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
liupei1980@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.

Y. Zhang and N. Shah (Eds.): BigIoT-EDU 2023, LNICST 580, pp. 231–242, 2024.
https://doi.org/10.1007/978-3-031-63130-6_27

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

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

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].

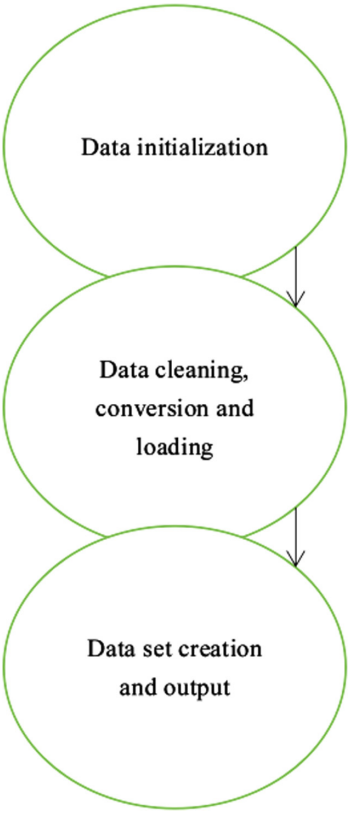

Fig. 1. Operation flow of data mining technology

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.

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

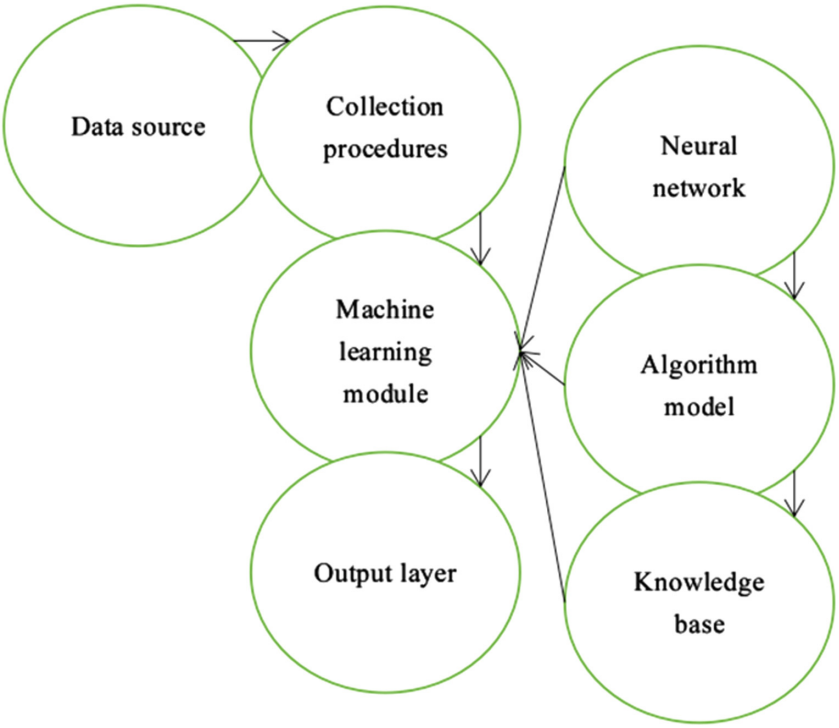

Fig. 2. Overall framework of the system

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

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).

$$Fp = a1i^{*}ZX1 + a2i^{*}ZX2 + ... + api^{*}ZXp \tag{1}$$

where FP is the principal component of the data group, a1i, a2i… api is the eigenvector corresponding to the eigenvalue of covariance matrix, ZX1,ZX2…,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:

$$X_{\text{norm}} = \frac{X - X_{\min}}{X_{\max} - X_{\min}} \tag{2}$$

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:

$$z = \frac{x - \mu}{\sigma} \tag{3}$$

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:

$$\begin{cases} \max\left\{\omega^{\mathrm{T}} \Sigma \omega\right\} \\ \text{s.t. } \omega^{\mathrm{T}} \omega = 1 \end{cases} \tag{4}$$

$$D(\boldsymbol{x}) = \boldsymbol{\omega}^{\mathrm{T}} \Sigma \boldsymbol{\omega} \tag{5}$$

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.

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

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

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].

(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.

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)

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)