

PAPER • OPEN ACCESS

Practical Exploration of Association Rules Mining Algorithm in Open Education

To cite this article: Shaozhen Huang 2019 *IOP Conf. Ser.: Mater. Sci. Eng.* **573** 012114

View the [article online](#) for updates and enhancements.

Practical Exploration of Association Rules Mining Algorithm in Open Education

Shaozhen Huang¹

¹ Hainan Open University, Hainan, Haikou, 570100

Abstract: Open education is a new type of education, which is produced along with the development of modern information technology. It is a teaching model based on modern information technology. With the continuous reform of China's education system, the autonomy of students in major colleges and universities is increasing. However, because students do not understand the course, it is very difficult to select courses. Therefore, it is especially important to excavate the university teaching platform database. This paper briefly introduces the concept of association rules mining algorithm, mainly studies the application of association rules mining algorithm in the open education learning system, in order to provide meaningful guidance for students' elective courses, thus improving students' academic performance.

1. Introduction

The characteristic of open education is that students learn independently, make full use of the various resources provided by modern distance education, change traditional educational concepts, and gradually transform into modern open and distance education. And from the teacher-centered passive learning to the student-centered autonomous learning. Students can learn at any time and any place without delaying their work, while learning knowledge and improving their quality. Organize teaching with "courses as a unit" to enable students to conduct courses more independently and flexibly. However, students' self-selected courses will inevitably lead to the problem of blind selection of students. We must improve them in order to provide meaningful guidance for students' electives and facilitate their selection. Data mining, as a new technology that can automatically and efficiently extract valuable information and knowledge from massive data to effectively support decision-making. Applying it to open education can provide useful guidance for student learning^[1].

Data mining is the use of pattern recognition, statistics, and mathematical techniques to filter out new and meaningful relationships, patterns, and trends from a large amount of stored data. It is the process of extracting hidden and potentially useful information and knowledge from a large number of incomplete, noisy, fuzzy, and random data. It relies mainly on artificial intelligence, machine learning and statistical techniques to inductively infer data and uncover potential patterns. Predict future trends and support decision making.

2. Association rule mining algorithm

The concept of association rules is first of all by R. Agrawal et al. propose a rule that describes the (potential) relationship between data items (attributes, variables) in a database. It has become a very important research direction in data mining.

There are four steps to discover the association rules: 1) Pre-processing the data related to the mining task. According to the requirements of the specific problem, the database is operated



accordingly to form a standardized database D. 2) For D, find all the itemsets that satisfy the minimum support, that is, the large itemsets. Since the database is generally large, this step is the core of the algorithm. 3) Generate a rule that satisfies the minimum confidence to form a rule set R. 4) Explain and output R.

Apriori, a classical association rule mining algorithm, is a basic algorithm for finding frequent itemsets. The core of the algorithm is mainly to find frequent itemsets, mainly based on the Apriori nature: all non-empty subsets of frequent itemsets must also be frequent. This property can be used to effectively compress the search space. The main idea of the algorithm: To find L_k , connect L_{k-1} with itself to generate a set of candidate k itemsets. The set of candidates is recorded as C_k , and then C_{k+1} is empty. When C_k , $k = 1, 2, \dots, k$ are generated, C_k is compressed using a pruning strategy. The use of any infrequent (k-1)-item set is unlikely to be the Apriori property of frequent k-item sets, deleting those k-candidate itemsets whose (k-1)-subsets are not in L_{k-1} ^[1].

The Apriori core algorithm is described as follows:

Input: database D, minimum support threshold min_sup

Output: frequent item set L in D

method:

$L1 = \{\text{find_frequent_1-itemsets}(D)\}$

For($X=2$; $L_{X-1} \neq \emptyset$; $X++$)do begin

$C_x = \text{Apriori_gen}(L_{x-1}, \text{min_sup})$

For all transactions $t \in D$ do begin

$C_t = \text{subset}(C_x, t)$

For all candidates $c \in C_t$ do begin

c. count++

end

3. Application of association rules technology in student score analysis

There will be more or less links between the course itself, between the course and the course, and between the course setting and the course grade. Through data mining technology, it is very intuitive to find valuable information hidden behind student achievement, so as to better guide teachers to complete teaching work. The implementation process of student achievement data mining is shown in Table 1:

Table 1: Mining implementation table

Serial number	Step
Steps 1	Determine the mining object
Steps 2	Selected model
Steps 3	Data collection
Steps 4	Data preprocessing
Steps 4	Data mining
Steps 6	Analysis result
Steps 7	Knowledge application

Specifically speaking, firstly, taking the scores of computer majors in a university as an example, we can find out the influence of the excellence of a certain course on other courses. Then, using the association rule mining algorithm Apriori analysis to explore the degree of influence between the students' grades. A similar conclusion is drawn: the excellent performance of the "C Language Programming" course directly affects the excellent performance of the "Data Structure" course. Then

collect the usual grades and test scores of the student's course, and the relevant teachers will provide relevant data.

Preliminary processing of the obtained data: 1) Integrated data: the data of the collected multiple databases are combined. Randomly select some course scores from students. 2) Clean up the data: When counting the data of the student's score source table, there will be some cases where the attribute lacks the attribute value. For these attributes lacking the attribute value, the cleanup technique is used to fill the vacant data value. 3) Convert data: use a unified format to represent grade data for data mining. This paper explores the excellent relationship between the various courses of computer majors. The use of association rules to mine performance data requires logical data, so the attribute value greater than 90 points is defined as "1", otherwise it is "0"^[3].

Data mining using Apriori algorithm: 1) Establish transaction data table D. According to the mining requirements, only excellent grades are retained, and excellent student records are not obtained for deletion. 2) Frequent 1-item set data table frequent-1-itemsets. The table is used to store the project name X1, X2 and the number of records containing this project. 3) Delete the record with the support count less than the minimum support min-sup in the data table frequent-1-itemset, and get the final frequent 1-item set. 4) Seeking succession of frequent itemsets. 5) Delete the records whose final frequent items are less than the minimum confidence threshold, thereby generating rule information^[4].

Finally, the analysis results, the transaction number of this paper is 11, set the minimum support degree of 30%, the confidence level is 50%. 1) The set L1 of the frequent -1 item set is {X1}, {X2}, {X3}, and all the frequent itemsets are obtained based on this. 2) The candidate set C2 is {X1, X2}, {X1, X3}, {X2, X3}. 3) The candidate set C3 is {X1, X2, X3}. 4) The final result: C3 is the frequent set L3, which is the maximum frequent set, deposited in the data sheet 2. X1, X2, X3 represent different computer courses

Table 2: Association rule data table

Xuehao	X1	X2	X3
0903020006	0	1	0
0903020027	0	1	1
0903020020	1	0	0
0903020016	1	1	1
.....

Calculate the final confidence, delete the record less than the minimum confidence threshold, and get the final association rule, which is stored in the data table 3.

Table 3: Result data sheet

Serial number	specific contents
X1 is excellent and X2 is excellent	X3 has more than 71% excellent possibilities
X1 is excellent and X3 is excellent	X2 has more than 100% excellent possibilities
X2 is excellent and X3 is excellent	X1 has more than 100% excellent possibilities
Excellent X1	X2, X3 have more than 71% and excellent possibilities
Excellent X2	X1, X3 have more than 50% and excellent possibilities
Excellent X3	X1, X2 have more than 83% and excellent possibilities

Implementation effect: Using the mining algorithm Apriori analysis to explore the degree of influence between the students' grades. In turn, as a reference, it is possible to mine valuable information between the course itself, the curriculum and the course, and the curriculum setting and

course achievement, so as to guide the teachers to achieve greater achievements in teaching reform, professional talent training program development, and teaching quality improvement^[5].

4. Conclusions

In general, the analysis of student achievement in open education is a complete data mining process. The user can automatically analyze the selected database as long as the user provides the necessary data. And return the information that the user needs, help the user to make decisions, thus helping the students to choose courses and further study, which plays a very good role in promoting and has certain practical value. The association rule mining algorithm is applied in open education, and the function of the distance learning system is improved through continuous data mining, so that learners have a clear understanding of their own learning. Complete the study in constant reflection and persistence, and truly achieve independent learning. At the same time, teachers have a global understanding of teaching to improve teaching.

Acknowledgements

Fund Project: Hainan Higher Education Research Funding Project

Project No.: Hnky2018-62

References

- [1] Bing L, Hsu W, Ke W, et al. Visually Aided Exploration of Interesting Association Rules[C]// Pacific-asia Conference on Methodologies for Knowledge Discovery & Data Mining. 1999.
- [2] Chen Q, Zhou K, Dan W. Exploration and Practice of the Computer Innovation Education in Mining Engineering[C]// International Workshop on Education Technology & Computer Science. 2009.
- [3] Zhong R, Wang H. Research of Commonly Used Association Rules Mining Algorithm in Data Mining[C]// International Conference on Computational Intelligence & Industrial Application. 2010.
- [4] Zhi L, Guoming S, Mingyu L. A Vector Operation Based Fast Association Rules Mining Algorithm.[C]// International Joint Conference on Bioinformatics. 2009.
- [5] Yan Y, Wang D, Ge Y U. Association Rules Mining Algorithm of Web Pageviews for Personalized Recommendation[J]. Computer Engineering, 2005, 31(1):79-81.