Research and application of association rules in practice teaching BBS

Zhenhong Lang.

Tianjin Electronic Information College, Tianjin, 300350, China

Keywords: association rules, practice teaching, Apriori algorithm

Abstract. Along with development and perfection of practice teaching BBS' platform, immediate interaction is achieved between teachers and students in the many coursers. Not only is all kinds of students' questions answered timely so as to improving quality of teaching, but also correlated relationship is mininged from the a large number of interactional information. The paper is merely citing practice teaching BBS' platform of computer software professional as example. An improved Apriori algorithm is proposed on the basis of overcoming shortcomings of traditional Apriori algorithm which it deals with enormous frequent itemsets. The association rules of correlative information are mininged, in order to guide future teaching reform.

Introduction

Being aimed at growing data information in the practice teaching BBS' platform and combining Data mining technology, if potential valuable information is excavated and found association relationship from them, which will be a useful attempt to apply data mining technology to the teaching field and mining results will be more effective in the future teaching. As a key technology of data mining, that association rules can be found. It can be found that the interesting relationship among the various data items by analyzing the information of the data set. Then the deep meaning of the relationship is analyzed, and the main contents of the hidden information are revealed. Among them, the Apriori algorithm is a classical algorithm for mining association rules. However, the traditional algorithm has the defects of repeated scanning transactional database, resulting in a large number of frequent itemsets, which leads to the large system overhead and low efficiency. So the information in computer software professional practice teaching BBS platform as the experimental data source in the paper, with discrete data preprocessing and pruning operation achieve improvement to the traditional Apriori mining algorithm to dig out the related information of value to the teaching.

The introduction of key technologies

First, Association Rules:

The core idea of data mining is that the possible relevancy and correlation is found out among a great deal of data itemsets.[1]While association rules are applied to practice teaching BBS, we can forecast that the courses students are interested in and the main purpose they use practice teaching BBS, by virtue of analysing information of student's questions, submitted homework and teacher's answer, in order to finding more valuable information and guiding improvability of practice teaching BBS platform.

If retrieval set is I={i1,i2,……im}, the transaction data set is made D. The itemset is expressed to T={i1,i2,……ik}, ij□I(j=1,2,3……,m) in the transaction. Using (ID,T) identifies specific transaction, where ID represents the transactional number. Three relevant association rules are defined as follows.

Definition 1: Correlation degree of association rules: It shows that the proportional relationship is between the transaction set which X and Y are simultaneously included in and the number of transaction sets. General Form of expression is "support（X⇒Y）". namely:

$$\text{support}（X{\Rightarrow}Y）=\frac{|\{T:X{\cup}Y{\subseteq}T,T{\in}D\}|}{|D|} \qquad （1）$$

Definition 2: Credibility degree of association rules: It shows that the proportional relationship is

between the transaction set which X and Y are simultaneously included in and the transaction sets which X is only included in. General Form of expression is "confidence（X⇒Y）". namely:

$$confidence（X⇒Y）= \frac{|\{T:X \cup Y \subseteq T, T \in D\}|}{|T:X \subseteq T, T \in D|} \quad （2）$$

Definition 3: If correlation degree of itemsets is more than minimal threshold "minsup", the itemsets are identified frequent itemsets.

Second, traditional Apriori algorithm:[2]

The core idea of the algorithm is based on the generated candidate set, then the database information is scanned and counted, and to judge whether a candidate item set is a frequent item set. The specific implementation process is: all transactions are scaned in the transactional database and the number of generated transactions is counted then candidate1-itemsets is generated. If the support degree of itemsets is less than the minimum support, it is regarded as a set of non frequent itemsets and will be removed. So, the residual information will produce the frequent1-itemsets according to filtering of the minimum support degree. Afterwards, the frequent1-itemsets are connected with their own, in order to the candidate2-itemsets is generated. And so on, frequent2-itemsets is confirmed. Repeated the above steps until the K- itemsets is generated. While (K+1)-itemsets is not satisfied with the demand of minimum support degree, the algorithm ends.

Third, Apriori algorithm basic idea with pruning function:

In the first place, the data information which is to be excavated is discretized in the preprocessing stage. According to the percentage of the total, the information which is less impact on the results of mining will be cut down, that will come into being a transactional database. What is more importance, minimum support degree data to be excavated——minsup is ascertained. While candidate1-itemsets C1 is set up, all transactions itemsets is scanned and frequent1-itemsets L1 is calculated and generated. Once again, the connection result of L1*L1 is regarded as candidate2-itemsets C2. At present, the branch which is less than minimum support degree will be cut by pruning. Next up, the branch which is greater than minimum support degree will be scaned instead of scanning the entire C2 to produce the frequent2- itemsets L2. At last but least, transaction itemsets are updated in the database. And so on, final frequent itemsets Lk is formed and association rules are excavated among various data items. Details of the pseudocode as shown below:

Input：transactional database D and minimum support degree's counting threshold——minsup;
Output：frequent itemsets L in the D.
L1 = {1-itemsets（D）}; // frequent itemsets including an item in the D
For(k=2;L_{k-1}≠φ;k++){C_k=apriori_gen(L_{k-1}, minsup);//generating candidate itemsets K
  For each transaction t∈D{C_t=subset(C_k,t);
   For each candidate item c∈C_t c.count++; }
  L_k={c∈C_k|c.count≥min_sup }} Return L=$U_k L_k$
Procedure apriori_gen(L_{k-1}, (k-1)-itemsets（D）)
 For each itemset l_1∈L_{k-1}
  For each itemset l_2∈L_{k-1}
   If(l_1[1]=l_2[1])＾(l_1[2]=l_2[2])＾...＾(l_1[k-2]=l_2[k-2])＾(l_1[k-1]<l_2[k-1]) then{c=l_1
l_2;//connected two itemsets in order to producing ulterior candidate itemsets
if has_infrequent_subset(c,L_{k-1})then{delete c;//deleted non frequent candidate itemsets
  determining whether the length of the item l_i is less than 1 or l_i is not in L $_{k-1}$
  updated database, deleted pruned information}
  else connected c andC_k, updated candidate itemsets of k} return C_k
 Procedure has_infrequent_subset(c, k-itemsets（D）,L_{k-1})
 For each（k-1）-itemsets s of c
  If s is not belong L_{k-1} then Return TRUE else Return FALSE

The simulation experiment of mining practice teaching BBS' association rules

Firstly, the information questioning by students will be discretized, which as shown in Table 1. The each comprehensive information is unceasingly discretized, which as shown in Table 2. Similarly, the information of submitted homework and teacher's answer is respectively discretized and all being dispersed information will be constituted transactional database.

Table 1 Discretization of students' question information

Type of question information	Total number of questions	Percent of total number
About database's course	510	19%
About C#'s course	927	34%
About HTML's course	465	17%
About ASP.NET's course	814	30%

Table 2 Discretization of comprehensive information

Type of question information	Refined subclass	Total number of questions	Percent of total number
		510	19%
About database's course	About Classroom teaching	180	7%
	About Submit assignments	230	8%
	About Course interest	100	4%
Type of question information	Refined subclass	Total number of questions	Percent of total number
		927	34%
About C#'s course	About Classroom teaching	465	17%
	About Submit assignments	388	14%
	About Course interest	74	3%
Type of question information	Refined subclass	Total number of questions	Percent of total number
		465	17%
About HTML's course	About Classroom teaching	187	7%
	About Submit assignments	96	4%
	About Course interest	182	6%
Type of question information	Refined subclass	Total number of questions	Percent of total number
		814	30%
About ASP.NET's course	About Classroom teaching	249	9%
	About Submit assignments	355	13%
	About Course interest	210	8%

Secondly, confidence level and support degree are respectively set 35% and 20% in the operation of data mining, as well as, transactional database is imported to data mining platform, which as shown in chart 1.
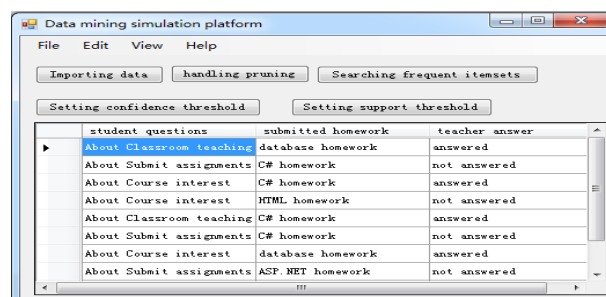

Fig.1 data interface of simulation platform

Third, frequent itemsets are orderly produced by pruning and connecting. After frequent itemsets-L3 is produced, the operation of data mining is over. The content of frequent itemsets-L3 is as shown in Table 3.

Table 3 The content of frequent itemsets-L3

Frequent itemsets	Support degree
student questions{About Submit assignment}—submitted homework{C# homework}—teacher answer{answered}	22%

Last, the association rule which obtained from frequent itemsets-L3 is that when students ask for about problem of homework using practice teaching BBS' platform, their aim is handing in homework. As well as, students are interesting in C#'s course in the many courses and after teachers explain questions, students finish favoringly and duly submit C#'s assignment.

Conclusion

As stated above, through improved Apriori algorithm is used to excavate computer professional data in the practice teaching BBS' platform, found the association rules among information, not only to obtain valuable information relevance, but also can better guide future teaching reform. In addition, the algorithm can be extended to the association rules mining of other course's information, so that it will help to improve the overall level of university talents training.

References:

[1]XU Hui-zeng. The Research of Association Rules Data Mining Algorithms[J.] Science Technology and Engineering, 2012,01:60-63.

[2]AGRAWALR,IMIELINSKIT,SWAMIA.Mining association rules between sets of items in large database[C]//Proceedings of the 1993 ACM SIGMOD International Conference on Managementof Da-ta.New York:ACM,1993:207-216.