

Application of Apriori Algorithm in Open Experiment

Li Mengshan^{1,+}, Liu Bingxiang² and Wu Yan³

¹Sub-Institute of Science&Technology and Art JingDezhen Ceramic Institute, jiangxi,333001,china

^{2,3}JingDezhen Ceramic Institute, Information Engineering Institute, Jiangxi,333001,china

Abstract. In this paper, we introduce the basic ideas of Association Rules and Apriori Algorithm, and analyse the relationship between the booking database of the open experiment and the implementation of Apriori Algorithm. After the pre-processing of the dataset we have accumulated, we apply the Apriori Algorithm of Association Rules to mine some related information in the booking database. So we can get some important association information both in the same type experiments and in some independent experiments. Then, we analyse the mined information, get the reasons of the association information, and give some advice to the administration of experiment teaching so as to improve the work efficiency.

Keywords: Apriori Algorithm; Association Rules; Teaching Reform; Open Experiment;

1. Introduction

With the rapid development of college education, the goal of higher education for student has been changed from application-oriented to research; inter-disciplinary investigation and innovation talent is one of the primary goals in the new era. The experiment teaching, as the important part of talent cultivation, plays a vital role in higher education. But in the traditional experiment teaching, the teaching methods are usually single and simple, and the contents are also relatively old. In order to overcome the shortcomings of traditional experiment teaching, we must consider the reform of experiment teaching.

The reform of the open experiment teaching is to cultivate the student's subjective initiative, at the same time, let the student's initiative to reveal the research by ourselves, and cultivate the student's ability of observation, thinking, experiment skills and the good scientific attitude of experiment.

In recent years, data mining has been recognized by more and more people, many mining researchers, techniques and systems have been pulled into our life. The mining technology of association rules is the process of extracting the previous unknown, incomplete, vague, random, valid and actionable information from large database or data warehouse.

In the process of the open experiment teaching reform, leaved the booking experimental database of each semester. What should we consider is that how to find out some important and useful information from the lager booking experiment database, which can use for the reasonable investment of experiment equipment and reasonable arrangement of experiment teachers.

From the characteristics and application fields of the algorithm of Association Rules, we can use the Apriori Algorithm of Association Rules to the booking experiment database, and dig out some useful, disciplined, available information which can provide scientific reference rule used for the set of the booking system. It is feasible from the theory and the technical.

2. ASSOCIATION RULES AND APRIORI ALGORITHM

2.1. Association Rule

The concept of the Association Rules was first proposed by Agrawal in 1993. The follow is the formal statement of the problem:

Set $I = \{i_1, i_2, \dots, i_m\}$ was a set of items, the elements i_j ($j=1,2,3,\dots,m$) called item.

Set transaction T is a subset of I , where $T \subseteq I$.

⁺ Li Mengshan
Email: limengshan123@163.com

Set D is a subset of T. If $X \subseteq T$, then transaction T contains X.

An Association Rule is an expression of $X \Rightarrow Y$ (X, Y is a collection or a regulation generally, $X \subset I, Y \subset I$ and $X \cap Y = \emptyset$). Its meaning is that the regulation or collection contain X tends to contain Y.

The formed conditions of the Association Rule $X \Rightarrow Y$ are:

(1) It contains a support S, the database D contains the least of S% records, which content the implication of $X \cup Y$, Support S is the probability of $X \cup Y$, that is $P(X \cup Y)$.

(2) It contains a confidence C, The probability of the database contains record X, which the least of C% record contains Y, Confidence C is the conditional probability of $Y|X$, that is $P(X|Y)$.

Support $(X \Rightarrow Y) = P(X \cup Y) = |T: X \cup Y \subseteq T, T \in D| / |D|$.

Confidence $(X \Rightarrow Y) = P(X|Y) = |T: X \cup Y \subseteq T, T \in D| / |T: X \subseteq T, T \in D|$.

The problem is to generate all Association Rules using user's specified minimum Support and Confidence.

2.2. The basic idea of Apriori Algorithm

Apriori Algorithm (AA) is invented as a way to compute rules from large databases. Generally speaking, AA uses an iterative method to search the rules layer by layer. That is using the k layer generates the k+1 layer. First, It must find out the 1-item-set which named 1 L, and then use 1 L to generate the set of 2-item-set which named 2 L, repeat this operations until it can't seek out the item-set when the set is NULL.

(1) Connection: L_{k-1} layer connect itself, generate a new set of candidate item-set which we named it C_k , which contains all of the frequent item-set L_{k-1} .

(2) Pruning: C_k which generate by step one is the super set of L_k . Actually speaking, it is scanty that all the subsets of C_k which fill the min-support. As a result, we must determine how many subsets of C_k are belongs to C_k by scanning the transaction database D.

According the repeated process, it should generate all Association Rules using user's specified minimum Support and Confidence.

3. Implementation of Apriori Algorithm

The open experiment is base on the open laboratory, which covered all kinds of laboratories; there are great differences between the experiment subject and experiment form. Now, give an example of the open physics experiment to discuss the application of Apriori Algorithm.

University physics laboratory open for the experiment content, experiment time and the place of experiment. Students according to their own free time, hobbies and interests book the physics experiments on line, and then go to the booked place to finish the experiment which they had booked on time after the success of booking. Physics experiment contains three types which are mechanical, optical and electrical experiment; the open object is all the engineering students. Physics experiment must complete in two semesters which are the second and the third semesters. Table 1 shows the experiment code we have summarized.

TABLE 1 EXPERIMENT CODE

code	experiment type
L_1	mechanical experiment in the second semester
L_2	mechanical experiment in the third semester
G_1	optical experiment in the second semester
G_2	optical experiment in the third semester
D_1	electrical experiment in the second semester
D_2	electrical experiment in the third semester
L_a	the measurement of sound velocity
D_c	the measurement of Planck's constant
X_3	the measurement of low resistance
G_1	the use of Oscilloscope
G_c	the use of Michelson interferometers

We have summarized the database of the mechanical, optical and electrical experiment, and converted the information in the booking database to the form which can be mined by computer. Then transform the booking data into the code which can be handled easily.

According to the experiment code, we scan the student booking database, then get the information which the student booking.

TABLE 2. INFORMATION OF EXPERIMENT DATA

Stuid	L ₁	L ₂	G ₁	G ₂	D ₁	D ₂	La	X ₃	D _c	G ₁	G _c
X01	2	3	2	3	2	3	T	T	T	F	F
X02	2	3	2	2	3	3	T	F	T	T	F
X03	3	3	3	1	3	2	F	T	F	T	T
X04	3	2	3	2	3	2	T	T	T	F	F
X05	2	3	2	2	3	3	T	T	T	F	F
X06	3	3	3	3	1	2	T	F	F	T	F
X07	3	3	2	3	3	1	T	T	T	T	F
X08	2	3	2	3	2	3	F	T	T	T	T
X09	3	3	1	2	3	3	F	T	T	T	F
X10	3	3	2	3	2	2	T	T	T	T	F
...
S01	3	3	2	3	2	2	T	T	T	T	F
S02	1	3	3	3	2	3	T	T	F	T	F
S03	2	2	3	3	3	2	T	T	T	F	F
S04	3	2	3	2	2	3	F	T	T	T	F
S05	3	1	3	3	3	2	T	F	T	F	F
S06	2	2	3	3	2	3	T	T	F	T	T
S07	3	3	1	3	2	3	F	T	T	T	F
S08	2	3	3	3	2	2	T	T	T	T	F
...
Sx1	2	3	1	3	3	3	T	T	T	T	F
Sx2	3	2	3	1	3	3	T	F	F	T	T
Sx3	2	3	2	3	3	3	T	T	T	F	F
Sx4	3	3	3	2	2	2	F	T	T	T	F
Sx5	1	3	3	2	3	3	T	T	F	T	F
Sx6	3	1	2	3	3	3	T	F	T	F	T
Sx7	1	3	3	3	2	3	T	T	T	T	F
...

In the table above, stuid is the student number which could only represents the student in school; the integer is the number of experiments that the student booked, the integer responses the true thing of student booking; the F and T in the table means someone booking the experiment or not, T is true, and F is false.

3.1. The association of the experiments in the same type

Using the Apriori Algorithm to mine in the large booking database of physics experiment, due to the higher degree of reliability in the same type of experiment, the generalized database is mined with the minimum support 85% and minimum confidence 80%. Therefore we get the association rules table when the mining process finished.

TABLE 3 ASSOCIATION RULES IN THE SAME TYPE

Association Rules	Confidence (%)	T or F
L ₁ →L ₂	98. 9	True
G ₁ →G ₂	99. 6	True
D ₁ →D ₂	99. 4	True
L ₁ →G ₁	49. 5	False
L ₁ →G ₂	51. 1	False
L ₁ →D ₁	49. 6	False
L ₁ →D ₂	50. 3	False
G ₁ →D ₁	50. 2	False
G ₁ →D ₂	49. 9	False
G ₂ →D ₁	51. 2	False
G ₂ →D ₂	48. 9	False

3.2. The association of the independent experiments

Scanning the database of student booked again, of course, we also use the Apriori Algorithm and get some interesting information in the experiments which are independent each other.

With the minimum support 75% and minimum confidence 60%. The table 4 shows the Association Rules of the independent experiment.

TABLE 4 RULES IN THE INDEPENDENT EXPERIMENT

Association Rules	Confidence (%)	T or F
$L_a \rightarrow D_c$	82.3	True
$L_a \rightarrow X_3$	78.6	True
$L_a \rightarrow G_1$	75.4	True
$D_c \rightarrow X_3$	80.5	True
$D_c \rightarrow L_3$	79.2	True
$D_c \rightarrow G_1$	75.9	True
$G_1 \rightarrow X_3$	79.9	True
$L_a \rightarrow G_c$	32.8	False
$D_c \rightarrow G_c$	39.9	False

4. Analysis of the Association Results

Trough the above tables of Association Rules, they show that:

1、 In the appointment of experiments in two semesters, students booking the same type of experiments are around 100%.

From the Association Rules we know that students who booked some type of experiments in the second semester book the same type of experiment in the third semester. For example, if the student who book the mechanical experiments in the second semester, he must book the mechanical experiments in the third semester also. The possibility is almost up to 98%.

Through this regulation we can see that student's book experiment according to their own interest. Students who like the mechanical experiments would book all the mechanical experiments, the same as the optical and electrical experiment.

2、 The probability booked in different type of physical experiment is around 50%, without obvious rules.

3、 Consistency among the independent experiments.

From the Association Rules table of the independent experiment, it shows that there are also some associations among the experiments which are independent each other. For example, the measurement of sound velocity, the measurement of Planck's constant, the measurement of low resistance, the use of Oscilloscope, the use of Michelson interferometers are experiments independent each other, the interesting information we get from the Association Rules when we mine the booking database is that the probability that students who booking some of the experiments independent each other while booking others experiments is up to 75%.

Through this phenomenon, it reflects that the students maybe favour one or some courses and neglect anther or others, they may be book the experiments they like and neglect the experiments they dislike, which are also have an important role in the course.

5. Conclusions and Future Works

Firstly, this paper introduces the current situation and background of open experiment, introduces the Association Rules and the Apriori Algorithm, then analyzes the feasibility of the Apriori Algorithm, the key content is that using the Apriori Algorithm to get information by mining the booking database, especially analyzing the data which base on the same type of experiment and the independent and similar experiment, we have got the practical value and information about association conclusion finally.

The data mining technology provides strong theoretical basis for the open experiment reform, meanwhile, the data of the open database provides reliable information to improve the management of laboratory and accelerate the pace of experiment reform.

Due to lack of time, it has many facts needed to be improved. For example, it is necessary to take a deeper information mining base on the existing basis; And it can also make some practical improvement base on the Apriori Algorithm. Because of the limited of knowledge and ability, the design thought and theory need to constantly update also. I hoped to be able to improve these in the future.

6. References

- [1] R. Agrawal, T. Imielinski and A. Swami, Mining association rules between sets of items in large databases, In proceedings of the international Conference on Management of Data, ACM SIGMOD, pp. 207–216, Washington, DC, May 1993.
- [2] R.Srikant, R.Agrawal. “Mining Generalized Association Rules”, proceedings of 21st VLDB conference Zurich, Switzerland 1995.
- [3] Kryszkiewicz, M.: Strong rules in large databases. In: Proceedings of 6th European Congress on Intelligent Techniques and Soft Computing EUFIT 1998, vol. 1, pp. 85 – 89 (1998)
- [4] William Gropp, Ewing Lusk, users Guide for MPICH a portable implementation of MPI Technical reports ANL-96/6, Argonne National Laboratory 1996.
- [5] Dean, J., Grzymala-Busse, J.: An overview of the learning from examples module LEM1. Technical Report TR-88-2, Department of Computer Science, University of Kansas (1988)
- [6] QI FAN, CHANG-JIE ZHU, An Application of Apriori Algorithm in SEER Breast Cancer Data, 2010 International Conference on Artificial Intelligence and Computational Intelligence, 2010
- [7] Parvinder S. Sandhu, An improvement in apriori algorithm using profit and quantity, Second International Conference on Computer and Network Technology, 2010
- [8] Chao-Tung Yang, Wen-Chung Shih, and Shian-Shyong Tseng, “A Heuristic Data Distribution Scheme for Data Mining Applications on Grid Environments”, Fuzzy Systems, 2008. FUZZ-IEEE 2008. (IEEE World Congress on Computational Intelligence). IEEE International Conference on 1-6 June 2008, pp.2398-2404
- [9] Agrawal, R., Imielinski, T., Swami, A.N.: Mining association rules between sets of items in large databases. In: Buneman, P., Jajodia, S. (eds.) Proceedings of the 1993 ACM SIGMOD International Conference on Management of Data, Washington, D.C., May 26-28, pp. 207 – 216. ACM Press, New York (1993)
- [10] Ya-Han Hu, Yen-Liang Chen .Mining association rules with multiple minimum supports: a new mining algorithm and a support tuning mechanism Decision Support Systems, 2006, pp.1 -24
- [11] Yanfei Zhou, Wanggen Wan, Mining Association Rules Based on an Improved Apriori Algorithm, 2010 IEEE, 2010