International Journal of Contemporary Research in Computer Science and Technology (IJCRCST) ISSN: 2395-5325
Volume 3, Special Issue 3 (September '2017)
Proceeding of - International Conference on Recent Trends in Computational Life Science and Information Technology,
Conference held at Tiruppur Kumaran College for Women,Tirupur, Tamilnadu,India.

MINING INFREQUENT WEIGHTED ITEMSET USING TRANSACTION MAPPING TECHNIQUE

S.Saranya,
M.Phil.,Research Scholar,
Department Of Computer Science,
Tiruppur Kumaran College for Women,
Tirupur, Tamilnadu,India.

P.Banumathi,
Assistant Professor,
Department of Computer Applications,
Tiruppur Kumaran College for Women,
Tirupur, Tamilnadu,India.

Abstract: Mining infrequent Itemset is fundamental method for mining association rules as well as for many other frequent Itemset mining tasks. In existing methods for Mining frequent and infrequent has been implemented using FP growth or Apriori algorithm. Also numerous experimental results have demonstrated that these techniques are scalable to the mining process. In this paper, we present a novel Transaction Mapping Technique for filtering the association rules for infrequent Itemset in the transaction dataset. Proposed technique produces improved performance for sparse data items. Furthermore, we present a open and closed Itemset extraction rules using optimization techniques. The experimental results prove that proposed technique highly scalable and consume less memory compared to the state of art techniques.

Keywords: Association Rules, Frequent Itemset, Data Mining, classification

I.INTRODUCTION

Mining Infrequent Item set is an exploratory data mining technique widely used for discovering valuable correlations among data which are weighted high but utilized in low numbers. In order to identify infrequent Itemset, Itemset mining was focused on discovering frequent item sets with valuable Correlations [1] i.e., patterns whose observed frequency of occurrence in the source data (the support) is above a given threshold [2]. However, many traditional approaches ignore the influence/interest of each item/transaction within the analyzed data. To allow treating items/transactions differently based on their relevance in the frequent itemset mining process[3], the notion of weighted itemset has also been introduced[4][5]. A weight is associated with each data item and characterizes its local significance within each transaction [6]. Reporting

an infrequent itemset which has an infrequent proper subset is redundant, since the former can be deduced from the latter. Hence, it is essential to report only the minimally infrequent itemsets. For instance, in [7], [8] algorithms for discovering minimal infrequent itemsets, i.e., infrequent itemsets that do not contain any infrequent subset, have been proposed. Infrequent itemset discovery is applicable to data coming from different real-life application contexts such as (i) statistical disclosure risk assessment from census data and (ii) fraud detection [9]. However, traditional infrequent item set mining algorithms still suffer from their inability to take local item interestingness into account during the mining phase. FP growth algorithm [10] is used to mine the frequent item .To address this issue, in [6] the analyzed transactional data set is represented as a bipartite hub authority graph and evaluated by means of a well-known indexing strategy, i.e., HITS [11], in order to automate item weight assignment. If the support threshold is too high, then less number of frequent itemsets will be generated resulting in loss of valuable association rules. On the other hand, when the support threshold is too low, a large number of frequent itemsets and consequently large number of association rules are generated, thereby making it difficult for the user to choose the important ones. Part of the problem lies in the fact that a single threshold is used for generating frequent itemsets irrespective of the length of the itemset.Weighted item support and confidence quality indexes are defined accordingly and used for driving the item set and rule mining phases. This paper differs from the above-mentioned approaches because it focuses on mining infrequent itemsets from weighted data instead of frequent ones. Hence, different pruning techniques are exploited. A related research issue is probabilistic frequent itemset mining [12], [13]. It entails mining frequent itemsets from uncertain data, in which item occurrences in each transaction are uncertain. The objective of discovering item sets whose frequency of occurrence in the analyzed data is less than or equal to a maximum threshold is considered as frequent Itemset. The Infrequent Itemset-support measure is defined as a weighted frequency of occurrence of an Itemset in the analyzed data. Occurrence weights are derived from the weights associated with items in each transaction by applying a given cost function. The Infrequent Itemset - support- min measure, which relies on a minimum cost function, i.e., the occurrence of an Itemset in a given transaction is weighted by the weight of its least interesting item. The Infrequent Itemset measure, which relies on a maximum cost function, i.e., the occurrence of an itemset in a given transaction is weighted by the weight of the most interesting item. Note that, when dealing with optimization problems, minimum and maximum are the most commonly used cost functions. Hence, they are deemed suitable for driving the selection of a worthwhile subset of infrequent weighted data correlations. The remainder of the paper is structured as follows. In section 2, we discuss related works of the frequent item set algorithms and technique. Section 3 describes the proposed based on infrequent weighted item set mining technique .Experimental evaluation of the proposed system is presented in the Section 4. Finally, section 5 concludes the work and points out future research directions.

II. RELATED WORKS

Frequent Item set mining

Frequent itemset mining is a widely used data mining technique for obtaining the association between items in the transaction dataset. Weighted association rules (WAR), which include weights denoting item significance. However, weights are introduced only during the rule generation step after performing the traditional frequent item set mining process to drive the association of items greater or lesser than threshold values .

Infrequent item Set Mining

Weighted item support and confidence quality indexes are defined accordingly and used for driving the itemset and rule mining phases. Probability of occurrence of an item within a transaction may be totally uncorrelated with its relative importance. A different item pruning strategy tailored to the traditional FP-tree structure to perform Infrequent Itemset mining efficiently. An attempt to exploit infrequent item sets in mining positive and negative association rules has been calculated through support count.

III. PROPOSED SYSTEM

3.1. Data Pre-processing of Transaction dataset

A data set is a collection of data, usually presented in tabular form. Each column represents a particular variables and a data set has several characteristics which define its structure and properties. These include the number and types of the attributes or variables. In these modules in this we are it is essential to understand how the data is represented in DB. We are going to collect a group of multidimensional data such as text, data. Which may have high dimensionality, usually in thousands or more. While this remains a challenging task, researchers have found that dimensionality reduction offers a middle ground, which generally results in faster computational time, while yielding reasonable accuracy.

3.2. Estimating the frequent weighted item set based on Correlations

Frequent Weighted items are Mined based on the Threshold conditions and correlation conditions .Cost function based on Support- Maximum is used to evaluate the frequent items in the transaction table. It is designed for pruning uninteresting infrequent and frequent itemsets discovered by FP model. One of the pruning measures used in FP model, *interest*, can be described as follows: to two disjoint itemsets A,B, if $interest(A,B)=|s(A\cup B)$ else $s(A)s(B)|<mi$, then $A\cup B$ is recognized as uninteresting itemset and is pruned, where $s(\cdot)$ is the support and mi a minimum interestingness threshold. This measure, however, is a bit difficult for users to set the value mibecause *interest* (A,B) highly depends on the values of $s(\cdot)$.

3.3. Estimating the Infrequent data correlations

Discovering item sets whose frequency of occurrence in the analyzed data is less than or equal to a maximum threshold is correlated into Itemset clusters,

3.4 Infrequent Weighted Item set using Transaction Mapping

FP-tree node pruning driven by the maximum -support constraint, i.e., early discarding of part of the search space thanks to a novel item pruning strategy.

- Cost function-independence, i.e., they work in the same way regardless of which constraint like minimum Support or Maximum support is applied.

3.5. Minimal Infrequent Weighted Item set Miner

Early stopping of the recursive FP-tree search in MIWI Miner to avoid extracting non- minima IWIs in the transaction item set through following strategies,

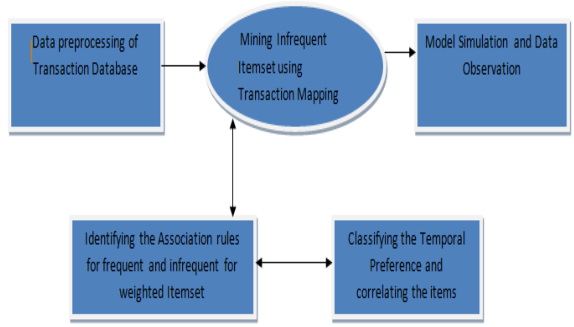

Figure 1: Architecture Diagram of the infrequent item set Mining using Transaction Mapping

Figure 1 explains the architecture of the infrequent item set mining process against the transaction dataset. It is an algorithm for frequent item set mining and association rule learning over transactional databases. It proceeds by identifying the frequent individual items in the database and extending them to larger and larger item sets as long as those item sets appear sufficiently often in the database.

The transaction tree is similar to FP-tree but there is no header table or node link. The transaction tree has compact representation of all the transactions in the database. Each node in tree has an id corresponding to an item and a counter that keeps the number of transactions that contain this item in this path. Here we can compress transaction for each Itemset to continuous intervals by mapping transaction ids into a different space to a transaction tree. Advantage of this algorithm is the performance can be improved compared to FP-Growth, FP-Growth* algorithms.

Algorithm:

Input: -Databse DB

Output: -all infrequent item sets

Scan the database and identify the infrequent item sets.

Construct the transaction tree with the count for each node.

Construct the transaction interval lists.

Else

Construct the lexicographic tree in a depth first order keeping only the minimum amount of information necessary to complete the search.

IV. EXPERIMENTAL RESULTS

In this section, we evaluated FP tree pruning and Transaction Mapping performance by means of a large set of experiments addressing the Following metrics,

- Support threshold
- Confidence

Support threshold

International Journal of Contemporary Research in Computer Science and Technology (IJCRCST) ISSN: 2395-5325
Volume 3, Special Issue 3 (September '2017)
Proceeding of - International Conference on Recent Trends in Computational Life Science and Information Technology,
Conference held at Tiruppur Kumaran College for Women,Tirupur, Tamilnadu,India.

The rule holds with support *sup* in *T* (the transaction data set) if sup% of transactions contain $X \cup Y$.

$$sup = \Pr(X \cup Y).$$

Confidence

The rule holds in *T* with confidence *conf* if *conf*% of transactions that contain *X* also contain *Y*. $Conf = \Pr(Y \mid X)$

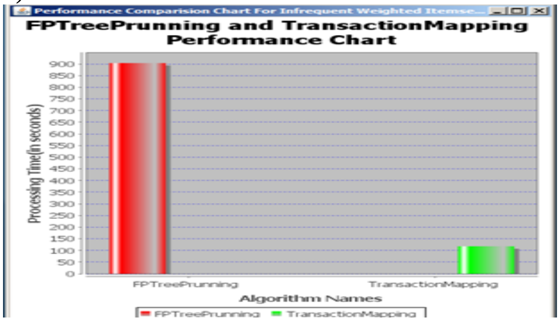

Figure 4.1 explains the performance of the system against the processing Time

Infrequent weighted mining, As it is observed in the figure 4.1. , which exhibits the performance of the transaction mapping and Fp tree pruning algorithm against the association generation

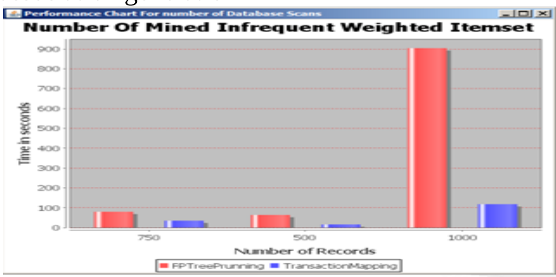

Figure 4.2. Performance analysis of the Infrequent weighted Item set against the processing Time

In figure 4.2., a weighted item set of infrequent items is obtained association rule as a pattern that states when *X* occurs, *Y* occurs with certain probability. Support count: The support count of an itemset *X*, denoted by *X.count*, in a data set *T* is the number of transactions in *T* that contain *X*. Assume *T* has *n* transactions.

$$confidence = \frac{(X \cup Y).count}{X.count}$$

$$support = \frac{(X \cup Y).count}{n}$$

As the support thresholds are reduced, the pruning condition becomes activated and leads to reduction in search space. Above the neutral point, the pruning condition is not effective.

V.CONCLUSION

We have designed and implemented infrequent mining method which has carried out using association rules and transaction Mapping algorithm by discovering infrequent item sets by using weights for differentiating between relevant items and not within each transaction. Existing algorithms evaluated on dense of the sparse dataset. The usefulness of the discovered patterns has been validated on data with multi level minimum support along the domain expert. As a future work, By Integrating the decision making algorithm for Association rules to mine the infrequent weighted dataset that supports domain expert's targeted actions based on the characteristics of the discovered Infrequent Itemset set with Missing values. Furthermore, the application of different aggregation functions besides minimum and maximum will be can analysed and implemented based n dynamic strategies.

VI.REFERENCE

[1]. RLuca Cagliero and Paolo Garza "Infrequent Weighted Itemset Mining using Frequent Pattern Growth", IEEE Transactions on Knowledge and Data Engineering, pp. 1- 14, 2013.

[2]. M.L. Antonie, O.R. Zaiane, and A. Coman, "A pplication of Data Mining Techniques for Medical Image Classification," Proc. Second Intl. Workshop Multimedia Data Mining in Conjunction with seventh ACM SIGKDD (MDM/KDD '01), 2001.

[3]. 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.

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

[5]. 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.

[6]. Ashish Gupta, Akshay Mittal and Arnab Bhattacharya "Minimally Infrequent Itemset Mining using Pattern-Growth Paradigm and Residual Trees", 17th International Conference on Management of Data (COMAD), 2011.

[7]. 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.

[8]. Diti Gupta and Abhishek Singh Chauhan "Mining Association Rules from Infrequent Itemsets: A Survey", International Journal of Innovative Research in Science, Engineering and Technology, ISSN: 2319-8753, Vol. 2, Issue 10, pp. 5801 – 5808, 2013

[9]. D.J. Haglin and A.M. Manning, "On Minimal Infrequent Itemset Mining," Proc. Int'l Conf. Data Mining (DMIN '07), pp. 141-147, 2007.

[10]. J. Han, J. Pei, and Y. Yin, "Mining Frequent Patterns without Candidate Generation," Proc.

ACM SIGMOD Int'l Conf. Management of Data, pp. 1-12, 2000.

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

[12]. C.-K. Chui, B. Kao, and E. Hung, "Mining Frequent Itemsets from Uncertain Data," Proc. 11th Pacific-Asia Conf. Advances in Knowledge Discovery and Data Mining (PAKDD '07), pp. 47-58, 2007.

[13]. T. Bernecker, H.-P. Kriegel, M. Renz, F. Verhein, and A. Zuefle, "Probabilistic Frequent Itemset Mining in Uncertain Databases," Proc. 15th ACM SIGKDD Int'l Conf. Knowledge Discovery and Data Mining (KDD '09), pp. 119-128, 2009.

[14]. C.K.-S. Leung, C.L. Carmichael, and B. Hao, "Efficient Mining of Frequent Patterns from Uncertain Data," Proc. Seventh IEEE Int'l Conf. Data Mining Workshops (ICDMW '07), pp. 489-494, 2007.

[15]. Asha Rajkumar and G.Sophia Reena "Frequent Item set Mining Using Global Profit Weight Algorithm", International Journal on Computer Science and Engineering, Vol. 02, No. 08, pp. 2519-2525, 2010..