A Decision Based Approach for Infrequent Weighted Itemset

Janani Sankari.M[1] Nirmala.R[2] MuthuLakshmi.P[3]

[1,2,3]Department of Information Technology

[1,2,3]Jeppiaar Engineering College, Chennai

Abstract— *Data mining process is used to analyze data from different perspectives and summarize it into useful information. It is used to extract information from datasets and transform it into an understandable structure for future use. Each data is to be represented by weights in order to extract the data quickly. An Item set consists of collection of data's that are given weights based on data features. Weights are assumed to the Item set by the frequent and infrequent use of the Item set. To find an infrequent weighted Item set is very complicate. In order to overcome the extraction of infrequent weighted Item set some algorithms and new methods should be used. This method will extract the infrequent weighted Item set by making some decision based approach.*

Keywords— Item set Mining, Data Sets, Data Mining

I. INTRODUCTION

Itemset mining is widely used for determining several data correlations. By means determining frequent itemsets the itemset mining is performed whose frequency is considered above the given threshold. These frequent itemsets can be used in several real life contexts.

E.g. Analysis of market basket and biological data, Processing of medical image. Each data item comes together with weight. Let us consider an example,

Tid	Usage readings of CPU
1	(w,0)(x,100)(y,58)(z,69)
2	(w,0)(x,42)(y,29)(z,69)
3	(w,43)(x,0)(y,43),(z,69)

It includes three transactions which compose of four distinct items. Here Tid 1 means that CPU x works at a large usage rate at fixed point of time (1).CPUs y and z have an intermediate usage rate, when CPU w will be temporary idle(weight 0).In upcoming years research community focuses on problem of infrequent mining of Itemset, whose frequency of occurrence is less than the given threshold. The infrequent itemsets is applicable for several real life contexts such as detecting fraud, assessment of risk from census data. This paper determines the infrequent weighted itemsets by means of weighted datasets transaction. Here we derive the weights from the weights which come together with items in every transaction by means of cost function.

II. RELATED WORKS

In earlier problem of itemset mining, items which belong to transactional data will be equally treated. Based on the interest of differentiating items, authors determine the association rule, which includes the weights. But introduction of weights happened during the rule generation step which is performed after earlier itemset mining process. The first trial of pushing. Item weights into mining of itemset process exploit the anti-monotonocity which drives the Apriority-based mining of itemset.

III. WEIGHTED TRANSACTION EQUIVALENCE:

The purpose of weighted transaction equivalence is to establish an association between a weighted transaction data set T, which is composed of transactions with arbitrarily weighted items and an equivalent data set TE where each transaction is merely composed of equally weighted items.

The proposed transformation is typically suitable for giving the original data set by means of the certain FP-tree index.

Table

A. *Weighting Function (Min)*

Tid	Weighted transaction(Equivalent)
1w	(w,0)(x,0)(y,0)(z,0)
1.x	(x,57)(y,57)(z,57)
1.y	(x,14)(z,14)
1.z	(x,29)

B. *Transac. Original*

(w,0)(x,100)(y,57)(z,71)

C. *Weighting function (max)*

Tid	Weighted transaction(Equivalent)
1.w	(w,100)(x,100)(y,100)(z,100)
1.x	(w,-19)(y,-19)(z,-19)
1.y	(w,-13)(y,-13)
1.z	(w,-57)

D. *Transac. Original*

(w,0)(x,100)(y,57)(z,71)

In the above table, the equivalent versions of the transaction with tid 1 which is obtained by using the minimum and the maximum weighting functions are given in the left-hand side , and for convenience purpose the original transaction and its equivalent versions are put side by side Every transaction in the equivalent data sets, includes only the equally weighted items. With the use of the minimum weighting function, the equivalence procedure first considers only the lowest weights among the weights which are occurring in the original transactions and it generates an equivalent transaction of equally weighted items.

If the maximum weighting function is adopted, the procedure here is analogous. Instead of lowest weight, the highest transaction weight is selected at each step.

If the item weights are reduced by the local maximum weight, it may yield negatively weighted equivalent transactions.

IV. THE MINER ALGORITHM FOR INFREQUENT WEIGHTED ITEMSET

This algorithm extracts all IWIs for a given weighted transactional data set and the maximum IWI support threshold.

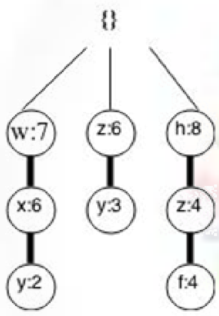
For either of the IWI-support max thresholds and IWI-support min thresholds, the IWI Miner mining steps are the same.

Instead of frequent weights, the IWI Miner discovers infrequent weighted itemsets. To perform this task, the modifications with respect to FP-growth have been performed:

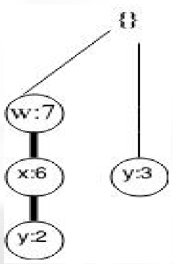
(i) The part of the search space to be pruned by pruning strategy and (ii) FPtree structure which is slightly modified, that allows storing the IWI-support value.

The IWI Miner adopts an FP-tree node pruning strategy to early discard items to reduce the complexity of mining process, which never belong to any itemset satisfying the IWI-support threshold. An item is said to be pruned if it appears only in the tree paths from the root to a leaf node which is characterized by IWI-support value greater than threshold.

A. *FP-Tree before Pruning*



B. *FP-Tree after Pruning*



In the above FP-tree before pruning diagram, the IWI-support max threshold value is given equal to 2.5.In this, the item *d* is given in the paths{d,c} and {h,d,f} whose leaf nodes will have IWI-support value equal to 3 and 4.

Here the item *d* can be pruned. The same consideration holds for *f* and *h*.The {a,b,c} is an IWI having IWI-support min equal to 2.

V. THE MINER ALGORITHM FOR MINIMAL INFREQUENT WEIGHTED ITEMSET:

This algorithm extracts all the MIWIs for a given a weighted transactional data set and a maximum IWI-support (IWI-support-min or IWI-support-max) threshold.

Both IWI Miner and MIWI Miner algorithms have same pseudo code. In the same way, IWI Mining procedure is same as MIWI Mining procedure. The MIWI Miner focuses on the generation of only minimal infrequent patterns.

So the recursive extraction in the MIWI Mining procedure is stopped as soon as infrequent itemset occurs.

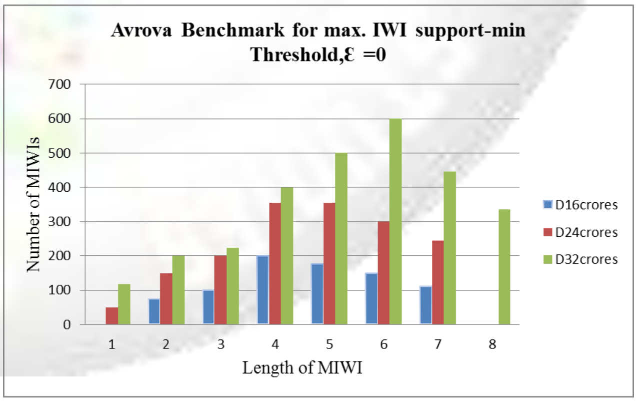
VI. EXPERIMENTING BY AVROVA BENCHMARK

Detecting the system malfunctioning, optimizing load balancing, resource sharing are some of the issues and we address these by analyzing and validating.

This is can be done by making use of *avrova benchmark*.

Data set	No. of original trans.	No. of equivalent trans.	No. of MIWIs with Ɛ=0	No. of IWIs with Ɛ=0
D16	106	823	135	55678
D17	106	823	156	65786
D18	105	824	189	23453
D19	110	825	179	21345
D20	110	867	180	1eo6
D21	110	893	186	2e06
D22	112	943	567	4e06
D23	112	947	432	16e06
D24	113	967	456	18e06
D25	113	1086	678	33e06
D26	113	1087	587	57e06
D27	114	1093	1523	67e06
D28	115	1096	1765	53e06
D29	116	916	446	1e09
D30	117	1051	867	2e09
D32	118	1052	456	4e09

In this table the number of IWIs and MIWIs has been mined from the avrova benchmark data sets by means of setting IWI support min to 0. The more detailed form is given by means of graphical representation by considering above benchmark values.



VII. PROPOSED SYSTEM

In this paper, only methodology and conditions are applied by form of algorithm to find the infrequent and frequent item sets. Infrequent items are mined by this technique. This system does not concentrate on advanced decision making for mining of infrequent items.

As a proposed work, an advanced decision making system is developed by the technique namely SVM. This technique usually have testing and training phase. This follows the machine learning approach. The datasets are inputted into the SVM and testing is done to mine the infrequent item sets.

VIII. Conclusion

This paper determines the infrequent itemsets by means of weights but not within each transaction. This we consider as an issue. We also accomplish two FP-Growth algorithms like IWI and MIWI efficiently. From the data in real-life context, the discovered patterns had been validated. Also validation has been performed with the help of domain expert.

Also this paper describes entirely with the two algorithms and also experiments which enhance the performance of the algorithms.

References

[1] K. Sun and F. Bai, "Mining Weighted Association Rules Without Reassigned Weights," IEEE Trans. Knowledge and Data Eng., vol. 20, no. 4, pp. 489-495, Apr. 2008.

[2] A. Manning and D. Haglin, "A New Algorithm for Finding Minimal Sample Uniques for Use in Statistical Disclosure Assessment," Proc. IEEE Fifth Int'l Conf. Data Mining (ICDM '05), pp. 290-297, 2005.

[3] R. Agrawal, T. Imielinski, and Swami, "Mining Association Rules Int'l Conf. Management of Data (SIGMOD '93), pp. 207-216, 1993

[4] G. Cong, A.K.H. Tung, X. Xu, F. Pan, and J. Yang, "Farmer: Finding Interesting Rule Groups in Microarray Datasets," Proc. ACM SIGMOD Int'l Conf. Management of Data (SIGMOD '04), 2004.

[5] W. Wang, J. Yang, and P.S. Yu, "Efficient Mining of Weighted between Sets of Items in Large Databases," Proc. ACM SIGMOD Association Rules (WAR)," Proc. Sixth ACM SIGKDD Int'l Conf. Knowledge Discovery and data Mining (KDD '00), pp. 270-274, 2000.

[6] F. Tao, F. Murtagh, and M. Farid, "Weighted Association Rule Mining Using Weighted Support and Significance Framework," Proc. nineth ACM SIGKDD Int'l Conf. Knowledge Discovery and Data Mining (KDD '03), pp. 661-666, 2003.