Application of Data Mining Technology in Pedagogy

Xinmei Zhao[1] and Juncheng Duan[2](✉)

[1] The College of Arts and Sciences, YunNan Normal University, Kunming 653000, China
[2] Yunnan University Secondary School, Kunming 650031, China

Abstract. With the continuous development of the information age, data mining technology is becoming more and more mature to meet the needs of people for a large number of information processing. At present, the application of data mining technology in finance, communication, transportation and other fields is more and more, but the application in the field of education is relatively less. In view of this situation, in the traditional analysis method, data mining algorithm is used to analyze and study the related applications of pedagogy.

Keywords: Data mining · Pedagogy · Association rules · Apriori · Database

1 Introduction

With the gradual maturity of tree database technology and the extensive application of database management system on the Internet, a large number of explosive data have been produced. However, these data are not well analyzed and mined, resulting in the situation of "too much data but forgotten". Therefore, in order to improve the utilization rate of information, data mining technology emerges as the times require and is widely used rapidly. Data mining is a process of extracting potentially useful information and knowledge from a large number of, incomplete, noisy, fuzzy and random data. Here, "data" refers to the collection of facts, records and the original information related to things. "Knowledge" is a more abstract description of the contained information and the process of analyzing a large amount of data, including data preparation, pattern search, knowledge evaluation and repeated modification and refinement [1]. Mining process requirements are extraordinary, that is to require a certain degree of intelligence and automation.

2 Data Mining

Data mining algorithm is a set of calculation methods to create data mining model based on data. The algorithm will first analyze the data proposed by researchers, and look up specific types of models and trends, and then create models according to requirements. The commonly used algorithms in research are association rule mining, decision tree

M. A. Jan and F. Khan (Eds.): BigIoT-EDU 2021, LNICST 391, pp. 438–447, 2021.
https://doi.org/10.1007/978-3-030-87900-6_51

algorithm and clustering mining algorithm. Association rule mining is to discover the association and correlation existing in a large number of data sets, thus describing the rules and patterns of some attributes appearing simultaneously in a thing. It is one of the most mature main technologies in data mining. The most classic association rule algorithm is Apriori algorithm.

Clustering mining is an important human learning behavior. It is a common phenomenon in nature [2–4]. Aggregation analysis is a mathematical analysis method based on this phenomenon. Its purpose is to divide a large number of data points into several categories, so that the larger the gap between the data in each class, the better, the more obvious, The data in different classes should be similar as far as possible. The smaller the difference is, the better l5.6. Dense and sparse data can be found by clustering, so as to find the global data distribution pattern and interesting relationship between data attributes. The data mining process is shown in Fig. 1.

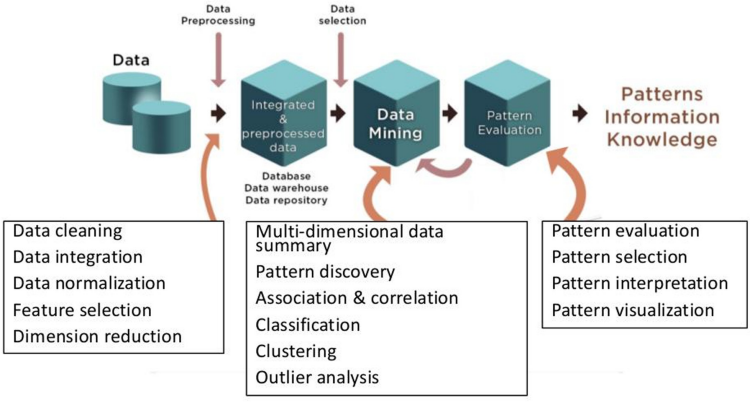

Fig. 1. Data mining process

3 Evolution of Association Rules Mining Algorithm Based on Data Mining Algorithm

The first algorithm of association rules, is the algorithm given by Agrawal and others when they proposed the association rule model. The basic idea is to generate frequency sets by scanning transaction database and count them. If the frequency sets in the previous step appear in the current transaction being scanned, the items in the transaction will be used to expand these itemsets to obtain new candidate sets [5, 6]. However, the major defect of the algorithm is that too many small candidate sets are generated. After that, cumulate and stratify, houstsma and others put forward an association rule algorithm called Setm algorithm, which uses s α L statement to calculate frequency set. The basic idea is to separate the generation and count of candidates, generate candidates by joi operation in SQL, and then save candidate copies and generate T of events in linear

structure_ This is a method to transform association rule mining into SQL statement execution.

Suppose that s% of transactions in transaction database T contain xuy, then S% is called the support degree of association rule $X \cup Y$, which is the ratio of the number of transactions containing X and y to the number of all transactions, that is, the probability value $P(X \cup Y)$, which can be recorded as:

$$\sup port(X \cup Y) = P(X \cup Y) = S\% \tag{1}$$

The confidence degree of $X \rightarrow Y$ is the ratio of the support degree including X and y to the support degree containing x, i.e. the probability value $P(Y|X)$, which can be recorded as:

$$confidence(X \rightarrow Y) \frac{\sup port(X \cup Y)}{\sup port(X)} = P(Y|X) \tag{2}$$

4 Application of Data Mining in Pedagogy

4.1 Data Electronization

In order to make data electronic, we should first build a suitable network platform, which is divided into two processes. If we need to collect data, we need to build an object-oriented network platform. In the research of College Students' social system and interpersonal relationship, the scale data is collected automatically on the web. Firstly, the scale is converted into a web page written in assembly language, and then the electronic version of the scale is connected with the table in the database by using aspnet. In this way, only the subjects log in the designated website can complete the scale on the Internet, and the data is directly stored in the database [7, 8]. The database management system uses SQL Server 2000 - and the same method is used in the study of Internet addiction and attention relationship bias.

If we analyze the existing data, we can skip the step of data collection and input the data into the data processing software directly. In the research on the relationship between youth and youth, a data warehouse was created to store the existing data, and sqserver2000 was selected as the construction platform of the data warehouse. Since there are not too many dimensions in the research of adolescent peer relationship and the dimension hierarchy is not complex, in order to consider the query efficiency and whether the user can easily understand, we decided to use star structure to create data warehouse. The star structure of adolescent peer relationship research is shown in Fig. 2.

4.2 Data Mining Simulation Analysis

After transforming the data into the required electronic text format, simple data processing is carried out. After eliminating some missing or obviously wrong data, data mining can be carried out [9]. The common methods of data mining are association rule mining decision tree algorithm and clustering mining algorithm. In this paper, association rule

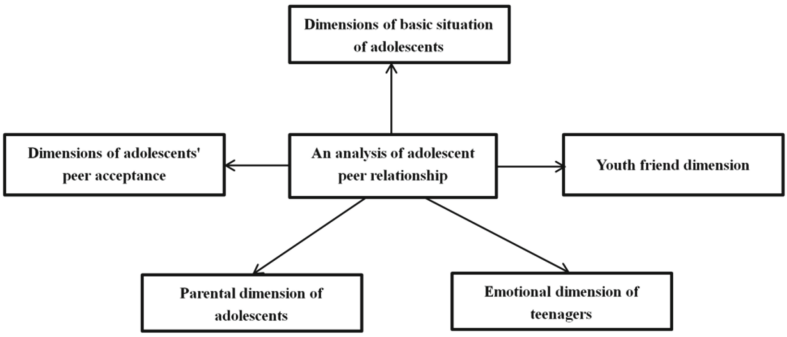

Fig. 2. Star structure of adolescent peer relationship research

mining algorithm is used to explain in detail. In the research of social network and interpersonal relationship of university students, the subjects' satisfaction with interpersonal relationship is selected as an example to mine association rules.

The steps of association rule mining are as follows:

(1) Select data according to the target of association rule mining, select the attributes shown in the above table, select the data through the following sq statement, and connect JE η Xinx table and zongle table through student number. Will you be satisfied with your relationship with your parents, whether you are satisfied with your relationship with your parents, Zongjie b where a student number = B student number.
(2) Save the selected data: first, insert the data selected in the first step into a table. Sq statement is as follows: SERT into guanlianguize1 select gender, whether you are an only child, are you satisfied with your interpersonal relationship, with your parents, with your partner, with your teacher, What kind of evaluation do you think people who know you better will give to your interpersonal relationship.
(3) Construct a transaction database that satisfies the mining of association rules, add an attribute item to the table guanlianguize 'and assign a value to TEM through the following sq statement. Note that the data mining mapping code of each attribute is shown in the above table. New table socia_ guanlian_ 1 as the event database of data mining, import item.

The results show that college students who have a high evaluation of their own interpersonal relationship have higher satisfaction with their own interpersonal relationship [10]. College students who have a low evaluation of their own interpersonal relationship have lower satisfaction with their own interpersonal relationship. Students who pay less attention to interpersonal relationship have lower satisfaction with interpersonal relationship. These data are related to each other, so we call on college students to pay more attention to interpersonal relationship and interpersonal rules, which is conducive to their early recognition of their shortcomings in interpersonal communication, so as to lay a good foundation for entering the society and creating greater social value.

The simulation analysis is shown in Fig. 3 and Fig. 4.

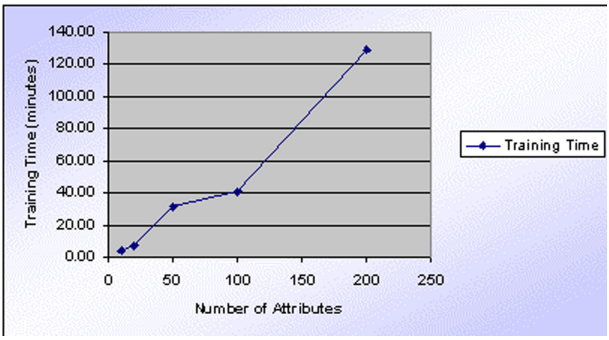

Fig. 3. Data mining simulation for teaching effect

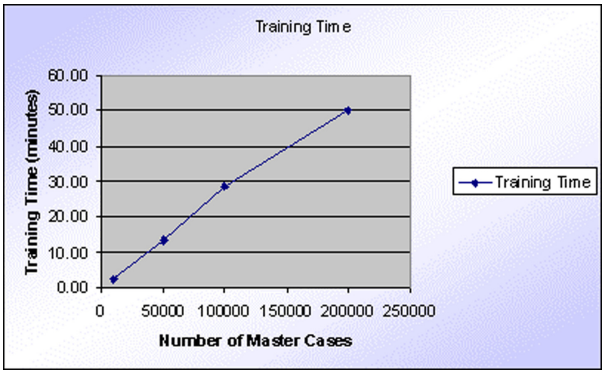

Fig. 4. Data mining simulation for teaching mode

5 Knowledge and Its Subject Attribute

5.1 Subject Attribute

Before discussing the knowledge of pedagogy, it is necessary for us to make a clear distinction between "subjects in the sense of academic research" and "subjects in the sense of teaching subjects". The "subject" mentioned in this paper is obviously not completely different from what we say everyday, such as Chinese, nature, society and other subjects in primary and secondary school curriculum [11, 12]. According to the explanation of Dogan. M. in the International Encyclopedia of social and behavioral sciences, the term "disc upline" refers to both the organizational units in various educational programs and the organizational units in knowledge production, Discipline is basically synonymous with academic classification or teaching subjects. For a long time, although there is little clear demarcation between them in academic circles, "there must be differences in the connotation of academic classification and teaching subjects.". The "discipline in teaching" should be called "teaching subject" in a strict sense, and its knowledge organization aims at promoting the development of teaching objects; while the "discipline in academic research" produces and organizes knowledge in the sense of epistemology.

"As a discipline in a research field, it does not have the meaning of knowledge transfer as teaching.

5.2 Subject Premise

The "discipline" here is based on the emergence of science. "Discipline" is a branch system of scientific knowledge with a fixed research object, while "science and discipline are the relationship between the whole and the part". Science is a branch of all disciplines, and discipline is a branch of local disciplines [13–15]. We can say that since the emergence of science, the classification of knowledge has been inextricably linked with the division of disciplines. From ancient Greece to the middle ages, knowledge has almost always been integrated into the "Virtue" and "wisdom" of philosophy. Aristotle's division of knowledge disciplines is only an attempt of theoretical systematization, which is fundamentally different from what we call "disciplines" today. "With the attention to higher learning, philosophers began to classify knowledge, From the late Middle Ages and the Renaissance, "when universities spread all over Europe, scholars wrote: science has established its own palace. this process of science establishing its own door to build its own house of knowledge, It is a process of knowledge differentiation, classification and reorganization, and the process of separation is based on the difference of knowledge itself. It has internal connection with the classical teaching "seven arts", but it has important differences. "It is worth mentioning that the branches of knowledge in the middle ages, twos and quadriviums, were named after the forks and crossroads in Latin.

5.3 Science Curriculum

In the field of pedagogy, the types of knowledge are very complex, but in many cases, we tend to simply think that the knowledge in the field of education is what we teach students, and that these contents are also the knowledge of academic disciplines, but we lack the understanding of the difference between the knowledge of teaching subjects and that of academic disciplines, It pays little attention to the knowledge of pedagogy itself [16]. The literature search before writing this paper also found that a large number of literatures about knowledge in the field of education mostly refer to the subject knowledge to be taught to students, but the discussion about pedagogy itself is rare, and in many cases, we often confuse "teaching content knowledge" with "subject knowledge". Shulman (L) and others call the subject knowledge as "pedagogic content knowledge", Teaching content knowledge represents the essence of content and teaching method. It exists in the intersection of "content knowledge" and "Pedagogical Knowledge". Learning content knowledge is a kind of special knowledge that distinguishes the understanding of subject content from that of subject experts. We can regard this kind of teaching content knowledge as teachers' reorganization and limited and appropriate transformation of relevant subject knowledge and its presentation form according to the logic of pedagogy knowledge system, so as suitable for teaching. This is also an important knowledge representation to distinguish "physics and physics research" as a middle school teaching subject [17–19]. As a teaching subject, we can organize part of the knowledge of physics into a "physics" course, and we can also organize the knowledge of physics, chemistry

and other subjects into a "science" course. This itself also shows that the logic of organizing knowledge of teaching subjects is different from that of "Chinese characters" in the sense of learning.

6 The Subject Characteristics of Pedagogical Knowledge

6.1 The Formation and Development of Education Discipline

In the process of the formation and development of almost all disciplines including mathematics, "differentiation" is a strategy used by disciplines to protect themselves from invasion and self doubt. The establishment of the academic terminology system of this discipline is often one of the results of the discourse practice of this discipline. This process of differentiation is manifested in the division and strengthening of subject boundaries by knowledge classification. In the process of discipline formation, in addition to the academic training and learning organization, the development of bibliography also played a key role. Historically, bibliography played an important role in the efforts of libraries, museums and archives to save the flood [20–22]. This effort was achieved by sorting books, utensils and documents and making them known to the growing community. As shown in Fig. 5. Up to now, the bibliographic arrangement of the library still has an extremely important influence on the classification of knowledge and subjects. In a certain sense, we can say that the classification of knowledge by professional academic training and bibliography is an important force to promote the specialization or differentiation of Chinese characters.

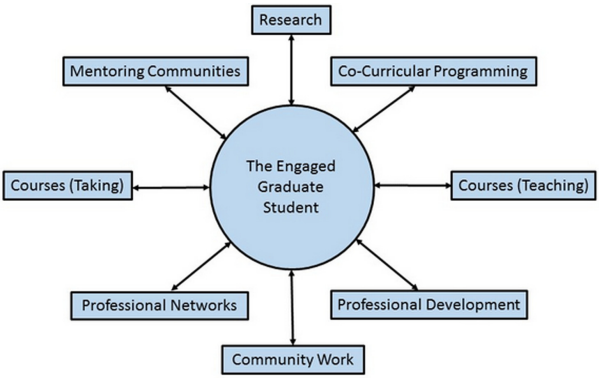

Fig. 5. Development of education discipline

6.2 Subject Classification

The subject classification of knowledge is closely related to knowledge and the social function of the subject producing such knowledge. "Every individual who performs a certain social role is considered by his social circle to have or is confident that he has

the necessary knowledge for normal role performance. Pedagogical knowledge is not limited to and sufficient for educators. In other words, teachers must have the knowledge of teaching phonetics, but not all teachers must have the knowledge of teaching phonetics [23]. We can take it as a criterion to judge that a kind of knowledge is teaching knowledge. The fundamental reason why a discipline can produce and exist independently is that it can produce some unique knowledge to meet the needs of society. The existence of teaching phonetics as an independent discipline indicates that there must be some special knowledge only belonging to pedagogy. Therefore, we can establish another criterion to judge the first knowledge of teaching, that is, only teaching is the discipline that mainly produces and provides this kind of knowledge. As the exclusive product of academic discourse practice of a discipline, the knowledge of the discipline is always closely related to the theme of the discipline and serves the main social functions of the discipline. This article can be used as the third criterion for us to determine whether a certain knowledge is pedagogical knowledge.

6.3 Subject Theme of Pedagogy

The subject purport of pedagogy has not been clearly differentiated, which is one of the important reasons why the subject characteristics of pedagogy knowledge are not clear and prominent. The indistinct theme of pedagogy is related to the particularity of pedagogy's own development history. From the pedagogy, which is one of the subjects of normal education and mainly focuses on classroom teaching, to the science of education, which focuses on all kinds of social phenomena in the whole field of education, to the comprehensive, holistic and independent pedagogy, which focuses on all kinds of educational phenomena, The theme of pedagogy has been in subtle changes [24–26]. Anyway, the theme of pedagogy must be about educational phenomena rather than other phenomena. Therefore, pedagogical knowledge must also be about educational phenomena rather than political, economic, legal and other social phenomena. However, it is still difficult for pedagogy to give a clear answer to this fundamental question. However, although there is no clear standard answer in theory, we can clearly distinguish education from experience. For example, when I pay for a shirt, the price of the shirt, the design of the shirt, the idea of guiding me to eliminate my wife, and so on, are obviously not educational phenomena.

7 Conclusion

With the continuous development of the information age, many researchers focus on Pedagogy from different aspects and different fields. In this process, a large amount of scientific data will be accumulated. Therefore, it is urgent to use computer means to conduct systematic data integration, which is conducive to deeper analysis of the data. The application of data mining technology in the field of pedagogy will have a broad prospect and can be further promoted as the research direction of pedagogy in the future.

Acknowledgements. Yunnan Provincial Education Department of the sixth batch of university science and technology innovation team project construction "Yunnan Province university wisdom tourism science and technology innovation team".

References

1. Ming, Z.: Data Mining. China University of science and Technology Press, Hefei (2002)
2. Wang, H., Wu, J.: Research on data mining algorithm based on SQL Server 2000. Comput. Eng. Des. **29**(3), 759–761 (2008)
3. Bai, Y., Gao, Y.: Research on the relationship between Internet addiction and attention bias based on data mining technology and database design. Learning science research center of Southeast University (2009)
4. Wang, X., Yu, L.: Construction of network education resource database based on data mining. J. Southwest Normal Univ. **30**(1), 163–167 (2005)
5. A brief introduction to the master program of adult education in Hebei University. J. College Adult Educ. Hebei Univ. **23**(01), 125 (2021)
6. Ke, L., et al.: Application and effect analysis of mixed teaching mode in quality training of public health talents during epidemic period. Health Voc. Educ. **39**(06), 114–116 (2021)
7. Announcement on the journal being recognized as an academic journal by the State Administration of press, publication, radio, film and television. Road Succ. **09**, 146 (2021)
8. Juan, Z.: Reflections on the discipline development of higher education. Teach. Educ. People (Higher Education Forum) **09**, 10–12 (2021)
9. Li, F., Pang, Y.: Supporting strategic decision with education research and serving reform and development with practice exploration – Summary of the fourth national high end Forum on theory and practice of graduate education discipline construction [J/OL]. Grad. Educ. Res. 1–7 (2021). https://doi.org/10.19834/j.cnki.yjsjy2011.2021.03.01
10. Wu, S.: "Five focuses" to improve the effectiveness of learning and using in Distance Education. Hefei evening news (2021) (A03)
11. Chen, M., Mo, J.: Building a high quality education system to make people satisfied with education. Nanning daily (2021) (005)
12. Yao, S., Gaofeng, L.: Focusing on the cultivation of the sense of Chinese national community, promoting the innovation and development of national ideological and Political Education – Summary of the Forum on "Ideological and political education and casting the sense of Chinese national community." Sch. Party Build. Ideol. Educ. **06**, 96 (2021)
13. Zhao, F., Li, Y., Liu, M., Xue, D., Wang, S.: Propositional guidance and performance support: a Study on the dual knowledge and ability cultivation structure of "future physical education teachers" in Sweden. J. Tianjin Inst. Phys. Educ. **36**(02), 125–133 + 150 (2021)
14. Guangfen, Y., Hui, S., Yuan, Y.: Interdisciplinary research and discipline construction of Vocational Education: context, return and change. China Vocat. Tech. Educ. (09), 11–17 (2021)
15. Qi, J., Gu, Y.: Research on the historical development trend of positivism paradigm from the perspective of Comparative Education. J. Yanbian Univ. (SOCIAL SCIENCE EDITION) **54**(02), 121–127+144 (2021)
16. Xiong, Y., Cheng, L.: Research on discipline development and curriculum construction of family education in the new era. Contemp. Youth Stud. (02), 5–12 (2021)
17. Yaming, Y.: The infiltration of moral education into the education of new music and Art: a review of research methods of music pedagogy. China J. Educ. **46**(03), 163 (2021)
18. Jie, S.: On the change of the category of mutual benefit between teaching and learning and its contemporary value. J. Hebei Normal Univ. (EDUCATION SCIENCE EDITION) **23**(02), 54–63 (2021)
19. Yukun, L.: Research on the innovation path of music education mode in Colleges and Universities under the multicultural background – a review of research methods of music pedagogy. Forestry Ind. **58**(03), 113 (2021)

20. Dong, Q.: Achievements and problems in the development of Ideological and political education since the reform and opening up. Forum High. Educ. **03**, 1–4+15 (2021)
21. Qianli, Y.: Dilemma and reflection of national defense education research in Colleges and universities. Forum High. Educ. **03**, 110–113 (2021)
22. Jianjun, C.: Reflections on Pedagogy Teaching from the perspective of teacher qualification examination. J. Liaoning Vocat. College **23**(03), 83–86 (2021)
23. "Vocational Education Forum" in 2020 National People's Congress reprinted newspapers and periodicals ranked again good results. Vocat. Educ. Forum **37**(03) (2021)
24. Chen, X.: Optimization and innovation of physics classroom teaching method in junior middle school. Math. Phys. Chem. Middle School Students (Teaching and learning) (03) (2021)
25. Introduction to cover characters. Art Educ. (03) (2021)
26. The editorial department of our journal. Continuous contributions in 2021: what kind of history education is needed in the new era. History Reaching Reference in Middle School, vol. 6 (2021)