ISSN: 2321-9653

INTERNATIONAL JOURNAL FOR RESEARCH IN APPLIED SCIENCE AND ENGINEERING TECHNOLOGY (IJRASET)

Weighted Association Rule Mining:A Survey

Shashi Chhikara[1], Purushottam Sharma[2]

Research Scholar(M.Tech)[1], Assistant Professor[2]
Department of IT, Amity School of Engineering & Technology
Amity University

Abstract- Association rule mining helps to extract large transaction databases for association rule. Without taking the weight of items into account, Classical Association Rule Mining (ARM) concludes that all items have the same significance. It also avoids the difference between the importance and transactions of all itemsets. In converse, WARM (Weighted Association Rule Mining) doesn't work on databases with only binary attributes, but also makes the use of importance of all transactions and itemset. It needs every item to be given weight to reflect their importance to the user. The weights may correlate to the benefit of different items. A number of weighted associative rule mining algorithms have been introduced in last few years such as WAR, WARM, WFIM, WIP, FWARM,WFP and many more. These algorithms engage different rule pruning, rule prediction, rule discovery, rule ranking methods. This paper targets on comparing and surveying the weighted associative rule mining techniques with regards to the above criteria.

Keywords— ARM,WAR, WARM, WFIM, WIP, FWARM,WFP

I. INTRODUCTION

Association Rule is an important type of knowledge representation, unfold the implicit relationships among the items present in a large number of transactions. Association rules are one of the most researched areas of information mining and have recently received much attention from the database community. They have turned out to be a rather useful in the marketing and retail communities as well as other more diverse fields. Association rule mining (ARM) is firstly proposed by R. Agrawal, T. Imielinski and A. Swam in 1993 [1]. Classical association rule mining employs support and confident measures which treat every transaction equally. On the contrary, different transactions have different weights in real-life datasets.

For example, in retail mining application, frequent item sets identified by the standard association rule mining algorithm may contribute only a small portion of the overall company profit because high profit and luxury items normally do not frequently appear in transactions and thus do not appear in rules with high support count values. Evolution of weighted association rule mining has solved this problem [2]. In the last few years, so many algorithms have been successfully proposed for mining association rules with weighted settings.

II. WEIGHTED ASSOCIATION RULE MINING

A weighted association rule (WAR) is an implication $X \rightarrow Y$ where X and Y are two weighted items. A pair (i_j, w_j) is called a weighted item where $i_j \in I$ and $w_j \in W$ is the weight associated with the item i_j. A transaction is a set of weighted items where $0 < w_j <= 1$. Weight is used to show the importance of the item. In weighted association rule mining problem each item is allowed to have a weight. The goal is to steer the mining process to those significant relationships involving items with significant weights rather than being flooded in the combinatorial explosion of insignificant relationships [3].

III. SOLUTION SCHEME

WARM problem can be divided into two steps:

1. Each item is given a weight, based on its significance. Weighted support is calaculated for each itemset. The weighted support is the fraction of the weight of the transactions that contains both A and B relative to the weight of all transactions. Find all frequent weighted

Vol. 2 Issue IV, April 2014

ISSN: 2321-9653

INTERNATIONAL JOURNAL FOR RESEARCH IN APPLIED SCIENCE AND ENGINEERING TECHNOLOGY (IJRASET)

itemsets (the average value of the weights of the items in the itemset) that have user specified minimum weighted support.

2. By exploiting large itemsets, association rules are generated that have above the user specified minimum weighted confidence.

Following are the algorithms introduced by different authors for mining association rules with weighted setting:

A. Association rule mining using weighted support and significance framework (WARM)

In the WARM context, an itemset is significant if its weighted support is above a pre-defined minimum weighted support threshold. As a matter of fact, the threshold values specified by the user are from the margin of significance of cost point of view. This method possibly will be more meaningful than only specifying relatively arbitrary support threshold. The weighted support of an itemset is defined as the product of the sum of the weight of its itemset and the weight of the fraction of transactions that the itemset occurs in.

Weighted support: Weighted support WSP of an itemset, A set of transactions T respects a rule R in the form A → B, where A and B are non-empty sub-itemsets of the item space I and they share no item in common[3]. Its weighted support is the fraction of the weight of the transactions that contains both A and B relative to the weight of all transactions. This may be formulated as:

$$\mathrm{wsp}(AB) = \frac{\sum_{k=1}^{|WST|\&A\cup B\subseteq tk} \mathrm{weight}(t_k)}{\sum_{k=1}^{|WST|} \mathrm{Weight}(t_k)}$$

The goal of the weighted association rule mining is then changed to determining all rules that are above a user specified minimum weighted support threshold holding a minimum user specified confidence. In order to calculate weighted support of an itemset, we requires a method to evaluate transaction weight. The transaction weight can be derived from weights of the items presented in the transaction. One may formulate it easily as the average weight of the items presented in the transaction.

$$\mathrm{Weight}(t_k) = \frac{\sum_{i=1}^{|WS_t(t_k)|} \mathrm{weight}(\mathrm{item}(i))}{|WS_t(t_k)|}$$

The itemset is then validated as important if its weighted support is above the pre-defined minimum weighted support.

B. Fuzzy Weighted Association Rule Mining (FWARM)

WAR uses a post-processing approach by deriving the maximum weighted rules from frequent itemsets. Post WAR doesn't interfere with the process of generating frequent itemsets but focuses on how weighted AR's can be generated by examining weighting factors of items included in generating frequent itemsets. Similar techniques were proposed for weighted fuzzy quantitative association rule mining[8,9]. In [7], a two-fold preprocessing approach is used where firstly, quantitative attributes are separated into different fuzzy linguistic intervals and weights assigned to each linguistic label. A mining algorithm is then applied on the resulting dataset by applying two support measures for normalized and un-normalized cases. The closure property is addressed by using the z-potential frequent subset for each candidate set. An arithmetic mean is employed to find the possibility of frequent k+1 itemset, which is not guaranteed to validate the downward closure property. Maybin and Sulaiman [10] proposed fuzzy weighted support and confidence framework (FWARM) algorithm belongs to the breadth first traversal family of ARM algorithms, developed using tree data structures **[14]** and works in a fashion similar to the Apriori algorithm, it was proposed to extract weighted boolean and quantitative data (by fuzzy means) to address the issue of invalidation of downward closure property and also it shows that using the proposed framework, rules can be generated proficiently with a valid downward closure property without biases made by pre- or post-processing approaches.

C. Efficient mining of weighted association rules (WAR)

Wei Wang et al. proposed an efficient mining methodology for Weighted Association Rules (WAR) [4]. The idea is a numerical attribute can be assigned for every item which in turn judges the weight of the item in a particular weight domain. WAR makes use of two-fold approach where the frequent itemsets are

ISSN: 2321-9653

INTERNATIONAL JOURNAL FOR RESEARCH IN APPLIED SCIENCE AND ENGINEERING TECHNOLOGY (IJRASET)

generated through standard association rule mining algorithms without considering the weight. Post-processing is then applied on the frequent itemsets during rule-generation to derive the maximum WARs. It focuses on how weighted association rules can be generated by examining the weighting factors of the items included in generating frequent itemsets. Therefore, we could classify this type of weighted association rule mining methods as a technique of post processing association rules.

D. Mining Weighted Interesting Patterns (WIP)

WIP (Weighted Interesting Pattern mining) based on mining weighted frequent patterns. Unil Yun and John J. Leggett defines the concept of a weighted hyperclique pattern that uses a new measure, called weight-confidence, to consider weight affinity and prevent the generation of patterns with substantially different weight levels [5]. The weight confidence is utilized to generate patterns with similar levels of weights and the h-confidence serves to identify strong support affinity patterns. The primary approach of WIP is to push weight confidence and/or hi-confidence in the weighted frequent pattern mining algorithm based on the pattern growth approach and prune uninteresting patterns. A level of weight and/or support is needed to reflect the overall weight and/or support affinity among items within the pattern. A new measure of weight confidence is seen in WIP. WIP also divides mining the FP-tree into mining smaller FP trees as in WFIM [5]. In WIP, an ascending weight order method and a bottom-up traversal strategy are used in mining weighted interesting patterns.

E. Mining Weighted Association Rules without Pre-assigned Weights

Item set evaluation by support in classical association rule mining is based on counting[1]. In this section, we will bring in a link-based measure called w-support and formulate association rule mining in terms of this new concept. In this paper, author bring in w-support, a new measure of item sets in databases with only binary attributes. The idea behind w-support is that a frequent item set may possibly not be as important as it appears, since the weights of transactions are different. These weights are entirely derived from the internal structure of the database based on the assumption that good transactions consist of good items. This assumption is exploited by extending Kleinberg‟s HITS model and algorithm to bipartite graphs[14]. Therefore, w-support is different from weighted support in weighted association rule mining (WARM)[6], where item weights are assigned. Moreover, a new measurement framework of association rules based on w-support is proposed.

The **w-support** of an item set X is defined as

$$\mathrm{Wsupp}(X) = \frac{\sum_{T:X\subset T \wedge \in D} \mathrm{hub}(T)}{\sum_{T:T\in D} \mathrm{hub}(T)}$$

where hub (T) is the hub weight of transaction T. An item set is significant if its w-support is larger than a user specified value. Experimental results show that w-support can be worked out without much overhead, and interesting patterns can be discovered through this new measurement. Compared with Apriori [15], the proposed mining algorithm requires an additional iterative procedure to compute the hub weights of all transactions. The database is scanned exactly once in each iteration. So, the convergence rate of the hub weights is critical to the performance. This method works at the cost of three or four additional database scans over the traditional techniques

F. Modified Weighted FP-Growth

In[11] the authors have introduced a new type of WARM (Weighted Association Rule Mining) algorithm. This algorithm takes into account the weights of items along with their support counts during discovering weighted association rules. The items of transactions are assigned weights to reflect their significance, in this concept. Weighted FP (WFP) works on the basis of making Frequent-Pattern tree in two ways. Firstly, it eliminates those items from the transaction database D whose weight is below the minimum weight threshold and secondly having more support count than the minimum support threshold. The experimental outcomes presents a series of six experiments in which Weighted Association Rule Mining (WARM) method defeats traditional Associated Rule Mining (ARM) method in all confidence levels.

The author examined the important issues of weighted association rule mining are – how good are the weights in comparison of support counts, and the problem of lowering the amount of association rules for achieving conciseness.

The experiments shows that WFP outperforms FP-Growth in almost all of the cases, but in some special case like the cases in

ISSN: 2321-9653

INTERNATIONAL JOURNAL FOR RESEARCH IN APPLIED SCIENCE AND ENGINEERING TECHNOLOGY (IJRASET)

which items having more support count values then their weights, simple FP-Growth can performs better.

IV. CONCLUSION

In this paper, we surveyed various weighted associative rule mining techniques.The limitations of traditional association rule mining model are identified by the authors,particularly its inability for treating units distinctly and proposed that weight may be integrated in the mining process to solve this problem.We come with our study with serval advantages and the problem formulation which can be implemented in future.

REFERENCES

[1] R.Agrawal,T.Imielinski,and A.Swami, "Mining Association Rules between Sets of Items in Large Databases", In: Proceedings of the 1993 ACM SIGMOD International Conference on Management of Data.New York: ACM Press, 1993, pp. 207-216.

[2] G.D.Ramkumar, Sanjay Ranka, and Shalom Tsur, "Weighted Association Rules: Model and Algorithm" KDD1998, 1998.

[3]. F. Tao, F. Murtagh, and M. Farid, "Weighted Association Rule Mining Using Weighted Support and Significance Framework," Proc. ACM SIGKDD "03, pp. 661-666, 2003.

[4] W. Wang, J. Yang and P. Yu "Efficient mining of weighted association rules (WAR)", Proc. of the ACM SIGKDD Conf. on Knowledge Discovery and Data Mining, 270-274, 2000.

[5] Unil Yun and John J. Leggett, "WIP: mining Weighted Interesting Patterns with a strong weight and/or support affinity" Texas A&M University College Station, Texas 77843, USA.

[6] C.H. Cai, A.W.C. Fu, C.H. Cheng, and W.W. Kwong, "Mining Association Rules with Weighted Items," Proc. IEEE Int"l Database Eng. and Applications Symp. (IDEAS "98), pp. 68-77, 1998.

[7] Gyenesei, A,"Mining Weighted Association Rules for Fuzzy Quantitative Items", Proceedings of PKDD Conference pp. 416--423 (2000).

[8] Wang, B-Y., Zhang, S-M.: A Mining Algorithm for Fuzzy Weighted Association Rules. In: IEEE Conference on Machine Learning and Cybernetics, 4, pp. 2495--2499 (2003).

[9] Shu, Y. J., Tsang, E., Yeung, Daming, S.: Mining Fuzzy Association Rules with Weighted Items, IEEE International Conference on Systems, Man, and Cybernetics, (2000).

[10] Maybin Muyeba, M. Sulaiman Khan, Frans Coenen,"Fuzzy Weighted Association Rule Mining with Weighted Support and Confidence Framework''.

[11] Ramesh Yadav, Snehal, and Aafaq Wali," Profit Maximizing Approach in Data Mining Using Modified Weighted Association Rule Mining Algorithm" IJCSITRE, Vol. 3, Issue 3, 2013.

[12] Arumalla Nagaraju, Yallamati Prakasarao, A.Veeraswamy," An Implementation of Mining Weighted Association Rules without Preassigned Weights", IJARCSSE,2012.

[13] C.H. Cai, A.W.C. Fu, C.H. Cheng, and W.W. Kwong, "Mining Association Rules with Weighted Items," Proc. IEEE Int'l Database Eng. and Applications Symp. (IDEAS '98), pp. 68-77, 1998.

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

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