Research Article
Intelligent Application of Data Mining Model in Chinese International Education

Feng Wu

College of International Education, Minzu University of China, Beijing 100081, China

Correspondence should be addressed to Feng Wu; 20170817347@mail.sdufe.edu.cn

Received 17 July 2022; Accepted 10 August 2022; Published 28 August 2022

Academic Editor: Hengchang Jing

Copyright © 2022 Feng Wu. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In order to improve the effect of intelligentization of Chinese international education, this paper puts forward the application method of the data mining model. Through the collection, transformation, processing, mining, and analysis of a large number of random data generated in education and teaching, this method provides learning suggestions for learners, realizes evaluation and prediction, presents knowledge information for educators, realizes educational decision-making, and realizes the continuous optimization of the teaching process. The experimental results show that the relationship between teachers’ teaching and students’ scores is obtained through the analysis of association rules in data mining technology. The support and confidence of students’ scores taught by teacher B are 28.57% and 72%, respectively, and the support and confidence of students’ scores taught by teacher C are 19.05% and 70.59%, respectively. The two rules are established to a great extent through data association. Data mining technology can effectively optimize the effect of intelligentization of Chinese international education.

1. Introduction

In recent years, Chinese international education has been booming at home and abroad. The number of Confucius Institutes and Confucius Classrooms around the world is increasing. Chinese has entered the national education system of some countries. Some private language education institutions teach Chinese courses. The number of practitioners in this field is growing. By December 31, 2016, 512 Confucius Institutes and 1073 Confucius Classrooms had been established in 140 countries (regions) around the world. As of September 2015, there were 363 colleges and universities nationwide offering undergraduate programs in Chinese international education, with 63933 students. There are 108 institutions offering professional master’s degrees, with 10133 students, 104 colleges and universities offer master’s degrees in Chinese Linguistics and Philology, and 95 colleges and universities offer master’s degrees in Linguistics and applied linguistics. There are 14 colleges and universities that enroll doctoral degrees in Chinese international education and 35 colleges and universities that train doctoral degrees in Chinese International Education at doctoral programs in linguistics and applied linguistics. Many graduates of these colleges and universities will find jobs in Chinese international education. On the surface, international Chinese education is developing rapidly from talent training to teaching research. The annual global Confucius Institute Conference has more than 1000 participants, and the number of members of the world Chinese language teaching association has also increased significantly. By 2013, the number of members has grown from more than 270 in 16 countries and regions at the beginning of its establishment to 4415 from 69 countries and regions at present. However, the basic research and applied research of Chinese international education are far from meeting the needs of diversified teaching. The demand for resources for Chinese teaching in the world is unprecedented, and the supply of resources obviously cannot keep up [1, 2].

The expansion of Chinese international education needs to innovate information technology, shift the focus from technological innovation to technological application, improve information construction, promote the Chinese language and culture internationally, give play to the important role of Chinese in international exchanges, and enhance the
national cultural soft power. In today’s information-based and globalized society, with the improvement of China’s comprehensive strength, the state attaches great importance to language development, and the information-based development of Chinese international education is increasingly valued. With the cooperation between linguistics, informatics, sociology, psychology, communication, and other related disciplines, we can make use of the research results of more disciplines to devote ourselves to the international education of Chinese and its promotion. At present, more and more overseas people realize that Chinese is an important tool to understand China and its language and culture. The international promotion of Chinese is developing rapidly, and Confucius Institutes are blooming everywhere. The number of overseas learners of Chinese as a second language is rising, and the gap among teachers in Chinese international education is growing. Therefore, it is of great strategic and practical significance to study and develop the technical theory of Chinese International Education and to think about its information construction methods and significance.

2. Literature Review

As human society enters the information age, education also enters the information age with human society, that is, the education information age. In the information age, the introduction of new educational technology has promoted the rapid growth of educational informatization, and a large amount of data have been generated and accumulated in the database. However, understanding these data has gone far beyond the scope of people’s ability. Finally, a large amount of data cannot be effectively used, forming the phenomenon of information island, data explosion but poor knowledge [3]. In response to this challenge, data mining technology came into being and showed strong vitality. It can find hidden and ignored rules and patterns from the vast ocean of data, so as to better support decision-making. The application of data mining in education informatization aims to find useful information from the massive data collected by the electronic education system. The ultimate purpose is to benefit all participants in the education system, provide a basis for solving the semistructured and unstructured decision-making problems in the field of education, and better promote the development of education informatization. Traditional data mining in the solution side is rarely considered from the perspective of users’ use. It pays more attention to professional issues, such as algorithms and system design. It is oriented by the data mining system. Such a system has very high requirements for professionals. It is generally only applicable to professionals, and it is not very applicable to nonprofessionals. Therefore, many enterprises need additional development costs for such a data mining system with high professional technology content [4].

3. Research Methods

3.1. Overview of Data Mining. Data mining is a process of discovering the relationship between models and data in massive data by using various analysis tools and revealing meaningful connections, patterns, and trends by carefully analyzing a large amount of data; it is a knowledge mining process that uses pattern cognition, statistics, and mathematics to extract unknown and operable information from large databases. Synonyms similar to data mining include knowledge mining, knowledge extraction, pattern analysis, and data fusion [5].

3.2. Research on Data Mining Model

3.2.1. Data Mining and Knowledge Discovery. Knowledge retrieval and data retrieval is a combination of intellectual, machine learning, and database technology. Knowledge exploration (KDD) is considered as the whole process of extracting the necessary knowledge from the data, and the data extraction is considered as a special stage in the KDD layer. Process uses a special process to delete patterns from a file. Discovery is a discussion topic that covers a wide range of areas, including technology education, style knowledge, information intelligence, knowledge, information representation, high performance, and expert [6]. KDD is an advanced process of extracting standards that is reliable, cost-effective, and understandable by big data. “Garim” can be considered as the foundation of knowledge that creates knowledge after examination and improvement [7].

As shown in Figure 1, the data mining process consists of the following steps:

1. Data cleaning: remove data noise and data obviously irrelevant to the mining topic;
2. Data integration: in the future, the related data in the white multi data sources will be combined together;
3. Data selection: extract data related to analysis tasks from the database;
4. Data conversion: convert the data into a data storage form that is easy for data mining;
5. Data mining: this is an important step in developing knowledge. Its function is to apply data structures or intelligent processes to the extraction of knowledge permanently;
6. Pattern evaluation: select meaningful patterns and knowledge from the mining results according to certain evaluation criteria;
7. Knowledge representation: its function is to use visualization and knowledge expression technology to show the relevant knowledge mined to users.

Data mining steps can interact with users or knowledge bases, provide interesting patterns to users, or store them in knowledge bases as new knowledge [8]. Data mining is used to discover hidden patterns, which is the most important step in the whole process of knowledge discovery.

Based on the data mining process shown in Figure 2, a typical data mining system is shown in Figure 2, which mainly includes the following main components:

1. Database, data warehouse, or other databases: this indicates that data mining objects contain one or
more databases, data warehouses, data tables, or other databases. In general, data processing and data aggregation should be done prior to the implementation of the equipment.

(2) Database or data storage server: this type of server is responsible for reading the affected data as per the user’s data extraction request;

(3) Basic knowledge: it is used to store the registry knowledge needed for data retrieval. This knowledge can help guide data mining processes or measure mining profits.

(4) File deletion: this is the most important part of file deletion. It usually consists of a group of mining functional modules to perform mining functions such as qualitative induction, organizational analysis, distribution induction, evolutionary arithmetic, and deviation analysis [9].

3.2.2. Task Model of Data Mining

(1) Concept Type Description: Qualitative and Comparative. A concept is often an overview of the overall situation of a data set containing a large amount of data. Data are associated with categories or conceptual models, a large number of data sets are summarized, and concise and accurate descriptions are obtained. Such descriptions are called conceptual descriptions. This description is called conceptual description. There are mainly two methods to obtain the concept description: one is to use more generalized attributes to summarize the analyzed data, which the analyzed data are called the target data set. The other compares the characteristics of two types of data analysis and provides a summary of the results, and the two types of data analysis are called plan data and comparison data.

(2) Association Analysis. Association analysis is used to discover association rules that show the condition that attribute values frequently appear together in a given data set. Association analysis is widely used in transaction data analysis. It is more formally described as follows: association rules are rules in the form of \( X \Rightarrow Y \), that is, 

\[
A_1 \land \ldots \land A_m \Rightarrow B_1 \land \ldots \land B_n,
\]

where \( A_i (i \in \{1, \ldots, m\}), B_j (j \in \{1, \ldots, n\}) \) is an attribute value pair. Association rules are interpreted as “data tuples that meet the conditions in \( X \) also meet the conditions in \( Y \)” Correlation analysis is widely used in marketing, transaction analysis, and other application fields. Association rule mining first finds out the frequent item sets in the data, for example, sets \( A \) and \( B \) meet the minimum support threshold and then generate association rules like \( A \Rightarrow B \) from them. These rules also satisfy the predefined minimum confidence threshold of the probability of satisfying \( B \) under the condition of satisfying \( A \) [10].

(3) Cluster Analysis. The methods of cluster analysis and classification prediction are obviously different. The classification prediction model uses the data of known categories, which belongs to the teacher-supervised learning method. However, the data processed by cluster analysis are non-categorical, which belongs to the non-teacher supervised learning method. In cluster analysis, first of all, according to the basic cluster analysis principle of “maximizing the similarity of data objects in each cluster, minimizing the similarity of data objects in each cluster,” and using the calculation formula to measure the similarity of data objects, the data objects are divided into several data groups. Therefore, the similarity of data objects in one group is greater than that of data objects in different groups. Each group obtained by group analysis can be considered as a
collection of data items that fall into the same category. From this similar data, we can obtain the appropriate patterns of distributed theory by studying classification. It is also possible to obtain hierarchical models of old data that are organized by extension to the receiving group as shown in Figure 3 [11].

(4) Heterogeneous Analysis. Heterogeneous analysis can be described as follows: given the set of N data points or objects and the expected number of outliers K, it is found that objects are incompatible and exceptional compared with the remaining data. The problem of heterogeneous analysis can be regarded as two problems: (1) define what kind of data in a given data set can be considered inconsistent; (2) find an effective method to mine such outliers [12, 13].

3.3. Structure and Implementation of Data Mining Model. Data mining is an integration that affects many sectors. As shown in Figure 4, it includes database procedures, statistics, technology training, training courses, data science research, and other disciplines. In addition, depending on the data extraction process used, technology in other places may be used, such as neural networks, dim/coarse light theory, representation, inductive logic programming, or high performance. Depending on the type of data to be extracted or the data to be extracted, the data can be combined and integrated with spatial data analysis, data extraction, data acquisition design, graphic analysis, signal processing, computer graphics, web technology, economics, and psychology and participated in data research [14].

3.4. Application of Data Mining Technology in Education

3.4.1. Selection of Teachers’ Teaching Methods. In the teaching process, teachers can use a variety of teaching methods to complete their teaching tasks such as teaching method, discussion method, experimental method, computer-assisted teaching method, visiting method, investigation method, and practice method. In general, one or several methods can be adopted. Therefore, it is possible to use data mining to determine which method to use in the next step to extract data from the data and meet the needs of learning, making it easier for students to understand knowledge [15]. For example, the evaluation of all learners in the teaching method, the evaluation of the results of learning from different teaching methods, and the process of regression linear analysis and the rules of correlation language will be used to determine which students and classes fit into these teaching methods so that further training scores can be achieved [16].

3.4.2. Learning Evaluation and Student Feature Mining. Learning evaluation is one of the important responsibilities of educators. Assessing students’ learning behavior not only plays a role in information feedback and stimulating students’ learning motivation but also means to check the curriculum plan, teaching procedures, and teaching purposes, or a way to examine students’ individual differences and facilitate teaching in accordance with their aptitude. The evaluation should follow the principle of “comprehensive evaluation contents, diversified evaluation methods, multiple evaluation times, and the organic combination of self-evaluation and mutual evaluation.” By using data mining tools to analyze and process students’ academic performance, behavior records, and reward and punishment databases, we can get students’ evaluation results in time and correct students’ bad learning behaviors in time. At the same time, it can reduce the workload of teachers and overcome the unfairness and objectivity of teachers’ subjective evaluation [17].

3.4.3. Intervention on Teacher-Student Behavior. The school teaching management database records the learning, work, social activities, rewards, and punishments of students and teachers in each session and uses the association analysis of data mining to find the internal relationship between teachers and students’ various behavior activities. For example, the rule “C can be deduced when there are a and B,” that is, when there are a behavior and B behavior, there will be C behavior. In the actual situation, if students or teachers are found to have a and B behaviors, they can immediately analyze the possibility of producing C behavior and timely formulate strategies to promote or stop the occurrence of C behavior [18].

3.4.4. Reasonable Course Setting. The course learning of students in school is gradual, and there is a certain relationship and order between courses. Before learning a higher-level course, you must take some precourses first. If the precourse is not well studied, it will inevitably affect the study of subsequent courses. In addition, different classes in the same grade learning the same course sometimes have great differences in the overall scores of students in the class due to different teachers and class cultures. By using the examination results of previous students in various disciplines stored in the school teaching database, combined with the related functions of data mining, such as association analysis and time series analysis, we can mine useful information from these massive data, help analyze the correlation, regression, and other properties of these data, get some valuable rules and information, and finally find the reasons that affect...
students’ performance. On this basis, we make reasonable arrangements for the curriculum [19].

3.4.5. Examination. Examination is the test of teaching and learning effect, and it is one of the indispensable links in teaching. Although we can generally evaluate the achievements made in a certain period from the point of view of score, it does not effectively explain in which factors the score is related to, and we cannot know the key factors of success and failure in teaching, which cannot promote teaching and learning. Moreover, the level of students’ scores is also closely related to the quality of test questions. Therefore, exploring effective methods to evaluate the quality of test questions (difficulty of test questions, comprehensiveness of knowledge points, etc.) is also of great significance in the actual teaching process. If the association rules in data mining are applied to the test paper analysis database, and then, indicators such as difficulty, discrimination, and relevance of each question are analyzed according to the student’s scores, and teachers can make a more accurate evaluation of the quality of the test questions and then can be used to check their own teaching situation and students’ mastery and provide guidance for future teaching.

4. Result Analysis

Classroom teaching evaluation not only regulates, controls, guides, and promotes teaching but also has strong guidance. It is an important part of school teaching management and the main means to evaluate teaching achievements [20]. The school conducts a classroom teaching evaluation survey every semester, accumulating a large amount of data. By using data mining technology and applying association rules to teaching evaluation data, we can mine some useful data, reasonably configure class teachers and enable students to better maintain a good learning attitude, so as to provide decision support information for teaching departments, promote better teaching work, and improve teaching quality.

Figure 4: Schematic diagram of close integration of data mining and related disciplines.

Figure 5: Statistical chart of final test scores.
For example, the final test scores of students taught by different teachers are shown in Figure 5. It is advisable to set “teacher A’s teaching ≥ student’s high score” as $X \Rightarrow Y$, and then, the support and confidence of the association rule “teacher A’s teaching ≥ student’s high score” are (1) and (2), respectively.

\[
\text{Support}(X \Rightarrow Y) = \frac{60}{315} = 19.05\%, \quad (1)
\]

\[
\text{Confidence}(X \Rightarrow Y) = \frac{60}{105} = 57.14\%. \quad (2)
\]

By analogy, calculate the support and confidence of other associations, as shown in Table 1. The support and confidence of “teacher B teaching ≥ students get high scores” are 28.57% and 72%, respectively, and the support and confidence of “teacher C teaching ≥ students get low scores” are 19.05% and 70.59%, respectively, indicating that these two rules are tenable to a large extent.

| Teacher A | Teacher B | Teacher C | Teacher A | Teacher B | Teacher C |
|-----------|-----------|-----------|-----------|-----------|-----------|
| Support   | ≥Maximum score | ≥Maximum score | ≥Maximum score | ≥Minimum score | ≥Minimum score | ≥Minimum score |
| Confidence level | 19.05% | 28.57% | 7.94% | 14.29% | 11.11% | 19.05% |
|            | 57.14% | 72.00% | 29.41% | 42.86% | 28.00% | 70.59% |

5. Conclusion

This article introduces the use of data mining technology to improve the learning experience. Gaining the necessary knowledge from a wide range of documents can not only support further refinement, improvement, and refinement of the study but also address issues. Of learning, analysis of the rules of the data mining technology organization provides a correlation between the faculty and the scores of the students. The support and confidence of students’ scores taught by teacher B are 28.57% and 72%, respectively, and the support and confidence of students’ scores taught by teacher C are 19.05% and 70.59%, respectively. After data association, the two rules are established to a large extent. Therefore, data mining technology can effectively optimize the effect of intelligentization of Chinese international education.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare no conflicts of interest.

References

[1] J. Gonondo, “Confucius institute and the development of Chinese language teaching in Cameroon,” Journal of Education and Practice, vol. 12, no. 3, pp. 34–39, 2021.
[2] A. Cun and M. Mcvee, “Experiences of Chinese international students within higher education: a narrative study,” Pедагогія, vol. 17, no. 3, pp. 1–22, 2020.
[3] C. Wei and L. Zhong, “Research on the problems and countermeasures of higher education management informatization development in the computer internet era,” Journal of Physics: Conference Series, vol. 1992, no. 2, Article ID 022126, 4 pages, 2021.
[4] A. A. Jalal and B. H. Ali, “Text documents clustering using data mining techniques,” International Journal of Electrical and Computer Engineering, vol. 11, no. 1, pp. 664–670, 2021.
[5] S. A. Wulandari, H. Kuswara, and N. Palasara, “Analisis penerapan data mining pada penjualan kerupuk rambak menggunakan metode naive bayes classifier untuk optimasi strategi pemasaran,” Jurnal SITEM: Sistem Informasi dan Teknologi, vol. 3, no. 2, pp. 83–94, 2021.
[6] Z. J. Wang, J. Wang, X. Zeng, and Y. Xing, “Knowledge discovery and data mining-based garment size selection for mass customization,” Journal of Fiber Bioengineering and Informatics, vol. 13, no. 3, pp. 113–128, 2020.
[7] B. Zhu, L. Zhang, and Y. Zhang, “The early warning method of drug adverse reaction monitoring based on data mining algorithm was studied,” Journal of Physics: Conference Series, vol. 1852, no. 3, Article ID 032052, 2021.
[8] M. Liao, J. Patton, R. Yan, and H. Jiao, “Mining process data to detect aberrant test takers,” Measurement Interdisciplinary Research and Perspectives, vol. 19, no. 2, pp. 93–105, 2021.
[9] B. S. Ahmed, M. L. Ben Maati, and M. Al-Sarem, “Predictive data mining model for electronic customer relationship management intelligence,” International Journal of Business Intelligence Research, vol. 11, no. 2, pp. 1–10, 2020.
[10] H. Lu, H. Shinzawa, and S. G. Kazarian, “Intermolecular interactions in the polymer blends under high-pressure co2studied using two-dimensional correlation analysis and two-dimensional disrelation mapping,” Applied Spectroscopy, vol. 75, no. 3, pp. 250–258, 2021.
[11] M. Hourqueig, G. Bouzille, M. Mirabel et al., “Risk of atrial fibrillation in hypertrophic cardiomyopathy: a clustering analysis based on the French registry on hypertrophic cardiomyopathy (remy),” Archives of Cardiovascular Diseases, vol. 13, no. 1, pp. 21–22, 2021.
[12] J. Wang and J. Yu, “Train performance analysis using heterogeneous statistical models,” Atmosphere, vol. 12, no. 9, p. 1115, 2021.
[13] V. Oleshko, “Phase evolution analysis during real-time solid-state chemical lithiation of crystalline thin window silicon membranes using low-loss stem-eels imaging,” Microscopy and Microanalysis, vol. 27, no. S1, pp. 2728–2730, 2021.
[14] Y. Li, H. Zhang, and S. Liu, “Applying data mining techniques with data of campus card system,” IOP Conference Series: Materials Science and Engineering, vol. 715, no. 1, Article ID 012021, 2 pages, 2020.
[15] Y. Zhang, X. Kou, Z. Song, Y. Fan, M. Usman, and V. Jagota, “Research on logistics management layout optimization and real-time application based on nonlinear programming,” Nonlinear Engineering, vol. 10, no. 1, pp. 526–534, 2021.
[16] H. Wang and Z. Du, “Research on application of teaching informatization construction based on cloud computing,” *International Journal for Innovation Education and Research*, vol. 9, no. 5, pp. 288–294, 2021.

[17] R. Huang, S. Zhang, W. Zhang, and X. Yang, “Progress of zinc oxide-based nanocomposites in the textile industry,” *IET Collaborative Intelligent Manufacturing*, vol. 3, no. 3, pp. 281–289, 2021.

[18] J. Hu, Y. M. Kang, Y. H. Chen, X. Liu, X. Li, and Q. Liu, “Analysis of aerosol optical depth variation characteristics for 10 years in urumqi based on MODIS_C006,” vol. 39, no. 8, pp. 3563–3570, 2018.

[19] D. Selva, B. Nagaraj, D. Pelusi, R. Arunkumar, and A. Nair, “Intelligent network intrusion prevention feature collection and classification algorithms,” *Algorithms*, vol. 14, no. 8, p. 224, 2021.

[20] M. Fan and A. Sharma, “Design and implementation of construction cost prediction model based on svm and lssvm in industries 4.0,” *International Journal of Intelligent Computing and Cybernetics*, vol. 14, no. 2, pp. 145–157, 2021.