International Journal of Engineering Inventions
e-ISSN: 2278-7461, p-ISSN: 2319-6491
Volume 4, Issue 3 (August 2014) PP: 33-38

A Novel Quantity based Weighted Association Rule Mining

S. P. Syed Ibrahim[1], J. Shanthalakshmi Revathy[2]
[1]*Professor, School of Computuing Sciences and Engineering, VIT University, Chennai -127*
[2]*Assistant Professor, Department of CSE, Velammal College of Engineering and Technology, Madurai*

Abstract: *Classical association rule mining algorithm discovers frequent itemsets from transactional databases by considering the appearance of the itemset and not other utilities such as profit of an item or quantity in which items bought. But in transactional databases large quantity of items is purchased may lead to very high profit even though items appeared in few transactions. Therefore the quantity of the item is considered as the most important components, lack of which may lead to loss of information. Here binary attributes of the Item is considered for calculating item weight using link based model. This paper provides novel framework, Quantity Based Association Rule Mining (QBARM) algorithm, considers quantity and item weight.*

Keywords: *Association rule mining, Weight, Quantity.*

I. Introduction

Data mining technologies have enabled organizations to get an insight into the huge amount of data to gain competitive advantages. Association rule mining [1, 2] is a data mining technique. It provides important information in assessing significant correlations of items by considering an item is present or not in the database. Association rule mining algorithm produces rules having support and confidence values greater than minimum threshold. The classical association rule mining algorithm gives the statistical relationship between items; it does not consider the semantic significance of the items [3].

Table1: Transaction Table

TID / Attributes	A	B	C	D
1	1	5	0	0
2	1	7	0	0
3	2	0	10	0
4	1	5	0	0
5	2	0	0	2

Table 2: Weighted items table

Item	Profit/Unit Sold
A	0.8
B	0.5
C	0.2
D	0.6

In the example database given in Table 1, sample association rule is given by A → B (support: 60%) and A → D (support 20%). If the minimum support threshold is 25% then one important rule A→D is not obtained but purchase of product A leads to purchase of highly profitable product D, but classical association rule mining ignores above difference. To meet the objective of user and business value weight of the item was introduced for association mining, which based on the some weight that are preassigned to items. Here the weight is based on quantity of an item such as profit.

If we consider Item A weight as 0.8 and Item D weight as 0.6 then the association rule A → D (Support based on weight - 28%). But most of the weighted association rule mining does not consider the quantity as important component. If an itemset appears in a few transactions but quantity is larger, then it is possible that buying of this itemsets may lead to get more profit, may not be a frequent itemset based on minimum support defined by user. This leads to a loss of information. In the above dataset purchase of item A in two quantity leads to purchase of 10 quantity of item C, but classical nor weighted association rule mining do not consider the above scenario.

To overcome the above problem, Sulaiman Khan et. al., proposed association rule mining based on utility [5]. In their paper they assume preassigned weights based on profit margin. But most data items do not come with such preassigned weights [6]. This paper deals with the weighted association rule mining based on quantity and weight is calculated hits model.

II. Related Work

A. Association Rule Mining

The most significant tasks in data mining are discovering frequent itemsets and association rules. There are many efficient algorithms are available for mining frequent itemsets and association rules. The frequent itemset mining was first introduced by Agrawal et al. [1]. Measure of support and confidence serve as the basis for association rule mining.

B. Weighted Association Rule Mining

Wang et al., proposed an efficient algorithm for Weighted Association Rules (WAR) [4]. WAR generates association rule according to the weight of individual item, which leads to downward closure property invalidation. The problem is solved by using an efficient model of weighted support measurements and exploiting a weighted downward closure property. A new algorithm called WARM (Weighted Association Rule Mining) [5] is developed based on the efficient model of Weighted Association rule.

C. Weighted Association mining without preassigned weight

Web site click-stream like data sets does not come with preassigned weights, so Sun et al., [6], proposed an algorithm for mining association rule. This algorithm calculates w-support and w-confidence from hub weight of transaction by extending HITS model Kleinberg's [7]. The item sets are generated by w-support, here binary attributes of item is considered. W-support is a frequent item set may not be as important as it appears, because the weights of transactions are different. The weights are calculated from the internal structure of the database based on the assumption that good items present in transactions.

D. Weighted Utility Mining

The Weighted Utility ARM (WUARM) [8] considers the significance and frequency values of individual items as their weights and utilities. Weighted utility mining focuses on identifying the item sets with weight as utilities with the user specified weighted utility threshold. The differences between items have a strong impact on decision making in many application unlike the use of standard ARM does not consider the quantity of the item in the transaction. Association rules are generated by w-support using transactional utility weight.

III. Proposed Work

The proposed novel algorithm called QBARM (Quantity Based Association Rule Mining) is developed for mining association rules based on quantity of the item. Here the item weight is manipulated from Link based model [7]. QBARM is the extension algorithm of weighted association rule mining that it considers items weight as their significance in the dataset and also considers the frequency of occurrences of items in transactions. Thus Quantity based mining of association rule is concerned with the frequency of item sets and significance of item sets.

Using transactional quantity weight and item significance, Quantity based association rules are extracted.

IV. Problem Definition

Here the concepts and related details involved in weighted utility mining are described.

A. Weighted Quantity Mining:

Let $T=\{t_1,t_2,t_3\ldots..\ ,t_n\}$ be the transaction which contains input data set D. Each transaction T have set of items $I=\{i_1,i_2,i_3,i_4\ldots i_n\}$ were n is the number of items in the transaction. The weight of the item is calculated from the internal structure of transactions that is link based approach is used. A set of positive real number is generated from the transaction $W=\{w_1,w_2,w_3\ldots.w_n\}$ for each item in I. Each transaction t_i has some subset of item {I} and Weight {W} for each item. It is represented as a pair (i,w) is called weighted item where iϵ {I} and w ϵ {W}. For example weight for the j^{th} item in the i^{th} transaction is given by $t_i[w_j(i_j)]$ with q as quantity of the item in a transaction from a set {Q} and are represented with real numbers. Weighted utility mining considers three factors item(I), weight(W) and quantity(Q).

B. Item weight **IW**

Item Weight is a set of positive real numbers $w(i_j)$ is generated for each item i_j considering quantity of the item.

C. Item Quantity

Item Quantity of the item i_j in the transaction t_i is denoted as $t_i(i_j,q)$.For frequency of the item, item weight is considered from each transaction which is dependent on it.

D. Item Quantity Weight

Item Quantity Weight is the combination of weight w and quantity q of the item i_j in the transaction t_i is denoted by $t_i[(w_j(i_j).q]$.

F. Weighted Quantity Support wqs

Weighted quantity Support of an item set X→Y is the transaction quantity weight that contains both X and Y relative to the transactional quantity weight of all transactions.

V. Quantity Based Association Rule Mining Algorithm

The Quantity Base Association rule mining algorithm is given below,

Input DB: Database; msupp: minimum support; mcon: minimum confidence; Output AR: Association rules

```
1.  Initialize auth(k) to l for each item k
2.  for(l=0;l<n;l++) do begin
3.  auth'(k)=0 for each item k
4.  for all transactions tϵDB do begin
5.  $hub(t)=\sum_{k:k\epsilon t} auth(i)$ for each transaction t normalize hub
6.  auth'(k)+=hub(t) for each item kϵ t
7.  end
8.  auth(k)=auth'(k) for each item i, normalize auth
9.  end
10. for all transaction tϵDB do begin
11. tqw(transactional quantity weight)
    for each transaction.
12. $wqc=\sum tqw(t_i)$
13. end
14. $wqsupp=\sum tqw(t_i)/ wqc$
15. wqcon(x,y)=wqsupp(x,y)/wqsupp(x)
16. $L_1$={Large one time set}
17. for(i=2; $L_{i-1}\neq\Phi$; i++) do begin
18. $C_i$=apriori($L_{i-1}$)
19. for all transactions tϵDB do begin
20.         $C_t$=subset($C_i$,t)
21.         for all candidates cϵ$C_t$ do
22.     c.wqsupp++
23.     c.wqcon++
24. end
25. $L_i$={cϵ$C_i$ | c.wqsupp >= msupp &&
c.wqcon>=mcon}
26. end
27. $AR_s=U_i\ L_i$
```

The line 1-9 item weight using hits model. Authority of item and hub of transaction is initialized to 1. First authority of item is calculated by adding hub value of transaction. Hub of transaction is calculated by adding the authority of item.And both values are normalized.These steps to be continued till the normalization value become equal. Line
10-11 presents the wq-support and wq-confidence.Line 12-
16 large item set is generated from that candidate item set is calculated. Line 17-20 calculates the wqsupp and wqcon for candidate item sets. Line 25 determines whether the item is frequent (above minsupport and minconfidence) then it is added to the rule set. The algorithm passes to k times and rules that are frequent are appended into rule set.

VI. Simulated Example

An example is given to compare the association rules generated by standard ARM, WUARM and QBARM algorithms.

Table 3: Sample Data set with Quantity

TID	Items				
	1	2	3	4	5
1	0	0	18	0	1
2	0	6	0	1	1
3	2	0	1	0	1
4	1	0	0	1	1
5	0	0	4	0	2
6	1	1	0	0	0
7	0	10	0	1	1
8	3	0	25	3	1
9	1	1	0	0	0
10	0	6	2	0	2

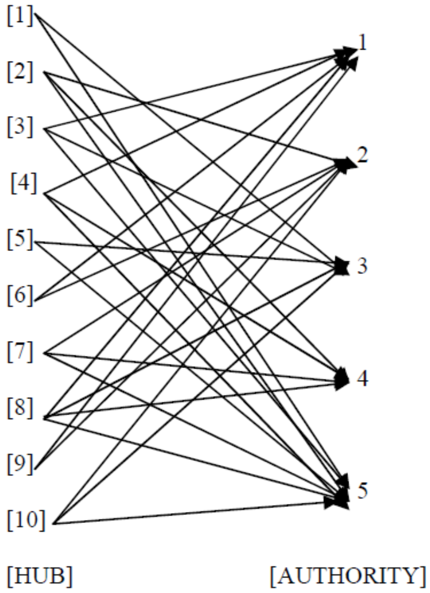

Fig 1 Bipartite Graph representation of sample data base. For calculating item weight binary attributes of the transaction is considered.

Table 4: Item Weight

Item	Authority Weight
1	0.36938354
2	0.33856377
3	0.4162621
4	0.361148
5	0.66725427

Above table shows the weight of each item generated using hits model. In case of WUARM algorithm weight of the item is preassigned depending upon the profit or many other criteria.

The following Table represents the frequent item set generated using QBARM algorithm. Table 5 shows the entire possible itemsets generated using standard ARM, WUARM and QBARM algorithm. The column 1 in the table 5 represents the itemset id, column 2 represents the standard ARM support, column 3 represents item set with WUARM support and column 4 represents the itemsets with QBARM. Support for standard ARM is calculated by considering only the occurrence of the item in the transaction. Support for WUARM is calculated from the item weight which is given externally and quantity of the item also considered. WUARM algorithm depends on the weight and quantity of the item, support and confidence of the frequent item set increases if the itemset weight increases and vice verse.

Table 5: Comparison between standard ARM, WUARM and QBARM,

Sl.No	**Standard ARM**	**WUARM**	**QBARM**
1	**1 0.5**	**1 0.25**	**1 0.34**
2	**2 0.5**	**2 0.70**	**2 0.31**
3	**3 0.5**	**3 0.43**	**3 0.74**
4	**4 0.4**	**4 0.57**	**4 0.43**
5	**5 0.8**	**5 0.90**	**5 0.95**
6	1,2 0.2	1,2 0.09	1,2 0.04
7	1,3 0.2	1,3 0.13	**1,3 0.26**
8	1,4 0.2	1,4 0.13	**1,4 0.25**
9	**1,5 0.3**	1,5 0.16	**1,5 0.30**
10	2,3 0.1	2,3 0.17	2,3 0.09
11	2,4 0.2	**2,4 0.43**	2,4 0.17
12	**2,5 0.3**	**2,5 0.60**	**2,5 0.26**
13	3,4 0.1	3,4 0.10	3,4 0.22
14	**3,5 0.5**	**3,5 0.43**	**3,5 0.74**
15	**4,5 0.4**	**4,5 0.57**	**4,5 0.43**
16	1,3,4 0.1	1,3,4 0.10	1,3,4 0.22
17	1,3,5 0.2	1,3,5 0.13	**1,3,5 0.26**
18	1,4,5 0.2	1,4,5 0.13	**1,4,5 0.25**
19	2,3,5 0.1	2,3,5 0.17	2,3,5 0.09
20	2,4,5 0.1	**2,4,5 0.43**	2,4,5 0.17
21	3,4,5 0.1	3,4,5 0.10	3,4,5 0.22
22	1,3,4,5 0.1	1,3,4,5 0.1	1,3,4,5 0.22

In Table 5 Consider the item set 1,3→5, the support value using standard association rule mining is 0.2, using WUARM algorithm the value is .13 whereas by QBARM algorithm the support value is .26. This is because the standard ARM generates the support value by considering the occurrence of the item in transaction.

In WUARM the support value depends on the weight of the item and quantity of that item in the transaction. So these two algorithms some time fail to generate important rule. In case of QBARM algorithm, occurrence and quantity of the item is considered and weight is calculated from that bipartite graph.

Suppose Support threshold is set to .25. Highlighted itemsets are frequent itemsets. This simulation illustrates the effect of quantity of the item's and its weight on the generated rules. Consider item set 1,4→5 whose sub sets {1→5,1→4 and 4→5}are also frequent. Thus the proposed model also satisfies downward closure property

VII. Conclusion

This is a novel framework for generating association rules. First, the Hits model algorithm is used to derive the weights from the transactions in a database by considering binary attributes of the item. Based on these weights of the item, a new measure wq-support and wq-confidence are defined to find the significance of itemsets. It differs from the traditional support by considering the quantity of the items in the transactions. The weight and quantity can be used together in mining rules and focus to the item sets with significant weight and high utility. The algorithm is developed by modifying WUARM with weighted quantity setting.

In future, the weight of the item in the transaction is calculated by using hybrid method. This may lead to efficient mining of rules from the data base**.**

REFERENCES

[1] R. Agarwal, T.Imielinski and A.Swami," Mining Association rules between sets of items in large databases", proc of the 1993 ACM SIGMOD International conference on Management of Data, Washington,DC,1993,pp.207.

[2] R. Agarwal and R.srikant," Fast algorithm for mining association rules in large data bases",Proceedings of the 20th international conference on very Large Data Base(VLDB'94), Santiago,chile,1994, pp 487-499.

[3] Hong Yao, H.J Hamilton," Mining item set utilities from transaction data bases", data and Knowledgs Engineering,pp.603-626,volime 59,issu 3 (2006).

[4] G. D. Ramkumar, S. Ranka, and S. Tsur, "Weighted Association Rules: Model and Algorithm," Proc. ACM SIGKDD, 1998.

[5] Tao, F., Murtagh, F., Farid, M.: "Weighted Association Rule Mining Using Weighted Support and Significance Framework." In: Proceedings of 9th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, pp. 661–666, Washington DC (2003).

[6] Ke Sun, Fengshan Bai,"Mining Weighted Association Rules without Preassigned Weights"Proceedings of IEEE Transactions on Knowledge and Data Engineering." VOL.20 No. 4 April 2008. pp. 489-495.

[7] J.M. Kleinberg, "Authoritative Sources in a Hyperlinked Environment," J. ACM, vol. 46, no. 5, pp.604-632, 1999.

[8] M.Sulaiman Khan, maybe Muyeba,Frans Coenen.: "A Weighted utility Framework for Mining Association Rules". In: Proceedings of second UKSIM European Symposium on Computer Modeling and Simulation. pp.87-92,2008.