



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.

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

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.

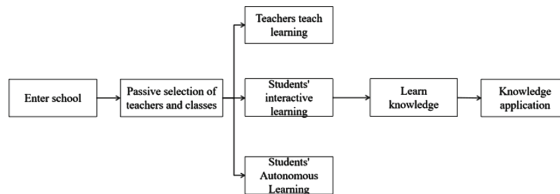


Fig. 1. School learning process

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.

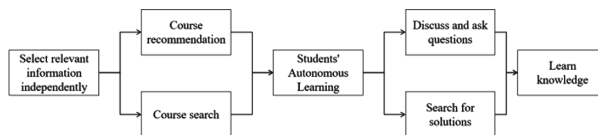


Fig. 2. Self regulated learning process of systematic education

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

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).

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

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.

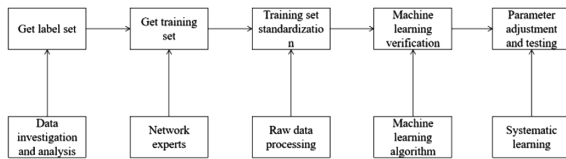


Fig. 3. Autonomous learning label standardization system

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.

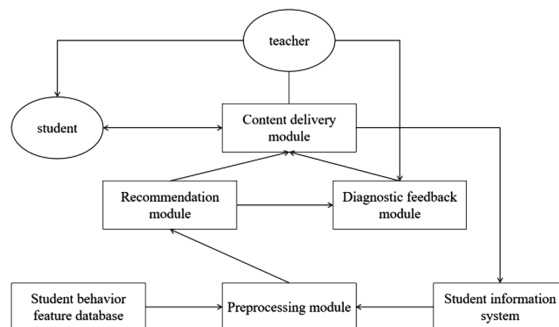


Fig. 4. Composition of adaptive learning system based on big data

(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

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.

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)

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), 133–134 (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), 219–220 (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)