of their learning characteristics and difficulties. At the same time, we will collect relevant data during the teaching process, including the use of teaching resources and the adoption of teaching strategies [2]. By using data mining algorithms, we can analyze the teaching process, evaluate the effectiveness of different teaching strategies, and provide suggestions for improving teaching.

10. Yang, J., Han, X., Gao, K.N., et al.: *Listeria monocytogenes* inoculation impedes the development of brain pathology in experimental cerebral malaria by inhibition of parasitemia. *ACS Infect. Dis.* **8**(5), 998–1009 (2022)
11. The assessment and exploration of forensic nursing concepts in undergraduate nursing curricula: a mixed-methods study. *Int. J. Nurs. Educ. Sch.* **18**(1), 16–21 (2021). Louisiana Tech University, 105 Wisteria Lane, Ruston, LA, 71272, USA
12. LIXianyu.: Analysis on the application value of experimental teaching method in the teaching of “Plant Cells and Animal Cells”. *外文版:教育科学* (5), 4 (2022)
13. Chen, X., Pan, L., Wang, X.: *Exploration on the Reform of Teaching Mode and Method of Product Design under Internet Mode* (2021)
14. Ye, P., Zhou, W., Zhao, C., et al.: Research on the forming method of lightweight body and parts. *IOP Conf. Ser. Earth Environ. Sci.* **632**, 032066 (2021)
15. Zhang, W.: *Exploration and online sharing network teaching platform based on information technology* (2022)

film reading assessment and group learning display. Through the examination of specimen microscopy, case analysis and defense, we can have a more in-depth and specific understanding of the extent to which students master practical operation skills.

5 Conclusion

With the development of science and technology and the mutual penetration of adjacent disciplines, some new fields and branches of animal pathology have emerged, such as molecular pathology, genetic pathology, comparative pathology, etc., and new technologies and new research fields will still appear in the future. However, in the future, animal pathology will still put the diagnosis of diseases in the first place, and the means and methods used should be advanced and modernized. In other words, in addition to extending the use of autopsy technology, we should also make full use of various ways to absorb some of today's high and new technologies, develop and establish new methods and technologies suitable for pathological diagnosis in combination with the actual situation of Pathomorphology, so as to improve the level of animal pathological diagnosis, so that some diseases that were difficult to be diagnosed by general methods of animal pathology in the past can be diagnosed in time and accurately. It should be the responsibility and task of contemporary Chinese animal pathologists to make the development of animal pathology in China keep pace with the development of animal pathology in the world and take the lead in some aspects.

References

1. Huang, L., Zhao, W.: Research on the Construction of Teaching Staff for Innovative Experimental Class in Applied Undergraduate Universities (2021)
2. Wang, W., Yin, G.: Analysis and research on the application of internet technology in sports track and field teaching. *J. Phys. Conf. Ser.* **1881**(4), 042026 (7pp) (2021)
3. Xia, L., Yu, J.B.: Research on the experimental teaching reform of "material analysis and test method" based on OBE engineering education mode (2021)
4. Meng, J.: The Necessity and Influencing Factors of Constructing a Diversified Cooperative Educational Mechanism for E-Commerce Major. Springer, Singapore (2022)
5. Manan, M.S.A., Wang, X., Tang, X.: Innovating animation teaching system: an experimental survey on the integration of design thinking and creative methods for animation education in China. *Open J. Soc. Sci.* **10**(3), 10 (2022)
6. Wang, S.Y.: Exploration and Practice of Quality Education for Undergraduates in Private Applied Undergraduate Colleges Based on TOPCARES, vol. 2. Francis Academic Press (2021)
7. Zhou L . The Experimental Research on Stylized Teaching Method Used in the Art Guidance of Chinese Undergraduates (Piano Accompaniment) — Take the Teaching of Haidun Piano Sonata as an Example[J]. *Arts Studies and Criticism*, 2021, 2(4)
8. Deng, X., Lei, J., Chen, M.: Application of VR in the experimental teaching of animation art. *Mobile Inf. Syst.* (2021)
9. Aoife, F., Ornaith, K., Long, N., et al.: P03 A qualitative exploration of antimicrobial stewardship and resistance education in undergraduate healthcare professional programmes. *JAC-Antimicrobial Resistance (Supplement_1):Supplement_1* (2022)

multiple levels and angles. In the above three stages of teaching, the TBL (team based learning) teaching mode can be introduced into the traditional animal pathology experiment teaching, as shown in Fig. 4. For example, the College of life sciences of Longyan University has set up more than ten comprehensive and designed experimental projects, such as the observation of clinical symptoms and pathological autopsy of Newcastle disease in chickens, the establishment of a carbon tetrachloride liver injury model in mice, the observation of tissue and organ damage in mice caused by acute heat stress, and the pathological autopsy and observation of *Pasteurella* infected rabbits, using good experimental platform conditions such as the Institute of animal medicine and the Fujian swine epidemic prevention and control engineering center, Students will be divided into free groups, with 5–6 people in each group. Each group will randomly select an experimental project to complete relevant learning tasks such as data collection, observation, analysis and drawing of pathological specimens, production and presentation of PPT, interactive Q & A, etc. In class, each group will show the learning situation to the whole class in turn, and lead everyone to learn the content of the experimental project together. After the speech, interactive Q & A will be carried out, and the student representatives and teachers of each group will score the whole group. Finally, the teachers will comment and make necessary analysis.

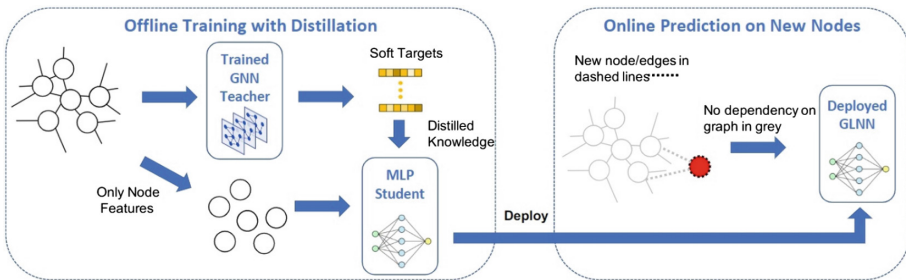


Fig. 4. Relying on the “industry university research” base to build a teaching mode map of industry education integration

During the practice week of pathological anatomy, based on some pathological cases submitted for clinical examination, students are required to diagnose the disease of animals from the clinical manifestations, signs and pathological changes (including ocular and microscopic examination) found during autopsy, discuss the evolution and development of the disease, and analyze the cause of death. For example, the pathogen of yellow diarrhea and white diarrhea of piglets is the same, both of which are acute and fatal intestinal infectious diseases of newborn piglets caused by pathogenic *Escherichia coli*. The difference lies in the age of onset, fecal color and mortality; Both transmissible gastroenteritis and epidemic diarrhea can cause diarrhea in piglets, but the mortality of piglets is different. The length of villi and the proportion of intestinal gland recess are different in pathological sections. In this way, through the analysis of some diseases with similar symptoms, students can draw inferences from one instance.

In addition, establish a comprehensive experimental performance evaluation system based on four indicators: experimental class performance, experimental report,

of animal pathology is divided into lectures, and 9 main teaching modules for Animal Pathology are designed. Theory is followed by practical training, alternating between theory and practical training, to deepen students' understanding of pathological changes. At the same time, the release of group tasks is used to promote students to actively seek for lesions, group students discuss lesions, group classroom reports, and inter group questions to stimulate students' understanding of the lesions. Finally, the teacher comments to deepen their understanding of the lesions (see Fig. 3).

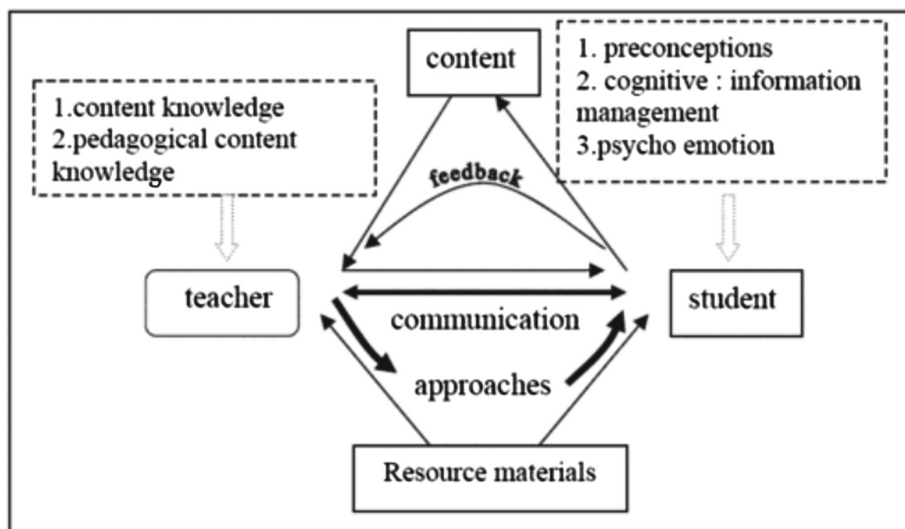


Fig. 3. Teaching and learning process

At present, many colleges and universities have less class hours for animal pathology experiments. Generally, 4–6 experiments are arranged in the second half of the second grade, with about 3 class hours for each experiment, a total of 18–20 class hours. The content of experimental courses is mostly the observation of specimen pathological tissue sections, and the teaching method is relatively simple, which often makes students in a state of passive acceptance of knowledge, and the classroom teaching effect is poor. In order to strengthen the practical teaching links and increase the corresponding class hours, the practical teaching is divided into three stages: the first stage, the pathology experiment course is opened at the same time as the pathology course is opened in the second semester of the second grade, and the pathological anatomy practice is increased, so that students can deepen their understanding of the pathological characteristics and changes of the disease, and enhance the basic operation and skill training; 2. In the second semester of the third grade, the veterinary skill experiment course is offered, including the autopsy and film reading of animal pathology; in the third stage, during the senior internship, the cases encountered in the clinical pathological diagnosis of animal diseases are described in writing, typical clinical cases are collected, and students are allowed to observe a large number of ocular and microscopic specimens, multimedia courseware, slides and wallcharts, so as to carry out pathological practice teaching at

enrich the teaching content of Pathology, activate the classroom atmosphere, stimulate students' interest in pathology, and enhance students' understanding and understanding of diseases.

For example, when describing the chapter of tumor, first introduce some pathological pictures of typical tumors to improve students' interest in learning, then introduce a case picture of clinical tumor, and ask students why they can judge whether the sick animal is inflammation or tumor, benign or malignant, early or late tumor by observing the tissue sent for examination, and then introduce the basic points of tumor diagnosis in combination with the content of books, Advance layer by layer to introduce the general characteristics and microscopic characteristics of the tumor, so that students can have a deeper understanding of the pathological changes of the tumor.

As teachers, they should pay attention to collecting and sorting out relevant clinical data in their usual work, and gradually establish and optimize animal pathology question bank, basic discussion question bank of clinical cases, medical history and case analysis, and picture Bank of histomorphological diagnosis. Organize students to use their knowledge of animal anatomy, histology and animal pathology to discuss relevant cases Medical history and means of diagnosis and treatment. It can not only cultivate students' clinical thinking ability in advance before they enter the clinical learning stage, but also improve students' interest, initiative and enthusiasm in learning animal pathology knowledge.

With the popularization of information technology, a large number of information technology and information database resources are used to assist teaching during the teaching process, such as course PPT, digital images, FLASH animations, hands-on videos, and course resource libraries, which can greatly increase the intuitive nature of the teaching process, save time in course teaching, simplify the teaching process, enable students to obtain various cognitive changes in pathology through various ways, and improve their ability to diagnose diseases, It is conducive to stimulating students' interest, mobilizing learning enthusiasm, and improving learning effectiveness. In addition, offline interaction between teachers and students is increased through QQ, WeChat, and other means to provide timely feedback on students' questions, breaking the gap between space and time, and extending the time and content of learning. With the help of information based teaching platforms, such as Learning Connect and Cloud Classroom, students can preview or review PPTs, lesson plans, and videos and pictures that accompany the course, both in and out of class, to deepen their understanding of pathological changes. It is also possible to continuously test the learning situation and effectiveness of students through the rational use of the test question library.

4 Exploration of Diversified Reform of Experimental Teaching Methods

Adopt a combination of traditional teaching methods and rational use of information based teaching methods, make full use of rich network information resources, and upload relevant course construction materials (such as lesson plans, PPTs, question banks, videos, photos, etc.) using network platforms (such as Xuetong or cloud course construction courses). Guided by the traditional teaching model, the abstract content

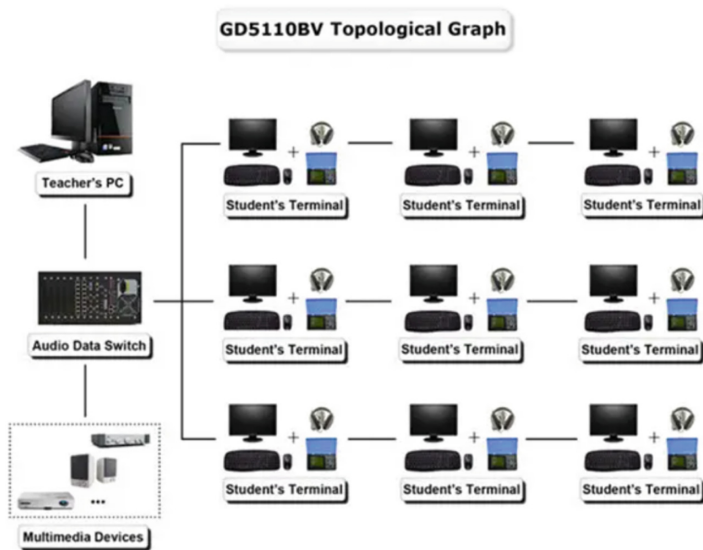


Fig. 2. Multimedia teaching mode

and students. The interactive teaching between teachers and students can also make up for the lack of typical specimen sources, and improve the teaching effect [9].

When applying multimedia teaching, it should be noted that multimedia teaching cannot completely replace the traditional teaching methods, especially the dominant position of teachers in the teaching process. When using multimedia teaching, teachers still need to impress and influence students through their body language and personality charm, a key. The difficult content should also be displayed in the form of blackboard writing, and the teaching outline should be kept on the blackboard, so that students can see the whole class, and have a general impression after class, so as to avoid the negative effects of fast switching of multimedia screen content, passing through the key points and difficulties as soon as they are put on the blackboard, and not easy to remember.

3.2 Combining Pathological Changes with Clinical Diagnosis and Clinical Manifestations, Case Teaching is Adopted

Animal pathology knowledge is an important basis for clinical diagnosis, and pathological changes determine clinical symptoms and signs. Combining pathological changes with clinical manifestations in teaching can enrich pathological changes and stimulate students' interest and enthusiasm in learning. During the teaching process, some cases can be arranged for discussion according to the contents of each chapter and common diseases. Students are required to diagnose the diseases of animals from the clinical manifestations, signs and pathological changes found during autopsy, discuss the evolution and development of diseases, and analyze the cause of death. Through case discussion, we can not only cultivate students' connecting, developing and overall thinking methods, promote the development of students' clinical thinking, but also It can also

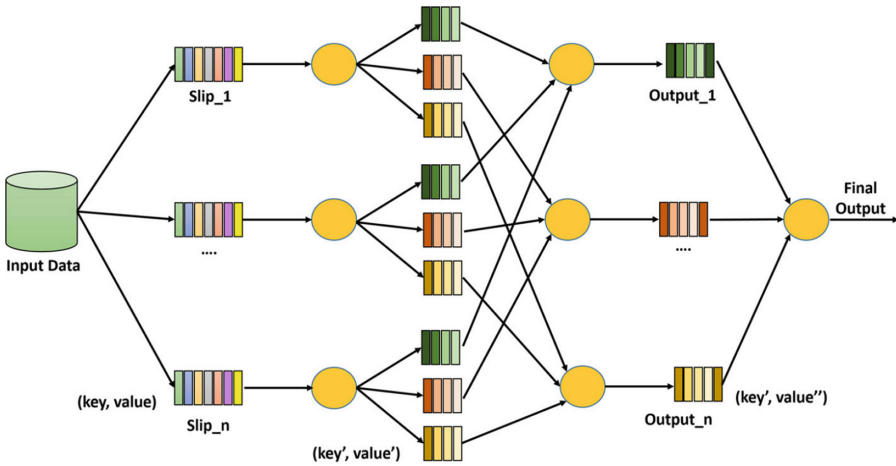


Fig. 1. The process of the MapReduce framework

3 Experimental Teaching Methods of Animal Pathology

3.1 Adopt Multimedia Teaching in Combination with the Characteristics of Animal Pathology

The fundamental goal of higher education is to cultivate college students into talents useful to society. This is not only the requirement of the country and society for higher education, but also the ardent hope of college students' parents. "That is, this function cannot be replaced by any institution." The content of the course of animal pathology is complex, and the current teaching model cannot meet the requirements of higher education. Therefore, it is imperative to reform the teaching of animal pathology. Through the reform of teaching content, teaching methods, and assessment models, students' learning interest and innovative thinking ability have been improved, making it easier for them to master boring knowledge.

Animal pathology focuses on understanding and mastering the essence, occurrence and development of diseases from the perspective of morphology. Multimedia teaching has the effect of image, text, sound and the combination of dynamic and static. It is clear and beautiful, vivid and informative. The formation, development and change of various diseases are vividly, vividly and concretely reproduced in the classroom by using the scene of sound and image synchronization and the picture of dynamic and static combination, forming a relaxed and pleasant teaching scene and atmosphere, which is conducive to stimulating students' interest in learning and enhancing the teaching effect, as shown in Fig. 2.

Using multimedia teaching can save the time of blackboard writing, so that teachers can provide students with as much information as possible in limited time and space; It can change the previous limited, one-sided and intermittent teaching mode, and show the occurrence and development process of a disease to students continuously, dynamically and comprehensively. It can also make up for the shortcomings that the microscope can only be observed by a single person, and it is not easy to communicate between teachers

teaching evaluation system, and the application results show that the division of weights conforms to the actual situation.

The numerical characteristics of rough sets represent the size of the set boundary region, but do not explain the structure of the boundary region; However, the topological characteristics of rough sets do not provide information about the size of boundary domains, but rather provide the structure of boundary domains. Therefore, in the practical application of rough sets, we need to combine the two types of information from the boundary domain, taking into account both the accuracy factor and the topological structure of the set.

Rough set theory also defines upper approximation and lower approximation for the classification of set classes on approximate spaces. Let $F = \{X_1, X_2, \dots, X_n\}$ be a classification or division of U that is independent of knowledge R , and the subset X_i is the class that divides F . The lower approximation and upper approximation of R for F are defined as:

$$\underline{R}F = \{\underline{R}X_1, \underline{R}X_2, \dots, \underline{R}X_n\} \quad (1)$$

$$\overline{R}F = \{\overline{R}X_1, \overline{R}X_2, \dots, \overline{R}X_n\} \quad (2)$$

There are two metrics to describe the imprecision of approximate classification. The first metric is the approximate classification accuracy based on knowledge R , F :

$$\alpha_R(F) = \frac{\sum_{i=1}^n |\underline{R}X_i|}{\sum_{i=1}^n |\overline{R}X_i|} \quad (3)$$

The second metric is the approximate classification quality based on knowledge R , F :

$$\gamma_R(F) = \frac{\sum_{i=1}^n |\underline{R}X_i|}{|U|} \quad (4)$$

The accuracy of approximate classification describes the percentage of correct decisions in possible decisions when using knowledge R to classify objects; The quality of classification represents the percentage of objects that can be accurately classified as F using knowledge R .

The reduce process then collects the output of the mapping task as input and combines these given key/value tuples into a smaller set of pairs to generate the final output. The MapReduce framework is shown in Fig. 1.

Comparing the concept of rough sets with ordinary set theory, it can be seen that the basic properties of rough sets, such as member relationships of elements, equivalence and inclusion of sets, are related to the knowledge of the universe represented by indistinguishable relationships. Therefore, whether an element belongs to a certain set is not an objective property of the element, but depends on our understanding of it; Similarly, the equality and inclusion of sets have no absolute significance, but depend on our understanding of the set in the problem under study.

From the current pattern and level of talent training in Higher Education in China, higher engineering and technical talents can be divided into engineering research talents, engineering application talents and engineering management talents. Applied undergraduate talents are engineering applied talents, mainly engineering technicians engaged in application research, technology application and development, and product trial production in a professional field in the front line of production [8].

The training objectives of applied undergraduate talents are: solid foundation, wide caliber, strong ability and emphasis on application. For application-oriented undergraduate talents, the most basic human qualities that best reflect the requirements of the times are as follows:

In terms of knowledge, we should not only have a broad range of knowledge, a solid mathematical foundation, strong theoretical skills and solid professional basic knowledge, but also have excellent applied knowledge.

In terms of ability, we should have strong research ability, practical ability, grasp flexible scientific methods, and have strong innovation ability. Have the ability to accept and apply new knowledge, and be able to apply professional knowledge and practical technology to comprehensively solve practical problems. Self learning and sustainable development.

In terms of quality, we should have a strong sense of responsibility, excellent psychological quality, strong will quality and good physical quality.

The connotation of the basic quality of Applied Undergraduate Talents provides a solid foundation for their future sustainable development.

2.2 Application Analysis of Rough Set Theory

Up to now, rough set theory has been widely used in various fields of data mining. Applying rough set theory to teaching evaluation is mainly aimed at the unreasonable weighting of teaching evaluation subjects. For example, there is no scientific basis, only through artificial determination or reference to the weight settings of other institutions, and the evaluation results obtained in this way often do not accurately reflect the teaching quality. At this stage, domestic experts have used data mining to conduct research on teaching evaluation. This kind of evaluation research usually requires a two-way approach, which involves assessing both the performance of students and the level of teachers, such as the quality of exam questions. The theoretical basis for evaluation is still teaching measurement and evaluation theory, and the technology used is data mining, while the method is the rough set algorithm of data mining technology. This algorithm is a new mathematical theory for analyzing incomplete and uncertain data. It is mainly used in teaching evaluation to make the data become the subject with decision-making power. In teaching evaluation, if the combination of clustering analysis and rough set theory is used, relatively better evaluation results can be mined. Because the improved methods (hard clustering methods) in the field of clustering analysis can more objectively adapt to practical applications, they have become a hot spot in clustering research. At the same time, based on the fuzzy C-means clustering method, clustering analysis is conducted for the data in teaching evaluation, and then the different weight coefficients of various factors in teaching evaluation are mined using rough sets. It can be seen that the combination of these two technologies provides a reliable theoretical basis for the

be well implemented through experimental classes (Wang Lu 2013). Students' learning evaluation is an indispensable part of the process of education and teaching. The general university biology curriculum standard (Experiment) points out that experimental ability is an important part of learning evaluation (Ministry of education 2003). However, at present, the evaluation of biological experiment courses is often single, mostly through paper and pencil tests or experimental reports, and a set of operable evaluation system has not been established [2]. Changing the traditional way of single evaluation of students, looking for an evaluation method that can improve students' learning interest and experimental results, and improve students' scientific literacy, is one of the important problems faced by the reform of basic education in China.

Animal pathology is a natural science based on experiments. Interest is the guide of learning, and biological science literacy is the core content of the new curriculum standard, which requires attention to the development of students. Therefore, the learning evaluation of biological experiment course should be based on improving students' learning interest and promoting the formation of biological science literacy, and the goal of teaching evaluation should establish an evaluation and examination system consistent with the concept of quality education. In experimental teaching, the purpose of evaluating students is not to grade students, but to better understand students and promote their development (Wang Lu 2013) [3]. The author's research on the diversified learning evaluation system of high school biological experiment class is to explore the mode of combining multiple subjects, multiple contents and multiple ways, improve the existing learning evaluation system of animal pathology experiment class, and provide reference for teachers' biological experiment class students' learning evaluation. At the same time, it is hoped that through this research, teachers can pay more attention to the learning evaluation of animal pathology experimental courses, promote the development of students with diversified learning evaluation methods, improve students' scientific literacy, and cultivate students' lifelong development ability [4].

2 Related Work

2.1 Educational Purpose of Applied Undergraduate Colleges

Educational purpose is the core of the whole educational and teaching activities. Educational purpose is the starting point and basis of educational and teaching activities, and it is also the final destination of school educational and teaching activities. It plays a guiding role in the determination of educational tasks, the establishment of educational systems, the selection of educational content and the organization of all educational activities. Therefore, clarifying the educational purpose of Applied Undergraduate Colleges and universities is the primary problem that colleges and universities should solve.

Application-oriented undergraduate not only has the generality of undergraduate education, but also pays more attention to practicality, application and technology. Its educational purpose is to cultivate applied talents suitable for social needs and high-level applied technical talents mainly composed of front-line engineers. These talents have a solid theoretical foundation, a wide range of professional knowledge, strong practical ability, high comprehensive quality, and strong ability to apply, promote and convert science and technology.



Exploration and Research on the Diversified Reform of Experimental Teaching Methods of Animal Pathology in Applied Undergraduate Colleges

Jing Li^{1,2(✉)}, Jiang Yurui^{1,2}, and Niqin Jing^{1,2}

¹ JiLin Agricultural Science and Technology University, Jilin City 132109, China
shengdiyu2009@163.com

² Department of Electronic Technology, Beijing Polytechnic, Beijing 100176, China

Abstract. Taking the implementation of the training of applied talents as the starting point, we should build a reform plan of animal pathology teaching curriculum that integrates production and education and pays attention to practice. By using the “production, teaching and research” base, new media auxiliary technology and other means, we should adopt the methods of “promoting learning through competition”, “double teachers and the same class”, and the main purpose is to improve the teaching methods of animal pathology and improve the knowledge level of students. The project will be implemented by creating an experimental teaching laboratory, which will allow teachers to use a variety of different models to create their own courses to test new ideas and technologies. The goal is to develop a more effective teaching method of animal pathology, which is applicable to all levels from preschool to university, and to provide a facility for researchers to study new technologies and tools.

Keywords: Applied undergraduate colleges · Diversification · Animal pathology · Experimental teaching

1 Introduction

The 21st century is the century of animal pathology. Life science plays an important role in many fields. Whoever masters life science will master everything. The university stage is the most important stage in students' life. It is the key period to form a scientific world outlook, outlook on life and values. It is also a period of rapid development of scientific literacy, which is the core content of the proposed curriculum standards. Animal pathology is an experimental science. Improving experimental ability helps to improve students' scientific literacy. The core content of the biology curriculum standard formulated by the Ministry of education in 2003 is to improve each student's biological science literacy (Dong Jing 2006) [1]. The curriculum concept advocated therein - facing all students, improving biological science literacy, and advocating inquiry learning can

8. Zhu, L., Zhang, S.: Study on university information management in the era of big data. *J. Changzhou College Inf. Technol.* (2017)
9. FANXinke. Study on application practice of intelligent control technology of coal mine electromechanical system. (7), 4 (2022)
10. Liu, H., Zhang, et al.: Application of data mining technology in the analysis and management system of scientific research achievements in universities. *Basic Clin. Pharmacol. Toxicol.* (2016)
11. Zhou, W., Yang, T.: Application analysis of data mining technology in ideological and political education management. *J. Phys.: Conf. Ser.* **1915**(4), 042040 (7pp) (2021)
12. Duan, Y., Huang, X.: Study on influence and application of data mining technology in university physical education. *Revista de la Facultad de Ingenieria* **32**(11), 440–445 (2017)
13. Li, J., Chen, H., Ren, F.: Application of data mining clustering technology in college and university political education optimization. *Boletin Tecnico/Tech. Bull.* **55**(8), 495–502 (2017)
14. Zhang, J., Jia, T.D., Fang, X.: Application of data mining in MOOCs for developing vocational education: a review and future research directions. *Int. J. Inf Educ. Technol.* **8**(6), 411–417 (2018)
15. Liang, B.: Application of network data mining technology in analysing the political education reform and development for higher vocational students. *Boletin Tecnico/Tech. Bull.* **55**(18), 114–120 (2017)
16. Chong, S., Mak, M., Loh, W.M.: Data-mining applications with the admission data of adult learners in higher education: a pilot study. *Int. J. Manag. Educ.* **10**(2), 131 (2016)
17. Liu, R., Zhu, Q.: Research on the implementation of data mining application in multimedia physical education and track and field teaching based on network data. *Boletin Tecnico/Tech. Bull.* **10**(4), 19–28 (2017)
18. Chen, X., Wang, Z.A.: Research on the higher education management system based on the data mining technology. *Basic Clin. Pharmacol. Toxicol.* **S2**, 125 (2019)
19. Sun, H.: Study on application of data mining technology in university computer network educational administration management system. *J. Intell. Fuzzy Syst. Appl. Eng. Technol.* **37**(3aPt1) (2019)
20. Ma, J.: Research on the application of data mining technology in educational management system. (8) (2020)

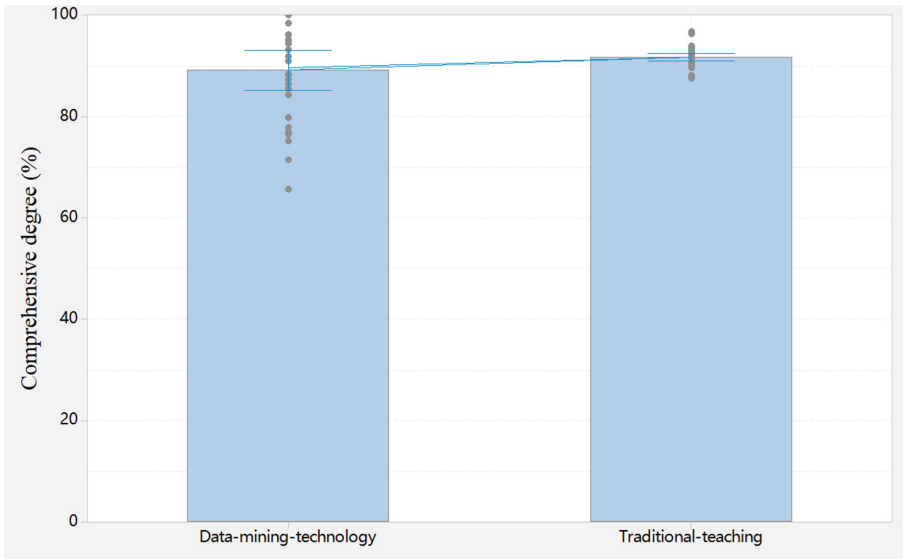


Fig. 4. Comprehensive management results of data mining technology

5 Conclusion

Under the conditions of civil aviation transportation, in view of the analysis situation, this paper proposes a data mining technology, and combines teaching data to optimize teaching management and manage data Conduct in-depth analysis. Studies show that data mining technology can improve the accuracy and stability of analysis and synthesize teaching content Analysis. However, in the process of data mining technical analysis, too much attention is paid to the analysis of management indicators, resulting in a relative decline in analysis accuracy.

References

1. Zeng, X., Li, S., Shi, T.: The Application of Data Mining Technology in the Education Management System (2018)
2. Zhao, J., Wang, J.: A design of college ideological and political education management system based on data mining technology. *Revista de la Facultad de Ingenieria* **32**(14), 73–78 (2017)
3. Dong, X., Huang, X., Lin, M.: Application of data mining technology in public welfare sports education in the era of artificial intelligence. Hindawi Limited (2021)
4. Teng, J., Polytechnic, B.: Application of data mining techniques in vocational education information system. Office Inf. (2019)
5. Bhoite, S., Pathare, S.: Students' academic performance analysis using data mining techniques in higher education: a case study of college (2022)
6. Yang, H., Zhang, W.: Data mining in college student education management information system. *Int. J. Embedded Syst.* **3**, 15 (2022)
7. Kovalchuk, Y.O.: Data Mining in Education: Current State and Perspectives of Development (2016)

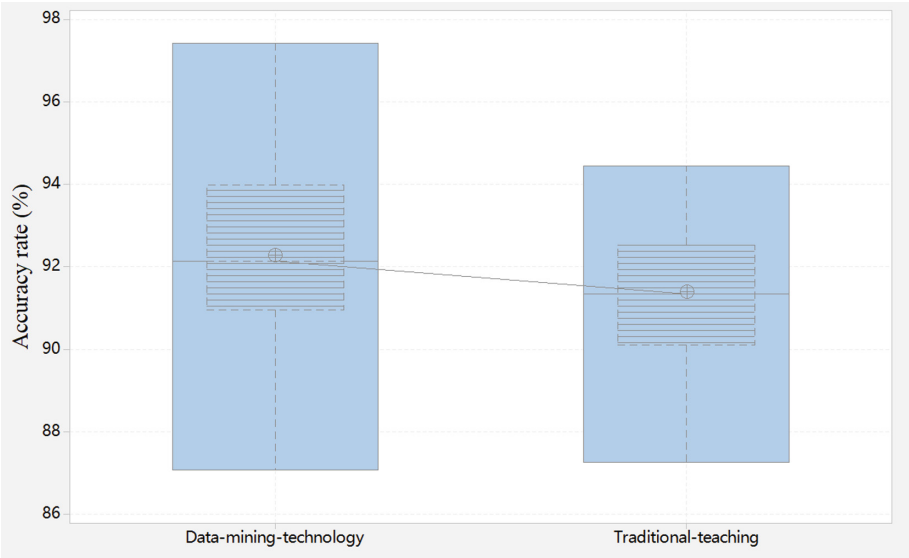


Fig. 3. The accuracy of different algorithms

Table 3. Comparison of analysis accuracy of different methods

Algorithm	Precision	Magnitude of change	Error
Data mining techniques	96.79	0.36	8.21
Traditional teaching management	85.71	1.07	1.43
P	28.57	50.00	58.93

Table 3 shows that the traditional teaching management method has shortcomings in accuracy and stability in terms of analysis, and the education plan has changed significantly, and the error rate is high [19]. The comprehensive results of data mining technology have high accuracy, which is better than traditional teaching management methods. At the same time, the accuracy of data mining technology is greater than 90%, and the accuracy has not changed significantly [20]. To further verify the superiority of data mining techniques. In order to further verify the effectiveness of the proposed method, different methods are used to analyze the accounting data mining technology comprehensively, and the result 4 is shown.

It can be seen from Fig. 4 that the management data of data mining technology is significantly better than the traditional teaching management method, and the reason is that data mining technology increases the teaching management coefficient and sets the corresponding standard. Eliminate non-compliant management data.

Table 2. Overall situation of teaching management

Time period	Teaching data	Manage data
1–4 weeks	18.95	16.84
5–8 weeks	21.05	16.84
8–12 weeks	18.95	8.42
Mean	20.00	20.00
χ^2	21.05	14.74
P = 0. 031		

integrate the NPC data and related technologies, use its functions to accurately formulate management objectives, accurately select management methods, accurately predict the management process and accurately evaluate the management results, and achieve the set objectives. Therefore, the accuracy of the precise management of college students' education in the era of big data is mainly reflected in four aspects: first, in terms of management objectives, Big data can accurately set ideological and political management objectives [16]. The long-term goal of college students' education and management is to train qualified builders and successors of the socialist future through a series of management activities, but the setting of medium and short-term goals is slightly insufficient. The main reason is that it is difficult to understand the students' interests and ideas, so it is difficult to formulate specific goals. Big data provides a certain method to solve this problem, which can analyze students' thoughts and trends on the basis of collecting a large amount of information, reasonably formulate management goals, and achieve accurate management. Secondly, in the selection of management methods, Be able to use big data to test the use effect of different management methods, and provide reference for the selection of future methods. In terms of management process prediction, big data can obtain reliable process prediction based on information capture [17]. For example, when conducting relevant education management activities, it can predict the problems faced in the process of this management activity, students' acceptance and other aspects based on the content data of similar management activities in the past. Finally, in terms of management effect, big data can achieve accurate evaluation. Human analysis of feedback results has certain subjectivity and may miss important information content [18]. Big data technology can record the information of management process, which is convenient for viewing relevant information on the one hand, and also can visualize the information of education management process on the other hand, providing visual information resources for evaluating the effect of education management, At the same time, it has also accumulated information resources for the future education and management of college students.

It can be seen from Fig. 3 that the accuracy of data mining technology is higher than that of traditional teaching management methods, but the error rate is lower, indicating that the analysis of data mining technology is relatively stable, while the analysis of traditional teaching management methods Uneven. The average management data of the above three algorithms is shown in Table 3.

Table 1. Related parameters

Teaching content	Time period	Manage effects	Digging effect
Course teaching	1–8 weeks	21.05	14.74
	8–12 weeks	7.37	17.89
Specialized teaching	1–8 weeks	8.42	16.84
	8–12 weeks	9.47	9.47
Teacher management	1–8 weeks	16.84	7.37
	8–12 weeks	12.63	11.58

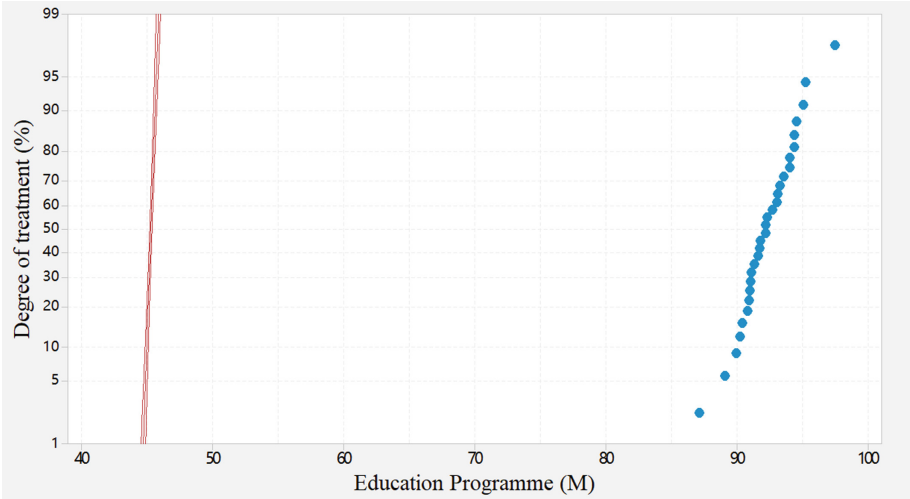


Fig. 2. Processing process of education management data

4.2 Management Situation

Management data includes unstructured information, semi-structured information, and structural information. After the pre-selection of data mining technology [15], the preliminary teaching management is obtained, and the feasibility of the accounting result management data is analyzed. In order to verify the effect of the analysis more accurately, select education programs with different time periods, as shown in Table 2.

4.3 Accuracy and Stability of Teaching Management

In order to verify the accuracy of the data mining technique, the management data is compared with the traditional teaching management method, and the management data is shown in Fig. 3.

The precise management of college students’ education in the era of big data mainly means that in the process of college students’ education management, we should fully

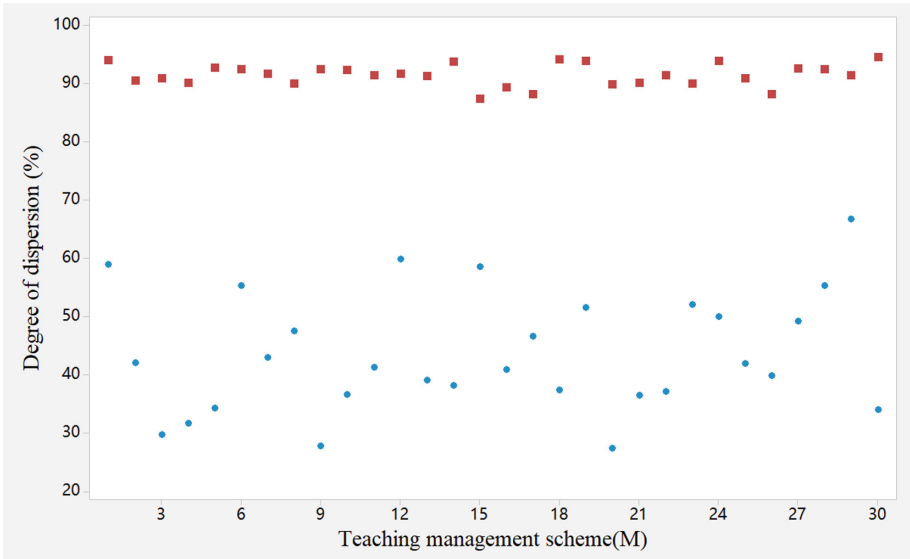


Fig. 1. Results of the selection of instructional management programs

3 Strategies for Instructional Management

Data mining technology adopts a random strategy for analysis and adjusts the corresponding parameters to optimize the analysis scheme. Data mining techniques divide the analysis into different periods and randomly select different strategies. In the iterative process, educational scenarios for different time periods are matched. After the matching process is completed, the level of teaching management analysis of different programs is compared to record the best teaching management.

4 Practical Cases of Teaching Management

4.1 Introduction to Teaching Management Data

In order to facilitate the analysis, the management data in complex cases are the research object, 12 are accounted for, and the test time is 12 weeks, and the specific analysis management data is shown in Table 1.

The processing of instructional administration data in Table 1 is shown in Fig. 2.

Table 1 shows that compared with the traditional teaching management method, data mining technology teaching management is closer to the actual requirements. It is better than the traditional teaching management method in terms of the rationality and fluctuation range of the selection of analysis indicators [14]. From the changes in the management data in Fig. 4, it can be seen that the stability of data mining technology is better and the analysis speed is faster. As a result, data mining techniques are more computationally faster, more accurate, and more stable.

example, the main body of colleges and universities should integrate resources, inject human resources, material resources and other resources into college students' education and management activities, and help to achieve college students' education goals. On the other hand, we should give full play to the role of the state, society and universities. For example, the coordination between the state, society and universities should be strengthened [10]. The state should make good policy deployment and planning, and the society should create a good management environment. Universities should systematically manage on the basis of responding to the national policies, and form a management situation in which the three parties work together. We should integrate all forces to play the management effectiveness, achieve the management objectives, and help the precise management of college students' education.

Third, dynamic innovation. On the one hand, college students' thinking is constantly changing. As the most active group, college students' thinking consciousness is prone to change, which requires that education management should analyze college students' thinking from a dynamic perspective, and reasonably repair the management plan to ensure the dynamic effectiveness of management. On the other hand, the management environment changes dynamically. There are many factors involved in college students' education management, such as management subject, management object, information, time, and so on [11]. The change of any element will cause the overall change. To sum up, college students' education management is dynamic. In the process of grasping the dynamic, the management plan is constantly adjusted, and new management concepts and management technologies are needed in the face of change. The innovation of management is also constantly highlighted.

2.3 Management of the Processing of Data

Before the data mining technology, it is necessary to conduct discrete analysis of teaching management, map the management data into a two-dimensional plane, and eliminate abnormal management data. First, the management data is comprehensively analyzed, and the standards and index weights of the management data are set to ensure the accuracy of accounting data mining technology. Management data is system test management data and needs to be standardized. If the management data is properly distributed, its teaching management will be affected, reducing the accuracy of the analysis [12]. In order to improve the accuracy of accounting data mining technology and improve the level of teaching management analysis, the specific prediction scheme selection is shown in Fig. 1.

The survey management data shows that the teaching management scheme shows a discrete distribution, which is in line with the objective facts. The education program is not directional, indicating that the teaching management program has a strong randomness, so it is used as an analysis study of 8–12 weeks [13]. The education scheme meets the normal requirements, mainly because the data mining technology adjusts the education program, eliminates the duplicate and irrelevant schemes, and supplements the default scheme, so that the dynamic correlation of the entire management data is strong.

After data import, data preprocessing is required. The data preprocessing is also the most time-consuming part of the whole system, because some data needs to be converted from the original rules to data that conforms to the system's calculation rules. This step of preprocessing requires a lot of time and relatively high-end hardware support, especially in the face of a large number of student data.

After the data is imported, the data can be preprocessed. The course score of the imported data is a hundred-point system, so the range of clustering will be large in the later data processing and association rule processing [8]. Therefore, the hundred-point system will be converted into a ten-point system. The decimal system represents the whole digit of the student's achievement, and the case when then code is selected for implementation. This classification can make the students' clustering rules more representative, and effectively reduce the amount of data engineering in the process of building the clustering model.

2.2 Choice of Educational Programs

Hypothesis 2: The scheme selection function is $z(d_i)$ and the weight coefficient is w_i [7], then the teaching scheme selection is shown in Eq. (2).

$$z(d_i) = \frac{z_i \cdot F(d_i, y_i)}{w_i \cdot \xi} + \quad (2)$$

Education management refers to the general term for the planned implementation of management activities for the educational objects under the guidance of certain educational objectives, following the management principles. Education and education management are different from each other, and are related to each other. The two are the relationship between generality and individuality. Education emphasizes the overall planning of individual people's thoughts, while education management highlights the practical approach in the process [9]. College student education management refers to the general term of various practical activities that focus on the advanced group of college students, and on the basis of insight into the ideological dynamics of college students, educators carry out in a planned way to promote the quality of college students according to their ideological characteristics. College student education management has the following three characteristics.

First, political orientation. The setting of management objectives needs to be guided by the correct political direction; the formulation of management plans needs to be guided by the correct political direction; The practice of management needs to be carried out closely around the correct political direction and unswervingly adhere to the correct political direction. In short, sticking to the correct political direction in college students' education and management is to strengthen the ideological status of socialist mainstream ideology in college students.

Second, strength integration. The realization of management objectives requires the integration of management forces. It is difficult to achieve the goal of college students' education and management only by relying on the role of one party. It is necessary to integrate the forces of all parties and give full play to the joint forces of education. On the one hand, we should give full play to the power of colleges and universities. For

in the learning process can also be used to effectively query the teacher's online understanding of students' mastery of different knowledge points by using big data technology. How long did it take to do the problem? Which knowledge points need to be explained repeatedly? Applying big data technology to daily teaching activities can effectively analyze students' individual learning behavior, and also discover students' thoughts and dynamic changes. Using electronic textbooks or online learning platforms to understand students' homework, teacher-student interaction, classroom communication and other data, and collecting all kinds of information data at the end of the term can add rich materials and data basis, and can also effectively grasp students' learning and growth process and characteristics so as to provide reference data for students' future growth and development.

Teaching management analysis is one of the important management contents of college teaching, which is of great significance to teaching management. However, in the actual teaching management process, there is a problem of poor accuracy, which has a certain impact on the management of colleges and universities. Some scholars believe that the application of data mining technology to teaching management can effectively analyze teaching management [4]. On this basis, this paper proposes to modify the genetic method to optimize teaching management accuracy and verify the method's effectiveness.

2 Related Concepts

2.1 Mathematical Description of Data Mining Techniques

Data mining technology is to use dynamic theory to optimize management data, and according to the indicators in the management data [5], find outliers in teaching management results, integrate the corresponding management data [6], and finally judge teaching management Feasibility of the result. The modified genetic method combines the advantages of data mining technology, using management data to quantify the results of teaching management, which can improve the accuracy of analysis.

Hypothesis 1: The management data is d_i , the teaching set is set_i , the teaching indicator is y_i , and the education function is $F(d_i \geq 0)$ as shown in Eq. (1).

$$F(d_i) = \sum x_i \cup y_i \cdot \xi \quad (1)$$

Although the content of imported information is extremely rich after data collection, some of it is useless information, such as whether to take an optional course, class schedule, etc., so you can eliminate this kind of garbage and useless information after importing. After data collection, the main work is to import data into the system for calculation, which is the process of system database establishment. In the process of importing student data, it is necessary to ensure that all student data are truly and accurately imported into the system [7]. Although some of the student's information is useless information, after the subsequent clustering rules and screening logic criteria are changed, some of the useless information can also be used as important parameters for clustering analysis. Here, because the system is in the research stage, some of the information is eliminated.

When will you finish the exam? Through data analysis, it is helpful to understand the students' recent learning situation and their way of thinking. This score is far more valuable than the score of the paper test paper. Whether in the classroom or in the exam, infiltration of big data through teacher-student interaction can help improve classroom teaching efficiency and also update teachers' educational concepts. In the traditional education process, it is mainly teachers and education authorities who effectively extend the classroom content with their teaching materials based on teaching experience, but some learning experience is not scientific, and even some common sense of experience will affect teachers' judgment. In the era of big data, the analysis of education data helps to mine the writing materials and learning content, effectively mine the teaching information according to the actual learning situation of students and the content of teaching materials, and formulate targeted teaching plans through feedback evaluation, so as to achieve targeted teaching and effectively improve the quality and efficiency of education and teaching.

Big data can realize personalized education, from group education to individual education. Using big data technology, teachers can focus on understanding the subtle performance of students, such as when students will read books actively, when they will feel happy to hear what they say, and when they will make mistakes in class... The generation of data is process, including students' autonomous learning process, classroom teaching process, and teacher-student interaction process. Through the reflection and recording of immediate sexual behavior, the effective integration of data is strengthened to further elaborate the learning status and performance level of students in the teaching process. The observation technology and equipment can be used to collect data without affecting students' daily life and learning. The data collection is very natural and real, and the real performance of students can also be obtained. Big data technology can provide teachers with real personality and student characteristics [4]. Teachers can also carry out teaching according to their aptitude in a targeted way. In the course of course explanation, students can focus on the part that students pay attention to and the basic part. Students can also be recommended to complete the exercises in which areas or read which kinds of books. After completing the homework assigned by the teacher, students can use big data information technology to effectively strengthen the analysis. When students do homework on electronic equipment after class, they can skip similar topics. If similar topics make mistakes, they can carry out multiple reinforcement to improve the efficiency of autonomous learning, and also enhance the application efficiency of fragmented time to reduce the learning pressure of students.

The use of big data technology in teaching evaluation can enrich the evaluation content, methods and channels and help improve the efficiency of teaching activities. In the traditional teaching evaluation process, the teacher only based on a paper or ordinary experience. This evaluation method is too simple, and it is difficult to improve the enthusiasm of students. Students can only passively wait for the evaluation. In the era of big data, teachers can use big data technology to summarize students' daily learning data, find out the rules of students' learning activities, and further optimize and innovate teaching content [3]. Through the Internet online platform, students' learning behavior records and learning guidance can be strengthened. The mouse click track of students



The Application Study of Data Mining Technology in Education Management

Chengyi Niu^(✉) and Zhang Jie

Student Affairs Office, Criminal Investigation Police, University of China, Shenyang 110854,
Liaoning, China
lcy_666888@163.com

Abstract. The role of data mining technology in teaching management is very important, but there is a problem of low management level. In the past, the teaching management method could not solve the problem of management level in teaching management, and there were few management level indicators. Therefore, this paper proposes a data mining technique to construct a teaching management model. Firstly, data mining technology is used to analyze the management data of teaching management and manage the data according to teaching standards. Divide and reduce interference factors in the management level. Then, the data mining technology manages the teaching management content, forms the teaching management results, and continuously analyzes the management data. MATLAB simulation shows that under the condition of certain teaching management requirements, the management accuracy and analysis time of data mining technology are better than before Teaching Management Method.

Keywords: intelligence · technology · teaching management · Analysis

1 Introduction

In the era of big data, education big data has reversed the traditional backward education concept and thinking mode. In the new era, the field of education is full of information and data. The words and deeds of teachers and students and all kinds of things in schools can be transformed into information or data. With the widespread popularity of intelligent equipment, every student can use computers for terminal learning, which can help improve students' learning enthusiasm and also help students use fragmented time for efficient learning [1]. It can be seen that big data is more meaningful and valuable than traditional digital teaching. For example, in the traditional examination, the paper test paper is used to evaluate the students' learning situation in the recent stage. The score of the test paper is only a simple traditional number. If you change the way of thinking and treat the number as data, you can find out all kinds of information hidden behind the number. For example, which option did the students choose for the score of each question? How long did you spend on each option? Have you modified the answer? Are there any jumps when doing questions? When will the inspection be carried out?

2. Paliy, I.: Probability Theory and Mathematical Statistics (2021)
3. Gao, J., Zhi, L., Sun, J., et al.: Research on the Application of Blended Teaching in Probability Theory and Mathematical Statistics Based on MOOC + SPOC + Flipped Classroom, vol. 6. Science Publishing Group (2021)
4. Sheynin, O.: Theory of probability and statistics as exemplified in short dictums (2021)
5. Hu, Y.: Fire risk assessment of urban utility tunnels based on improved cloud model and evidence theory. *Appl. Sci.* **13** (2023)
6. Jonis, M., Monnier, M.C., Schmid, O.: Analysis of regulatory framework and standards applied to organic wine-making in Europe (2022)
7. Lu, L., Zhou, J.: Research on mining of applied mathematics educational resources based on edge computing and data stream classification. *Mob. Inf. Syst.* **2021**(7), 1–8 (2021)
8. Schmidt, J.H., Mccann, C.: ESG challenges in the construction of UK balanced portfolios for private investors: an analysis of the availability and performance of ESG funds across various asset classes. *J. Appl. Financ. Banking* **12** (2022)
9. Ummarino, G.A.: Mathematical and physical properties of three-band s^+ -eliasberg theory for iron pnictides (2023)
10. Xuan, D., Di, Z., Yikai, C., et al.: Characteristic mode analysis: application to electromagnetic radiation, scattering, and coupling problems. *Chin. J. Electron.* **33**, 1–14 (2023)

of samples is too small or too large, it will lead to a large number of computational costs. Moreover, there are irregularities such as random error, noise and outliers in real data processing. Therefore, how to find low probability events with typical significance in massive data is a very important problem. Because of the real data test, we assume that there are enough small probability events, because there are very few samples, but when the number of data reaches some extent, these data will change. This complex problem cannot be well solved by traditional statistical analysis methods. In addition, due to the increasing amount of data, the traditional mathematical statistical analysis method can no longer meet the needs of large capacity, and the traditional mathematical statistical analysis method also has many deficiencies, such as the need for manual conduct. However, if the above problems are not solved well, it is difficult to deal with a large amount of data effectively. In this way, statistical methods can be better adapted to changes in the data, thus improving the efficiency of data processing.

If Chinese enterprises want to improve the level of economic development in the era of big data, they must make good use of existing database technology. Although the complicated data information is obscure and difficult to understand, if you can understand the knowledge content hidden behind the data information, it will help enterprises make correct choices in the future, avoid possible risks, and promote the overall improvement of the national economic level. In the face of massive data information, data mining technology is born. In the early days, people developed data mining technology and focused on analyzing data through artificial intelligence. With the continuous exploration of data mining technology, people gradually shifted the focus of data mining technology from artificial intelligence to probabilistic statistical methods. Only by combining probability and statistics with data mining technology can we accurately control the future trend in the era of big data. First of all, data mining and statistical methods should be organically combined to promote the progress of statistics, so that statistics can still play the role of induction and data collation in massive data statistics. Secondly, the data mining work should be monitored in real time, and adjusted quickly when the data mining work is abnormal to ensure the accuracy of the data results. Finally, we should establish a visual data mining prototype system to promote the development of statistics.

5 Conclusion

In short, data mining, as an emerging technology, is increasingly expanding, especially in the field of economic management. Data mining refers to extracting useful knowledge and rules from massive amounts of data and performing a series of operations to draw the expected conclusions. The application of mathematical statistics and probability theory in data mining plays a great role in promoting the development of data mining, especially in data processing and data analysis. By learning statistics and probability theory, we can have a relatively complete understanding of the characteristics and change laws of the data.

References

1. Discussion of three paradoxes in the teaching of probability theory and mathematical statistics. *Creative Educ. Stud.* 09(3), 617–620 (2021)

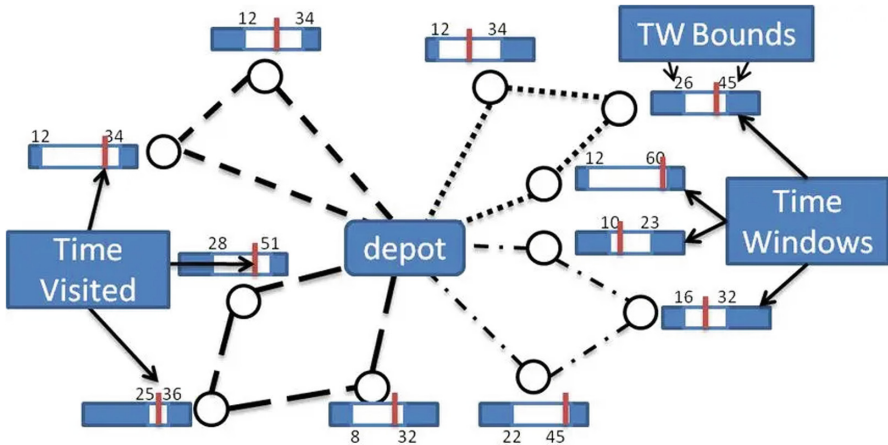


Fig. 2. Link lock

data. Therefore, the staff should always pay attention to the calculation process during the implementation of genetic algorithm, and select excellent data to replace the data reorganization during the data reorganization to ensure the smooth completion of genetic algorithm. The introduction of genetic algorithms in data mining can help staff solve a large number of problems, effectively deal with massive and complex data, and obtain the general rules of data. In the actual operation process, staff can also use compressed genetic algorithm to assist in solving, which can reduce the difficulty of calculation to a certain extent and improve the speed of genetic calculation. Nevertheless, there are still many problems in the process of probabilistic evolutionary computation, such as the selection requirements and selection proportion of data at the initial stage of genetic algorithm, which are important issues that need to be solved urgently.

4 Provide New Research Directions

The development of mathematical statistics and probability theory often comes from the processing of actual data and the full use of human needs to explore the changes between massive data. Therefore, extracting useful knowledge from a large amount of data and turning it into available information has become an important concern in the current statistical community. As an emerging technology, data mining, its application scope is expanding, especially in the operation of enterprises. Therefore, based on the real situation, through the analysis of a large number of data, it is found that the current statistical method in China is not perfect, that is, how to determine the data allocation mode and sampling size. The above problem is an important problem facing the current statistical community. This is because when doing data processing, the classification according to the “sample” and “overall” will constantly change to meet the needs of data processing. Due to the large amount of data, it is difficult to accurately define the sample and the whole, and the data change is also diversified; is the gradual characteristics of the data consistent with the advance prediction. At the same time, because the number

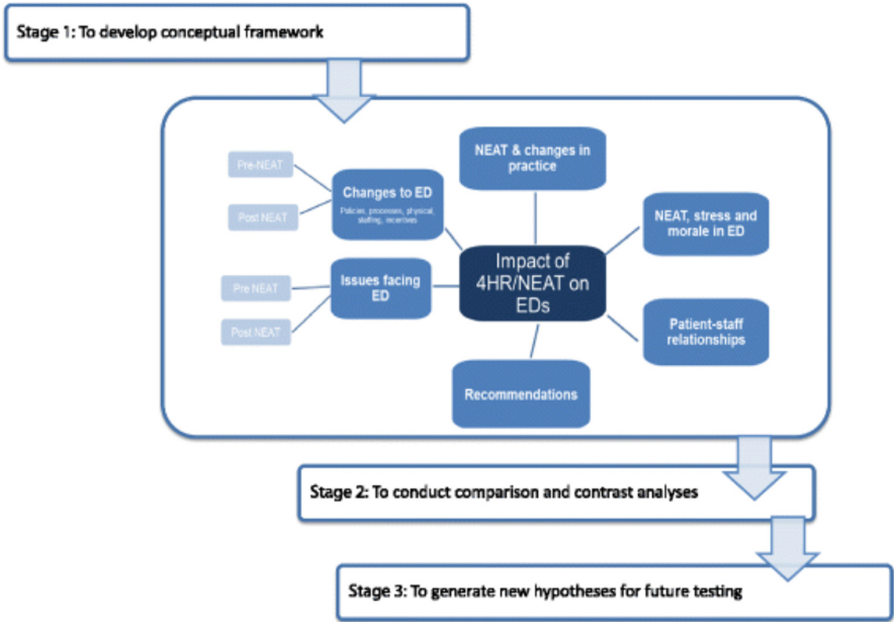


Fig. 1. Application of Bayesian network in data mining

large problems. The proposed algorithm can construct the corresponding classification mode according to the different decision tables. The method is based on the basic ideas and basic principles of PMEA, and it is used for stochastic evolution, by expanding the method through its analysis to reduce the operational complexity and make it more efficient. The experimental results show that this algorithm is very effective. Probabilistic evolution method is widely used in various industries because of its good convergence performance and good robustness. It can quickly find the hidden laws and laws, and analyze some inaccurate or uncertain information; so its application prospect is very broad. However, in PMEA, import the predicted information into PMEA to solve the difficulties encountered in the learning process; and enhance the efficiency and accuracy of data mining. Therefore, we need to construct a precise, reasonable mathematical model. Which are the most critical questions in the evolution of possibilities. Traditional stochastic evolution algorithms do not solve these problems very well, and therefore do not yet serve as a general, universal collection of data available for multiple complex systems. This paper conducts an in-depth study on this issue. On this basis, combining the stochastic evolution methods, such as: population size, selection mechanism, ratio, and other problems, the current probabilistic evolution methods are the main [4].

Compared with other statistical methods, the accuracy of statistical results obtained by genetic algorithm will be higher. However, because genetic algorithm is a interlinked process, once there is a mistake in a calculation link, it will lead to a deviation in the final statistical results. In the actual operation process of the staff, the data block is often lost due to data reorganization, and the local data block is abnormal, resulting in serious deviation of the data result, which can not effectively summarize a large amount of

methods have good robustness and good convergence, so they have attracted wide attention from domestic and foreign scholars. I often encounter some problems in practice. For example, whether the sample data has a certain reliability and uncertainty, or there is some unknown information, and then using the Bayesian method for abnormal analysis, usually use a close way to analyze the internal connection of each data; to realize the processing of data.

The work steps of data mining are divided into the following steps: First, the staff should clean up the massive data, eliminate some irrelevant data, and retain valuable data for analysis. Secondly, it is necessary to analyze, summarize and summarize the retained data, and transform the massive data through statistical means to reveal the universal laws of data. Finally, the transformed data results should be represented by statistical charts or other forms of expression to make their images easy to understand and understand. It is worth noting that the staff should finally analyze the data results to judge the accuracy of the results. In the whole data mining work, we need to use neural network technology, Bayesian algorithm and other technologies. Through these statistical technologies, we can help staff quickly obtain the universal rules between data. Bayesian network plays an important role in the whole development of statistics [10]. With the passage of time and the progress of theoretical and technical level, Bayesian network has gradually become a calculation method of uncertain knowledge and data reasoning, which is inextricably linked with probability theory and mathematical statistics. People need to use Bayesian network in the process of studying machine learning. By using Bayesian network, staff can summarize and analyze massive data, get the general rules between massive data, and promote the rapid progress of machine learning research, as shown in Fig. 2. In addition, the data disorder diagram in Bayesian network solves the problem of complex and obscure data statistics results. By combining the data statistics results with the probability diagram model, the staff can easily obtain the rule information of the data results. It is worth noting that in the process of using Bayesian networks, approximate methods are generally used to analyze data and obtain the general rules of data (Fig. 1).

(3) Data mining based on the probabilistic evolution algorithm

Genetic Analysis (GA) is a commonly used algorithm in data mining, which is also commonly used when processing some data. The groups were then divided into offspring of different sizes. With variation as assistance, data reconstruction and other optimization, so that GA can establish modules in a large number of nonlinear data, data processing, after processing, after screening, recombination, gene composition; before completing the data structure optimization, get the best data, get the best data, get the best results. Using this method, the redundancy of multiple data groups can be reduced, and the data processing accuracy can be improved. However, in practice, because of the data recombination, it often leads to the data fragmentation, so that the routine algorithm can not effectively process the data. In the case of data change, in order to retain the original information, it must be redesigned to avoid the original knowledge loss, resulting in a reduced operational efficiency of the system. This type of data destruction is called a link locking problem (as shown in Fig. 2). The paper makes the traditional changes to the random distribution to make it closer to the real situation. Based on this idea, this paper gives a new solution idea, and uses it in data mining, so as to obtain a number of

ANN is a widely used data mining algorithm that is composed of a series of networks called data nodes. The processing method of input, output, numerical value and other nonlinear data is adopted to adjust the positioning of each node, so as to realize the data analysis. Data mining requires a large amount of qualitative data, quantitative data and other data, which are data lost, so it is necessary to combine data mining and data mining technology to mine the rules behind the data. Based on this, the application of probabilistic statistical network in data mining is deeply discussed. Probabilistic analysis network is a data analysis model based on mathematical statistics. Its basic principle is to allocate the association between the sample points, so as to obtain the overall trend change. Probabilistic analysis networks are a general framework for statistical inference in a variety of complex systems. Probabilistic networks are used for data mining, including data pattern recognition, optimization problems in non-linear regression, and other data utilization and processing. Probabilistic analysis method is a new method that can accurately judge the different data distributions in the existing situation, so as to draw uncertain conclusions. This method also uses the traditional weight-threshold neural network (BP) as a new weight-threshold network, with a learning speed of more than 100 times faster than BP, and with higher accuracy. At the same time, the association of the original data is probability to achieve the purpose of prediction. This fully proves that in a sense, in a sense, the computational rate is faster than the weight-threshold network. Meanwhile, the model can be predicted based on the input information, so the model has high accuracy and high reliability. In practice, because the characteristics of the network are the specific structure of its nodes and the high randomness, it needs to be emphasized in practice. The Markov Chain Table (Markov) is an algorithm that can synthesize and quantify a large number of nonlinear data comprehensively, and deeply discusses the transmission probability and convergence between different states. This method focuses on solving the optimal parameters of the network model [9]. By studying the relationship between various data, and analyzing the transfer matrix in the uncertain data network, we conclude that the average convergence and complement, and the evolution trend of the data is studied.

(2) Application of Bayesian networks in data mining

A series of work on data mining, data cleaning, transformation and visualization was conducted, and the results of data mining were verified and analyzed. The intrinsic correlation and logicity between the data are studied by using decision tree, neural network and Bayeses methods, and they are also described intuitively by using data mining technology.

Bayesian neural networks are widely used in many fields, such as machine learning, Bayesian, machine learning, etc. Due to the accelerating speed and speed of the computer operation, the operation amount of the Bayesian network is getting bigger and larger, and it is more and more difficult to meet the demand of massive and high precision. Using Bayesian principle and probability allocation (physical or Bayesian) method, logical connection can be easily solved between data; such as prediction, analyzing data analysis, mining clustering, etc. In bioinformatics theory, a large number of complex gene expression data are often achieved by computer. Probabilistic evolution

further improve the company's operation and management capabilities. Data mining is the in-depth analysis, summary and mining of a large number of data.

Table 1. Data processing and analysis

$k = \{0,1\}$	$n \geq 0$ $0 \leq p \leq 1$	$\lambda \geq 0$
$p > 0$	$k \in \{0, \dots, n\}$	$k \in \{0,1, 2, \dots\}$
$\frac{q-p}{\sqrt{pq}}$	$\frac{1-2p}{\sqrt{np(1-p)}}$	$\lambda^{-1/2}$
$\frac{6p^2-6p+1}{p(1-p)}$	$\frac{1-6p(1-p)}{np(1-p)}$	λ^{-1}
p	np	λ
pq	$np(1-p)$	λ
$-q\ln(q) - p\ln(p)$	$\frac{1}{2}\ln(2\pi nep(1-p)) + O(\frac{1}{n})$	

- (1) Statistics, like data mining, extract the structural features of the data from a large number of data, and analyze the internal connections and attributes between the data.
- (2) Data mining technology is a major direction of the current development of statistical technology, which puts forward a new idea for statistical analysis and data processing.
- (3) As an important link in statistical analysis, data mining has been widely used in database, intelligent processing, machine learning and other aspects.
- (4) Statistics and probability theory is a relatively mature and widely used technology in the field of data mining. Its application will promote the in-depth and effective development of data mining.

Both data mining and statistics are used to sort out, summarize and analyze huge data, find common rules in the data, and predict and make decisions in the future. As for the relationship between the two, data mining is an important branch of statistics. As a new statistical method, data mining provides statistics with a new idea of data statistics, that is, using advanced computer technology, database technology, etc. to carry out statistical analysis on huge data. Probability theory and mathematical statistics, as the important content of traditional statistics, have been developed for many years. Their accurate and efficient processing methods will also promote the level of data mining work to be improved again [8]. The two statistical methods should learn from each other and make progress together to promote the rapid development of statistics.

3 Application of Statistics in Data Mining

- (1) Probabilistic analysis network

million new microblogs will be generated every day; China Mobile's 700 million users will contribute 360 billion minutes of voice calls, 69.6 billion MB of network data traffic and 62.4 billion SMS messages every month on average in 2012. These users' personal information, communication records, network traffic records, SMS records, and business customization records constitute almost the largest user behavior information database in the world; As the world's largest bank by market value, ICBC provides services to 4.11 million corporate customers and 282 million individual customers through its 16648 outlets. Every deposit and withdrawal data, every online banking login or transfer data, will be accurately recorded by ICBC. According to IBM's report, 2.5 EB (1EB = 1024 PB) of data was generated every day in 2012, and this number is doubling every 40 months—whether you accept it or not, the era of big data has arrived [6].

The data generated by all walks of life is not only huge in quantity but also diversified in type. In addition to traditional structured data, we are exposed to more and more types of unstructured data: system logs, user complaint letters, SNS user friends, microblogs, pictures, audio; Video, sensor data... Not all data can be applied to the relational database model we are familiar with in the past. New technologies are needed to meet the needs of big data.

In recent years, column-stores have received widespread attention. In short, the column number database is a database that stores the data table columns separately. The attribute values belonging to the same column are continuously stored on the disk, while the traditional database stores data item by item. Because all attribute values of a column belong to the same data type, a columnar database is easier to compress than a traditional database. Unlike a traditional database, which cannot omit null values in a record, a columnar database does not need to store null values. These characteristics make a columnar database occupy less disk space. The columnar database is more suitable for OLAP (online analytical processing) than the traditional database, because OLAP generally only needs to read a part of the columns in the table. If the columnar database is used, only the data of the corresponding columns need to be read, but if the traditional database is used, the entire table needs to be read [7]. The ParAccel Analytic Database database adopts column-oriented storage technology and MPP technology, which combines the characteristics of fast query speed and compression of columnar database with the scalability of large-scale parallel processing technology.

2.2 Relationship Between Statistics and Data Mining

Statistics is a discipline that studies the basic principles and methods of statistics, mainly including the relevant knowledge of mathematics and probability, which is easy to collect, organize and process. In these three aspects, data processing is the most important and commonly used link in statistical work (as shown in Table 1). The data analysis mainly adopts analysis of variance, correlation analysis and regression analysis. Because of many defects in traditional data processing methods, they have been widely used in modern enterprise operations. With the rapid development of Internet and computer technology, people's demand for information is also increasing, so data mining has become a necessary means. Data mining refers to the use of knowledge discovery capabilities in databases or data warehouses to deal with practical problems, so as to

development, all industries need better data analysis and utilization capabilities to make timely and accurate judgments, so as to promote the development of the industry. Data mining mainly uses computer technology to analyze complex data through advanced algorithms, explore the internal relationship between data, and find some rules in the data; Provide users with decision-making basis [4].

In the era of big data, probability theory, mathematical statistics and data mining all play an important role in data management. In data management, probability theory and mathematical statistics are the most commonly used methods of induction and summary, which can make messy data orderly and facilitate technicians to predict future data. Data mining is a new technology, which generally uses computers with high computing power to process data. Through data mining, massive and complex data can be sorted out to make these data present regularity and uniformity, so as to facilitate the interpretation and analysis of these data by technicians. Both data processing methods can find the laws existing in huge data and predict and analyze the future in advance.

2 Related Work

2.1 Bottleneck of Data Analysis

Mathematical statistics, namely statistical theory, is a methodological science that studies how to collect data, sort out data, and analyze and infer data. Its purpose is to explore the internal quantitative regularity of the data of things, and to achieve a scientific understanding of objective things. Mathematical statistics plays an important role in all walks of life in reality. The solutions to a series of problems such as seed selection, fertilizer selection and farming conditions in agricultural activities are all related to mathematical statistical methods. Generally, the investigation results are obtained through proper design and statistical analysis of field experiments. The statistical methods of mathematical statistics, such as experimental design, regression design, variance analysis and multivariate analysis, play a wide role in the trial production of new products, improvement of old products, reform of technological processes, use of raw materials and search for appropriate formulations in industrial production [5]. The significance of mathematical statistics in social and economic fields can not be ignored, for example, in the investigation and prediction of social population, and in the analysis of ability in psychology, mathematical statistics knowledge should be used. In a word, in modern society, people's life, economic development, social policies and so on can not be separated from mathematical statistics to pave the way.

More and more applications are constantly producing a large amount of data: from medical image processing, gene sequencing to weather observation data, satellite data: from financial industry, insurance industry, communications industry to pharmaceutical industry, e-commerce, traditional retail industry; From the system log, microblog, SNS user relationship network, website click stream... we are producing a huge amount of data that could not be imagined in the past at any time.

Let's take a look at the following set of data: Tencent currently has more than 783.9 million QQ active accounts (the maximum number of simultaneous online accounts reached 167.3 million), 592.8 million QQ space active users and 469 million microblog users; The registered users of Sina Weibo have also exceeded 400 million, and 100

thought is the essential understanding of mathematical facts and theories after generalization. The basic mathematical thoughts are the foundational, summative and most extensive mathematical thoughts that are embodied or should be embodied in basic mathematics [1]. They contain the essence of traditional mathematical thoughts and the basic characteristics of modern mathematical thoughts, and are developing historically. Through the cultivation of mathematical thinking, the ability of mathematics will be greatly improved. To master mathematical ideas is to master the essence of mathematics.

There are both differences and connections between mathematical thought and mathematical method. The main difference is that mathematical thought is the soul of mathematics and plays a guiding role in the practice and exploration of mathematics. In other words, mathematical thought is the guiding principle of mathematical work. Any research on mathematical work is guided by mathematical thought, while using mathematical methods to solve daily mathematical problems, Mathematical methods, like tools used in daily life, generally focus on the practical work of mathematics. They are the manifestation and realization means of mathematical ideas, and the concrete manifestation of the application of theory to practice; Compared with mathematical methods, mathematical ideas are more abstract, universal and general, which shows that mathematical ideas are implicit. That is to say, mathematical ideas are often hidden in daily mathematical practice, but mathematical methods are often concrete and operable [2]. The mathematical methods used in the solution of a mathematical problem or the exploration of mathematical theory are both concrete and available.

At the same time, mathematical thinking and mathematical methods are closely related, interdependent, and inter-dependent. Mathematical thinking guides mathematical workers to flexibly use mathematical methods to solve mathematical problems and explore mathematical knowledge, and mathematical methods are more to verify mathematical thinking and enrich mathematical thinking in mathematical practice. Mathematical thinking and mathematical methods are inseparable in the daily mathematical exploration work, and the boundary between them is not clear and fuzzy. Therefore, in the actual mathematical learning and research, we always refer to these two concepts as mathematical thinking methods. There are many commonly used mathematical thinking methods, such as reduction thinking method, limit thinking method, number and shape combination thinking method, classification discussion thinking method, function and equation thinking method, analogy thinking method, statistical thinking method, construction thinking method, and so on [3].

Although probability and mathematical statistics are used for data statistics, they are also widely used for data mining. Data mining is to discover potential patterns and rules in the database and apply them to better understand and use data. Data mining is a means to obtain useful information from databases or other databases through computer technology and modern communication technology. By collecting, sorting and processing a large amount of data, we can mine the knowledge and rules hidden in the data for users to achieve higher work efficiency. Data mining is a new science. It is a data classification with large, incomplete and discrete attributes. It can extract data and make it more reasonable; New, unified and valuable information, which provides data support for public decision-making, is an effective means to comprehensively analyze and solve various practical problems. In today's increasingly mature social and economic



Applied Analysis of Probability Theory and Mathematical Statistics in Data Mining

Jidong Zhao¹(✉), Xiaoxuan Gong², and Cheng Zhenhua³

¹ Basic Teaching Department, ShanDong JiaoTong University, Weihai 264209, Shandong, China
zhaojidong0913@163.com

² Yunnan College of Business Management, Yunnan 650106, China

³ Sports Department, Modern College of Northwest University, Xi'an 710130, Shaanxi, China

Abstract. With the improvement of the national scientific and technological level, China has entered the era of big data. If we want to achieve better development in the era of big data and make rapid progress in the domestic socio-economic level, we need to find the universal laws of data in massive, complex and low correlation data, which requires the use of data mining and probability theory and mathematical statistics. Probability theory is a branch of mathematics that studies probability, random variables and random functions. It has been applied in many different fields such as engineering, computer science, economics and so on. Data mining is the summary and analysis of a large number of data, while probability theory and mathematical statistics are the more detailed analysis of data based on data mining. The two complement each other and achieve each other. Probability distribution is a function that assigns probability to the result or event according to the occurrence or non-occurrence of the result or event in the experiment or observation. Apply probability theory and mathematical statistics to data mining to improve the accuracy and efficiency of data mining. Therefore, in order to improve the quality of data mining, it is necessary to apply statistical methods effectively. On this basis, the specific application of statistics in data mining is analyzed through the connection of probability theory, mathematical statistics and data mining. Combined with specific algorithms, the application of statistics in data mining is discussed. This paper studies and analyzes the characteristics of data mining, probability theory and mathematical statistics, and discusses the specific application of statistics in data mining.

Keywords: data mining · probability theory · mathematical statistics · application analysis

1 Introduction

The so-called thought refers to the result of objective existence reflected in human consciousness through thinking activities, which is the basis of all human behaviors. People are great because of thought, and people are noble because of thought. Mathematical thinking refers to the result of thinking activities that the spatial form and quantitative relationship of the real world are reflected in people's consciousness. Mathematical

of evaluation work, reduce the burden on evaluators, but also more accurately evaluate students' physical fitness, classroom performance, competitive performance, and comprehensive abilities. At the same time, data mining simulation algorithms can also provide important references and guidance for educators, improve teaching plans and methods, and promote students' physical fitness and comprehensive development. Therefore, data mining simulation algorithms have broad application prospects and profound significance in physical education teaching evaluation.

References

1. Xia, L.: Design of efficient search system for English Translation Terms in massive Internet data. *Modern Electron. Technol.* **40**(13), 134–136 (2017)
2. Fang, L.: Design of high similarity English word selection system. *Modern Electron. Technol.* **40**(23), 147–150 (2017)
3. Bin, L., Xiaopeng, L.: Design and application of meteorological information query system based on Silverlight. *Meteorol. Sci. Technol.* **44**(3), 369–373 (2016)
4. Liping, L., Xiaofeng, M., Lihui, L.: Design and implementation of computer aided system based on code automatic generation. *Aerosp. Control* **33**(4), 68–72 (2015); Xiangrui, H.: *Reliability Engineering*, pp. 2–15. Tsinghua University Press, Beijing (1990)

4.2 Exploring and Constructing the Evaluation Process of Physical Education Teaching in Colleges and Universities

The evaluation of physical education teaching in universities is conducted to improve teaching quality and ensure the comprehensive development of students’ physical fitness. The evaluation process mainly includes the following steps: first, clarify the evaluation objectives and indicators, and determine the focus and content of the evaluation; Then determine the evaluation methods and methods, commonly used methods include questionnaire survey, on-site observation, and skill testing; Next, collect and organize evaluation data, and conduct data analysis and processing; Finally, based on the evaluation results, evaluate the existing teaching quality and students’ physical fitness, and propose corresponding improvement measures and suggestions to achieve the goal of improving teaching quality and students’ physical fitness. Throughout the entire evaluation process, it is necessary to attach great importance to feedback from teachers and students, pay attention to the scientific and practical nature of evaluation indicators, and ensure the fairness and reliability of evaluation results. The specific process is shown in Fig. 2.

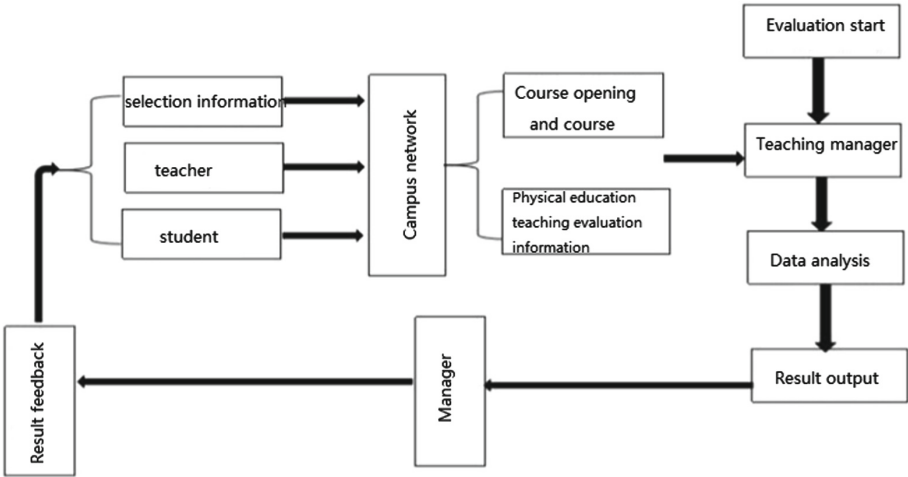


Fig. 2. Evaluation process of physical education teaching

5 Conclusion

The application of data mining simulation algorithms in physical education teaching evaluation can help educators evaluate students’ physical fitness and comprehensive abilities more comprehensively and scientifically. Simulation algorithms can analyze and process a large amount of data from students’ learning process, extract effective information from it, discover correlation and regularity, and predict and simulate during the evaluation process. This method can not only reduce the complexity and subjectivity

roles and responsibilities need to be clearly defined. The second step is to determine the evaluation index system. This system can include multiple aspects such as physical fitness, classroom performance, competitive performance, comprehensive ability, etc. It needs to be scientifically reasonable and take into account national and industry standards. The third step is to apply appropriate evaluation methods and tools. Different evaluation indicators require different evaluation methods, such as questionnaire surveys, field visits, and skill tests. Finally, this article points out the practical significance and application prospects of constructing a framework for the evaluation of physical education teaching in universities. Practice has proven that a correct evaluation subject framework can improve the quality of teaching and the comprehensive development of students' physical fitness, promoting their diversified growth. At the same time, the construction of the evaluation framework for physical education teaching in universities can also provide useful experience and reference for the reform and development of the evaluation system.

Therefore, when constructing the evaluation system for physical education teaching in universities, we should recognize that different evaluation subjects have different roles, and clarify the commonalities and personalities between each subject through evaluation indicators and indicator weights, namely physical education teachers, students, peers, and physical education teaching leaders. The specific framework is shown in Fig. 1.

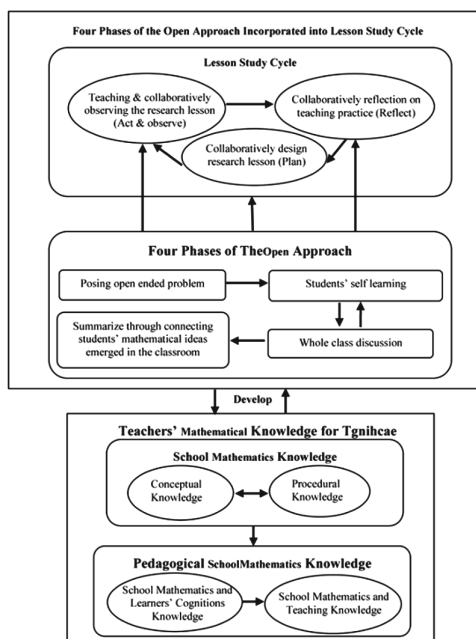


Fig. 1. The main frame of physical education teaching evaluation

Others' evaluation refers to the evaluation conducted by others other than the evaluated person, also known as "external evaluation". For example, students' evaluation of teachers, teachers' mutual evaluation, etc.

3.2 Principles of Constructing the Evaluation System of Physical Education Teaching in Colleges and Universities

The design of college physical education teaching evaluation system under the background of big data application should be based on theory and from all aspects of physical education teaching process, and mainly meet the following principles:

- (1) The principles of scientificity and objectivity. In combination with the background and characteristics of the big data era, and on the basis of objective laws and practice, the construction of college physical education teaching evaluation is carried out scientifically and objectively. For example, the selection of evaluation indicators should be based on the collected and investigated data to make reasonable screening, so as to ensure the scientificity and objectivity of evaluation indicators.
- (2) Integrity and comprehensiveness. That is, when constructing the evaluation system of physical education teaching in Colleges and universities, the selection of evaluation indicators should be comprehensive, not repeated, well-organized, widely representative, and can basically show the whole process of physical education teaching.
- (3) Feasibility and testability principles. Feasibility means that the way of evaluating physical education teaching has practical operability. Measurability means that the evaluation of evaluation indicators has testable and quantitative standards. For example, the selection of indicators must be easy to understand and measure without ambiguous indicators.

4 PE Teaching Evaluation Based on Data Mining Simulation Algorithm

4.1 Exploring the Construction of the Main Body Framework of Physical Education Teaching Evaluation in Colleges and Universities

The evaluation of physical education teaching in universities is an important component of the national education evaluation system, and also an important means to ensure the quality of physical education teaching in universities and the comprehensive development of students' physical fitness. This article aims to explore the methods and principles of constructing a framework for the evaluation of physical education teaching in universities. Firstly, the evaluation of physical education teaching in universities requires diversified and scientific evaluation indicators. In this context, this article proposes the basic principles for constructing a framework for the evaluation of physical education teaching in universities, including scientificity, diversity, practicality, and adaptability. These principles will help evaluate the multifaceted performance of the subject. Secondly, specifically regarding the method of constructing the main framework, this article proposes the following steps: the first step is to clarify the evaluation subject. The evaluation subject includes the evaluator, the evaluatee, and the evaluatee, and their respective

2.2 Domestic Research Status

At present, domestic theoretical research is mainly based on absorbing and utilizing foreign theories. The project team of the national basic education curriculum reform “Research on the for promoting teachers’ growth and development” took the lead in conducting research on developmental evaluation and achieved some results. There are more and more researches on diversified teaching evaluation.

In terms of teaching evaluation software, there are “online student evaluation system” jointly developed by Nanjing Xunji Technology Co., Ltd. and the Academic Affairs Office of Nanjing Forestry University, and “Jinzhong network teacher evaluation system” and “Jinzhong network student evaluation system” developed by Beijing Jinzhong Network Information Engineering Co., Ltd. These evaluation systems are all based on Web, which realize efficient teaching evaluation without time and space constraints [4]. However, their evaluation model is single, the evaluation subject and evaluation object are few, and they are closely coupled with the teaching system, so they do not have universality and can not achieve diversified evaluation. Recently, the design and development of “E-Learning archives” software system has risen, which has preliminarily realized the support of information technology for process teaching evaluation, but it is still very imperfect and its application is very narrow.

3 Physical Education Teaching Evaluation

3.1 Classification of Teaching Evaluation

Relative evaluation refers to the evaluation in which one or several objects are selected as the standard in the set of evaluation objects, and then the rest of the evaluation objects are compared with the standard, or the evaluation in which all the evaluation objects are arranged in order in some way. Through comparison, the the evaluated object in the set can be determined. Relative evaluation is mainly used in selective and competitive activities, and is often used in network teaching evaluation.

Absolute evaluation refers to the evaluation that determines an objective standard in advance outside the set of evaluation objects, compares the evaluation objects with the evaluation standard, and judges the degree of reaching the standard. It is mainly used for qualification and compliance activities.

Intra individual difference evaluation is to compare the past and present of each individual in the overall set of evaluation objects, or compare several aspects of an individual with each other. It is often used in two situations. One is to compare the past and present of the evaluation object. One is to compare some aspects of the evaluation object. The comparison between the past and the present of an individual can enable the evaluation object to understand their own development, and the comparison of various aspects of an individual can enable the evaluation object to understand their own advantages and disadvantages and make self-regulation.

Self evaluation is that evaluators evaluate themselves according to certain evaluation criteria. For example, teachers’ evaluation of their teaching ideas, contents, methods, attitudes and effects.

artificial intelligence and other technologies can find useful knowledge and rules in the evaluation results, and further guide teaching.

As one of the important tools to measure the process and effect of PE teaching, PE teaching evaluation can not only reflect the shortcomings of PE teaching; It can also help students find problems in the efficiency. Exploring the evaluation has a positive significance both in theory and practice.

2 Related Work

2.1 Research Status Abroad

At present, there mature systems in the theoretical of teaching evaluation abroad. Such as multiple intelligences theory, humanism theory, developmental teacher evaluation theory and “the fourth generation of education evaluation theory”. As mentioned above, the theory of multiple intelligences classifies people’s intelligence into eight types, and holds that each type of intelligence represents different abilities, each type of intelligence needs to be developed, and each type of intelligence needs different evaluation methods. Humanistic theory emphasizes the value and potential of human beings, pays attention to the self realization of human beings, and believes that evaluation must be people-oriented [3]. The theory of developmental teacher evaluation thinks that teachers should be aware of teaching problems and their own shortcomings through the evaluation of experts, students and themselves, and promote teachers’ teaching work and self-improvement through the feedback control mechanism. “The fourth generation of educational evaluation theory” has absorbed the reasonable elements of the above theories. Its main points are as follows: (1) it emphasizes that the complete and individualized people are the object of evaluation, and through evaluation, it promotes the full development of the educatees’ personality, and pays attention to qualitative analysis; (2) It is advocated to evaluate their own development process from the internal needs and actual conditions of students’ development, generally using the individual reference evaluation method; (3) Adhere to the humanitarian spirit, organize evaluation activities in a good interpersonal atmosphere of mutual trust and respect, and fully reflect respect for personality, trust in ability and concern for development; (4) Pay attention to self-evaluation, adhere to the democracy of evaluation, and emphasize the habit and ability of self-analysis, self-evaluation and self-regulation in the process of evaluation; (5) Considering the needs of all participants, it is considered that the evaluation is a coordination process and the evaluation result is an inseparable isomorphic process for both sides of the evaluation; (6) Emphasize “value diversity”, advocate fully listening. Therefore, it also advocates diversified evaluation.

In terms of teaching evaluation software, “question mark” and “rubric star” are well-known in foreign countries. With the education and teaching evaluation theory, these software continue to release new versions. In some aspects, they can be used as teaching evaluation tools, but their ability to support the integration of teaching process and evaluation process is quite poor; However, teaching evaluation software systems such as “smart” have lost the flexibility to support different teaching theories and teaching contents due to their overemphasis on the integration with specific teaching processes.



Application of Data Mining Simulation Algorithm in Physical Education Teaching Evaluation

Zhenhua Cheng¹(✉) and Jidong Zhao²

¹ Sports Department, Modern College of Northwest University, Xi'an 710130, Shaanxi, China
251721357@qq.com

² Basic Teaching Department, ShanDong JiaoTong University, Weihai 264209, Shandong, China

Abstract. The application of data mining simulation algorithm in physical education teaching evaluation is a process of using computer simulation model to evaluate. The use of simulation models is not new, but researchers have only recently used them in this way. Simulation models are often used to evaluate educational programs or courses (for example, to test whether the program improves outcomes). In these models can be used to assess the effectiveness of Instructional Technology (e.g., how much time should be spent in each class) and to determine which instructional strategies work best under what conditions (e.g., what type).

Keywords: Data mining · Physical education teaching evaluation · Simulation algorithm

1 Introduction

In the 20th century, educational science research developed rapidly, forming a huge educational science system. Research on basic theories of education have become the three major fields of modern educational scientific research [1]. Educational measurement and evaluation is an applied discipline with the characteristics of comprehensiveness, technicality, practicality and applicability in the educational science system. It is a skill and method that people usually rely on to guide educational practice according to the basic theories and laws of education. Its core part is teaching evaluation.

Teaching evaluation. It takes teaching objectives as the basis, uses effective evaluation techniques and means, measures, analyzes, compares, and gives. Teaching rose in the United States in the 1930s. It flourished in the western world and spread to China in the 1960s. It widely exists in various practical and is an indispensable work.

The of information education great to education. The way of teaching evaluation has also shifted from traditional symposiums, questionnaires and expert lectures to online evaluation. Online evaluation the network, which is not, convenient and fast, to collect the teaching evaluation data, and use the powerful information processing ability of the computer to statistically process the teaching evaluation result data, which can improve the fairness, effectiveness and credibility [2]. At the same time, using data mining,

15. He, J., Liu, B.: Reform and practice of university computer basic course teaching mode based on smart learning[C]. 2020 15th International Conference on Computer Science & Education (ICCSE) (2020)
16. Hui, L.I.: The significance and strategies of offering drama creation course in western colleges and universities [J]. J. Guangxi Sci. Technol. Normal Univ. (2016)
17. Jianrong, L.I.: On the humanistic mission of creative writing teaching reform in universities [J]. J. Guangxi Sci. Technol. Normal Univ. (2017)
18. Xiao, H., Ma, Y.N., University, T.: Teaching present situations and thinkings of creative writings based on application: take tongling university as an example [J]. J. Tongling Univ. (2016)
19. Zhang, X.: Exploration on teaching reform of contemporary Chinese literature under the major field of creative writing [J]. J. Guangxi Sci. Technol. Normal Univ. (2019)
20. Zhang, X.: exploration on teaching reform of literature course under the major field of creative writing [J]. Writing (2018)

tools to collect information from the data has become a challenge, and text data mining helps to solve this problem to a large extent.

5 Conclusion

This study aims to find out the best teaching mode of creative writing course in Colleges and Universities Based on data mining technology. The main purpose of this paper is to study the influence of teaching mode on students' learning process. This paper is divided into two parts; The first part contains literature review, and the second part contains research methods. Let the students exchange their ideas with each other, carry out creative collision, and then write the draft and finalize it. Such a teaching method makes students become the masters of the classroom, helps to cultivate their own innovative spirit, and helps to break out their inspiration.

References

1. Wang, X.: Construction of civil engineering teaching system based on data mining algorithm and big data technology [J]. *J. Phys. Conf. Series* **1852**(3), 032022 (6pp) (2021)
2. Kong, Y.: Research on reform intelligent effect evaluation system (RIEES) of teaching methods of ideological and political theory course in colleges and universities[C]. 2018 International Conference on Smart Grid and Electrical Automation (ICSGEA). IEEE (2018)
3. Luo, Y., Wan, J., She, S.: The construction and research of the platform of intelligent sharing laboratory based on big data[C]. International conference on Big Data Analytics for Cyber-Physical-Systems. Springer, Singapore (2022)
4. Jamil, A.: Reflections on the teaching of creative writing at the American universities [J] (2016)
5. Li, Z.: Paths to realize informationization teaching in the psychological course of tourism service based on intelligent tourism[C] (2019)
6. Han, X., Hongwen, Y.U.: Construction of Characteristic Thematic Database of Record Room in Colleges and Universities – Discussion on the Construction of Chinese Language Thematic Database (2016)
7. Liu, X.: The application of data mining technology in the teaching evaluation in colleges and universities [J]. *J. Comput. Theoret. Nanosci.* **14**(1), 7–12, 14 (2017)
8. Zhang, Z.L.: Research on construction technology of prefabricated structure based on BIM technology [J] **6**(5), 14–19 (2022)
9. Shen, L.H.: Research on innovation of Chinese writing teaching in colleges and universities based on creative perspective [J]. *J. Qiqihar Junior Teachers' College* (2018)
10. Zhi-Chun, L.I., Zhang, Y., Bai, H.W. et al.: Teaching design of “environment science” course based on intelligent classroom in colleges and universities [J]. *Sci. Technol. Vis.* (2019)
11. Jia, S., Zhang, X.: Teaching mode of psychology and pedagogy in colleges and universities based on artificial intelligence technology [J]. *J. Phys: Conf. Ser.* **1852**(3), 032033 (2021)
12. Sun, B.: Analysis of empirical research model for computer teaching in colleges and universities based on data mining learning analysis [J]. *Modern Electron. Technique* (2019)
13. Sun, C.: Construction and implementation of intelligent classroom teaching model in colleges and universities [J]. *Wirel. Int. Technol.* (2019)
14. Qian, Z.: Innovation of the school computer teaching model based on the information network and the intelligent campus[C] (2020)

improve the soft power of Chinese culture. Therefore, setting up diversified creative writing teaching courses in Colleges and universities has become a necessary choice on the road of innovation.

By using certain terms and associating them, you can classify one group of documents from another. In addition, this correlation helps to identify patterns and frequencies of data in a given document set. Another common document classification method is document clustering [18]. As described above, in this method, documents of similar sets are identified and clustered into similar groups. Document classification can form some rules because it is different and difficult compared with other data. These documents will be huge and the background will be different. The general workflow of the text creation and review process is shown in Fig. 4 below.

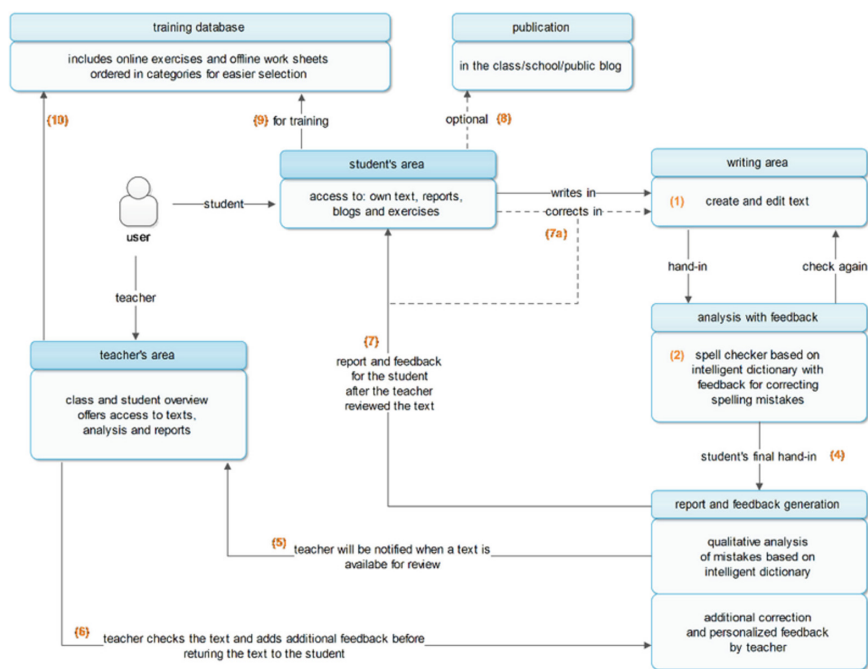


Fig. 4. General workflow for the text creation and review process

When some words or phrases appear frequently, it is best to form a relationship between words in order to form a correlation. This is helpful for data association mining and process automation. Data must be preprocessed to avoid confusing words by parsing and identifying terms. It can form an algorithm to identify common words, and can easily identify patterns. This automated process reduces human effort and produces relevant results to validate the data. Association mining is based on terms in data, because every repeated word is recognized as a term or item.

Most of the global data are unstructured, and the importance of text data mining is self-evident. If information is not collected from unstructured data, the collection of data itself becomes irrelevant [19, 20]. Storing large amounts of data and using traditional

poems and songs, but also contains all other text content based on human creativity - such as the writing of a wechat official account or the creation of a song lyrics, which belong to the category of “creative writing”.

This is the best way to identify the data you need whenever there are many documents, whether online or offline. Automatic document classification helps you easily identify data with few keywords. WordNet or expert knowledge tools help you find relationships between terms or clusters for easy automation. In addition, this can be achieved by pairing attributes and values with each other and creating entity relationships for words, although the repetition of words in documents is not common.

4 Research on Teaching Mode

In constructivist theory, learning is not a passive acceptance, but an active process that requires learners to use their sensory organs throughout the body to absorb and construct their own meaning in a certain social and cultural environment. According to the basic content of constructivism theory, during the entire teaching process, teachers usually play a leading role, coordinate and take charge of the entire teaching process, organize the teaching process and specific implementation, guide students to actively participate in the entire teaching process, fully consider the individual differences of different students, and develop targeted teaching content [16]. It should be noted that the concept advocated by constructivism is “active learning by students”, which is obviously different from the traditional teaching of “passive absorption” by students. In constructivist theory, teachers mainly assume the dual roles of guidance and supervision, and students’ knowledge and skills are acquired through their own independent learning, rather than being imparted forcefully by teachers.

Constructivist learning theory emphasizes the importance of taking students as the center, constructing an ideal teaching environment in teaching, providing students with a system that can build their own knowledge, while online intelligent scoring systems provide students with an unconstrained learning platform that can be based on their own learning level and style [17]. Therefore, teachers must fully understand the differences in the actual ability levels of different students, and maximize their cognitive structure by implementing differentiated teaching to help them achieve better development. Students need to rely on their own knowledge to conduct second language writing, using online scoring systems to modify their compositions, using machine evaluation to modify their compositions, making full use of the learning resources on the platform, and receiving timely guidance and feedback from teachers, actively participating in writing learning activities to fully tap their potential.

Diversified choices of creative writing teaching courses in Colleges and universities. As for the current situation of the simplification of creative writing teaching courses in Colleges and universities, colleges and universities should set up diversified courses to expand the writing content into special topic writing, novel writing, memoir writing [8], TV program writing and advertising creative writing, so as to stimulate students’ writing thinking, innovate their ideas, enrich their creative writing content, and broaden students’ writing path. In this way, whether it is the accumulation of literary knowledge or the expansion of employment choice channels, It is also of great significance to

suitable for students to learn from social life. These content can be the same as or opposite to the “core”. The purpose is to broaden the field of students’ learning [14]; The third is “integration”, which integrates the selected teaching content, clarifies the sequence, enriches students’ emotional experience, and broadens their horizons for learning Chinese. Therefore, intelligent Chinese education is formed. For example, when Mr. Xue was teaching Bai Juyi’s “Light Rain Walking at Night”, he selected and edited famous sentences about “light rain” around the description of “light rain”, forming a teaching content plate with “rain” as the core. On the basis of learning the original text, he guided students to feel the difference between rain in different seasons, light rain and rainstorm, and the difference between vernacular and ancient poetry. In this way, Students’ emotional experience of “rain” is based on their own experience and thinking, which makes their emotional experience more rich and detailed, thereby gradually expanding their horizons in learning Chinese, and thereby generating wisdom in language learning. Text vocabulary data generation is shown in Fig. 3 below.

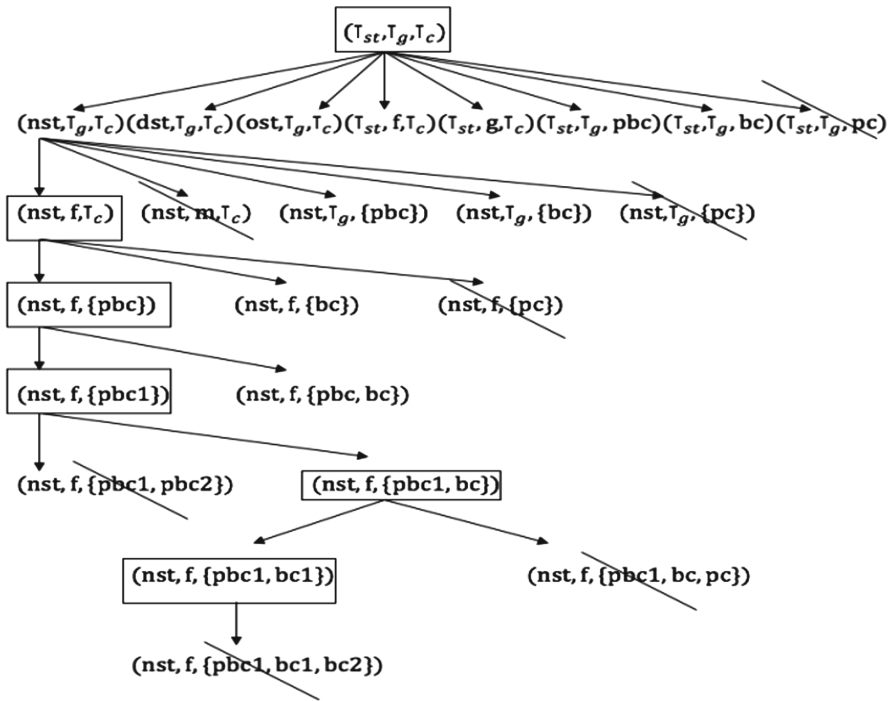


Fig. 3. Text vocabulary data generation

Such a definition seems too abstract and academic. However, once it is pointed out that “creative writing as a writing activity includes literary writing in our traditional sense, as well as many non literary writing oriented to cultural and creative industries, emerging technologies and daily applications. These writing activities, including writing education activities, are based on literary writing”, we can clearly realize that [15], The “writing” recognized by creative writing is not only the traditional “literary writing” of

sources to form accurate information from the data, extracting the same content as a large number of original documents [10]. Different techniques are used in the text summary because this process requires different inputs. Neural network and regression model are commonly used in text summarization. The long text is shortened so that the entire text forms a summary. This makes it easy for users browsing the summary to retrieve the data and understand it.

3 Research on Intelligent Teaching Mode of Creative Writing Course in Colleges and Universities Based on Data Mining Technology

When many people see “writing”, they will associate it with “writer” and “Literature” at the first time [11]. However, creative writing is quite different from traditional writing. In the research of Professor gehongbing ° of Shanghai University, creative writing is defined as “the general name of all creative activities in which human beings take writing as their activity style and work as their final outcome”. At first glance, this seems nothing, but in fact, this definition is quite different from the traditional writing.

First of all, the understanding of creative writing for writing is not limited to the category of “literary creation” [12]. Creative writing has a new interpretation of the two concepts of “creativity” and “writing”. Professor gehongbing pointed out that in the perspective of creative writing, creativity is a fundamental practical activity for people to realize themselves as subjects; Writing is an aesthetic practice based on human’s ontological practice.

Data mining teaching focuses on integrating scattered and trivial language training into a comprehensive language practice activity, forging rich language with distinct images and sincere emotions into a “language alloy” that integrates images, emotions, and language in language learning, with strong vitality and aggregation functions, and can integrate new language information and transform students’ own language [13]. Therefore, data mining teaching scientifically and reasonably arranges various Chinese practical activities such as listening, speaking, reading, and writing based on the characteristics of students’ age stages, integrating scattered Chinese activity projects in the past into a comprehensive practical activity block, and constructing three levels of Chinese practical activity blocks: first, a sequence of integrated reading and writing activity blocks, with the aim of improving students’ reading and writing abilities; The second is the sequence of exploratory learning activity blocks, which aims to stimulate students’ enthusiasm, initiative, and creativity in learning; The third is the sequence of comprehensive practical activities, which is based on social life, selects the problems faced by students, guides students to solve problems, and closely communicates with various disciplines. Over time, in such integrated language practice activities, students’ language abilities gradually improve imperceptibly.

Data mining teaching also focuses on the reorganization of teaching content, integrating teaching content into one section. Select and combine teaching content blocks suitable for students’ learning through the “trilogy”. One is “subtraction”, which filters out important parts of the textbook that are worth learning as a “core” of the teaching content. The other is “linkage”, which focuses on the core “core” and selects content

So What is a Case Study analysis?

- "separation into parts"
- "examination of a thing to determine its parts or elements"
- ".....with a view toward prescription"
- Analytical processes used in medicine, auto repair, business turn around
- Analysis more than simple description

Fig. 2. Creative writing scheme

$$J(\Pi, W) = \sum_{k=1}^K \sum_{x \in \pi} \sum_{d=1}^D (x_{id} - v_{kd})^2 \quad (2)$$

$$O_{v,j} = h \left(\sum_{i=1}^{f_{k-1}} \sum_{u \in N[v]} w_{i,j,u,v} x_{u,i} \right), (j = 1, \dots, f_k) \quad (3)$$

$$\max \sum_I \left[U^I(X^I) - C^I(X^I) \right] \quad (4)$$

From the perspective of data scientists, this process belongs to supervised learning technology. This is also called naturallanguageprocessing. Based on the data in the database, predefined topics are selected and provided to the data. Collect and analyze the documents, provide the correct title for the data, and process them for further analysis [9]. This process is very simple. Just as users know relevant topics, they can search for topics, so there is no need to analyze the entire data. This saves time and the work of data analysts. In addition, this process can be easily automated in different environments. In fact, this process has been automated, so spam filtering, web page classification and other data science and technology can be easily applied to the NLP process.

$$\arg \min_{SC} \sum_{i=1}^k \sum_{x \in C_i} |X - \mu_i|^2 \quad (5)$$

As the name suggests, the abstract of an article is formed by this technique. Provide valuable information to users by automatically compressing data. Examine different text

emotions and ideas (Zemal, 1985). Writing is a powerful tool for thinking, learning, and communicating. For high school students, writing can cultivate their thinking awareness and language output ability, and can also focus on reflecting a high school student's comprehensive language use ability. Therefore, in the teaching of English writing, teachers should not only provide students with formal assistance in language, but also attach importance to guiding students in thinking, cognition, discourse, semantics, and other aspects [6].

To understand what creative writing is, what it is for and against, we must first know where it comes from. As an independent subject, creative writing was born with its specific historical background. The concept of "creative writing" was first put forward. It is generally believed that it came from Ralph Waldo Emerson's speech to the National Association of university students honors in 1837. In this speech, Emerson clearly used the concepts of creative writing and creative reading.

The historical background of the concept of "creative writing" can be summarized as "the reform of educational thought" and "the development of English Literature" [3]. The historical context behind Emerson's speech was the reality that academic writing, literary works and literary education in American higher education were not valued at that time. At that time, teaching practice overemphasized the status of linguistics, especially ancient language and rhetoric, which restrained the development of literary originality. Since its birth, creative writing has clearly defended the independent status of literature and protected the original value of literature. Creative writing, as we know it today, evolved on the basis of English writing courses in American colleges and universities in the 19th century. It itself emerged to get rid of the tedious classical Chinese education and change the situation that literature itself is not valued [7]. Its initial goal is to bridge the gap between academic research on literature and social practice. Figure 2 below shows the creative writing scheme.

Obviously, the traditional Chinese language education has been unable to adapt to the development background of the times, and the rise of creative writing is natural.

2.2 Text Data Mining Technology

This is the process of extracting useful information from unstructured data, identifying and recording the relationships and entities between related data. In addition, you can record data entities and attributes from unstructured or semi-structured data. This information is stored in the database. Whenever necessary, the relational schema is examined and used for business or any other requirements [8]. Check the efficiency of stored data so that data stored procedures can help meet future needs.

This is another process of extracting required patterns and collections from data according to user needs. For example, users can search for data using a specific set of words or patterns to easily understand the data. The algorithm is used to identify the user pattern of search data to provide relevant information faster. In addition, these algorithms help the system identify the user behavior to be tracked to find the same pattern of relevant data. Google search is the best example of data retrieval technology.

$$\arg \min_{SC} \sum_{i=1}^k \sum_{x \in C_i} |X - \mu_i|^2 \quad (1)$$

their self-confidence and making them more passionate about English learning [4]. The architecture is shown in Fig. 1 below.

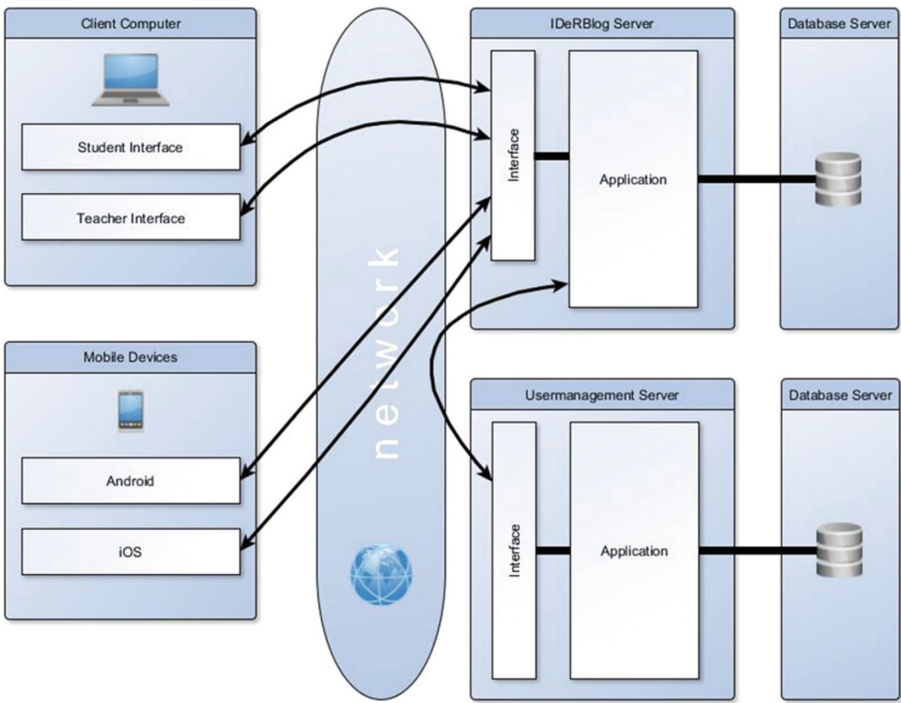


Fig. 1. Architecture

It is hoped that more people will know creative writing, its existence, its past and its value. Perhaps, compared with the efforts made by experts and scholars to promote creative writing, a push does not play much role, but as a creative writing enthusiast, learner and practitioner, it is difficult to curb his desire to write such an article to introduce his creative writing to more people [5]. Based on this, this paper studies the intelligent teaching mode of creative writing course in Colleges and Universities Based on data mining technology.

2 Related Work

2.1 Creative Writing

As an important way of language output, second language writing is an emotional cognitive activity. Zemal pointed out that writing can essentially be seen as a way for authors to communicate with others and an effective way for authors to think and create. In his view, writing is a process of continuous reflection, modification, and elaboration by the individual's brain, as well as the output and expression of the individual's own inner

It is well known that students' learning cannot be separated from the five skills of "listening, speaking, reading, reading, and writing", and among these five skills training links, writing is clearly the most important and difficult to conduct training. In the teaching of English writing, to ensure a good quality of composition writing, students are required not only to have a solid knowledge of English, but also to have a high comprehensive ability to use English, and be able to effectively output their basic knowledge of English. It can be seen that the teaching of English writing involves not only many aspects, but also a challenging task. From the perspective of English writing teaching, constructing an effective writing evaluation mechanism can not only help students achieve positive feedback on writing issues, but also help teachers conduct teaching effectiveness analysis. In the teaching of English writing, if only the teacher corrects the composition and gives scores, the workload for the teacher is heavy, and the feedback cycle for the students is long, which can easily cause students to lose interest and hinder the development of English writing ability (Qin Chaoxia, 2009) [2]. However, in the traditional writing teaching model, English teachers' teaching tasks are heavy and mainly focus on exam oriented education. In addition, correcting compositions takes time and effort, and teachers often rush through the process. As a result, students' writing training is very scarce, and there are few writing opportunities. They often only complete writing in written English expressions in tests, which only give a score, and few students revise their compositions again.

From the perspective of students' current writing level, students generally master the basic skills of English writing. However, in daily teaching, the author has conducted several limited time writing exercises and found that students' English writing is not satisfactory and their English writing ability is low. The main problems include: confusion in structure and unclear paragraph division; Empty and poor content; Vocabulary and grammar errors. According to daily conversations with students, their previous high school writing feedback model was traditional and single teacher feedback, which means that the teacher corrects and scores them all at once. Some students, even if they receive the teacher's correction and scoring, simply do it in a hurry and do not make any changes or reflections on their compositions, which means they lack the awareness of autonomous revision. Even if students want to revise their compositions repeatedly, they will be affected by the timeliness of teacher feedback, which seriously affects students' enthusiasm and initiative in English writing. In addition, for some students who want to learn independently and receive timely writing evaluations and feedback, teacher evaluations cannot be timely fed back to students [3]. Over time, English writing will not receive the attention of students, and students will lose enthusiasm for writing. This will also reduce confidence in writing, which is not conducive to improving students' writing skills. As for online intelligent evaluation systems, their advantages are mainly manifested in the following aspects: First, they can achieve real-time feedback; Second, comment on sentences; Third, it has online modification function. Using this system to develop English composition writing, students can not only stimulate their desire to learn writing, but also enable this part of students who want to learn independently to receive timely feedback during the writing process, and modify their compositions. Students' errors are gradually corrected, ultimately achieving continuous improvement in their writing skills. Not only that, repeated training is also conducive to improving



Intelligent Teaching Mode of Creative Writing Course in Colleges and Universities Based on Data Mining Technology

Xiaoci Yang^{1,2(✉)}, Janaka Low Chee Kong¹, and Zhang Liu¹

¹ Malaysia University of Science and Technology, Hefei, China
yangxiaoci1978@163.com

² Xi'an Fanyi University, Xi'an 710105, Shaanxi, China

Abstract. The research on intelligent teaching mode of creative writing course in Colleges and Universities Based on data mining technology is a paper aimed at studying the effectiveness of intelligent teaching mode of creative writing course. The research will be based on data mining technology, and the research methods will be qualitative methods using content analysis, investigation methods and case study methods. The researchers conducted this study by collecting data from different sources, such as journals, books, and the Internet. This paper also introduces the results of the research work, which shows that there is no significant difference in performance between students who use intelligent teaching mode and students who do not use intelligent teaching mode.

Keyword: Data mining technology · Writing · teaching model

1 Introduction

When we have a deeper understanding of the subject of creative writing, we have become more aware that this is a major that best meets our interests and needs. However, it is also found that although the theoretical research on creative writing in domestic academic circles is becoming more and more in-depth, and the teaching of creative writing has also shown an obvious trend of sinking to primary and secondary schools [1]. Some creative writing courses have also appeared on the Internet, but creative writing itself seems to be confined to the space of colleges and universities to a large extent, and has not been recognized, loved and applied by more and more people.

Of course, university scholars have also noticed this problem and are working hard for it. In the field of new media (mainly wechat official account), there is not much promotion and introduction of creative writing itself. Many platforms related to it are to display works and publicize activities. Even if there is only a word of introduction, it is often too academic and not easy to be widely understood by netizens with different levels of education. Previously, I also promoted the theoretical books on creative writing on the official account (click to read), but if I have no concept of creative writing, I believe that not many people are willing to read it.

11. He, F.: A study on the teaching of Chinese culture and the evaluation of foreign students' Chinese application abilities [J]. *J. Huangshan Univ.* (2017)
12. An, L., Dong, J., Ma, C. et al.: Application of flipped classroom mode in teaching of database and data mining technology course [J]. *Logist. Technol.* (2019)
13. Wu, Q.F.: Distance teaching method of accounting informatization course based on big data mining[C]. *International Conference on E-Learning, E-Education, and Online Training*. Springer, Cham (2021)
14. Gao, Y., Li, Z.: Application of the micro course based on we chat platform in the practice of respiratory medicine in foreign students [J]. *China Continuing Med. Educ.* (2017)
15. Wang, Z.: Technology. Research on experimental teaching method of “data mining” course based on Aliyun big data platform [J]. *Res. Explor. Lab.* (2018)
16. Zhang, Y.R., Ouyang, Y.M., Xiao, F.: On the teaching of chinese culture course for English majors in application-oriented universities based on MOOC network platform a case study of Pingxiang university [J]. *J. Pingxiang Univ.* (2018)
17. Long, Y.: Research on art innovation teaching platform based on data mining algorithm [J]. *Clust. Comput.* **22**(2), 1–7 (2019)
18. Luan, H.B., Wen, F.A.: The application research of data mining in university english scores forecasting [J]. *Comput. Eng. Softw.* (2016)
19. Jiang, J., Dong, X.L., Wang, X.E.: Analysis of students' Elective course behavior based on frequent itemsets [J]. *Sci. Technol. Vis.* (2018)
20. Cai, L., Library, H.: Application of data mining technology in characteristic database [J]. *Comput. Technol. Automat.* (2018)

results, which requires the presentation of diverse information content. Therefore, we can choose to use the knowledge atlas method to analyze and present it. The knowledge atlas is to convert complex information into an information structure that can be structured for display through calculation and processing, The knowledge content that needs to be represented can also be presented through graphical rendering, providing a valuable basis for readers and observers.

5 Conclusion

The teaching platform of characteristic culture courses for foreign students based on data mining no longer takes Book assignments as the evaluation method. It is hoped that foreign students can truly enter the local social life, experience the local cultural characteristics, and reduce the sense of conflict caused by cultural differences. Through participation and cooperation to complete practical tasks, international students can also enhance their interpersonal skills and awareness of collaboration, and guide them to look at, analyze and solve problems from a cross-cultural perspective; Help them build a global vision and improve their ability to deal with conflicts and uncertainties; Encourage them to dare to challenge and innovate, so that all overseas students can spread Chinese culture with a more positive attitude and become messengers of friendship between the Han nation and other nations.

References

1. Na, Z.: Database design of regional music characteristic culture resources based on improved neural network in data mining [J]. *Pers. Ubiquit. Comput.* **24**(1), 103–114 (2020)
2. Hui, D., Jin, W.: Design of real-time data analysis system for physical training based on data mining technology [J]. *J. Phys. Conf. Series* **1982**(1):, 012206 (2021)
3. Lu, Z., Zhao, Y.: Research on teaching evaluation of college physical education based on data mining and network survey (2016)
4. Lei, Y.R., Lei, L., Liu, L.Q.: Application of fuzzy association rules in the analysis on higher vocational college students' performance [J] **28**(1), 1–12 (2017)
5. Han, X., Hongwen, Y.U.: Construction of Characteristic Thematic Database of Record Room in Colleges and Universities -- Discussion on the Construction of Chinese Language Thematic Database (2016)
6. Wu, D., Guo, M.: Application of data mining in traditional benchmark evaluation model for buildings energy consumption [J]. *Scientific Programm. (Pt.8)*, 2021 (2021)
7. Guo, C.: Application of computer technology in painting teaching in colleges and universities [J]. *Tobacco Regulatory Sci.* (2021)
8. Gao, L.: Research on application of data mining in virtual community of foreign language learning [J] (1), 3 (2018)
9. Na, Z., University, F.Y.: Construction and implementation of network course platform for foreign language teaching based on MOOC of Chinese Universities [J]. *Electron. Design Eng.* (2017)
10. Feng, Y., Zhang, J., Xie, P., et al.: Discussion on characteristic and categorization of traditional Chinese medicines for activating yang based on data mining method [J]. *J. Hunan Univ. Chin. Med.* (2017)

After putting all the characteristic cultural curriculum data together. Then solve each set again, we can be certain not achieve the classification, it will change, and the above steps must be repeated again [18]. Then put the test dataset into model validation. In this experiment, test dataset. Because the training method using the random gradient descent method ensures the rationality of parameter selection, we use the multiple times to find the optimal learning rate. In order to ensure that the optimal learning rate can be obtained, we will compare the results of multiple experiments. Each experiment will take a different learning rate for experiments. Due to the characteristics of the random gradient descent method, the selection of the learning rate is particularly important. The learning rate is equivalent to determining the speed of gradient descent. If the learning rate is too large, it the decline speed being too fast and missing the optimal value. If the selected learning rate is too small, “It is possible to obtain a local minimum instead of a global minimum. We can determine whether the selected learning rate is appropriate based on the corresponding model prediction accuracy when selecting the learning rate.”. We conducted comparative experiments with learning rates of 0.1, 0.5, and 0.8 by constructing models multiple times, and the results in Fig. 4 below:

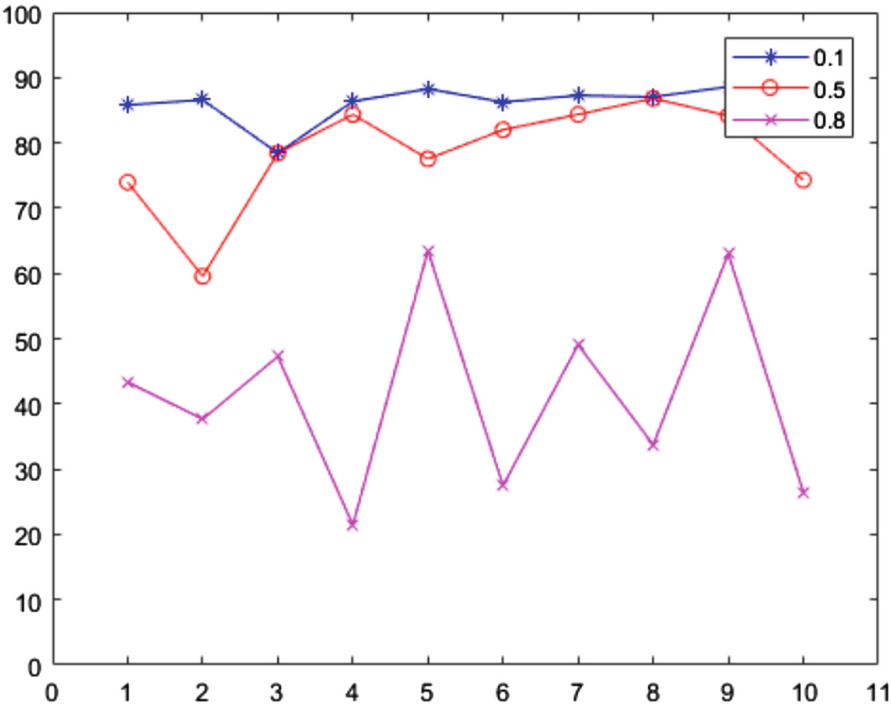


Fig. 4. Course selection accuracy

In the above Fig. 4, it mining studied in this article has the highest accuracy rate for setting up characteristic cultural courses.Because the results of mining using association rules have a certain degree of complexity, which is reflected in the diversity of mining

background to observe and understand, forming a comprehensive and correct understanding. In turn, these theoretical understandings will better guide practical activities and form a virtuous circle [14]. The practice link various the school, including venues and personnel, to achieve resource integration and optimization, and open up a variety of experience ways for foreign students, so that foreign students can feel the regional cultural characteristics of the region in a close and all-round way.

4 Big Data Course Selection Algorithm

The offering of Chinese cultural experience courses in a certain university is at the forefront of universities. Although there are many categories, it cannot be ignored that they are closely linked to Chinese cultural courses in the implementation of specific courses, although these courses are actually featured as Chinese cultural experience courses. From the perspective of the methods and teaching models offered, the teaching model of most Chinese cultural experience courses is very single, even breaking away from the four basic words of “cultural experience” in the textbook of cultural experience. The main reason is that there are limitations in teaching venues and class hour arrangements, and specific issues are reflected in the interview feedback from the teaching director.

Cluster analysis refers to grouping based on the information involved in the characteristic cultural curriculum data that describes objects or their relationships. There are various clustering algorithms, but not all of them are suitable for this experiment. Considering the efficiency of the algorithm, we give priority to selecting efficient clustering algorithms. One of the greatest advantages of clustering methods based on definition partitioning is their fast convergence rate and speed, and they are also simple, efficient, and low in time and complexity for some large characteristic cultural curriculum datasets. K-means algorithm is a typical clustering algorithm based on the distance between targets, using the distance between targets as an evaluation indicator of the similarity between targets, that is, the closer the distance between two targets is, the greater the likelihood that they will be clustered together.

Firstly, we need to determine the size n of a known selected sample featured cultural curriculum dataset, and the characteristic dimension of each selected sample featured cultural curriculum dataset is m . Recorded as:

$$D = (X_{1i}, X_{2i}, \dots, X_{mi}), i = 1, 2, 3, \dots, n \quad (1)$$

Then we need to set the required number of clusters, K . And we need to randomly select the final clustering center point that is the same as the set clustering number k in a characteristic culture curriculum dataset.

$$C = (C_{1j}, C_{2j}, \dots, C_{mj}), j = 1, 2, 3, \dots, k \quad (2)$$

Calculate the Euclidean distance between each centroid and each point in the data set of the characteristic cultural curriculum. According to the criterion of the closest distance, they are divided into corresponding classes that are closest to the cluster center (most similar). The calculation formula is as follows:

$$D(i, j) = \sqrt{(x_{1i} - c_{1j})^2 + (x_{2i} - c_{2j})^2 + \dots + (x_{mi} - c_{mj})^2} \quad (3)$$

in the characteristic culture, such as the writing method of words, music singing, dance movements, national handicraft production and other practical operations. It is opened for students of specific majors, focusing on the inheritance and education of characteristic culture for students.

Therefore, the digital teaching platform of foreign students' characteristic culture must be developed and designed in combination with general education and practical teaching courses to meet the teaching needs.

The curriculum is the core and concrete embodiment of the overall planning for cultivating talents in a school. The curriculum is divided and constructed according to the type of school, school functions, teaching objectives, learning methods, and requirements, as well as the organization and arrangement of school courses, as well as the division of labor and cooperation between courses. It mainly reflects the overall planning for the training objectives of the school, covering teaching objectives and disciplines, scope, content, progress, tasks, and activities. In addition, curriculum design is a dynamic fine-tuning process, and its changes are also affected by political, economic, cultural, and other aspects. Therefore, the author believes that curriculum setting refers to the setting and arrangement of various courses selected by schools according to market needs the goal of talent cultivation. In summary, this study believes that it is more reasonable to divide the curriculum content.

3.2 Application of Teaching Platform for Foreign Students' Characteristic Culture Courses

The teaching of foreign students' characteristic culture course based on data mining is a new experiential teaching method gradually accepted by everyone at present. It requires to first collect the representative cultural elements in the local area, and then compile the "guide book for the practice of foreign students' Cultural Courses" (hereinafter referred to as the "guide book") according to these factors to guide practice with theory. The curriculum can be divided into two parts: Theory and practice, which can be carried out alternately. At ordinary times, classes teach the cultural elements of the region, and then let foreign students choose a certain point to carry out practical activities, and use practice to test the theory. This teaching mode allows foreign students to participate in practice from the place where they study. Starting from the familiar places around them, on the one hand, it can eliminate their sense of strangeness and timidity, on the other hand, it is also easy to perceive cultural elements and feedback practical information. The general procedure of this experiential teaching mode characterized by regional culture is: teachers select regional culture practice projects (no more than 4 per semester) k. Before each project, foreign students preview book. In class, teachers explain relevant cultural projects and assign practical tasks. Foreign students carry out practical activities in groups, Then report the project results in the form of a written report or presentation, and the teachers and other team members will comment and give feedback together.

After the implementation of the characteristic culture course teaching for foreign students based on data mining, two obvious results will be produced. First, foreign students' understanding of regional cultural factors is no longer fragmented and unsystematic, but will put the fragmented individual cultural representation into a whole cultural

3 Teaching Platform of Foreign Students’ Characteristic Culture Courses Based on Data Mining

3.1 Design of Teaching Platform for Foreign Students’ Characteristic Culture Courses

Since the characteristic culture digital teaching platform is to carry forward and inherit the characteristic culture, the teaching platform constructed should have teaching characteristics, and the most important thing is to reflect the teaching design. Teaching design is the teaching strategies, other elements in an orderly and optimized manner, and forming a teaching plan, as in Fig. 3.

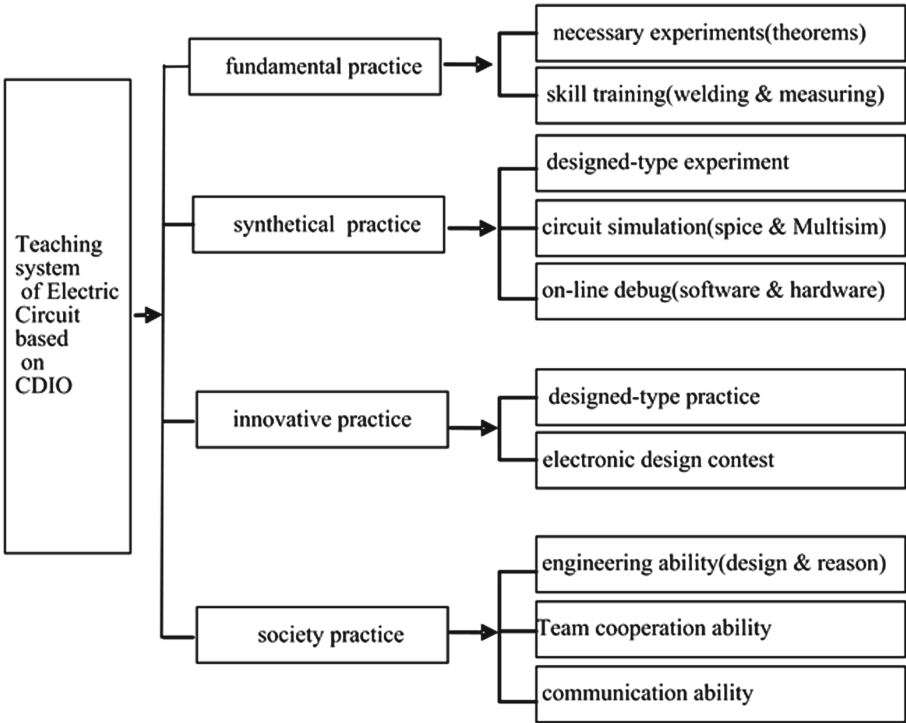


Fig. 3. Structure of course teaching system based on Data Mining

The courses generally offered can be divided into:

- (1) Special cultural general courses. It mainly introduces the relevant local characteristic culture in the form of words, pictures and videos, such as national characteristics, distribution, use of words, music and dance characteristics, so as to broaden the knowledge of characteristic culture, open it to all professional students, and focus on the education of humanistic spirit and scientific spirit for students.
- (2) Practical teaching courses of characteristic culture. It mainly focuses on skill learning, and uses video and teacher demonstration to learn some skills and culture

Firstly, the content of characteristic cultural courses for international students is mainly reflected in the following two aspects: firstly, it emphasizes the basic knowledge of traditional Chinese culture. Traditional Chinese studies, classic works, literature and art are key areas that international students need to focus on, helping them better understand the cultural background and history of China, and thus better understand and integrate into local society [14]. The second is to emphasize modern Chinese trend culture. For example, Chinese sports, movies, music, etc., allow international students to understand the development and trends of contemporary China, and feel the vitality and modernity of Chinese culture.

Secondly, the teaching forms and methods of characteristic cultural courses for international students are constantly being updated and improved, if a combination of online and offline teaching is adopted. Online teaching utilizes advanced means of digital technology, such as educational videos, online cultural courses, online reading, online quizzes, etc. It extensively utilizes the Internet and information technology to facilitate the learning of international students. Offline teaching includes various forms such as cultural relics inspection, cultural tours, literary and artistic activities, and on-site experiences, which can help international students better access and understand Chinese culture and society [15].

According to the survey, at present, Chinese culture courses are mainly offered to middle and senior foreign students. Although there are many kinds of textbooks, there are not many high-quality textbooks, and they mainly focus on theoretical introduction, and the design of practical links is less. There is no unified standard for the length and grade of Chinese culture courses in each school. In addition, the course content is mainly the introduction of Chinese mainstream traditional culture, while the regional cultural characteristics are less 3. Through the investigation of Chinese culture courses offered by foreign student teaching institutions in Shanghai, Beijing, Tianjin, Jiangsu, Anhui and so on.

It mainly introduces the theoretical knowledge of textbooks, with simple practical activities. This mode mainly focuses on the explanation of traditional cultural knowledge in textbooks, and students' activity space is centered on the classroom. Teachers intersperse one or two practical activities in the whole teaching activities. The classroom emphasizes the teachers. Students passively accept knowledge [8]. Its advantage is that teachers can and teach basic theoretical knowledge systematically and completely in unit time; Students can get and master more theoretical knowledge. However, this teaching mode overemphasizes the dominant position of teachers, which is suitable for the knowledge structure with strong theory, and is not suitable for the cultural knowledge that requires a lot of operation and participation. In the cultural courses taught in this teaching mode, foreign students generally do not have any cognitive experience after learning, which is prone to the problem of generalizing, making them think they know Chinese culture very well, and are prone to "cultural shock" in the process of practice.

China’s research and attention, mainly because in recent decades, the stage of transformation and development, and various new technologies and new ideas can not play a good the process of development of education. However, through the relevant research results published by many private teaching institutions or big data research institutions, we can find that, Big data has also achieved good results [11]. The knowledge structure of course selection is shown in Fig. 2 below.

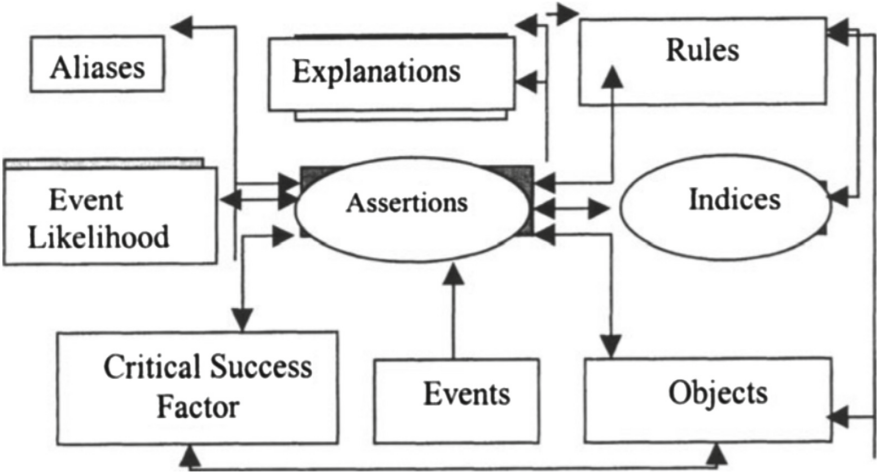


Fig. 2. Course selection knowledge structure

However, through the research on the of big data, this paper finds that the in China’s and marketing industry has been in the forefront of the world. Taking Alibaba as an example, Taobao under the group is the leading brand of China’s and marketing big data application, including, supercomputer and other technologies in the marketing industry, which has injected a strong shot into the education industry, China can integrate and transplant the particularity of the education industry with the advanced big data application experience of other industries in the form of industry transplantation to form China’s education big data platform [12].

2.2 Teaching Status of Characteristic Culture Courses for Foreign Students

Teaching characteristic cultural courses for international students has always been one of the key tasks of domestic universities. After arriving in China, international students not only need to learn professional knowledge, but also need to understand China’s traditional and folk culture in order to better integrate into Chinese society.

Most colleges and universities have cultural courses for foreign students, including language, culture, art, history and other courses [13]. With the increasing demand for international students in universities and the diversification of their needs, the content and form of characteristic cultural courses for international students are also constantly being updated and improved.

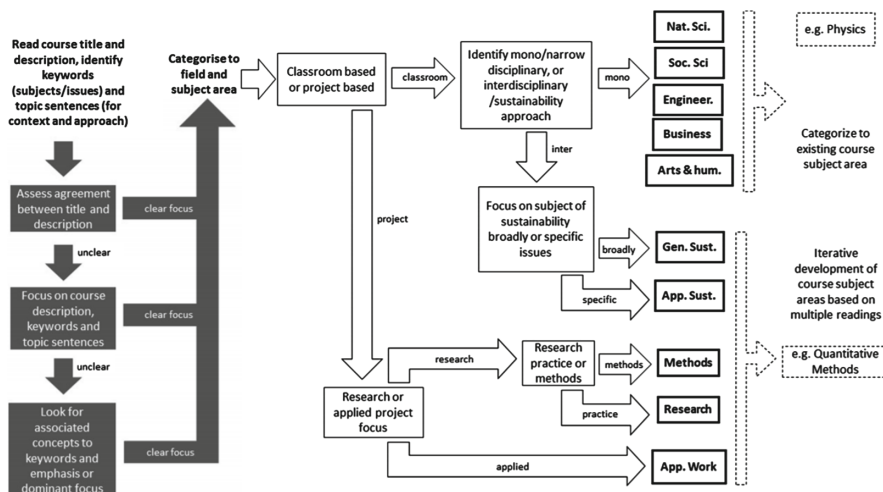


Fig. 1. Classification and Analysis of Curriculum Teaching Subjects

2 Related Work

2.1 Data Mining and Education Big Data

Data mining and education Big data analysis are popular research directions in modern education technology. Data mining is a technology to discover knowledge, information and patterns from a large amount of data, and education Big data analysis is to apply data mining to the education field, so as to realize the comprehensive mining, in-depth analysis and fine application of education data resources.

Education Big data analysis plays an increasingly important role in Education reform and Educational assessment. By analyzing, mining, and applying a large amount of data, comprehensive monitoring and analysis of students' learning behavior can be achieved, insight into students' learning trends and patterns can be provided, more accurate feedback and evaluation can be provided for teachers, and the improvement of educational fairness and quality can be promoted [9].

At the same time, education Big data analysis also helps to promote the realization of personalized teaching. By analyzing detailed data on student learning, personalized teaching plans can be developed based on different student characteristics and needs, achieving precise matching, and thereby enhancing students' interest and effectiveness in learning. In addition, education Big data analysis can also find existing education problems and improvement space, providing support for education decision-making and policies.

In 2000, the began to cooperate with its computer development center to out primary big data related on the U.S. education industry [10]. After 12 years, the released the Research Report on the promotion of education and analysis in 2012, which covers the research results of U.S. educators and big data researchers on American advanced, intermediate The cases of the combined and education in primary institutions, and summed up the relevant laws, have made a guiding and instructive role for the application worldwide.

and mining students' learning data, we can better understand the learning characteristics and needs of international students, and provide them with personalized teaching services and learning resources [2]. Therefore, the application prospects of data mining technology in the teaching of cultural courses for international students are very broad.

Curriculum setup is the most fundamental part of curriculum development and training programs. Many scholars often take curriculum setup into account when researching courses, linking it to training goals and requirements. The content of relevant courses varies depending on the teaching object [3]. The specific setting of course content also determines the form of teaching organization to be adopted, the venues, and application of methods. Therefore, the significance of curriculum design is significant, which shows that the rationality of curriculum design can directly affect the entire process of training. Once the rationality of the curriculum cannot be achieved, the school's goal of cultivating cultural awareness cannot be achieved [4]. In this context, the main significance of conducting research on Chinese cultural experience courses as follows:

Theoretical significance: The Chinese cultural experience course is a relatively new attempt in the cultivation and at present. This type of course can the foreign China for Chinese culture. Starting from the curriculum setting, this article explores the and advantages of the curriculum setting of the Chinese Culture Experience Course in Guizhou universities, the curriculum structure of the Chinese Culture Experience Course, which has great significance for the cultivation of cultural awareness of foreign students from Guizhou universities in China, and can also supplement the current relevant content of the curriculum setting [5].

Practical significance: From a small perspective, the cultural experience curriculum is conducive to the scientific standardization of the Chinese culture curriculum for international students in China. In the general direction, it is conducive to improving the quality and level of training and education for foreign students from universities and colleges in China, and it is also conducive to the steady [6]. The cultivation of international students in China not only includes their professional learning abilities and work and life abilities in China, but also increases the corresponding cultural knowledge to enable them to understand China and its culture, and become an ambassador and bridge for China to communicate with countries around the world in the future [7]. The classification analysis of curriculum teaching disciplines in Fig. 1 below.

Nowadays, Chinese culture curriculum has become an indispensable part of the curriculum for middle and senior foreign students, and the teaching mode of culture curriculum has gradually changed from teaching textbook knowledge to "walking on the ground and seeing with your own eyes", and practicing by yourself. Based of Chinese culture courses offered by foreign students' teaching institutions, and the experience centered research paradigm, this paper summarizes two mainstream modes of Chinese culture teaching, analyzes their advantages and disadvantages, and puts forward an experiential culture teaching mode and specific operation mode with regional culture as the entry point and "practice guide book for foreign students' culture courses" as the blueprint [8]. Therefore, mining, this paper designs and applies the teaching platform of foreign students' characteristic culture courses.



Design and Application of Teaching Platform for Foreign Students' Characteristic Culture Course Based on Data Mining Technology

Yan Xu^(✉) and Hui Yuan

Shandong Institute of Commerce and Technology, Jinan 250103, Shandong, China
yanyanxu1983@163.com

Abstract. This course aims to introduce the characteristics and culture of Chinese students. This course will help foreign students understand their own culture, which is a good reference for their study in China. Data mining technology is used to collect data from three sources: (1) questionnaire; (2) Observation and analysis; (3) Interview teachers of foreign students. SPSS software is used to analyze the collected data, and then TF-IDF algorithm is used to establish the model. The platform utilizes data mining technology to analyze and mine the learning data of cultural courses for international students. By studying the learning characteristics and needs of international students, we provide them with personalized teaching services and learning resources. The platform not only digitizes traditional offline teaching resources, but also integrates interactive methods such as socializing, Q&A, and resource exchange into its course platform, achieving online and offline communication, comprehensively meeting the teaching needs of international students' characteristic cultural courses. Practice has proven that this platform can effectively enhance the learning initiative and enthusiasm of international students, and enhance their understanding and cognitive level of Chinese culture.

Keyword: Data mining · Overseas student · Characteristic culture courses · Teaching platform

1 Introduction

With the intensification of globalization and the trend of internationalization, the number of international students is constantly increasing, and the teaching of cultural courses for international students is also receiving increasing attention. However, due to the background and cultural differences of international students, traditional cultural curriculum teaching methods are difficult to fully meet their needs. Therefore, in the teaching of cultural courses for international students, how to innovate teaching methods and improve teaching quality has become an important research direction [1].

This article designs a teaching platform for characteristic cultural courses for international students based on data mining technology. Data mining is a computer science technology that can find patterns and patterns from a large amount of data. By analyzing

9. Deng, J., Qing, H.: Practice teaching method of electronic commerce major based on data mining algorithm [J]. J. Phys. Conf. Ser. **1852**(2), 022039 (7pp) (2021)
10. Wang, X.: Construction of civil engineering teaching system based on data mining algorithm and big data technology [J]. J. Phys. Conf. Ser. **1852**(3), 032022 (6pp) (2021)
11. Chen, Z.: Using BIG DATA FUZZY K-means clustering and information fusion algorithm in english teaching ability evaluation[J]. Complexity **2021**(5), 1–9 (2021)
12. Han, Y.: Evaluation of English online teaching based on remote supervision algorithms and deep learning[J]. J. Intell. Fuzzy Syst. Appl. Eng. Technol. **4**, 40 (2021)
13. Hong, J., Qin, X.: Signal processing algorithms for SSVEP-based brain computer interface: state-of-the-art and recent developments[J]. J. Intell. Fuzzy Syst. Appl. Eng. Technol. **6**, 40 (2021)
14. Cobian, D.: Comprehensive efficiency analysis of machine learning algorithms for developing hardware-based cybersecurity countermeasures [J] (2022)
15. Du, Y., Zhao, T.: Network teaching technology based on big data mining and information fusion [J]. Secur. Commun. Netw. **2021**(9), 1–9 (2021)

in terms of teaching content, methods and effectiveness would be effective and quick to improve the teaching level of teachers in this category.

Through clustering, all teachers can be divided into categories, and the overall characteristics of teachers in each category can be analysed to obtain the performance of teachers in different categories on each indicator, so that teachers can focus on their own lower level indicators according to their performance in each category. At the same time, teaching managers can use the results to develop different teaching improvement programmes for different categories of teachers, such as organising for teachers in the lower categories to attend the classes of high level teachers.

5 Conclusions

Traditional assessment in the classroom usually includes peer assessment, teacher's peer assessment, teacher's self-assessment and student assessment. However, most of the existing assessment indicators are very general and focus mainly on aspects such as attitudes to teaching and IT. When these indicators are assessed from the students' perspective, they are often formal and do not truly reflect the strengths and weaknesses of teaching effectiveness, nor do they take into account the specific factors that affect teaching effectiveness. At the same time, the analysis of existing teaching outcomes is mainly in the form of statistical reports on teaching effectiveness, which is not only a heavy burden but also makes it difficult to find hidden information in the data base, which is why new data management techniques are needed to solve this problem." The 'Internet + Education' has become a trend that presents new opportunities and challenges for English at university. The use of DM technology to analyse and model student learning data and to explore the relationship between English exams and various TAL tasks through an English learning platform can be very beneficial to student learning and teacher education.

References

1. Chen, H.: A Data mining-based model for evaluating tennis players' training movements [J]. *Discrete Dyn. Nat. Soc.* (2022)
2. Xin, Y.: Analyzing the quality of business english teaching using multimedia data mining [J]. *Mobile Inf. Syst.* (2021)
3. Zhou, Q., Jiang, W.: Implementation of openairinterface-based real-world channel measurement for evaluating wireless transmission algorithms [J] (2021)
4. Huang, X., Wang, Q., Lu, S. et al.: Evaluating FFT-based algorithms for strided convolutions on ARMv8 architectures [J]. *Performan. Evaluat.* (2021) (152-Dec.)
5. Iftikhar, U., Asrar, K., Waqas, M. et al.: Evaluating the performance parameters of cryptographic algorithms for IOT-based devices[J] (2021)
6. Li, Z.: Elliptocyte detection technology based on super-resolution algorithms for a lensless imaging system [J]. *Measur. Sci. Technol.* **32**(2) (2021)
7. Zhang, A.: Influence of data mining technology in information analysis of human resource management on macroscopic economic management[J]. *PLoS ONE* **16**(5), e0251483 (2021)
8. Wang, A., Yu, H.: The construction and empirical analysis of the company's financial early warning model based on data mining algorithms [J]. *J. Mathematics* (2022)

Table 4. Mean values of assessment scores for each category

	Category 1	Category 2	Category 3
Indicator 1	−0.73	0.75	−2.22
Indicator 2	−0.7	0.74	−2.4
Indicator 3	−0.74	0.73	−1.8
Indicator 4	−0.7	0.73	−2.26
Indicator 5	−0.68	0.73	−2.38
Total Assessment Score	−0.73	0.74	−2.28

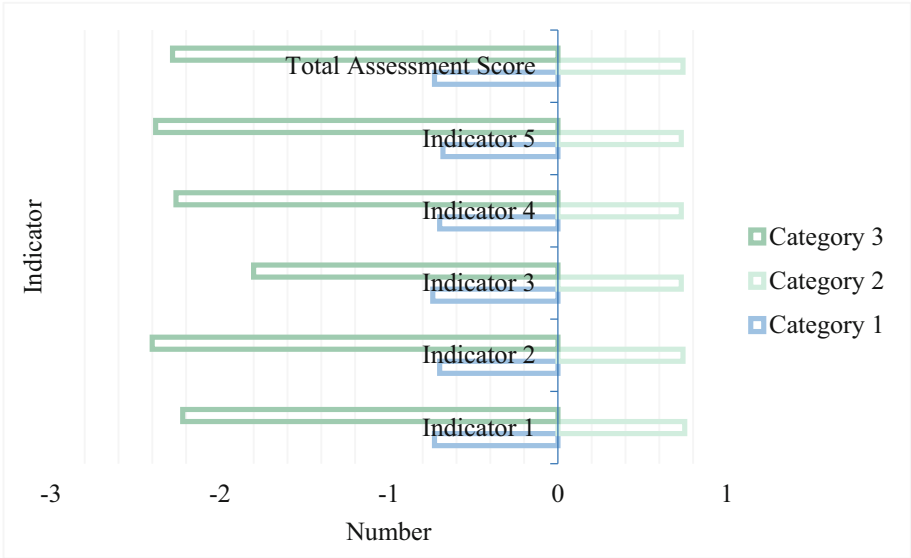


Fig. 4. Analysis of the mean scores for each category of assessment

categories on each indicator, with teachers in category 2 outperforming teachers in all indicators, followed by teachers in category 1, followed by teachers in category 3, who performed the worst. Category 2 teachers scored relatively evenly on all indicators, indicating that good teachers tend to be very demanding in all areas, and reminding teachers that mastering every teaching skill is a must, and that improving their teaching skills in all areas is something that all teachers should strive for. Teachers in categories 1 and 3 had below-average averages in all scores, and their teaching needs to be further improved. In particular, teachers in category 3 scored very low on all indicators, which means that teachers in this category urgently need to adjust their teaching methods, etc. Administrators should also combine student questionnaires and individual teacher interviews to improve the teaching standards of teachers in this category. Teachers in category 3 have the worst scores on indicators 2 and 5, which means that some measures

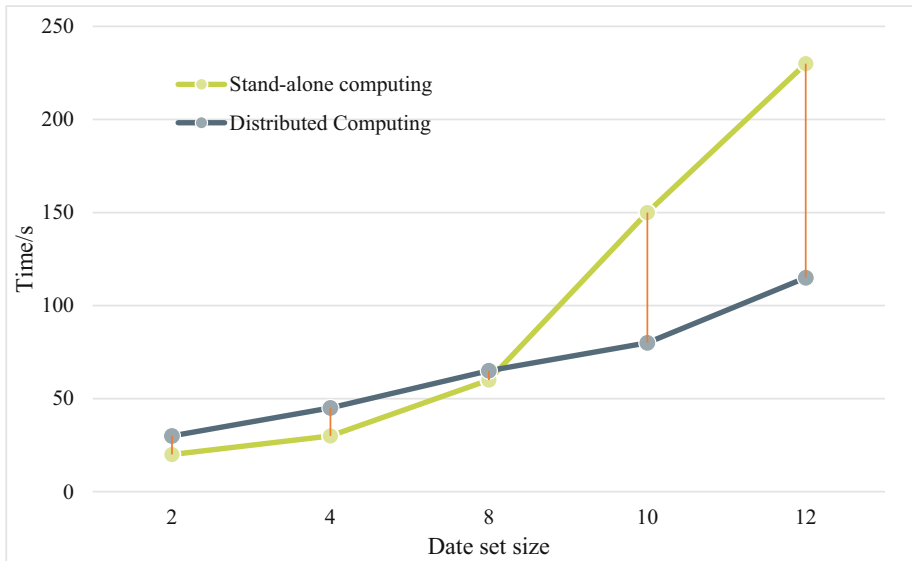


Fig. 3. Performance test chart of distributed DM algorithm

The comparison in Fig. 3 reveals that when distributed-based DM correlation analysis is performed on TT process, when the amount of data is small, the advantage of single-computer computing is more obvious in terms of running speed, and when the amount of data gradually increases, the advantage of the computing rate of distributed starts to be obvious because it is running in parallel on multiple hosts at the same time.

4.2 Clustering Analysis of TE Data

The k-means cluster analysis is used to automatically group professional teachers with the same or similar teaching ability, which is convenient for teaching managers to make grouping arrangements for the subsequent supervisory listening system and to guide teachers to carry out targeted teaching skill enhancement training. As there are many teachers in the college, it is impossible to give feedback on teaching analysis for each teacher. It is necessary to cluster teachers with the same characteristics into one category through clustering and then provide targeted advice measures for each category. The k-means clustering method of the SPSS statistical analysis software was used to cluster the student assessment scores of each teacher, and the number of clusters, K , was specified by the user in the experiment. The K value of 3 is therefore appropriate. The results of the analysis are as follows Table 4.

After the k-means cluster analysis by SPSS, the system will automatically give the category to which each teacher has been assigned. The final clustering centre, the mean of each category of teachers on each indicator and total score, will also be obtained, and the mean performance of teachers in each category on each indicator and total score will be plotted as a visual bar chart, as shown in Fig. 4. The clustering results overall appear to show a large difference in the mean scores of teachers in different

(3) MTP.

The MTP involves selecting an appropriate machine learning model to extract and analyse the processed ‘clean’ data, such as a classification problem, and selecting an appropriate classification algorithm. During the modelling process, parameters are continuously adjusted to achieve optimal results, and the training model often requires a return to the previous phase to reprocess previously trained data and features.

(4) Model validation and evaluation phase.

Once the training of the machine learning model is complete, the final search results need to be analysed and evaluated, which means that the model generated during the training process is tested against the data to validate the results. For example, accuracy and recall can be evaluated for the classification task and mean squared error for the regression task.

(5) Online use phase

At this stage the entire DM process is completed and the resulting model can be used for online development or to solve more complex problems. Figure 2 illustrates the five stages of the DM process.

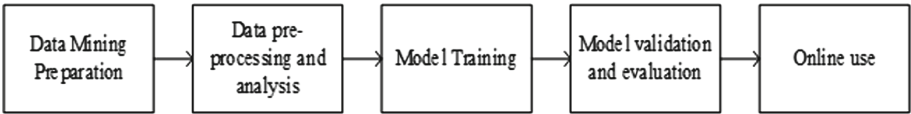


Fig. 2. The five stages of DM

4 Experiment Analysis

4.1 Experimental Comparison and Analysis

A region was selected to conduct an experimental comparative analysis of the performance of all teachers’ evaluation data over the last five years, as shown in Table 3.

Table 3. Distributed DM algorithm performance tests

Calculation/Time	2	4	8	10	12
Stand-alone computing	20	30	60	150	230
Distributed Computing	30	45	65	80	115

Table 2. TE data sheet

Field Name	Data Type	Field Length	Description	Remarks
Id	Int	4	Data Number	Primary Key
Course	Varchar	20	Course Name	Non-null
Major	Varchar	20	Major	Non-null
Sex	Varchar	4	Gender	Non-null
Nation	Varchar	8	Ethnicity	Non-null
A	Int	4	Indicator 1	Non-null
B	Int	4	Indicator 2	Non-null
C	Int	4	Indicator 3	Non-null
E	Int	4	Indicator 4	Non-null
F	Int	4	Indicator 5	Non-null
G	Int	4	Indicator 6	Non-null
I	Int	4	Indicator 7	Non-null
J	Int	4	Indicator 8	Non-null
L	Int	4	Indicator 9	Non-null
N	Int	4	Indicator 10	Non-null
P	Int	4	Overall Rating	NULL

(1) Preparation phase.

The DM preparation phase focuses on understanding the project objectives and business and user requirements, defining the problem and initially identifying search and data collection objectives. There are many methods of data collection, such as questionnaires, interviews, observations and discussions. Initial process definition and DM definition. Each method has more or less unique characteristics. One of the main sources of massive data in the Internet industry is log files, which record information about users and different behaviours and activities, or data acquired through technologies like data loggers. The data source for this study was the data produced when students logged into the ELT system for practice and examinations in English, which was stored in a database.

(2) Pre-processing and data analysis stage.

In the pre-processing and data analysis stage, the raw data collected is processed and incomplete, inconsistent and noisy data is processed and integrated using pre-processing techniques and methods to obtain 'clean' data for information extraction; in this stage, information extraction tasks such as classification, regression, etc. are identified. At this stage, information extraction tasks, whether classification, regression, etc., must be identified. Feature extraction methods generate features in the sample data.

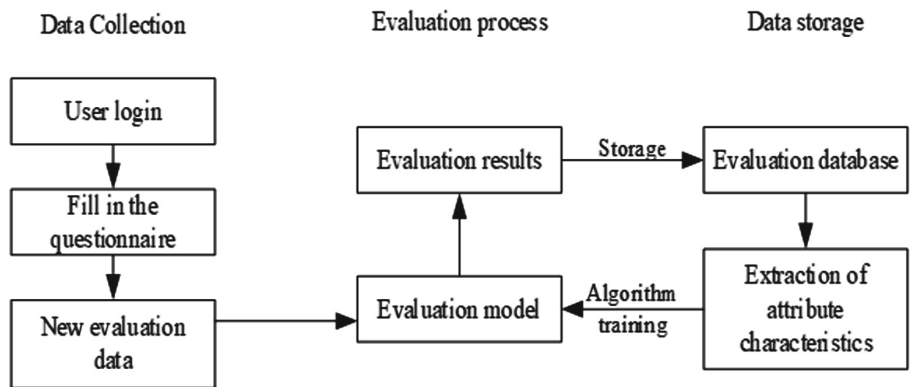


Fig. 1. Components of the Student Assessment Module

(3) Database design

In this paper, MySQL 5.7 is used as the relational database to store data. The two data tables designed in the database are the user information table and the student assessment data table.

Firstly, the basic information data table is shown in Table 1 below.

Table 1. User information table

Field Name	Type	Name	Password
Data Type	Varchar	Varchar	Varchar
Field Length	8	10	10
Description	User type	User name/account number	Password
Remarks	Not null	Primary key	Non-null

Next is the student assessment data table, which records the evaluation subjects, basic information about the evaluator, the results of the 10 evaluation indicators and the overall evaluation value. The specific fields are shown in Table 2 below.

3.2 DM Process

In general, the ABM process can be divided into five phases: preparation phase, data pre-processing and analysis phase, Model training phase (MTP), model validation and evaluation phase, and online development phase.

After the above calculation steps, the order relationship of this indicator set can be derived as

$$m_1 > m_2 > m_3 > \dots > m_n \quad (1)$$

Step 2: Determine the relative importance between neighbouring indicators w_{k-1} and w_k . Denoted as

$$q_k = w_{k-1} / w_k, k = n, n-1, n-2, \dots, 3, 2, 1; \quad (2)$$

This allows the relative importance of each neighbouring indicator to be determined according to the ordinal relationship derived in the first step, and to facilitate comparison we can set the least important indicator $q_k = 1$.

Step 3: Calculate the weight values for each indicator. When the group of experts are all measured objectively, it is clear that there are

$$q_{k-1} \geq q_k, k = n, n-1, n-2, \dots, 3, 2, 1; \quad (3)$$

$$w = \{1 + \sum_{k=2}^n \prod_{i=k}^n q_i\}^{-1} \quad (4)$$

$$W_{k-1} = p_k W_k, k = n, n-1, n-2, \dots, 3, 2, 1. \quad (5)$$

3 Experimental Study

3.1 Design and Implementation of the Teaching Evaluation (TE) Module

(1) Overall design of the module

TE system usually includes several modules such as expert evaluation, teachers' mutual evaluation, teachers' self-evaluation and students' evaluation. This paper mainly implements the student evaluation module from the perspective of student learning, and other modules can be evaluated according to their corresponding evaluation indexes, using the same evaluation algorithm to construct models and complete the evaluation.

The student evaluation module implemented in this paper is divided into two parts: the front-end visualization interface and the back-end data processing. The front-end interface includes the collection of data, the display of evaluation results, etc.; the back-end data processing includes the storage of data, the invocation of algorithms, etc.

(2) Main function analysis

The main functions of the student assessment module designed in this paper include: user login and registration, enquiry of assessment records, collection of assessment questionnaire data and application of assessment algorithms, etc.

The main components of the module are shown in Fig. 1 below.

who are the main source of knowledge for students, can also reflect the TQ by evaluating the TT performance. Past evaluation approaches have also focused on these two aspects, but the drawback is that either the students' perspective or the teachers' perspective is evaluated, and less research has been done to evaluate multiple perspectives together [12].

2.2 Monitoring Content

Based on the many factors that influence the quality of TAL, such as the school environment and conditions, teacher conditions, management levels, educational and teaching activities, and academic quality, it constitutes outcome monitoring and process monitoring. The content of monitoring will be determined as follows [13].

(1) School-level monitoring content

The content of school-level monitoring includes: teachers' classroom teaching plans and teaching processes; SL behaviour and moral quality; subject tests and stage examination results and test paper answers.

(2) District and county level monitoring

The content of monitoring at the district and county levels includes: the completion of the school curriculum plan and curriculum objectives; the structure of teachers; the teaching environment and school management; the teaching process of teachers; the learning behaviour and moral quality of students; and the results of subject tests and stage general examinations and the answers to examination papers.

(3) Construction of the indicator system

The establishment of a system of monitoring indicators is a key element of monitoring. An organic set of indicators and criteria for monitoring at all levels will facilitate quantitative and qualitative analysis of technical quality. The document defines the structure of the monitoring indicator system based on four aspects of monitoring: the quality of school management, the teaching and learning process of teachers, the learning behaviour of students and academic quality [14].

2.3 Steps of G1 Method to Calculate Weights

Step 1: Determine the sequential relationship.

Let the set of assessment indicators be noted as $m_1, m_2, \dots, m_n$, the group of experts to compare the importance of these indicators to judge, and in these indicators they think the most important one, recorded as m_1 ; and then in the remaining $n-1$ indicators selected as long as the most important one, recorded as m_2 ; and so on, in the remaining $n-(k-1)$ indicators in this way selected the most important indicators, respectively level as m_k ; until the selection, the indicator until only the last indicator remains in the set, which is recorded as m_n [15].

DM technology is a product of the rapid development of information technology. The use of DM methods can capture more finely the data of changes at all levels and their complex relationships, helping decision makers to have a clearer grasp of the current situation, monitor TQ more comprehensively, and effectively regulate the factors affecting TQ, so as to achieve the purpose of safeguarding the quality of education. In this work, the SPSS and Weka DM tools are used to conduct an in-depth study and analysis of various data on the current state of the teaching model, to explore the complex relationships behind the data and their value and meaning, and then to propose additional methods and measures on how to effectively use DM to implement TQ monitoring and to explore how to optimize the use of student assessment data in conjunction with the results of the analysis To explore how to conduct [8].The study also explores how the results of the analysis can be used to optimise the use of student assessment data, improve the student academic monitoring and curriculum early warning system, and innovate academic evaluation and career guidance methods, so as to provide meaningful references for the implementation of TQ monitoring.

2 Design Research

2.1 Direction of Evaluation

The Microsoft SQL Server 2000 database can scientifically organize and manage a large amount of data, and can analyze and count these data. And has excellent sharing and manageability. The selection of a database should fully consider the actual needs of the system, not only to meet the needs of system operation, but also to facilitate the user and programming habits of developers. In this system, I have selected the Microsoft sQLserver 2000 database. As a database management system, Microsoft SQL Server 2000 provides stable data storage and management, which is fully suitable for the needs of this system. These functions are sufficient for this system [9]. Microsoft SQL Server 2000 has powerful network processing capabilities, support for data warehouses, and online transaction processing capabilities.

The assessment of the TQ and learning (TAL) is also based solely on the good or bad results of students' examinations or the high or low evaluation of teachers to make a simple assessment, the data collection process is cumbersome, such assessment lacks systematization and comprehensiveness, and in today's complex educational situation, the results are mostly neither practical nor accurate.

Nowadays, the value implied by BD has gradually become the focus of attention in all walks of life, and the processing technology of BD is becoming increasingly mature, ushering in the best development opportunities for education [10]. Using big data (BD) to serve education comprehensively and adopting evidence-based teaching allows students to be more aware of their learning, allowing teachers to adjust their education methods in a timely manner and allowing the education system to be more rationally transformed.

Teaching quality (TQ) is an important criterion for measuring teaching and guiding teaching reform, and can objectively reflect the level of educational attainment and the extent to which educational outcomes are good or bad [11]. The evaluation of TQ can be based on two aspects: firstly, students, who are the most important subject in teaching, can objectively reflect the TQ by analysing their learning effectiveness; secondly, teachers,

In a related study, Evangelia et al. explored the effectiveness of a pedagogical approach using information and communication technology (ICT), based on Van Hiele's level of geometric thinking and Hoffer's skills in describing geometric thinking, for teaching the English alphabet [2]. The results of the study showed that the impact of the intervention on SL of the English alphabet was statistically significant and that this impact varied according to students' skills, which could be attributed to a number of factors. The Sedat study investigated Turkish pre-service English teachers' perceptions on the FCM and LMS platforms Google Classroom [3]. The findings show that service learning teachers have a very positive view of FCM in terms of student motivation, effectiveness, engagement and overall satisfaction. They also found that classes required more time to prepare for lessons, that lesson preparation had a positive impact on motivation, that each participant was given the opportunity to learn at their own pace, and that learning by doing was encouraged through increased practice. It was also noted that participants were satisfied with the teachers' attitude towards the online platform [4].

The integration of the global economy has made English a universal language as early as the early 1990s. The deepening of international scientific, economic, and cultural exchanges has set higher requirements for the college English teaching level of Chinese universities and educators. The country urgently needs to cultivate high-quality practical English talents with solid basic knowledge. The typical problem with English teaching quality evaluation methods has always been that the data is not comprehensive enough and the evaluation is not objective enough [5]. The emergence of data mining technology, and the improvement and maturity of massive data classification and recognition algorithms provide the possibility to solve these problems. Yu Chaoyang pointed out the idea of combining data analysis technology with teaching quality evaluation system in his "AHP-PCA-SVM Evaluation of Higher Mathematics Teaching Quality", Expounded the theoretical significance and practical attempt of teaching evaluation of big data technology application "Chen Liangdi, Yang Bo, and others have made discrete improvements to the kernel function of the SVM model for teaching quality monitoring and evaluation, which can effectively improve the objectivity of teaching quality monitoring. Yang Zhao and other experts proposed applying the Logistic algorithm to the evaluation of English teaching quality, selecting 0.47 as a cross-sectional area, and conducting segmentation experiments to confirm that the accuracy of artificial statistical recognition has greatly improved [6]. Based on previous research experience, due to the Logistic regression model Type is closely related to the parameter values of boundary points, and unreasonable parameter settings can lead to a greater probability of miscalculation. This paper proposes an improved Logistic algorithm and introduces SVM support vector machines to combine with this algorithm to reduce the impact of boundary points on the algorithm, improve the accuracy of classification recognition and feature extraction, in order to improve the level of college English quality teaching in China, Objectively, it provides an objective, scientific, and reasonable evaluation method to solve the problems existing in the process of English teaching evaluation, such as the unreasonable formulation of teaching content, the single teaching method, the uneven level of teachers, and the incomplete teaching evaluation system [7].



Data Mining Technology-Based Algorithms for Evaluating English Language Teaching Indicators

Wang Qian^(✉), Wang Jiaxin, and Zhou Huixing

Beijing Polytechnic, Beijing 100176, China
stephaniewq@163.com, wangjiaxin@bpi.edu.cn

Abstract. TE is a judgement on the value of teachers' teaching (TT) and students' learning (SL), and has become an important part of the teaching management and teaching process in universities. There are many common TE systems, most of which evaluate the behavioural performance of teachers, while the learning process and effectiveness of students are rarely mentioned. At the same time, the workflow of implementing TE is tedious and often requires the completion of a large number of data calculation tasks. Therefore how to use modern science and technology to establish a sound, objective and feasible classroom TE system and optimise the evaluation process is an important issue that needs to be addressed urgently. The main objective of this paper is to conduct a study on the evaluation algorithm of English teaching indicators based on data mining (DM) technology. Starting from the construction of a learning-centred university TE system, this paper optimises student TE indicators by using data correlation analysis and association rules. At the same time, machine learning algorithms are introduced into the TE process to build TE models and automate the TE process. Through clustering, the experiment can divide all teachers into corresponding categories, analyse the overall characteristics of teachers in each category, and obtain the performance of teachers in different categories in each indicator, and teachers can focus on their lower level indicators according to the performance of each indicator in their respective categories.

Keywords: Data Mining · English Language Teaching · Teaching Indicators · Indicator Evaluation

1 Introduction

Currently, many forms of TAL have emerged in the education sector. Although, with the rapid development of the Internet, online education is becoming more and more popular, traditional classroom teaching still has an important and fundamental position in campus teaching. At the same time, classroom TE, as a necessary part of teaching management, is of great importance to the development of schools. TE mainly studies the value of TT and SL, and its evaluation includes both the evaluation of TT work and the evaluation of SL outcomes [1].

8. Hou, N., Safwan Samsir, M.: Application and exploration of virtual reality technology in the teaching of sports anatomy. *International Conference on Information Systems and Intelligent Applications*. Springer, Cham (2023). https://doi.org/10.1007/978-3-031-16865-9_28
9. Xie, C., Zhang, W., Dang, Z. et al.: Chestnut pellicle as a new potential source of antioxidants: antioxidant activity, inhibition on cooking oil oxidation, and identification of phenolic compounds by HPLC–ESI–MS/MS. *Europ. Food Res. Technol.* **249**(6), 1589–1598 (2023). <https://doi.org/10.1007/s00217-023-04239-4>
10. Gul, F.S., Bahader, A., Nawaz, M. et al.: Evaluation of the impact of alkyl silane entity (APTES) on photocatalyst (TiO₂) and their photocatalytic degradation of organic compounds for environmental remediation. *Environ. Monitoring Assess.* **195**(6) (2023). <https://doi.org/10.1007/s10661-023-11363-x>
11. Kuchmenko, T.A., Dorovskaya, E.S., Menzhulina, D.A. et al.: Evaluation of a possibility of the application of “electronic nose” to monitoring deviations in the functional obesity of children in hospital conditions: observations of changes during treatment. *J. Analyt. Chem.* **78**(3), 390–401 (2023). <https://doi.org/10.1134/S1061934823020053>
12. He, C., Jin, Y., Zhang, Y. et al.: The pharmacokinetics and safety of lidocaine in liver cancer patients undergoing hepatic resection. *Europ. J. Clinical Pharmacol.* **79**(6), 829–839 (2023). <https://doi.org/10.1007/s00228-023-03498-0>
13. Jeong, S.Y., Kim, S., Lee, C.L. et al.: Sea urchin removal and effectiveness evaluation from the long-term monitoring data in Dokdo, Korea. *Ocean Sci. J.* **58**(1) (2023). <https://doi.org/10.1007/s12601-022-00094-8>
14. Qu, L., Chai, F., Liu, S. et al.: Comprehensive evaluation method of urban air quality statistics based on environmental monitoring data and its application. *J. Environ. Sci. (China)* **123**, 500–509 (2023). <https://doi.org/10.1016/j.jes.2022.10.003>
15. Jiang, J., Zhu, W., Xie, X., et al.: Indoor environment quality monitoring and evaluation system based on lora communication. *J. Comput. Commun.* **10**(4), 15 (2022)

according to the data differences in different environments. According to the design, it can extract relevant original data from the business, and transfer it to the data integration area, waiting for the next step of data processing and processing. The monitor module is mainly responsible for automatically monitoring the data changes in each data source and notifying the data integration module of the changes. After receiving the change message, the integration module extracts the relevant data into the data integration area.

4 Conclusion

Data mining technology has important application value in the quality in universities. This technology can analyze and mine a large amount of teaching data, thereby helping educational institutions understand information about students' learning situation, teachers' teaching quality, and the utilization, providing scientific basis and decision-making support. At the same time, data mining technology can also help educational institutions establish a teaching quality evaluation index system and prediction model, thereby achieving fast and accurate teaching quality evaluation. In practical applications, data mining technology can also discover and extract knowledge through machine learning algorithms, mining valuable teaching knowledge and teaching quality. The exploration of data mining technology in China is relatively limited, but the selected data sources and volumes, as well as the knowledge obtained through mining, may not meet the needs of university managers. What I did was just a beginning, a wise decision. The size of the data is smaller than, and the complexity of time and space is relatively large. The current can only the next stage. Our goal is to find a fast and effective algorithm. In summary, the application in the monitoring and teaching quality in universities has high practical value and application.

References

1. Li, Z.: Research on the application of data mining in the quality analysis of college English teaching. EAI International Conference, BigIoT-EDU. Springer, Cham (2023). https://doi.org/10.1007/978-3-031-23944-1_43
2. Wang, S., Bin Nazarudin, M.N.: Research on the application of virtual reality technology in physical education in colleges and universities. Springer, Cham (2023). https://doi.org/10.1007/978-3-031-16865-9_29
3. Wei, J., Guo, J.: Research on the application of data mining technology in physical training (2023). https://doi.org/10.1007/978-3-031-23944-1_44
4. Qi, N., Deng, F.: Research on teaching development and application of data mining algorithm in Japanese corpus (2023). https://doi.org/10.1007/978-3-031-23944-1_22
5. Wang, P.: Research on the application of computer technology in the dissemination and promotion of folk art culture (2023). https://doi.org/10.1007/978-3-031-23944-1_42
6. Chong, N.: Research on the evaluation of college English classroom teaching quality based on triangular fuzzy number. EAI International Conference, BigIoT-EDU. Springer, Cham (2023). https://doi.org/10.1007/978-3-031-23947-2_5
7. Peng, Y., Yi, J.: Research on the application of big data technology in the process of cross-border E-commerce product selection. International Conference on Decision Science & Management. Springer, Singapore (2023). https://doi.org/10.1007/978-981-19-2768-3_3

3.2 Overall Architecture of Application System

The quality in universities is an important component of higher education. Currently, in the context of digitization and informatization, from the lack of data to the data explosion, there is already enough data in university education scenarios to support a large number of excellent decisions. Researchers advocate the use of data technology for refined evaluation of teaching quality.

In general, the teaching will continue to develop iteratively and integrate new Technological convergence such Big data into the system, which will of teaching quality and the popularization of teaching management methods with science and technology as input, so that the teaching and universities will truly move towards refinement, intelligence and visualization. The system in this paper takes the data warehouse as the core, integrates data collection, modeling, management, analysis and presentation, and is designed according to the hierarchical method. The overall framework of the in Fig. 2.

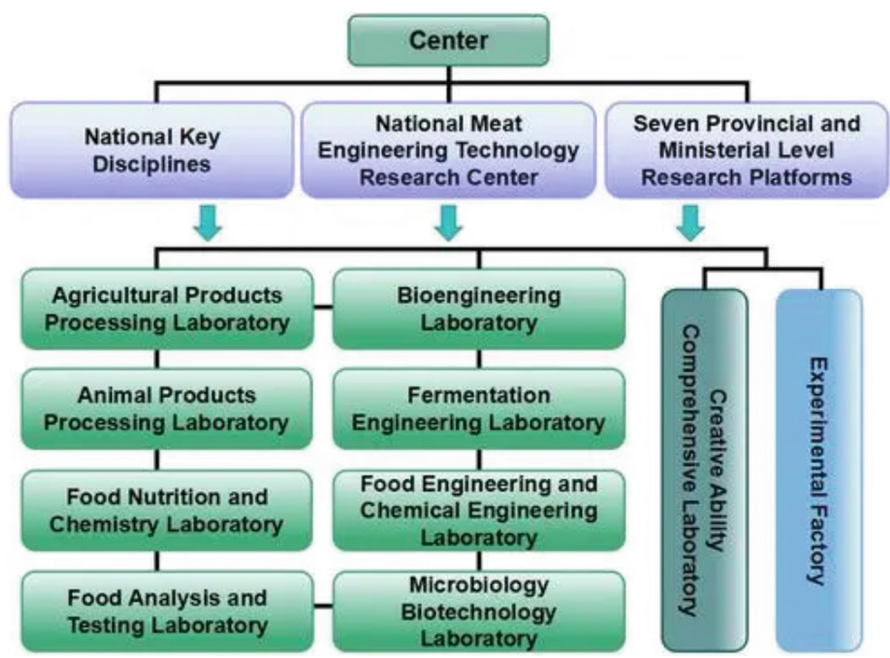


Fig. 2. Structure of teaching quality monitoring system in colleges and universities

The underlying architecture of the system. The data source has the characteristics of distribution, heterogeneity, autonomy, etc. It can the basic situation of the classroom, the situation of teachers' learning and relevant government regulations; The format of these data can be text, database, or report and file. The data related to teaching quality analysis in the data source is copied by the data extraction module and transferred to the data warehouse to provide a consistent data support platform for high-level query and analysis. The data extraction module can set different intermediate processing processes

3 The Application of Data Mining in the University Teaching Quality Monitoring

3.1 Establishment of Data Mining Model

Data mining is the from a large amount of data, and its core is to establish a data mining model to achieve the goal of converting data into useful information. The establishment of a data mining model can generally the following steps:

(1) Data collection and preprocessing

First, we need to collect sets according to the needs of mining. In the, we need to consider the integrity and quality of the data. The data may come from different data sources, and we need to Data cleansing the data, remove abnormal data and some useless data.

(2) Feature selection

The features in data play the accuracy and efficiency of data mining models. Choosing appropriate features is crucial. For feature selection, certain statistical testing methods can be used to determine the corresponding weights of each feature and filter out low weight features.

(3) Model selection and establishment

Before selecting a model, it is necessary to first consider the specific situation and characteristics of the target [15]. In different problems and scenarios, different models need to be selected, such as classification, prediction, clustering, association rules, etc. At the same time, different model parameters also need to be adjusted and optimized based on actual situations, in the accuracy and robustness of machine learning.

(4) Model training and testing

The training process of the is to make the model predict or classify the required targets by inputting the original data, so as to obtain the prediction results. In the process of model training, factors such as data set division, batch training, training rounds and Learning rate need to be considered. After the training is completed, independent data needs to be tested to obtain the evaluation results of the model.

(5) Model Application and Update

In the actual application process, the constructed model needs to be applied to the actual data to get the actual results. Meanwhile, due to the dynamic nature of the data, the model also needs to be constantly updated and adjusted.

In summary, the establishment of data mining models is a core part of the entire data mining process. In this process, it comprehensively consider the data, the needs of the problem, and the selection of models, in order to ultimately generate efficient and reliable models, and improve the efficiency and accuracy of data mining.

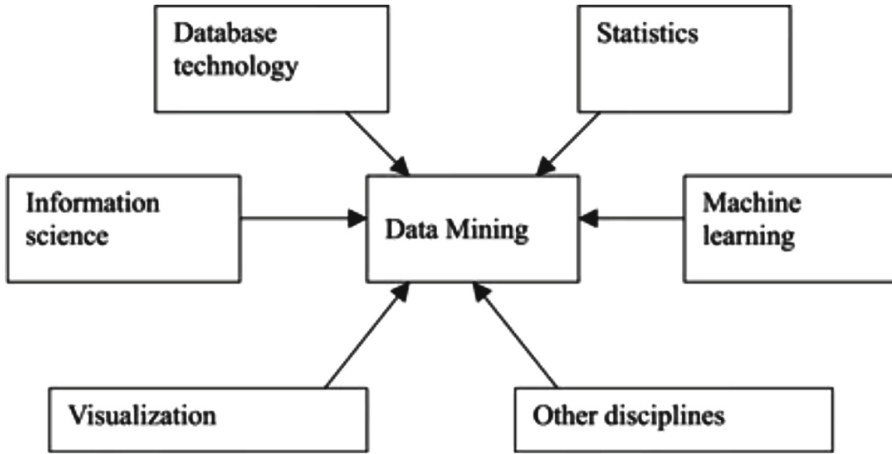


Fig. 1. Several main functions of data mining

- 1) Statistical method: This method can help determine the predicted value of property and estimate the distribution of attribute values based on similar data analysis. Regression analysis is a common prediction method. Regression can not only describe the also predict based on this relationship.
 - 2) Association Rule Method: association rules can be used as target attribute values of budgets. This is the property of the target attribute described by the conditional attribute test. The rule can be used to predict the value of the target value.
- (2) Describe function

The descriptive function is mainly used to find out the common properties of data, highlight the common characteristics of such data and discover the relationship between data [12].

- 1) Data classification: Data classification is to find a group of objects and create a class group of different things from the same properties and characteristics in the database. Classification is a supervised learning method to establish a classification model. Some sample data will be selected as the sample database of the training set, and some sample data will be used as the test set.
- 2) Data summary and induction: The concept level is a structure that often exists in the database and describes the subordination between data [13]. There are two main. The data cube method, which pre calculates the frequently queried and expensive operations and stores them in the array [14]. The second is the attribute oriented induction method, which expresses that data mining uses a query language similar to SQL to query requests and the use of property, so as to eliminate the functions of concept hierarchy tree, threshold control, communication and a group of data convergence, and finally becomes various rules.

artificial intelligence, the teaching process is analyzed, and quantitative teaching quality indicators are proposed to achieve refined management of education.

(2) Multidimensional teaching quality monitoring

In order to comprehensively monitor teaching quality, the recent research trend is to develop the monitoring and evaluation of teaching quality systems from a single perspective to multiple dimensions [6]. In addition, multidimensional standards help promote better education for students, not only focusing on their academic performance, but also taking into account factors such as teacher effectiveness, satisfaction with the teaching process, and student feedback.

In the process, the teaching quality are developing and dynamic, not constant. Therefore, in the we should use a dynamic perspective to look at this key factor. The evaluation mostly adopts the evaluation by students. Generally, the educational administration department sends the teacher's teaching quality evaluation card to students at the middle or end of the semester. Teachers are scored according to the evaluation items in the evaluation card [7]. After the statistics of the educational administration department, the teacher's evaluation according to the scoring results. However, due to the randomness of students in the process of scoring, different majors have different tendencies towards different courses, and some human factors interfere, students' scoring cannot truly reflect the actual teaching effect [8]. Under this evaluation method, we can only obtain simple and one-sided evaluation results, but can not analyze the evaluation data, so that the teaching evaluation can the guiding role of teaching. How to express, analyze, explain, publish and use the results of teaching evaluation is a new topic worth studying at present [9].

In short, with the continuous strengthening and application of digital technology, the universities is entering a new era of diversification, with accurate and multi-directional coordinates, and will lead the development trend of future changes in research on monitoring quality in universities [10]. Therefore, strengthening the scientific and data-driven management of education, achieving the effectiveness and accuracy of education monitoring and evaluation, is one of the current development directions for universities.

2.2 Data Mining Functions

The while the latter refers to the discovery that patterns. Figure 1 shows the data mining.

The introduction to these two types of functions.

(1) Prediction and verification functions

Prediction refers to the prediction of unknown data from known data. For example, some workers' wages are predicted based on the wages of other workers in the same unit. Trend analysis is to analyze the trend of spatio-temporal data based on past behavior and predict possible future events [11]. Change of things can be obtained through trend analysis. Trend analysis and prediction, named some of them.

The monitoring and evaluation system for teaching quality in universities utilizes data mining technology and is generally divided into the following three steps:

(1) Data preprocessing

Collect and organize teaching data related to various disciplines over the years, including student grades, teacher evaluations, inductive question types, teaching progress, etc., to eliminate sampling errors such as abnormal data and missing data, making it more accurate and reliable.

(2) Data mining

Using data mining techniques, perform low dimensional dimensionality reduction, classification/clustering, association/sequence analysis on the preprocessed dataset, and determine quality evaluation indicators through data correlation and various types of analysis modeling, thus obtaining quantitative and scientific results of teaching quality evaluation.

(3) Analyze the results and make decisions

By comparing data mining results with teaching quality objectives, identifying problems and their causes, providing feedback and measures, and making targeted adjustments and improvements to achieve the goal [3].

In summary, the broad application prospects in the monitoring and evaluation system of teaching quality in universities. In practical applications, it is necessary to carefully select data and technologies based on the characteristics and development needs of various disciplines, analyze corresponding data, greatly improve the scientific level of teaching quality monitoring and evaluation, help universities achieve scientific, systematic, and standardized teaching management, and promote innovation [4].

2 Related Work

2.1 Research Status of Teaching Quality Monitoring and Evaluation System in Colleges and Universities

The monitoring and evaluation quality is an important component. Currently, in the digitization and informatization, from the lack of data to the data explosion, there is already enough data in university education scenarios to support a large number of excellent decisions. Researchers advocate the use of data technology for refined evaluation of teaching quality [5]. Driven by the development of digitalization, the current research status and new trends in the monitoring quality in universities are as follows.

(1) Leading Research in Data Technology

In recent years, with the acquisition, processing, analysis, mining and evaluation of teaching quality data has become the core issue of the current construction and evaluation system. Based on relevant technologies such as data mining, machine learning, and



Research on the Application of Data Mining Technology in College Teaching Quality Monitoring and Evaluation System

Miaomiao Xu^{1,2(✉)}, Xie Suofei^{1,2}, and Xie Ping^{1,2}

¹ Anhui Xinhua University, Hefei 230088, Anhui, China
xmm0109232@163.com

² Xianyang Normal College, University of the Cordilleras, Xianyang 712000, Shaanxi, China

Abstract. With the advent of the era of knowledge and information, the web-based management system has been widely used, which not only promotes the innovation of higher education management system, but also improves the management level of teaching management departments. Because there are a lot of data in the university educational administration database, it is very important to find useful information from the database in the evaluation system. This paper makes a quantitative study on the technology in the monitoring and college teaching quality. This research is technology. The first stage of this study aims to determine the main characteristics of teaching quality, which come from different sources, such as students' evaluation, teachers' evaluation and students' views on teachers. The second stage is to use these characteristics.

Keyword: Teaching quality · Data mining · Monitor · Assessment

1 Introduction

The monitoring and in universities is an important link in educational management. With the development of Big data technology, data an indispensable part of the education quality monitoring and evaluation system, providing scientific and digital means for college teaching quality monitoring, improving the efficiency and accuracy of guiding decision-making and improving teaching [1].

Data mining technology can analyze and mine a large amount of educational data (such as students' academic performance, exam time, course selection, course evaluation, etc.) to discover problems or information in quality monitoring [2]. It not only saves manpower, material resources, and time costs, but also greatly improves the monitoring level and evaluation effect, and has great application value.

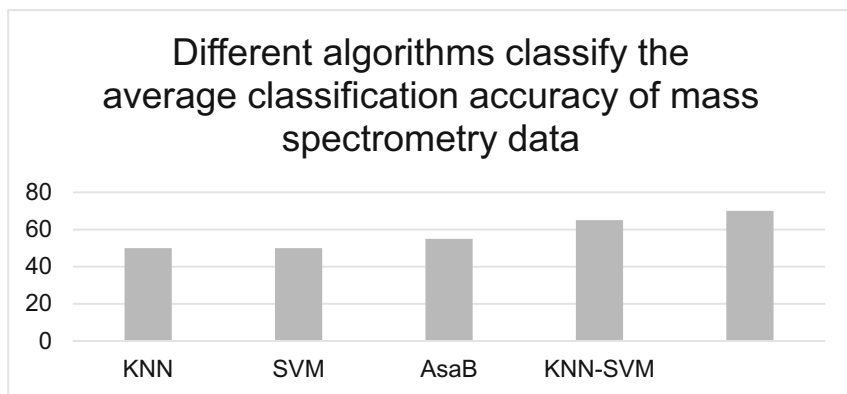


Fig. 2. Statistical graph of the average classification accuracy of mass spectrometry data classified by different algorithms

present, Boosting algorithm has been applied to the research of QSAR, but it focuses on solving the classification problem, and applying Boosting algorithm to QSAR research to solve the regression problem, which proves that this algorithm is widely used in drug molecular efficacy prediction.

5 Conclusion

To sum up, Boosting algorithm is the focus of attention in machine learning, and has received close attention from a wide range of industry insiders. Although this algorithm highlights some advantages, there are still some problems to be solved and studied. However, how to use Boosting algorithm to process large-scale chemical detection data is also a key problem to be solved in chemical data mining.

References

1. Dong, L., Cui, X.: Intelligent data mining algorithm for picking robot execution system optimization. *Agric. Mech. Res.* **45** (7), 224–227 + 237 (2023)
2. Qiu, D., Cai, X., Wu, L., Liang, S., Li, J.: Design of command ticket visual system based on data mining algorithm. *Electron. Design. Eng.* **30**(21), 61–65 (2022)
3. Li, J.: International road cost accounting method based on data mining algorithm Sankei, China **19**, 117–119 (2022)
4. Cai, L., Wei, X., Wang, J.: Analysis and application of data mining algorithm in big data. *J. Honghe College* **20**(05), 154–157 (2022)

approaches, As can be seen from the investigation and study, Although the learner was not significant in terms of learning performance alone, And in many cases, learning performance will be a significant leap, The combination is more suitable for solving large-scale data sets than the learners alone, This also, to some extent, represents some potential in data mining (Fig. 1).

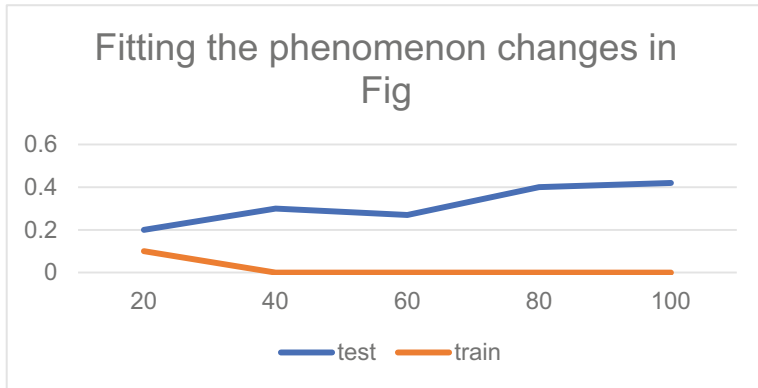


Fig. 1. Fitting the trend of phenomenon

From an overall perspective, the following Boosting combination learning methods have been applied in many fields:

4.1 Atlas Analysis

Integrating Boosting and Bayes optimization rules has developed a brand new algorithm that is used to handle the detection data classification of 215 patients with thyroid disease, saturated alcohol and ether mass spectrometry data classification, and mass spectrometry data investigation substructure classification. According to relevant investigations and studies, the two algorithms have been greatly improved to different degrees, and this method is used to carry out second-class prediction and multi-class prediction for many mass spectrometry data. Using Boosting combined classification regression tree as a classifier, can fundamentally on the green tea, wheat and cream, HIV carriers such as medical detection data retention index classification, found that part of the data classification effect is significant, but in the green tea data classification during overfitting, found Boosting results in chromatographic retention index classification is not ideal, shows that Boosting is a kind of excessively sensitive to data, noise, etc. (Fig. 2).

4.2 Prediction of Drug Efficacy

The Boosting method can be used to deal with the toxicity prediction such as nitrobenzene and lipooxidase inhibitors. After multiple linear regression and support vector regression comparison, the Boosting is slightly better in the prediction accuracy, which can provide more practical information and data for the diagnosis and treatment of tumor patients. At

implementation of combined classifier usually includes weighted stock, unauthorized stock and gate network, etc. Boosting uses weighted stock. For classification problems, the weights are often obtained by measuring the accuracy of each member classifier in the training set and other levels, and the weight and accuracy hold the proportional relationship. 5.5 The combustion combination method mainly adopts the naive Bayesian method to learn and master the weight-important points. The weight of the classifier is usually obtained by the assumed accuracy measured during the weighted training distribution [3].

4 The Application and Research Direction of Boosting in Data Mining

At present, the difficulties of machine learning and knowledge discovery in data mining are reflected in the large-scale and high-dimensional characteristics of data. In order to fundamentally meet the diversified requirements of massive data, many machine learning algorithms should be improved and improved according to the actual situation.

Learning in large-scale data. Three different decision tree algorithms can be extended to the large-scale data processing work. First, it is based on the intelligent sampling of training data during the growth of the decision tree, and the stage of selecting and testing for the tree often includes the analysis of training data and the prediction of output class characteristics. Among the many redundant dataset areas, it is possible to choose from data-based sampling. The second is based on developing smart data structures to prevent the storage of other training data in memory space. The SPRINT method can completely disperse the training data and integrate different examples or feature values into independent disk files. SLIQ usually consumes more memory than SPRINT, but the overall speed is relatively fast. The third is the use of decision tree integration, training data can be reasonably divided into multiple disjoint subsets, each subset can also independently generate independent decision tree, these trees can also be jointly voted to predict, although the accuracy of the individual tree is relatively low, but the overall combination of performance value is good, and can speed up the efficiency and quality [4].

Learn about high-dimensional data. The current C 4.5 and BP algorithms do not have good scale characteristics, However, at a statistical level, There are numerous examples of irrelevant or polynoised features providing the corresponding informative data, Therefore, we need to reasonably screen the characteristics according to the actual situation, There are three methods in this case: First, Provide a learning algorithm for performing an initial analysis of the training data and selecting a subset of features; second, Try to differentiate feature subsets based on learning algorithms, And estimate the algorithm performance supported by these different types of subsets, Ensure that well-performing subsets are available for subsequent use; third, The screening and weighting of features are all integrated into the learning algorithm. Although existing methods can be learned using scientific and reasonable computational time among many training examples, However, the existing prediction learning accuracy should not be adopted when the quality of the predictor cannot be greatly improved, The idea of a combination of learning came into being, It often gives matching results to the learner through multiple

3.2 Training Error Analysis

Schapire and Singer highlighted hypothesis training error boundaries during generalization of Freund and Schapice, If each weak hypothesis is somewhat better than the random conjecture, Constrained γ is the farthest from 0, Training errors will decrease exponentially, There are also some similarities between the Ada Boost and the Boosting algorithm, However, the previous adopted algorithms need to obtain a known lower bound γ before practical learning, There is relatively little information associated with such boundary knowledge in practical life, While Ada Boost can fundamentally adjust the error rate of weak assumptions, Therefore, highlight the advantages of adaptation.

3.3 Margin Theory

$\sqrt{\frac{TD}{m}}$ Freund and Schapire use VC dimension to further explore the generalization error of Boosting from the perspective of training error. VC dimension is the measure of learning algorithm complexity and learning ability, which can fundamentally derive the generalization error value. The main formula is $\Pr(H(x) = y) + O(\cdot)$. During this period, m mainly represents the sample number, d is VC dimension, T is the training rounds, P is the empirical probability of training set, this formula fully shows that if the training wheel value T is too large, Boosting will lead to overadaptation, but after a series of research can be seen that Boosting does not lead to overadaptation, in this case, Schapire pair training case margin for multi-level analysis, large positive boundary represents high credibility accuracy prediction, large negative boundary credibility error prediction. However, the smaller boundaries represent the predictions with low confidence. The interpretation of Schapire is mainly reflected that after the training error is reduced to 0, Boosting can still maintain the original state lifting boundary, further increase the minimum boundary, increase the reliability of classification, and reduce the overall error problem.

3.4 Combination and Classification Technology

Combined classifiers are a unified integration of classifiers, whose individual decisions are often combined through a pathway approach, and they expand the classification processing of new samples in this case. In the supervised learning link, the research field of high activity is reflected in the well-constructed classifier combination, according to the relevant experiments can show that the single classifier opinion difference during the combination of the accuracy is higher than the separate classifier, if the individual classifier finally obtained assumption error rate exceeds 0, then represents the final voting results and errors will increase. Therefore, the final assembled combination method uses individual classifiers with an error rate below zero, and these individual classifier errors are not correlated at some level. Combination classifier involved in special use and general due to two types, general technology combined with its own characteristics can be divided into change on the training case distribution, control the output target, introduce randomness, etc., for the combination of special application involves using targeted measures to implement training neural network B P algorithm or decision tree produce combination, special application need to analyze and explore the actual situation. The

and then focus on the classification activities of the samples with high classification difficulty in the next classification learning stage. In the Boosting, By modifying the data information by applying the weights to the training observations, The more the weight, the stool represents the greater effect of the observation on the classifier, Initial classes of samples were also given the same weights, To implement the first step of training the classification, Classifier construction by implementing a classification algorithm on samples in each cycle stage, And using the weights to measure the error of the classifier, Increase the weights of the samples misclassified by the classifier, Reducing the weights of the samples correctly classified by the classifier, The modification of the weights also highlights the corrected data to some extent, Promote the next study to focus on difficult to classify, The final classifier was obtained by weight comparison [1].

For the characteristics of Boosting, it highlights the advantages of simplicity, high efficiency and easy programming. In addition to the iteration number T , it can reduce the adjustment of parameters and other links, without the prior knowledge of weak learners. Therefore, diversified methods can be used to explore weak assumptions according to the actual situation. At the same time, each training set of Boosting is not in an independent state, and its own choice is greatly correlated with the previous learning results. The prediction functions of Boosting have high weights and often can only be generated in the specified order. In addition, the Boosting method requires a small change in the classifier without stability, namely, in the training set, which can cause significant changes to the classification model, such as support vector machines, neural networks and decision trees. Although Boosting presents many advantages, but there are also many shortcomings. The actual performance of Boosting at some level and in specific problems often over-relies on weak learners and data content, which is consistent with theoretical knowledge. In insufficient data sets, Boosting cannot play its own real power on the too weak or too complex level, and is too sensitive to the noise generated by the data set.

3 Basic Content of the Boosting Algorithm

3.1 Main Algorithms

The Ada Boost algorithm is a typical algorithm of the Boosting family. Its basic goal is to find out the hypothesis $H(x)$ according to the provided example x and the prediction tag y by using boosting. Its pseudo-code is reflected as the algorithm input: the training set $S = (x_1, y_1), (x_2, y_2), \dots, (x_m, y_m)$. In this case, each member is a labeled training example, $x \in X$, X represents a domain or an instance space, $y \in Y$ a label set Y , learners can accept (x, y) is randomly selected from $X \times Y$ distributed to P , assuming two types of problems, $Y = -1, +1$, the basis of a diversification problem, Ada Boost repeatedly call a given weak or basic learning algorithm, the main idea is usually the training set of weight distribution, distribution weights on t can be recorded as D_t . In the initial case, the weight of all examples is set to be equal, but the instance and weight of each misscore will continue to increase, thus reducing the difficulty that the learner is forced to focus in the training set. In general, the fundamental task of weak learning is embodied by combining the distribution D_t to find the matching weak hypothesis h_t [2].



Boosting Algorithm and Its Application in Secondary Chemistry Data Mining

Yingshuang Wu^{1(✉)}, Shijie Zhang¹, Yingshuang Wu¹, and Rong Zhang²

¹ College of Chemistry and Materials Science, Jiangsu Normal University, Xuzhou 221116, Jiangsu, China

wuletttetu@163.com

² Xingtai University, Xingtai 054000, Hebei, China

Abstract. In order to fundamentally improve the accuracy of data mining algorithm, efficiency and quality of the algorithm, the introduction of condition mode and depth priority strategy optimization data mining algorithm, please improve the data mining algorithm in various fields, by deleting invalid or using low frequency method to accelerate the resource positioning work, and improve the performance of the system resource positioning. This paper mainly combines the practical application of Boosting algorithm in middle school chemical data mining for in-depth analysis and discussion for reference.

Keywords: Boosting algorithm · Chemical data mining · Application

1 Introduction

In data mining, classification and prediction is a basic task. The traditional practice usually uses pattern recognition and statistical methods to obtain prediction rules. However, the proposal of Boosting algorithm can effectively solve the problems of blocked predictor accuracy.

2 Basic Overview of the Boosting Algorithm

The basic goal of Boosting algorithm is to use circular iteration to transform weak classifier science into classifier with low classification error rate, and then produce multiple rough estimates with good accuracy and low difficulty. Finally, these rules are processed comprehensively processed to form high-precision estimation, which belongs to the category of learning model. Boosting is a loop-class algorithm that can retain the weights of the samples in the training set to match it and reflect its own importance, and correct the data information according to the actual situation, to ensure that the learner concentrates in the differentiated samples, thus producing classifiers of different types. In addition, the algorithm can combine the classification errors of the generated single classification model in each cycle stage, adjust and change the weights of the training samples in time,

2.6 Predict Students' Behavior

The school roll management system makes statistics on the annual tuition fee and academic status, and uses data mining technology to classify, analyze and sort out the existing files, so as to better forecast the tuition fee of the new semester. Each class can formulate corresponding countermeasures according to this situation. Similarly, the use of the classification method can be used to assist the analysis of the admission situation of the current year, and formulate the admission plan, so as to reasonably arrange the admission of students. It will also be helpful to the students' enrollment strategy. In-depth analysis of students' shopping behavior can guide other plans to make corresponding changes, and conduct data analysis according to the meal consumption data of each restaurant in the school, and guide the location arrangement of the restaurant.

3 Conclusion

In a word, in the daily management of the school, the student management is a complex and tedious work, which is related to the students' school status, academic performance and employment. In order to do a good job in student management, many universities have established a relatively complete teaching information management system, and accumulated a large number of data and data. Therefore, the data mining technology is applied to the management of college students, to find the relevant important links from a large number of existing data, and to find some potential laws, so as to improve the scientific, targeted and effective nature of college student management.

References

1. Wenwei, C., Diao, H., Xinyu, Z.: The Data Mining Technology [M], pp. 1–5. Beijing University of Technology Press, Beijing (2002)
2. Lujie, J., Jing, Z.: The application of data mining in university academic affairs administration [J]. *China Sci. Technol. Inf.* (12), 225–226 (2007)
3. Liu, M., Xi, L., Yongsheng, L.: U1. *Comput. Eng. Design* **5**, 1130–1133 (2010)
4. Feng, P.: The application of data mining technology in university teaching management [J]. *J. Chongqing Univ. Sci. Technol. (Soc. Sci. Edn)* **4**, 100–101 (2008)
5. Guanghong, W., Ping, J.: Data mining review [J]. *J. Tongji Univ.* **32**(2), 1–7 (2004)

2.2 Improve the Teacher-Student Relationship

In universities, both teachers and students are learning, and their work, social activities and rewards are recorded in the school database. Using the correlation in data mining, we find out the various internal connections between teachers and students. Carefully observe and analyze the relationship between a teacher and a classmate, whether it helps them in their studies, or whether some of their behavior will work on their thinking mode, and if beneficial, it will continue. If the relationship between teachers and students is not good, we can change teachers and change teaching methods to promote the teacher-student relationship, so as to promote the development of the teacher-student relationship. For the students at school, their mental state and learning hobbies are very good.

2.3 Reasonable Arrangement of Courses

Students' learning is step by step, and there is some connection between various disciplines, which requires colleges and universities to follow a certain order when opening courses. Courses, from simple to complex, and based on pure courses, can we learn more in-depth courses well, otherwise it will affect the learning and absorption of subsequent courses. In addition, different classes, different teachers, the same course and homework performance are also very different. This paper discusses how to use data mining technology to conduct scientific and efficient planning and management, so as to achieve better teaching effect. According to the previous teaching database, each course of students of each grade and the teaching situation of each teacher is analyzed, and data mining technology is used to analyze the relevant analysis and time series, from which useful information is extracted [5].

2.4 Correct Evaluation of the Learning Situation

Making a proper evaluation of students' learning activities and achievements can not only enable students to get better feedback and motivation, but also play a good role in promoting teaching programs, teaching programs, teaching procedures and other aspects, so as to achieve the purpose of teaching students in accordance with their aptitude. Using data mining technology, the academic performance database and behavior record database were built, and the reward and punishment database was analyzed and processed, which can obtain students' scores in real time and correct them, so as to help teachers accurately understand each student and reduce the learning burden.

2.5 Formulate a Student Training Plan

The student data information database includes graduates' business adaptation, employment status and other information, and collects positive and negative views from different groups, such as: the importance order of each subject, the implementation of teaching and practice links, strengthening and expansion, etc. Finally, a student information management system suitable for the school. Use data mining technology to classify, summarize and process students' feedback, so as to provide students with a most suitable training program for themselves, so that students' management is more clear.

Table 2. Converted data tables

xh	k1	k2	k3	k4	...
4106005001	79.8	67.2	86.8	87	...
4106005002	75	79.2	92.4	90	...
4106005003	77.4	78.4	97.4	91	...
4106005004	77.4	72	97.2	96	...
4106005005	81.4	70.4	94.2	91	...
...	...	...	...	...	...

2.1.3 Data Conversion

Conform the relevant data into a standard data storage mode. Based on this solution, one or more sets of data are used to obtain the corresponding set of optimization criteria, thus in the corresponding results. Learning the mining of score correlation analysis requires logical data, so to convert the score table data into a Boolean type, then what we need to do is to explore the good connection between the majors, so the value “1” of 90 points or higher means that there is no such item in the transaction. Form of information obtained in Form 3:

Table 3. Student Achievement Affairs Data Sheet

xh	k1	k2	k3	k4	...
4106005001	0	0	0	0	...
4106005002	0	0	1	1	...
4106005003	0	0	1	1	...
4106005004	0	0	1	1	...
4106005005	0	0	1	1	...
		...	...	...	...

2.1.4 Association Principle

In data mining, if the Apriori algorithm: 0–2 and confidence 0–6 is applied, the data shown in Table 3. is input into the SQLSever2000Analysis service, and the Apriori algorithm is used for data mining. Statistics on each piece of data in the database and the minimum amount of support leads to valuable conclusions. Table 3. lists certain relevant rules:

Table 1. .

Student ID	Surname and personal name	Course number	Curriculum name	Mark	Credit	Class hour	Attributes of course selection	School year semester
4106005001	Feng Xiao	b05011	Sports 1	80	2	32	optional	Autumn 2006
4106005001	Feng Xiao	b05015	College English	79.8	4	64	Optional	Autumn 2006
4106005001	Feng Xiao	b05016a	College English II	67.2	4	72	Obligatory	Spring 2007
4106005001	Feng Xiao	b05017a	College English III	76.8	4	72	Obligatory	Autumn 2007
4106005001	Feng Xiao	b05020	Higher mathematics I	86.8	4	64	Obligatory	Autumn 2006
4106005001	Feng Xiao	b05021	Higher mathematics II	61	6	108	Obligatory	Spring 2007
4106005001	Feng Xiao	b05022a	Iatrochemistry	87	5	80	Optional	Autumn 2006
4106005001	Feng Xiao	b05024	Linear algebra	73	2	36	Obligatory	Spring 2007
4106005001	Feng Xiao	b05025	General physics I	75.27	4	72	Obligatory	Spring 2007
4106005001	Feng Xiao	b05026a	General physics II	95	4	72	Obligatory	Autumn 2007
					...	...	...	...

2.1.2 Preprocessing of the Data

In the process of data mining, data preprocessing is a very critical link, occupying 70% of the whole data mining work. Only after the necessary processing of the data can we obtain meaningful conclusions. Practice shows that a lot of time will be saved during the process of model construction. Change the data storage mode, use irrelevant entries as the optimal algorithm, and replace the text name with marks, such as k1 for English Ik2, and English Iik3 for advanced math Ik4 for medical chemistry. We standardized the original data with the above parameters as variables and converted them into a one-dimensional matrix type. Obtain the data format as shown in Table 2 [4]:

be an appropriate evaluation way, in order to effectively identify them. The so-called prediction, is the use of the obtained model, to classify the data objects forecast [3].

(4) Clustering

Clustering is the division of data sets with the same characteristics or meeting the similar requirements into several groups. Compared with the categories, the method is an unbootstrap treatment, and the sample category and sample number are not understood.

(5) Decision-making tree method

The decision tree algorithm uses the relevant information to find the maximum amount of information, establishes a node in the decision tree, and uses multiple nodes to construct a new node, and repeatedly generates a node and branch on that node.

2 Application of Data Mining in Student Management

This paper takes the results of the Department of Biomedical Engineering of the 2020 medical school as the main data, using SQL Sever 2000 Analysis Service and Apriori technology to calculate and process the data:

2.1 Assist Students in Characteristic Analysis

First, the students' basic information is collected, including learning history, learning preferences, knowledge structure, etc., and the ability of data mining is used to comprehensively analyze the data, and provide corresponding guidance for each person according to the results of the analysis. Finally, the students' bad learning behavior and habits are corrected, so that they are consistent with the established code of conduct, so as to improve their learning effect, and develop the comprehensive quality of the students in an all-round way.

2.1.1 Data Collection

The results of the first three semesters of the two classes in 2020 were selected as the original data, and were based on the original database. Through the analysis of the name, class and grade of each major, the relevant relationship between each major, and the relationship between each major and other related majors are obtained. This database has 51 students, 21 subjects, and 1,071 records. See Table 1:

study of correlation law has received increasing attention. Among them, the relevant guidelines mean that there is some possible connection between the data items in the database, so that the use of resources can be strengthened by mining the possible rules, so as to improve the quality of services.

The main application object of association rules is the database of object type, setting $\{I = i_1, i_2, \dots, i_m\}$ as a collection of projects. Where a transaction was filed up to the server. There will be one or more actions, including about events, relationships, time, producing a series of results. Task-related information D is where each T is a collection of items that meet the TI requirements. Each element in the transaction is a one-to-one counterpart. Each transaction has an identification code, called a TID . When a transaction occurs in a particular area, the transaction is either of a table. Suppose that A is a collection of items containing A , and only in the case of AT . The association rule is like $A \rightarrow B$, AI , BI , and $ANB = \Phi$. In the case of some connection between two sets, a set of association rules, namely association rules, can be constructed. Rules are established on the transaction set $DA \rightarrow B$, Yes, in D , S is the firm's AUB , or AU , or AUB . In this way, the association rule can be determined based on this proportion. If the D transaction set D has a trust c , then $A-B$ is reliable, and the conditional likelihood $P(B/A)$ if the transaction containing A also includes a B ratio is c . So, we can tell the truth from the perspective of the rules:

$$\text{support}(A \rightarrow B) = P(AUB)$$

$$\text{confidence}(A \rightarrow B) = P(B/A)$$

Rules that can satisfy both the minimum support threshold (min_up) and the minimum trusted threshold (min_conf) are called a strong rule. After an association rule is determined, the properties of the rule are first reduced, and then the maximum similarity and the corresponding matrix are solved. There are many ways to determine whether it is more rules. The former provides for the minimum amount of support required for the association rule. The latter is a minimum amount of information, requiring the user's degree of association and the most critical set of association rules. The latter refers to whether the association rule has reached the maximum credibility.

It contains the k -term, and there are an infinite number of association rules in the transaction database without considering the support and trust of the association rules. These association rules have an important application value in data mining. But we often focus that these association rules meet a degree of reliability requirements. Mining association rules usually employ the Apriori algorithm.

(3) Classification and prediction

There are generally two steps: building a classifier and using the classifier to classify the data. The first approach is to create new libraries of training samples, and the second approach is to apply them to the classification. The so-called taxonomy is conducted under guidance, where the training set of examples is the already determined category, and the number of species can be determined. However, in practice, because of the training samples obtained and the data to be identified are very different, so there must



The Application of Data Mining Algorithm in College Student Management

Yikun Mu¹(✉) and Maomao Cheng²

¹ Yunnanyiyaojiankanzhiyexueyuan, Kuming 650033, China
muyikun1213@163.com

² Jiangxi Medical College, Shangrao 334000, Jiangxi, China

Abstract. As the network technology is more and more widely used in learning, life and other fields, the data mining algorithm is also gradually integrated into the management work of colleges and universities, and it is applied to a large number of transaction activities. Through the analysis of management data, the use of data mining technology, improve student management, improve the quality of teaching.

Keyword: Data mining algorithm · College student management · Application

1 Overview of Data Mining Algorithm

(1) Definition of data mining

Data mining (DataMining) is to extract and mine useful data from a large number of data, so as to find the correlation between the data. This is a deep data processing technology, but also a deep data processing technology, database, neural network, model identification, mathematical statistics. Integrate fuzzy mathematics with other related technologies, apply rough set technology to various industries, and achieve good results. When performing specific operations, certain steps or steps are often taken repeatedly. On this basis, combined with the practical work of college graduates, a set of simple and feasible data collection procedures and application model are proposed. Data mining technology for association analysis, classification, prediction and clustering methods, decision tree methods and other techniques are more suitable for the management of college students [2].

(2) Association analysis

Association rule mining is an important part of data mining. This paper analyzes and discusses the relevant problems in the data mining in detail, and points out the relevance of these data, which is the basis of further deepening the data mining. This law is generally applied in other ways. Among the data mining methods, the most important and the most active is the connection law method. With the increasing demand for knowledge, the

4 Conclusion

Previously, it was based on work experience and the thinking methods of pedagogy and education management. At present, it adopts the scientific research method of systematic thinking based on data mining technology to think about the constraints of various influencing factors in college students' community education management by constructing the influence model of College Students' community education management system and using the feedback basic model research method, Scientifically put forward corresponding management policies to make up for the fragmentary and subjectivity of previous relevant research. In the process of student behavior mining and analysis, due to the limitation of data sources, only some student data are collected. The evaluation indicators of students' behavior can't fully reflect the students' situation at school. For example, if the clustering results are used to assist the work of the school funding department, the rest of the students' situation needs to be analyzed. Thinking about the influencing factors of student community management systematically, and putting forward a scientific and suitable student community management model for local universities at present, will help the socialization reform of university logistics to better serve the education and teaching reform of universities, and can be transplanted to other work areas of university logistics service guarantee, which will help to improve the connotation quality of universities.

Acknowledgements. Project: Hainan Provincial Natural Science Foundation of 2021 High-level Talents Project "Research and application of Student Service group clustering and community model based on big Data mode" (No. 621RC611).

References

1. Jianping, G., Dong, D.: University data mining analysis based on clustering and sequential clustering [J]. *Comput. Knowl. Technol. Acad. Edn.* **16**(25), 3 (2020)
2. Xiao, L., Zhihong, L., Yunjiang, X.: Modeling and analysis of user innovation community knowledge network based on clustering [J]. *Comput. Simul.* **4**, 5 (2016)
3. Jingxin, T., Yajun, W.: Application of data analysis technology in community correction information system [J]. *Comput. Eng. Appl.* **53**(20), 6 (2017)
4. Kai, H.: Exploration on quantitative learning methods of community in the era of big data [J]. *Indus. Sci. Technol. Innov.* **23**, 2 (2020)
5. Fei, X., Xiaoxu, S.: Community health information space construction model and service design [J]. *Library Theory Pract.* **3**, 4 (2019)
6. Jun, W., He, X.: Construction and research of higher vocational student behavior analysis model based on data mining [J]. *Modern Inf. Technol.* **4**(18), 3 (2020)
7. Mei, W.: Reflections on the construction of the "six in one" service-oriented student work system [J]. *Sci. Educ. J.* **22**, 2 (2018)
8. Jie, Z.: Reconstruction of the organizational framework of service-oriented student sports associations in colleges and universities [J]. *Sports Sci. Technol.* **38**(6), 3 (2017)
9. Zhizhong, C.: Application of data mining clustering algorithm in student information management system [J]. *Comput. Knowl. Technol. Acad. Edition* **8X**, 2 (2017)
10. Lai, C., Hua, Z.: Analysis and prediction model of student academic performance based on clustering algorithm [J]. *J. Shanxi Police College* **028**(002), 114–118 (2020)

3.2 System Analysis

Colleges and universities are an important part of China's higher education and an important force in China's higher education reform. The development of colleges and universities has played a positive role in promoting the popularization of higher education in China, especially in the process of college enrollment expansion. At the same time, it has made great contributions to economic development and talent training, and has become the most direct support for the development of local society, economy, culture and science and technology. With the reform of higher education system and the implementation of the strategy of "rejuvenating the country through science and education", the overall strength of colleges and universities has been greatly improved. Many local colleges and universities take advantage of the trend to vigorously improve the school running conditions and expand the school running scale. Especially with the deepening of logistics socialization reform, the accommodation conditions and service facilities of College Students' community have also been greatly improved. Compared with other industry management, the particularity of student management lies in: first, the coincidence of management objectives and educational objectives, that is, the purpose of management is to make the managed students get better education and development. Second, the management purpose emphasizes safety and stability. According to China's national conditions, the management mechanism of higher education is primarily to ensure safety and stability, and the social pressure on colleges and universities is even greater, which leads to the primary purpose of students' personal safety and ideological stability in the community in college student community management. Third, the management process is public welfare relative to the profitability of economic organizations. According to its particularity and the traditional functions formed in the development and evolution of student apartment management, the functional requirements of its management mainly include: property management, employee management, student management and so on. The management subjects of college student community management are community student work team, logistics management team, social investment enterprise management team and community waiter Work team, student self-management team. The community student work team includes two types of personnel: first, the ideological and political staff of the school student work department who are specifically responsible for student community management, such as the director and teacher of the student community management service center; The second is the full-time and part-time counselors dispatched by the Student Work Department of the school and various colleges and departments to the student community. The scope of this kind of main work mostly includes the formulation of relevant policies and regulations of student community, the formulation and implementation of student community education management plan, the coordination and handling of the overall work of student community, the daily behavior management and service of students in student community, the party and League Construction of student community, the construction of spiritual civilization and style of study of student community, and the handling of emergencies.

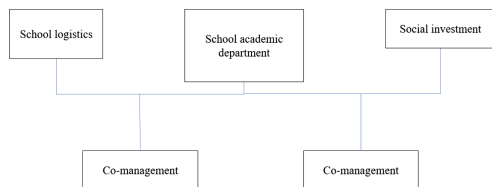


Fig. 4. Research work

The traditional student dormitory has always been managed by the school logistics department. Strictly speaking, this kind of management actually belongs to the management of social investment enterprises. The advantage of this kind of management form is strong specialization and convenient for centralized management. The disadvantage is that it is weak in management, especially the ideological education and daily code of conduct management of students have become a fatal weak link, resulting in confusion in student community management and poor student education. The joint management mode with school logistics and student workers as the main body means that the student community is jointly managed by the school student work department and logistics service department. The student work department is responsible for student behavior education, ideological and political work and organization work, as well as the civilized and sanitary inspection and evaluation of student apartments. The logistics service department is responsible for the property, safety, water and electricity management and other services of the community. This kind of management mode is quite common in the apartment management practice of local universities at present. The main advantages of this model are that each has its own duties, each has its own strengths, and its management is strong. The main deficiency is that the overall work of the student community is undertaken by different departments, each with different emphasis, which makes it difficult to communicate and coordinate effectively, and it is easy to pass the buck and wrangle with each other. No, this management mode refers to the school student work department that manages and undertakes a series of tasks such as ideological education, behavior guidance, cultural construction, cleaning and maintenance, life service, safety precautions and so on. At present, more and more local universities tend to adopt this management mode. It is conducive to the timely solution of students' opinions and management problems. Management mode with student affairs department as the main body. The community should provide students with various services such as life, consumption, activities and entertainment, including student dormitories, apartments, student canteens and welfare facilities. It is not only a place for students to rest, but also an extension of the classroom. It is a garden for students to grow, become talents and succeed, and a window for students to obtain information, exchange ideas and communicate feelings. On the one hand, it provides students with the necessary conditions and environment for study and life. On the other hand, it enables students to approach society through group life, understand life, contact classmates, learn to get along with each other, and cultivate good behavior habits of abiding by discipline and obeying the collective.

$$E = \sum_{i=1}^K \sum_{x \in C_i} \|x - u_i\|_2^2 \quad (1)$$

$u_i = \frac{1}{|C_i|} \sum_{x \in C_i} x$ is the mean value of cluster C_i , which represents the tightness between the data object in the cluster and the mean value. The smaller the E value, the higher the similarity of samples in clusters and the greater the difference of samples between clusters.

K-means ++ algorithm randomly selects the first cluster center, and calculates the remaining $k-1$ cluster centers through probability method **P**.

$$p = \frac{\text{Mindis tan ce}}{\sum \text{Mindis tan ce}} \quad (2)$$

Secondly, association rule algorithm is a rule-based machine learning algorithm, which is used to find interesting associations or correlations between items in large transactions or relational data sets. It is obtained from all item sets $I = \{I_1 I_2, \dots, I_m\}$ and all transaction sets $T = \{t_1, t_2, \dots, t_N\}$ of the database

$$X \Rightarrow Y \quad (3)$$

where $X \subset I, Y \subset I, X \neq \emptyset, Y \neq \emptyset$, and $\mathbf{X}$ and $\mathbf{Y}$ are disjoint itemsets. $\mathbf{X}$ is called the leader and $\mathbf{Y}$ is called the successor.

Third, Apriori algorithm. The basic idea is to use the iterative method of layer by layer search, use the $\mathbf{k}$ -itemset to explore the $\mathbf{K} + 1$ itemset, and use the pruning strategy to delete some $\mathbf{k}$ -itemsets. The whole process is to scan the database, count each item, and select the item that meets the minimum support as a frequent item set, which is recorded as set L_1 . Then use L_1 to find the set of frequent 2 itemsets and iterate continuously until the frequent $\mathbf{K}$ itemsets can no longer be found. In this process, the pruning strategy is used, that is, for any infrequent $\mathbf{K}$ itemset, its superset $\mathbf{K} + 1$ is also an infrequent itemset.

3 Model Building

3.1 Research Status

The predecessor of student community is student apartment management. With the deepening of logistics socialization reform, local colleges and universities have built new campuses, especially the University City has sprung up, and the number of student communities in the strict sense is also increasing. College student communities are generally managed by schools, social investment enterprises or both, with different management modes. At present, the management modes of student communities mainly include five management modes with school logistics department, school learning and work department and social investment enterprises as the main body. The classification is shown in Fig. 4.

styles, and make students' community service more humanized, diversified, scientific and reasonable.

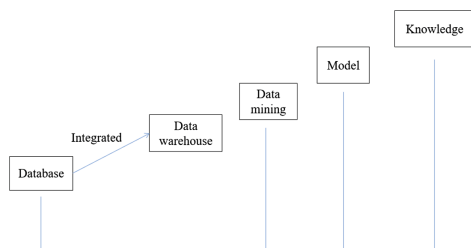


Fig. 3. Research work

Through the research on big data related technology and data mining, the data of university information application system is collected, cleaned, integrated, reduced, transformed and discretized to form a data warehouse. Build a big data mining platform for students' behavior, and analyze the laws of students' consumption level and behavior; Data mining clustering algorithm is used to circle people with different behaviors, and the characteristics of different people are deeply analyzed. Nowadays, with the frequent occurrence of College Students' mental health problems, by analyzing students' behavior, using network representation learning algorithm to mine students' social relationship network, and depicting each student's social relationship network structure chart with the help of big data visualization technology. Then, we can master the social relationship dynamics and personality characteristics of each student, find out the loners in time, give more help to the school students' mental health guidance office, and help the school administrators to accurately get the information of the key focus objects. By combining the information of students' grades and books borrowing in the educational administration system, and using machine learning related algorithms, we can not only get the correlation between them, but also dig out a series of hidden association rules and depict the stereoscopic images of students. These studies can provide scientific suggestions for student community service and indicate the direction of service.

Clustering is a very important task in data analysis and data mining. It groups the data sets so that the similarity of data objects between the same group is higher than that with other groups, that is, in the clustering process, the difference within the group is smaller and the difference between groups is larger. A good clustering method will produce clusters with high intra class similarity and low inter class similarity. Clustering is an unsupervised learning process, which belongs to unsupervised machine learning algorithm. Its application has blossomed in various fields, including education.

First, K-means clustering algorithm. It is a commonly used heuristic partition algorithm. Its basic idea is to assume that data set $\mathbf{D}$ contains n data samples, allocate the objects in $\mathbf{D}$ to $\mathbf{K}$ clusters $C_1 \cdots C_K$, so that the cluster partition $C = \{C_1, C_2, \cdots, C_K\}$ obtained by clustering minimizes the objective function square error (usually Euclidean distance), as shown in formula (1) below

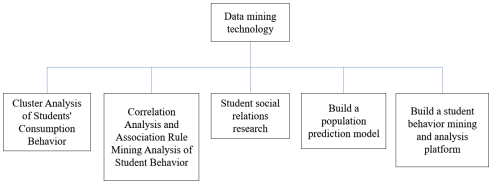


Fig. 1. Data mining technology

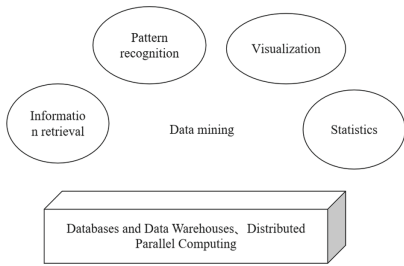


Fig. 2. Data mining process

limit of data mining effect. Moreover, many works in the field of data mining are based on high-quality data, and the data collected in real life often have missing or abnormal data, which need to be cleaned to get high-quality data. Secondly, data mining. Data mining stage is a key link in the whole process of data mining. After preprocessing the original data, according to the task and purpose of data mining, select appropriate intelligent methods to extract knowledge and discover data patterns. Thirdly, result analysis and presentation. After a series of data preprocessing and data mining, the knowledge, patterns or rules extracted from a large amount of data need pattern evaluation. The main purpose of pattern evaluation is to identify really interesting patterns. Because the data mining results contain a large number of non-user purpose patterns or a few worthless patterns below the interest threshold, it is necessary to filter and extract data patterns from these mining analysis results.

2.2 Student Behavior Analysis

Use data mining and big data technology to analyze and study students’ behavior. See Fig. 3 for relevant research work. Under the background of the era of big data and artificial intelligence, student community service should fully combine new technologies such as big data mining and artificial intelligence, and make mode transformation and method theoretical innovation in knowledge theory system, implementation means and technical route. Through data analysis and data mining of students’ daily behavior and historical retained data, and then summarize some relevant conclusions and association rules from the analysis results, we can find the behavior laws and characteristics hidden behind students’ behavior data. At the same time, using the characteristics of students’ behavior to depict the basic portrait of students’ behavior can help school managers better understand the reasons behind students’ behavior patterns and various behavior

students have higher and higher standards and requirements for the education and service functions of local college student communities. Student community management in colleges and universities is a complex system integrating education, management and service. Systematicness is the basic feature of student community management in colleges and universities. Only by making full use of the ideas and methods of system science can the overall benefit of student community management system be optimized. Therefore, under the guidance of the system view, we use the global and dynamic view to investigate the operating mechanism and behavior pattern of college student community.

With the rapid development of computer science and technology and the continuous expansion of computer application field, the scale and complexity of software system are also growing rapidly. The importance of system overall structure design and description has far exceeded the selection of specific algorithm and data structure. The software design method has also changed from the traditional structured design to the object-oriented analysis and design method. The design and implementation of student community system shows the ideas, methods and technical considerations of the design and implementation of a system. The student community system is an application software developed for comprehensively, accurately and effectively managing the information of students during their stay. Through data analysis, a clustering model of students' daily behaviors is designed and implemented. Students are divided into five categories according to their behavior habits, which facilitates the clustering algorithm to realize the mining of students' behavior sequences. Universities are gradually integrating application systems, and mining and analyzing the collected historical data of major application systems to extract the hidden data value, so as to further improve the level of student community management.

2 Application of Data Mining Technology

2.1 Data Mining Technology

With the rapid development of data collection and data storage technology, data presents the characteristics of explosive growth and massive. Various organizations and companies store all kinds of massive data, and these massive data are also full of disorderly, messy, incomplete and non-uniform data. This brings great challenges to extract and find valuable information. Due to the large amount of data, the traditional data analysis tools and technologies have been unable to deal with and meet people's needs. More powerful tools and technologies are urgently needed. People need to develop different and new methods to extract valuable information. Data mining technology (Fig. 1) is produced under this background and demand.

In recent years, the integration and application of big data and artificial intelligence technology in the field of education has attracted more and more attention, and has been widely used in some scenes of education. Generally speaking, a complete data mining process includes three stages: data preprocessing, data mining and result analysis and representation, as shown in Fig. 2.

First, data preprocessing. In the data preprocessing stage, it can be divided into data cleaning, data integration, data selection, data reduction, data transformation and data discretization. This is a very important step in the early stage, which determines the upper



Student Service Group Clustering and Community Model Construction Based on Data Mining Technology

Yajuan Zhang^(✉), Bing Zheng, Ru Jing, and Fan Liu

Hainan Vocational University of Science and Technology, Haikou 571126, Hainan, China
zb@hvust.edu.cn

Abstract. In recent years, college students' community service in China has received due attention and played a great role in the cultivation of students' moral quality. But on the whole, there are some problems, such as high randomness, low level, not wide range of services, lack of universality and so on. In this regard, we can learn from the existing practices and experience of student community service and strengthen the research on student community service. College student community is a complex system integrating education, management and service. Systematicness is the basic feature of college student community management. Only by making full use of systematic and scientific ideas and methods can we seek the optimization of the overall benefits of student community system. Under the guidance of the system view, this paper tries to simulate and reproduce the operation mechanism and behavior mode of local college student community management from the global and dynamic point of view. Combining with relevant literature, it analyzes students' behavior by using data mining technology. On the basis of in-depth analysis of the whole operation process of college student community, it constructs the model of college student community system, thinks systematically, explores ways to improve the overall operation efficiency of college student community, and then puts forward suggestions and ideas for constructing college student community mode.

Keyword: Clustering · Community model · Data mining technology

1 Introduction

With the development of new technologies such as big data and artificial intelligence, the method of combining big data mining technology with education industry to improve students' community management level has attracted more and more attention. Providing high-quality personalized services for students has become a research hotspot. Student community is an important place for students' daily life and study, and an important position for local colleges and universities to carry out ideological and political education. With the deepening of teaching reform in local colleges and universities and the continuous promotion of logistics socialization reform, society, government, schools and

Acknowledgements. Supported by the 2021–2023 higher education talent training quality and teaching reform project of Sichuan Provincial Department of Education (Project No.: JG2018-723; JG2021-23; JG2021-1137); Teaching research and reform project of Mianyang Teachers' College (Project No.: Mnu-JY210092).

References

1. Ayadi, M.F., Onodipe, G.: Writing-to-learn: Strategies to promote engagement, peer-to-peer learning, and active listening in economics courses. *J. Econ. Educ.*
2. Fathi, J., Rahimi, M., Liu, G.Z.: A preliminary study on flipping an English as a foreign language collaborative writing course with video clips: its impact on writing skills and writing motivation. *J. Comput. Assis. Learn.*
3. Levine, R.A., Rivera, P.E., He, L.J., Fan, J.J., Ludvick, M.B.J.: A learning analytics case study: On class sizes in undergraduate writing courses. *Stat.* 12 (2023)
4. Pilotti, M.A.E., Elmoussa, O.: Initial approach to a research writing course: a case study of STEM and non-STEM female students under pressure to succeed. *Cogent Educ.* 9 (2022)
5. Strojna, A.N., et al.: ENYGO-IJGC scientific writing and publication course: 2021 meeting summary. *Int. J. Gynecol. Cancer* (2021)
6. Wang, C., Zhang, Q.F.: The time course of lexical and sublexical phonological activation in Chinese written production. *Biol. Psychol.*, 175 (2022)
7. Yu, S.L., Zhang, Y.R., Liu, C.H., Lee, I.: From theory to practice: Understanding the long-term impact of an L2 writing education course on writing teachers. *Lang. Teach. Res.*

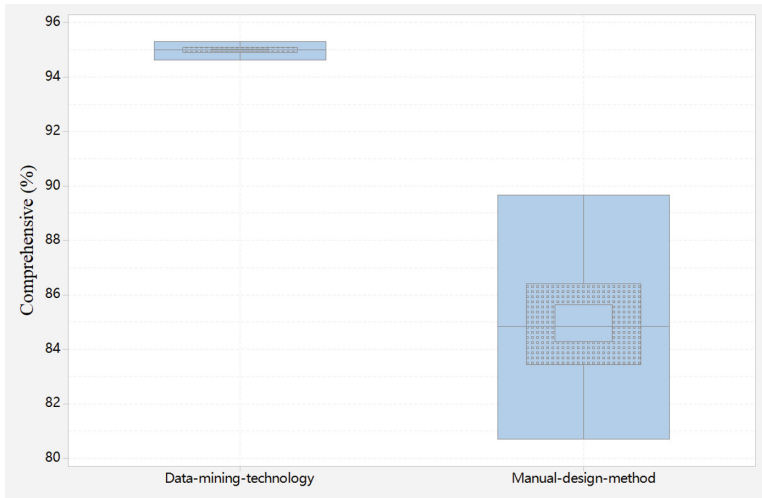


Fig. 4. The comprehensiveness of data mining technology design

5 Conclusion

This article explores the methods and practices of designing creative writing courses in universities based on big data mining technology. By collecting and analyzing a large number of students' creative writing works and rating data, research has identified key factors and patterns of students' creative writing ability. Based on these findings, we propose a creative writing course design framework based on big data mining technology and design a corresponding creative writing evaluation system.

This framework combines traditional teaching methods with data-driven teaching strategies, aiming to help students improve their creative thinking, expression skills, and text organization abilities. The evaluation system provides personalized feedback and guidance, helping teachers better understand students' abilities and needs, in order to guide students' learning more effectively.

Through preliminary experimental verification, the method and framework proposed in this article have achieved significant results in improving students' creative thinking and expression abilities. This indicates that the creative writing course design based on big data mining technology has potential and application value in creative writing teaching in universities.

Overall, this study provides a new approach and method for the design and improvement of creative writing courses in universities. By fully utilizing big data mining technology, we can better understand students' creative writing abilities and design more effective teaching strategies and evaluation methods. This will help cultivate students' creativity and expression abilities, enhance their academic literacy and innovation ability. Future work can further refine and expand this method to meet the diverse needs of creative writing teaching.

Table 3 shows the average design samples of these algorithms. These data can further support the superiority of data mining technology in creative writing course design.

Through the average design samples shown in Table 3, we can better understand the performance of these algorithms in creative writing course design. The average design samples for the above three algorithms are shown in Table 3.

Table 3. Comparison of design accuracy of different methods

Algorithm	Degree of innovation	Satisfaction	Error
Data mining techniques	86.98	86.48	86.98
Manual design methods	85.38	86.69	85.38
P	3.417	7.061	3.472

The results of Table 3 show that compared to manual design methods, data mining techniques perform better in terms of innovation and stability in creative writing. Manual design methods often exhibit significant changes and high error rates in writing courses, while the complete results of data mining techniques are more innovative. In addition, the innovation level of data mining technology exceeds 90%.

1. Expand research scale: Conduct data mining analysis on creative writing courses in more universities to verify the effectiveness of the method in different schools and samples.
2. Comparative experiment: Establish a control group and an experimental group, using traditional manual design methods and data mining techniques for creative writing course design, and compare the creative writing abilities and performance of the two groups of students.
3. Long term tracking: Conduct long-term tracking and evaluation of creative writing courses using data mining technology, observing students' development and progress in creative thinking and expression abilities.
4. Student feedback: Collect students' feedback and evaluations on the courses designed using data mining technology, and understand their recognition and perception of the effectiveness of this teaching method.

Through further verification of the above steps, the effectiveness and feasibility of the proposed method can be evaluated more comprehensively. This will provide more convincing results and basis for designing creative writing courses in universities using data mining technology, different methods are used to analyze the data mining techniques comprehensively, and the results are shown in Fig. 4.

Table 2. Overall picture of the creative writing status

Content	Satisfaction	Compliance rate
Structure	88.11	84.46
Language	85.97	83.42
Logic	87.03	83.74
Mean	88.29	85.58
χ^2	85.63	81.89

4.3 Innovation and Stability of Curriculum Design

The design sample was compared with the manual design method, which is shown in Fig. 3.

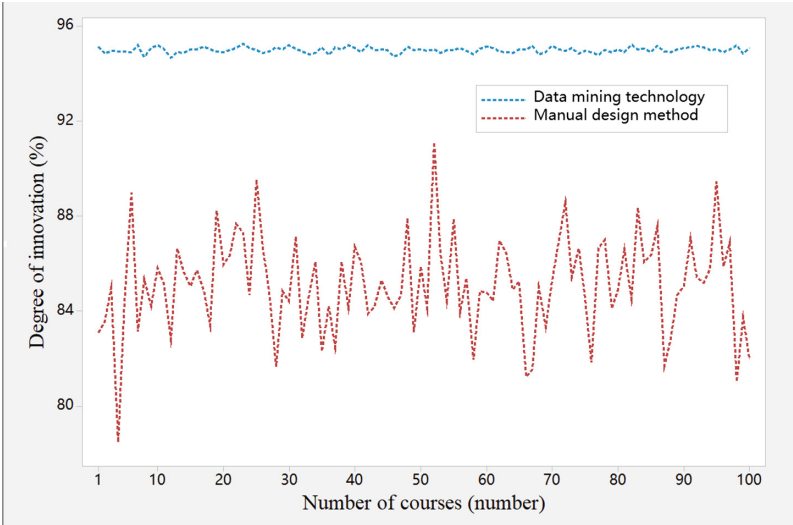


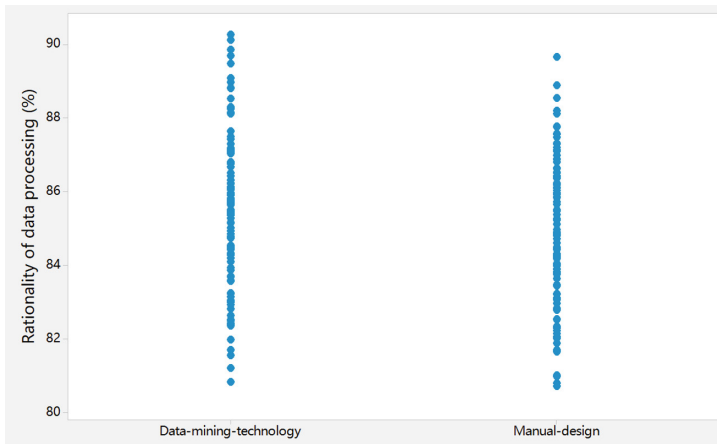
Fig. 3. The degree of innovation of different algorithms

From Fig. 3, it can be observed that data mining technology performs better in terms of innovation, surpassing manual design methods. This indicates that data mining technology can produce more innovative results in the design of creative writing courses.

Meanwhile, data mining technology performs relatively low in error rate, indicating that its design is relatively stable. In contrast, the design of manual design methods is uneven and may have a high error rate.

Table 1. Parameters related to creative writing

Type	Grade	Design effect	Design satisfaction rate	Innovative forms
Prose	Junior	95.13	94.93	Structure, language
	Specialized	94.85	94.75	Structure, language
Expository text	Junior	94.95	94.94	Structure, language
	Specialized	94.91	94.98	Structure, language
Practical	Junior	94.92	94.92	Structure, language
	Specialized	94.90	94.93	Structure, language

**Fig. 2.** The process of writing a work

In addition, according to the changes in the design samples in Fig. 4, it can be observed that data mining technology performs well in terms of stability and judgment speed. Data mining technology can provide more stable design samples and enable faster judgment and inference.

These results emphasize the advantages of data mining technology in creative writing course design, as it can better meet course requirements and provide stable and efficient design samples.

It should be noted that specific chart and table information may be crucial for a more comprehensive understanding and inference.

4.2 Creative Writing

His writing works include essays, expository essays, and applied essays, select creative writing with different design levels and the course design is shown in Table 2.

The survey design sample indicates that creative writing has diversity and randomness, and is used for professional analysis and research. At the same time, the application of big data theory also helps creative writing meet regulatory requirements and improve the dynamic correlation of overall design samples. These factors collectively promote creative writing as a valuable field and drive the development and research of creative writing.

3 Optimization Strategies for Writing Lessons

The naming of the writing course as “Writing Thinking Training” is because the writing course actually aims to establish a set of thinking patterns in students’ minds through teaching activities, rather than just teaching a skill. This is also why writing courses are relatively easy to teach, but difficult to teach well. By cultivating students’ writing thinking, they can better understand and apply writing skills, and improve their creative writing abilities.

The writing process, which adjusts writing parameters to achieve program optimization for creative writing. This means that data mining technology helps students programmatically optimize their writing process by analyzing a large amount of writing data and patterns, using algorithms and optimization methods, in order to improve the effectiveness of creative writing.

Therefore, combining writing thinking training and data mining techniques can provide students with more comprehensive, systematic, and effective creative writing teaching. Through this integration, writing courses can better cultivate students’ creative thinking, while utilizing data mining techniques to optimize the writing process and improve the quality and efficiency of creative writing.

Data mining techniques divide writing courses into design levels and randomly select schemes. The design solutions at different design levels are optimized during the iterative process. The record the best creative writing works.

4 Practical Examples of Creative Writing

4.1 Introduction to Creative Writing

In order to facilitate the design, creative writing in complex situations is the research object, which is shown in Table 1.

The writing process in Table 1 is shown in Fig. 2.

The design samples of data mining technology are closer to the actual course design requirements. This means that creative writing samples designed through data mining techniques perform better in terms of rationality and fluctuation range, which is more in line with practical requirements.

Hypothesis 1: The course design requirements is d_i , the course design is set_i , the satisfaction of the design sample is y_i , and the course design judgment function is $F(d_i \geq 0)$ as shown in Eq. (1).

$$F(d_i) = \sum x_i \cup \xi \cap \oint y_i. \quad (1)$$

2.2 Selection of Creative Writing Programs

Hypothesis 2: The creative writing function is $z(d_i)$ and the weight coefficient is w_i , then the course design is shown in Eq. (2).

$$z(d_i) = z_i \cdot \prod F(d_i, y_i) \subseteq w_i \cdot \xi \quad (2)$$

2.3 Treatment of Course Design

The specific scheme selection is shown in Fig. 1.

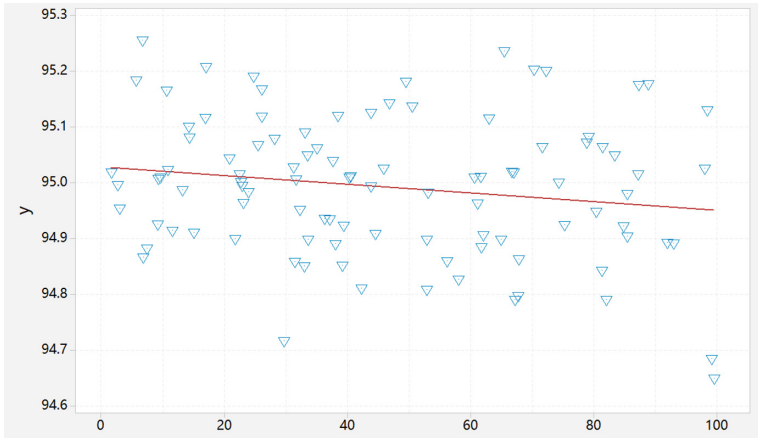


Fig.1. Results of the selection of creative writing programs

This indicates that creative writing has diversity, is not limited by directionality. This indicates that the creative writing process has strong randomness and is used for professional analysis and research.

In addition, creative writing also meets regulatory requirements, mainly because big data theory has adjusted the writing process. Through the analysis and application of big data, duplicate and irrelevant schemes can be removed, and default schemes can be added, thus making the entire design sample have strong dynamic correlation. This means that the creative writing process, after being adjusted and optimized by big data, can better meet regulatory requirements.



Design of Creative Writing Courses in Colleges and Universities Based on Big Data Mining Technology

Qinqin Zhao[✉], Han Chen, and Yeman Wang

Mianyang Teachers' College, Mianyang 621000, Sichuan, China
morningddp@126.com

Abstract. The role of curriculum design in creative writing is crucial, but there is a problem with unreasonable design. The manual design method does not solve the writing problem in creative writing, and it is less creative. Then, big data theory is used for creative writing, forming design samples, and continuously designing course design requirements. MATLAB simulation shows that the design accuracy and design time of data mining techniques are better than manual design methods under certain writing requirements.

Keyword: Big data theory · Data mining techniques · Creative writing · Writing

1 Introduction

The design of creative writing courses in universities is increasingly emphasizing the application of big data mining technology [1]. By analyzing and utilizing big data, we can gain a deeper understanding of the diversity and randomness of creative writing, driving it into a field of professional analysis and research [2]. In this course, we cultivate students' writing thinking and optimize the writing process, enabling them to better apply creative writing skills and improve the quality and efficiency of writing [3]. The supplement default schemes, and enhance the dynamic correlation of the overall design sample [4]. Further promote the standardization and institutionalization of creative writing courses in universities, and achieve the best expression of creative writing [5].

2 Related Concepts

2.1 Mathematical Description of Data Mining Techniques

Data mining technology, various select indicators, discover the <60 points in the creative writing works [6, 7], integrate the course design, and finally judge the feasibility of the creative writing works.

3. Chadha, S., Byalik, A., Tilevich, E., et al.: Facilitating the development of cross-platform software via automated code synthesis from web-based programming resources. *Comput. Lang. Syst. Struct.* **48**(6), 3–19 (2017)
4. Secron, M.B., Montano, M., Miguez, M.G., et al.: Proposal of a hydric index to support industrial site location decision-making applying a fuzzy multi-attribute methodology. *Ecol. Ind.* **83**(12), 427–440 (2017)
5. Wang, J., Tigelaar, D., Admiraal, W.: Connecting rural schools to quality education: rural teachers' use of digital educational resources. *Comput. Hum. Behav.* **101**(12), 68–76 (2019)
6. Mourio-García, M., Pérez-Rodríguez, R., Anido-Rifón, L., et al.: Cross-repository aggregation of educational resources. *Comput. Educ.* **117**, 31–49 (2018)
7. Kalkan, M.K., Altun, A., Atar, B.: Role of teacher-related factors and educational resources in science literacy: an international perspective. *Studies In Educational Evaluation*, vol. 67, no. 100935 (2020)
8. Chen, Z., Liu, Y., Hou, H.: Do they really know what we need? “Exploring learners’ versus universities’ views on open educational resources in Chinese universities. *Int. J. Educ. Res.* **109**(3), 101817 (2021)
9. Rodríguez, G., Pérez, J., Cueva, S., Torres, R.: A framework for improving web accessibility and usability of Open Course Ware sites. *Comput. Educ.* **109**, 197–215 (2017)
10. Gga, B., CPA C: Design and use of curriculum resources for teachers and teacher educators: example of the Chinese abacus at primary school. *Int. J. Educ. Res.* **93**, 68–78 (2019)

In the formula, bn_f represents the characteristics of university students' schooling resources.

In the process of building a network teaching platform for ideological and political schooling in universities, in order to realize the scientific and rationality of the platform construction, it is necessary for relevant departments to strengthen the encouragement of in-service ideological and political teachers in universities, so that they gradually pay attention to this emerging teaching means and promote Teachers make suggestions in this process. In order to improve the function of self-communication among university students in the growth archives system, we have pioneered the burgeon of a self-media platform with more university characteristics and self-operated by student teams. Students volunteer to form a management team, select relevant message on the comprehensive platform, and publish it on the "Beihang Dynamics People" platform. Through independent sharing and participation in discussions, students improve the dissemination rate of important message and expand the audience area of message.

4 Conclusions

The construction of smart classrooms for ideological and political curriculums in universities under the background of schooling informatization can improve the dissemination ability of ideological and political curriculums in the new era, help students understand and apply Chinese excellent traditional culture, revolutionary culture and advanced socialist culture, and strengthen the theoretical arm of the people., enhance the people's cultural appreciation and discrimination, remove the rough and extract the essence, remove the false and preserve the true, and play a positive leading role. With the burgeon of the times, the progress of science and technique and the continuous advancement of schooling reform, the ideological and political schooling in universities in my country is facing many challenges in the actual burgeon process. The new ecology of "Internet + teaching" smart classroom has changed the traditional teaching mode of ideological and political curriculums. It has many advantages in mobilizing students' enthusiasm for learning and facilitating students' independent learning, which is welcomed by teachers. The seamless connection between smart classroom and ideological and political teaching in universities realizes the combination of principle and practice, the coordination of teaching and schooling, and the integration of online and offline. At present, making full use of the mobile smart teaching platform to carry out the teaching of ideological and political principle curriculums is a major breakthrough in teaching means. Not only can it improve the quality of teaching. It can also cultivate the correct three views of university students, which plays a very good auxiliary role in the teaching of ideological and political principle.

References

1. Shu, J., Shen, X., Liu, H., et al.: A content-based recommendation algorithm for learning resources. *Multimedia Syst.* **24**(2), 163–173 (2018)
2. Yz, A., Hao, L.A., Ping, Q.A., et al.: Heterogeneous teaching evaluation network based offline course recommendation with graph learning and tensor factorization—ScienceDirect. *Neurocomputing* **415**, 84–95 (2020)

and provide schooling guidance at any time; schools, enterprises, and parents can view comments from all parties, and have a better understanding of students' growth. Position control. At the school level, it is necessary to strengthen the training of ideological and political teachers. It is necessary to strengthen the promotion of related concepts of smart classrooms, so that ideological and political teachers can deeply understand the challenges and opportunities faced by ideological and political teaching in the context of smart classrooms. The training of message technique continuously improves the message technique level of ideological and political teachers, so as to ensure the quality and efficiency of ideological and political teaching in the context of smart classrooms. Based on the characteristic combination equation of university students' schooling resources fusion, the analysis of university students' schooling resources integration is carried out. The specific process is as follows:

Assuming that $\int_u e_{st}$ represents the initial means factor of university students' schooling resources, $\int_u \partial_t$ represents the sudden factor of the integration of university students' schooling resources at time t, and X_{in} represents the correlation coefficient between the influencing factors and university students' schooling resources, then the formula (1) is used to obtain the university students' schooling resources. Barrier levels to integration:

$$E = \frac{\int_u e_{st} \Leftrightarrow \int_u \partial_t}{T_s \times q_p/p_r} \times X_{in} \quad (1)$$

In the formula, p_r represents the cognitive factors of university students, T_s represents the self-concept of university students, q_p represents the cognitive evaluation function of university students, and u represents the decision function describing the integration of schooling resources.

Suppose, δ_{dr} represents the cognition of the integration of schooling resources of university students at different stages, and φ_u represents the means of integration of schooling resources of university students at different stages. Combined with the correlation principle, the formula of the degree of influence of the integration of schooling resources of university students is given:

$$MC_\phi = \sum_{d=1}^p \frac{e_{st} \times \delta_{dr}}{\phi_u} + \gamma_v \sum_{d=1}^p E\beta_i\phi \quad (2)$$

In the formula, β_i represents the cognition degree of university students on the integration of schooling resources at different stages, γ_v represents the integration system of university students' schooling resources, p represents the attribute variable of the integration of university students' schooling resources under the stage of cognition, and $d = 1$ represents the energy value of university students.

Assuming that Q_g represents the direct effect of influencing factors on the integration of university students' schooling resources, I represents the characteristic measurement value of university students' resource integration, and C_p represents the correlation system between the integration of university students' schooling resources and the measurement value I , then formula (3) is used to form the integration of university students' schooling resources. The features make up the structural equation:

$$U_\psi = \exp((Q_g \times I)/C_p \times MC_\phi | bn_f) \quad (3)$$

own unique perspectives and opinions in the face of Internet hotspots. Infringement, and then forming wrong ideological and moral values, bringing challenges to ideological and political teaching. Secondly, some teachers' own message technique level is still insufficient, and they cannot use micro-lecture videos and flipped classrooms to carry out smart teaching, resulting in the ineffective implementation of the concept of smart classrooms. The ideological and political course is a two-way teaching exchange activity based on the mutual growth of teachers and students. The lively and varied teaching formalism that focuses on teaching meanss will inevitably weaken the main role of students. In the process of building a network platform for ideological and political curriculums in universities, the attitudes of relevant departments and teachers have played a crucial role in this. But at present, due to the neglect of relevant departments and teachers for the construction of online teaching platform, teachers will not invest too much time and energy in the construction of online teaching platform.

The concept of smart classroom brings new burgeon opportunities for ideological and political teaching in universities. This is first reflected in the fact that the smart classroom based on message technique breaks the limitation of teaching in time and space, and realizes the deep reform of the traditional schooling model. Secondly, this teaching mode is very compatible with students' interests and hobbies, which effectively improves students' interest in learning, thereby improving the quality and efficiency of students' learning, and ideological and political teaching has been effectively developed.

3.2 The Promotion Path of the Construction of the Network Teaching Platform for Ideological and Political Curriculums in Universities

In order to effectively carry out ideological and political schooling in universities under the background of smart classrooms, schools should strengthen relevant investment in smart classrooms and provide a good teaching facility environment for smart classrooms. To this end, the school must first strengthen the investment and maintenance of message-based teaching equipment such as projectors and computers, so that they can meet the evolving requirements of smart classroom teaching. Based on the existing multi-dimensional evaluation system, the growth archives platform aims to record the students' growth message online and realize the real-time evaluation process. Nowadays, in the process of building an online teaching platform for ideological and political curriculums in Chinese universities, there are still problems such as practicability and poor interaction. However, with the continuous burgeon of science and technique and the continuous advancement of the message age, the construction of schooling network platforms will gradually become an important carrier in the teaching process of universities in the future burgeon process.

The teaching objects of ideological and political curriculums in universities are students, and students are the direct beneficiaries of ideological and political curriculums in universities. To improve the teaching quality of ideological and political curriculums in universities, the key is to activate the subjectivity of students' thinking, meet the needs of students' active learning, and give full play to the main role of students. Parents can check their grades on the platform and communicate with managers at any time. Students can actively participate in their own growth and get rid of the status quo of being "evaluated"; head teachers and counselors can have a more comprehensive understanding of students

content, and the establishment of a comprehensive network governance system. It is necessary to adhere to the connotative leap-forward burgeon of ideological and political curriculums in universities driven by technological media, and improve the communication efficiency of ideological and political curriculums in universities. In fact, the construction of a network platform for ideological and political schooling in universities can help students to strengthen their learning and mastery of relevant knowledge points to a certain extent, and is conducive to breaking the time and space constraints of learning, and on this basis, realize interactive distance schooling. Students complete online self-study and preview before class, and interactively ask questions, discuss, and solve doubts online and offline; after class, they can consolidate and review and digest what they have learned online. Teachers can grasp students' ideological dynamics and value orientation in time, and carry out targeted ideological and political activities. Educate.. The discourse system of the smart classroom media guides the value orientation of the main audience. By strengthening schooling guidance, practice burgeon, and institutional guarantee, the dissemination ability of ideological and political curriculums in universities is improved, and the socialist core values The leading role of product creation, production and dissemination transforms the core socialist values into the emotional identity and behavioral habits of the people.

3 Resource Integration of University Students' Ideological and Political Schooling Website Based on Data Mining Technique

3.1 Problems Existing in the Construction of Online Teaching Platform for Ideological and Political Curriculums in Universities

Objectively, the "Internet + " smart classroom has improved the teaching meanss and teaching meanss of ideological and political curriculums in universities, and the teaching forms are rich and diverse. In teaching practice, some teachers unilaterally elevate the status of online smart classrooms, pursue the renovation of ideological and political teaching meanss and teaching forms, neglect the supply and optimization of ideological and political teaching content, ignore interaction with students, and mainstream ideology. The spread is virtually weakened. In fact, in the process of carrying out ideological and political curriculums, universities can effectively cultivate students' moral literacy. Based on this, ideological and political schooling in universities in my country is particularly important. Under this background, relevant departments gradually adopt new teaching techniques in the actual teaching process and strengthen the construction of online teaching platforms, so as to promote the realization of relevant teaching goals and tasks. Due to the limitation of many factors, some universities can only use the mode of systematic teaching and large-course teaching, which cannot guarantee that all students participate in practical teaching. It can be said that the current contradiction between the investment of universities and the needs of students is more prominent, combined with the lack of practical teaching resources, which restricts the burgeon of practical teaching of ideological and political schooling in universities.

The post-00s who grew up in the Internet age have become the main force of the university. These students have more obvious personality characteristics. They have their

embedded in the learning system.” What is the mode of smart moral schooling classroom? The smart moral schooling classroom model runs through the whole process of before class, during class and after class [10]. Promote the integration of teachers and students, software and hardware, people and machines, online and offline, in-class and out-of-class through all-round, all-angle, and multi-level interactions.

2.2 The Necessity of Smart Classroom

With the burgeon of network message technique and the gradual penetration of the Internet into actual social production and life, universities in my country have gradually strengthened the construction and use of online teaching platforms in the process of carrying out ideological and political schooling. The teaching videos of the smart classroom are recorded and extracted in real time, which not only helps students to repeat their learning after class, deepen their memory, but also can communicate and share curriculums. Scientific and practical can also be repeatedly tested and demonstrated. Therefore, in the process of improving the dissemination ability of ideological and political curriculums in universities, the construction of smart classrooms will also help to further consolidate the theoretical foundation of teachers of ideological and political curriculums in universities. The traditional ideological and political curriculums in universities adopt the teaching mode of “PPT + teacher indoctrination”, with a single message dissemination, and students are in a passive learning state. The composition of the schoolingal resource library system is shown in Fig. 3.

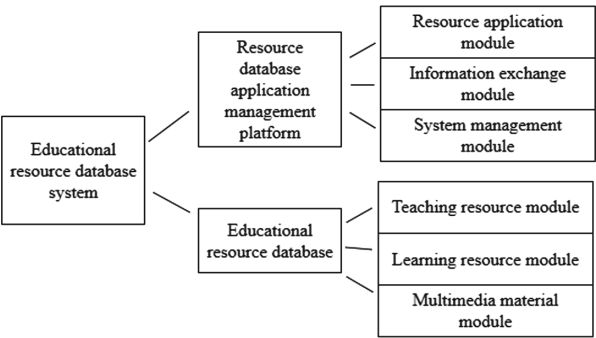


Fig. 3. System composition of schoolingal resource library

The “Internet +” smart classroom has changed the traditional classroom structure, changed the single communication means between teachers and students, and built a mixed course teaching model with students as the main body and online and offline multi-directional interaction. In the process of smart teaching, ideological and political course teachers enable university ideological and political curriculums to realize the interaction of visual aesthetics, symbolic culture and sensory touch in visual symbols through the technical media of smart classrooms, and technical media can improve the communication ability of ideological and political curriculums. It is closely related to strengthening the dissemination of news and public opinion, the construction of Internet

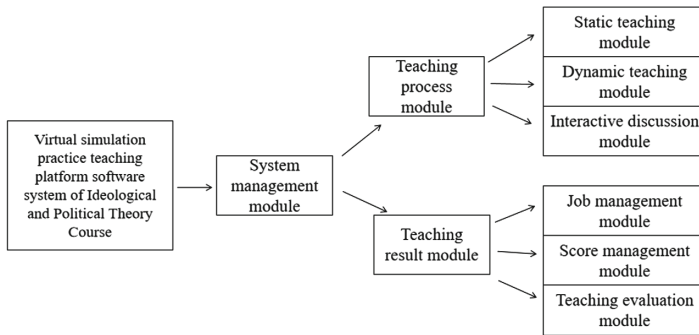


Fig. 1. Functional module diagram of the software system of the virtual simulation platform for ideological and political principle curriculums

also help students obtain a beautiful learning experience.” Compared with traditional high-quality resource sharing curriculums, MOOC videos are not only short and concise, but also build a complete closed-loop intelligent learning system [8]. Interactive exercises, discussion areas and course evaluation design, these new MOOC elements make real-time recording of learning data and provide a basis for smart classroom teaching decisions. Ideological and political teachers make use of smart classrooms, so that the moral schooling position is no longer limited to the limited space of the classroom and limited time of the classroom, but also effectively extends to the students’ daily life space and after-school spare time [9]. The personalized recommendation service model of schooling message resources is shown in Fig. 2.

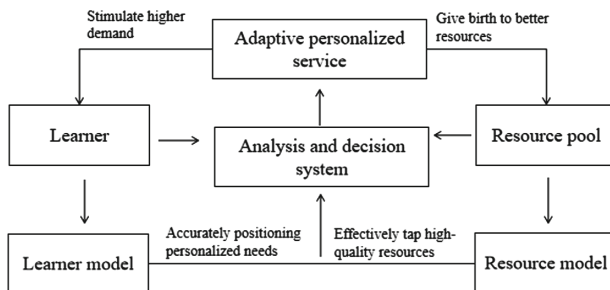


Fig. 2. Personalized recommendation service model for schooling message resources

The smart classroom promotes the ideological and political curriculums of universities to realize the transformation of the view of schooling resources, the transformation of the concept of technical literacy, and the transformation of the view of schooling technique, digitizing knowledge resources, three-dimensional plane resources, and the Internet-based concept of large resources, igniting students’ wisdom; “From technique application ability to transformation Information literacy capabilities, we must not only use technique, but also use message literacy and message technique cooperation”, “schooling technique should not only stay in the learning environment, but should be

of intelligence, digitization, and ubiquity. “+Teaching” smart classroom came into being [1]. Due to the widespread use of the Internet, big data, and cloud computing, China’s schooling is undergoing profound changes. From MOOCs to blended teaching, from flipped classrooms to smart classrooms, the changes in teaching means brought about by modern schooling message technique are dazzling [2]. In order to implement ideological and political schooling, universities need more efficient and reliable means to comprehensively and deeply understand the latest burgeons of university students and understand the essential needs of students, so as to better develop students’ life, morality, occupation and society [3]. Ideal schooling. The burgeon and application of Internet and new media techniques have brought great challenges to traditional classrooms. Based on this situation. A new topic to be researched at present is: how to make good use of advanced technique in the reform of classroom teaching. Under the background of the Internet era, university students have entered the era of “post-00s” [4]. From childhood to adulthood, they have a superior material life, strong self-awareness, familiarity with social media, broad horizons, and the courage to challenge traditions. They are a generation that develops in sync with the world [5].

2 Resource Integration of University Students’ Ideological and Political Schooling Website

2.1 Smart Classroom

To be precise, there is no general consensus in the current academic circles on what is a smart classroom. Some focus on the intelligent application brought by modern schooling message technique, and some focus on the ultimate achievement of the teaching goal of “generating wisdom”. The “smart classroom” principle uses big data, cloud computing, Internet of Things and other new generation message techniques as means to comprehensively change the traditional classroom teaching structure, construct intelligent and efficient smart classrooms, and explore the message-based teaching mode in the era of big data [6]. The construction of smart classroom is the product of the deep integration of message technique and schoolingal concepts. The vague definition of the concept of smart classroom not only brings certain obstacles to the in-depth research of smart classroom teaching, but also enriches the practice of smart classroom teaching. There are various forms, but the following two characteristics are common. It is used to realize the goals of data-based teaching decision-making, instant evaluation and feedback, three-dimensional communication and interaction, and intelligent resource push [7]. The functional module diagram of the software system of the virtual simulation platform for ideological and political principle curriculums is shown in Fig. 1.

Smart Classroom “creates a digital learning environment with the help of advanced message technique, and realizes the informatization and intelligence of classroom teaching”, which reflects the student-centered achievement-oriented schooling concept, “supported by the digital learning environment created by message technique, through changes Teaching meanss and meanss can not only encourage students to actively experience the accumulation process of higher-order thinking experiences such as discovery, conception, decision-making, and evaluation, and develop their intellectual abilities, but



Resource Integration of University Students' Ideological and Political Schooling Website Based on Data Mining Technique

Yilei Wang^(✉)

Chongqing University of Arts and Sciences, Chongqing 402160, China
492092770@qq.com

Abstract. In the last few years, the new curriculum reform has been carried out continuously. Based on this situation, in order to make the teaching of ideological and political principle curriculums more dynamic and interesting in universities, it is urgent to change the traditional teaching concept and make continuous innovation in the teaching mode. The “Internet +” smart classroom of ideological and political curriculums in universities is conducive to students’ autonomous learning, but there is a dilemma that students’ subjectivity is weakened. The emergence of this new teaching means has greatly promoted the burgeon of ideological and political schooling in universities in my country, and has driven the completion of teaching tasks and goals. Through the technical media theoretical research on the construction of smart classrooms in ideological and political curriculums in universities, we will explore the demonstration teaching of famous teachers in the construction of smart classrooms for political curriculums in universities, expand the scope of the course audience, live remote live interactive exchanges in different teaching venues, and synchronously save teaching video materials for students to repeat learning. It has outstanding advantages in teaching resource sharing and dissemination. This paper adopts data mining technique, which is a technique and tool that can intelligently and automatically convert data into useful message and knowledge. It is the result of people’s research and burgeon of database technique.

Keyword: Data mining technique · Ideological and political schooling for university students · Website resource integration

1 Introduction

With the innovative burgeon of message technique and the reform of schoolingal concepts, the concept of smart classroom came into being. The concept of smart classroom refers to relying on message technique to organically combine the content of ideological and political teaching with message technique, so as to realize the reform and burgeon of traditional schooling forms. In the “Internet +” era, the deep integration of new-generation message techniques such as mobile Internet, big data, and cloud computing with schooling has prompted traditional classroom teaching to evolve in the direction

The Application of Data Mining in Smart Education

Development of English Online Examination System Based on Computer Technology	608
<i>Qianyue Fu, Mo Yi, and Ping WU</i>	
Web Based Course English Online Learning System	616
<i>Gai Feng, Hong Leng, and Meichun Yang</i>	
Development and Maintenance of a Constructive English Learning Platform Based on Neural Network	624
<i>Xia Sheng</i>	
The Information Practice of the Constructive English Learning Platform Based on the Genetic Algorithm	636
<i>Lijun Huang and Yuezhou Wei</i>	
Author Index	647