

Frequent Pattern Mining and Customer Retention using Vertical Data Set and Boolean Matrix

S.Tejaswi^{#1}, Ch.Sita Kameswari^{#2}

^{#1}Student, M.Tech(CST), Baba Institute Of Technology and sciences , vizag,

^{#2}.Professor, Dept of CSE, Baba Institute Of Technology and sciences , vizag

Abstract: Association Rule mining is one of the most important functionality in data mining, by using which we can find out the frequent patterns. By the association rule mining either we can find out the frequent patterns of item sets or we can find out the customers who visits our shopping mall frequently. In this paper we propose a vertical data set which consists of two rows where we have the Customer-ID's and the items which they purchased. Here we use Boolean matrix technique to represent the data and by using it we can easily find out frequent items purchased by the customers, so that we can improve our sales by this frequent pattern mining as well as we can find out which customers purchased our frequent item sets, so we can give some exclusive offers to those customers, so that we can improve our sales.

Keywords: Boolean Matrix, Logical And, Transaction Reduction, Vertical Data Set.

INTRODUCTION:

Association Rule Mining is one of most important functionality in data mining. Association Rule Mining mainly used to find out the associations between the item sets. A rule which satisfies both support and confidence(interesting measures) is called as a strong association rule.

Association rule mining mainly used for Business Intelligence. We have various strategies to improve our business. They are..

Strategies:

- (i) To get profit in our business we need to purchase more items at a time. So we have to attract the customer to purchase more items by providing some exclusive offers.
- (ii) Customer retention is one of the most important strategy in business. To retain a customer we need to provide some exclusive offers to those customers who purchase more items(frequent itemsets) at a time and who visits shopping mall more frequently.

METHODOLOGY:

Here we propose a novel methodology to find out frequent patterns by using Boolean Matrix and vertical data set mining.

Here we consider the transactional dataset which contains two columns. In the first column we have the "Customer_ID" and in the second column we have the "List of items" which they have purchased.

Now, we convert the given transactional data set to the vertical dataset, here in rows we have the items and columns we have the "C_ID"s

Now the vertical data set which will be converted into Boolean matrix.

Now in the first iteration we have the single item-set in rows and in columns we have "C-ID". If the item purchased by the customer we put '1' else we put '0'.

After the boolean matrix is prepared for the first item set, in the second iterations we will find out the combinational pairs of every item with every other item in the vertical dataset, which makes a 2-item sets again we put '1' when the customer purchase this 2-item set else we put '0'.

Now we find out combinational triples for every 2-item sets by following the minimum support count. A 2-itemset which is frequent is considered as frequent subset. Now we form 3-item set which contains all these 2-item sub sets in that

We iteratively repeat this procedure until no frequent pattern is generated.

TRANSACTIONAL DATA SET:

CID	List_of_item_IDs
C100	Bread,Butter,Honey
C200	Butter,Donuts
C300	Butter,Jam
C400	Bread,Butter,Donuts
C500	Bread,Butter,Jam
C600	Butter,Jam
C700	Bread,Jam
C800	Bread,Butter,Jam,Honey
C900	Bread,Butter,Jam

I1:T-Bread

I2: Butter

I3:Jam

I4:Donuts

I5:Honey

Boolean Matrix :

Here Min_Sup_Count=2

	C100	C200	C300	C400	C500	C600	C700	C800	C900	S.C
Iteration-1										
I1	1	0	0	1	1	0	1	1	1	6
I2	1	1	1	1	1	1	0	1	1	8
I3	0	0	1	0	1	1	1	1	1	6
I4	0	1	0	1	0	0	0	0	0	2
I5	1	0	0	0	0	0	0	1	0	2
Iteration-2										
I1^I2	1	0	0	1	1	0	0	1	1	5
I1^I3	0	0	0	0	1	0	1	1	1	4
I1^I4	0	0	0	1	0	0	0	0	0	1
I1^I5	1	0	0	0	0	0	0	1	0	2
I2^I3	0	0	0	0	1	0	1	1	1	4
I2^I4	0	1	0	1	0	0	0	0	0	1
I2^I5	1	0	0	0	0	0	0	1	0	2
I3^I4	0	0	0	0	0	0	0	0	0	0
I3^I5	0	0	0	0	0	0	0	1	0	1
Iteration-3										
I1^I2^I3	0	0	0	0	1	0	0	1	1	3
I1^I2^I5	1	0	0	0	0	0	0	1	0	2

RESULTS:

Here the frequent patterns are (I1,I2,I3) and (I1,I2,I5).

The candidates who purchased frequent pattern (I1,I2,I3) are:- C500,C800 and C900

The candidates who purchased frequent pattern

(I1,I2,I5) are:- C100 and C800

So we can provide some exclusive offers to these customers

ALGORITHM:

STEP1: Convert the Transactional Data Set to Boolean Matrix by placing 1 in the transaction where the item is existent. Place 0 when the item is not exist in transaction.

STEP2: At the End of Iteration-1 count the sum of binary 1's(i.e.. support count of each item).ignore the transaction which does not support Min_Sup_Count and consider those transactions for next iteration which satisfies Min_Sup_Count.

STEP3:Find the combinations of 2-item sets for the items which satisfy Min_Sup_Count in 1-item sets and check whether these 2-items together exists in every candidate or not. If they exist put binary 1 else put binary 0. Count the sum of all binary 1-s which gives the support of 2-item sets and ignore the items which does not support 2-item sets.

STEP4: Find the combinations of 3-item sets for the items which satisfy Min_Sup_Count in 2-item sets and check whether these 3-items together exists in every candidate or

not.If they exist put binary 1 else put binary 0. Count the sum of all binary1-s which gives the support of 3-item sets and ignore the items which does not support 3-item sets. Suppose if we want to have a three item set I1^I2^I3 it should satisfy its subsets in 2nd iteration and they are I1^I2 , I2^I3 and I1^I3.

STEP5:We repeat iterations until no further frequent pattern is generated. Then the item sets in the last iteration considered as frequent item sets.

CONCLUSION:

This paper proposes a new methodology which is a refinement to apriori, here we find the frequent item-sets along with the candidates who purchased this frequent items. This will help us for customer retention. So we can provide some exclusive offers to those customers who purchase these frequent patterns and moreover the association rules generated by these frequent patterns considered as strong rules which helps us to improve our business.

REFERENCES:

[1].http://paginas.fe.up.pt/~ec/files_0506/slides/04_AssociationRules.pdf
 [2].<http://arxiv.org/ftp/arxiv/papers/1403/1403.3948.pdf>
 [3]. <http://www.csejournals.com/docs/IJCSR10-01-01-010.pdf>
 [4].<http://www.ijeat.org/attachments/File/v4i2/B3627124214.pdf>
 [5].<http://airccse.org/journal/ijdms/papers/4512ijdms05.pdf>