ISSN 2229-5518

Identifying Weak Subjects using Association Rule Mining

Anuradha.Tadiparthi, Satya Prasad.R, Tirumala Rao S.N

Abstract— Many educational institutions in India today are concentrating on identifying the weak students and the subjects in which those students are weak in the current semester for improving their student results. They are even appointing a faculty member as a counselor to identify the weak students and to know in which courses the student is weak. After identifying this, this information will be given to the faculty who is teaching those courses so that he/she can take a special interest on those students or even conduct special classes to those students. In this paper we propose that the data mining technique called association rule mining can be applied to identify the subjects in which the students are weak in the current semester using previous semester's results.

Index Terms— Association Rule mining, Apriori algorithm, Confidence, Data mining, Strong association rules, Support, Weak subjects.

1 INTRODUCTION

Association rule mining, one of the most important and well researched techniques of data mining, is first introduced in [1]. It considers a set of items I ={I_1,I_2,…I_m} and a set of database transactions where each transaction T is a set of items such that $T \subseteq I$. Let A be a set of items. A transaction T is said to contain A if and only if $A \subseteq T$.An association rule is an implication of the form $A \Rightarrow B$, where $A \subset I$, $B \subset I$ and $A \cap B = \Phi$. The rule$A \Rightarrow B$ holds in the transaction set D with support S, where S is the percentage of transactions in D that contain $A \cup B$. This is taken to be the probability, P(AUB).The rule $A \Rightarrow B$ has confidence c in the transaction set D, where c is the percentage of transactions in D containing A that also contain B. This is taken to be the conditional probability ,P(B/ A). That is

$$\text{Support}(A \Rightarrow B) = P(A \cup B) \quad (1)$$

$$\text{Confidence}(A \Rightarrow B) = P(B/A) \quad (2)$$

In general, association rule mining can be viewed as a two-step process:

1. Find all frequent itemsets: By definition, each of these itemsets will occur at least as frequently as a predetermined minimum support count ,min_sup.
2. Generate strong association rules from the frequent itemsets: By definition, these rules must satisfy minimum confidence.

1.1 The Apriori Algorithm

There are many techniques for finding frequent itemsets. Apriori algorithm is a simple and popular algorithm for finding frequent itemsets. It is based on the apriori property that all nonempty subsets of a frequent itemset must also be frequent. It is a two step process.

1. The join step: To find L_k , the set of k-itemset that satisfy the minimum support count,a set of candidate k-itemsets is generated by joining L_{k-1} with itself. This set of candidates is denoted by C_k.
2. The prune step: C_k is a superset of L_k, that is, its members may or may not be frequent, but all of the frequent k-itemsets are included in C_k. A scan of the-database to determine the count of each candidate in C_k would result in the determination of L_k.

1.2 Generating Association Rules from frequent itemsets

Once the frequent itemsets are identified, we can generate the strong association rules from them.Strong association rules must satisfy both minimum support and minimum confidence.

$$\begin{aligned} confidence(A \Rightarrow B) &= \Pr obability(B/A) \\ &= Support\_count(A \cup B) / Support\_count(A) \end{aligned} \quad (3)$$

2 IDENTIFYING WEAK SUBJECTS

2.1 Finding the frequent weak course sets from the result database

Weak subjects are the courses in which the probability of failure is more for a student in the external exams. This can be identified based on internal test marks or if the student is feeling more difficulty in understanding that course. That is, it can be identified only after the course is started. But in this paper we are proposing a method to identify the weak course before that course is started.

In this paper, we have considered the result database of 300 students in 5 different courses as the transactional database. In these 5 courses , 3 courses belong to the previous semester and 2 courses belong to the current semester. Transaction_ID is

- *Anuradha Tadiparthi is a Research Scholar in Computer Science, Acharya Nagarjuna University, India, PH-+91866-2440298. E-mail: atadiparty@yahoo.co.in*
- *Dr. Satya Prasad R is Working as Associate Professor in Department of Computer Science, Acharya Nagarjuna University, India.*
- *Dr. Tirumala Rao S.N is Working as Professor in the Dept of Computer Science in Narasaraopet Engineering College, India.*

taken as student_ID and the courses in which either the student got F grade or E grade as the item set. For example if he got the grades. For example if he got the grades like this-EM211–B,EM212-C,EM213-E,EM221-F,EM222-D, then the itemset for that student consists of {EM213,EM221}. Then the apriori algorithm is applied on the result database for finding the frequent weak course set.

2.2 Identifying the weak courses of the current semester using association rule mining

Association rules are generated from the frequent weak course set .From this we have considered only the rules which consisted of courses from both the semesters. These rules can be applied to a new student to identify the weak courses of the current semester using previous results. For example if we have a rule like

$$(X,\text{EM211}) \Rightarrow (X, \text{EM221}) \text{ confidence } 85\%$$

That means when ever a student fails in EM211 course in the previous semester, 85% of the times he may also fail in the current semester course EM222 .

3 EXPERIMENTAL RESULTS

The following are the experimental results for the result database of 300 students in 5 –different subjects .Here we have considered min_sup count as 15.So all course sets with sup_count greater than or equal to 15 will be frequent course sets.

3.1 Frequent Course Sets

Result database D is scanned for finding count of each candidate .

C1

1-courseset	Sup.count
{ EM211 }	114
{ EM212 }	132
{ EM213 }	151
{ EM221 }	196
{ EM222 }	210

Compare candidate support count with minimum support count

L1

Frequent 1-courseset	Sup.count
{ EM211 }	114
{ EM212 }	132
{ EM212 }	151
{ EM221 }	196
{ EM222 }	210

The set L1 is used for finding C2 there by L2.

L2

Frequent 2-courseset	Sup.count
{ EM211 , EM212 }	33
{ EM211 , EM212 }	65
{ EM211 , EM221 }	58
{ EM211 , EM222 }	55
{ EM212 , EM212 }	48
{ EM212 , EM221 }	82
{ EM212 , EM222 }	74
{ EM212 , EM221 }	82
{ EM212 , EM222 }	83
{ EM221 , EM222 }	123

The set L2 is used for finding C3 there by L3.

L3

frequent 3-courseset	Sup.count
{ EM211 , EM212 , EM212 }	22
{ EM211 , EM212 , EM221 }	15
{ EM211 , EM212 , EM221 }	29
{ EM211 , EM212 , EM222 }	27
{ EM211 , EM221 , EM222 }	24
{ EM212 , EM212 , EM221 }	29
{ EM212 , EM212 , EM222 }	16
{ EM212 , EM221 , EM222 }	50
{ EM212 , EM221 , EM222 }	41

The set L3 is used for finding C4 there by L4

3.2 Association Rules

In C4, no courseset is having sup-count greater than min-sup count.So we have used set L3 for generating association rules. From the generated association rules , we have considered only the rules with courses from previous semester on the left side and courses from current semester on the right side.

{EM211,EM212} ⇒ {EM221} confidence=0.45

{EM211,EM213} ⇒ {EM221} confidence=0.45

{EM211,EM213} ⇒ {EM222} confidence=0.41

{EM211} ⇒ {EM221,EM222} confidence=0.21

{EM212,EM213} ⇒ {EM221} confidence=0.60

{EM212,EM213} ⇒ {EM222} confidence=0.33

{EM212} ⇒ {EM221,EM222} confidence=0.38

{EM213} ⇒ {EM221,EM222} confidence=0.272

We have considered min_confidence as 0.45. So all rules with confidence greater than or equal to 0.45 are considered to be strong.

3.3 Weak course-set

From the strong association rules we can identify the weak courses of a student in the current semester using his previous results. So from the generated strong rules above, we can iden-

tify that

1. When ever a student fails in courses EM211 and EM212 in the previous semester , he /she will in the course EM221 in the current semester.
2. When ever a student fails in courses EM211 and EM213 in the previous semester, he /she will in the course EM221 in the current semester.
3. When ever a student fails in courses EM212 and EM213 in the previous semester, he /she will in the course EM221 in the current semester.

We consider the probable failure courses as the weak courses for a student in the current semester.

4 CONCLUSIONS

Association rule mining initially developed for market basket analysis has more applications. In this paper we have used it for identifying the weak courses of a student in the current semester based on previous semester results.

REFERENCES

[1] Agrawal, R., Imielinski, T., and Swami, A. N. 1993. Mining association rules between sets of items in large databases. In Proceedings of the 1993 ACM SIGMOD International Conference on Management of Data, 207-216.

[2] Han, J. and Kamber, M. 2000. Data Mining Concepts and Techniques. Morgan Kanufmann page no:230-240

[3]. R. Agrawal and R. Srikant, "FastAlgorithms for Mining Association Rules," Proc. 20th Int'l Conf. Very Large
Data Bases (VLDB '94), pp. 487-499, 1994