A Novel Algorithm for Temporal Infrequent Weighted Item Set using Frequent Pattern Growth

S. Priyanka\textsuperscript{1} R. Rajasekaran\textsuperscript{2}
\textsuperscript{1}PG Scholar\textsuperscript{2}Assistant Professor
\textsuperscript{1,2}Sri Subramanya College of Engineering and Technology

Abstract— The Infrequent Weighted Itemet system focuses on the issue of discovering infrequent itemsets by using weights for differentiating between relevant items and not within each transaction. To reduce the complexity of the mining process in high dimensional databases, the temporal infrequent weighted itemset TIWI Miner adopts an FP-tree node pruning strategy to early discard items (nodes) that could never belong to any itemset satisfying the TIWI-support threshold. The SMA, a split and merge algorithm for infrequent item set mining, which can easily be extended to allow for “fast data” mining in the sense that dynamic data. Other distinguishing qualities of the method are its exceptionally simple processing scheme and data structure, it very easy to implement, convenient to execute on dynamic and external storage. The algorithm integrates a novel strategy named EUCI (Estimated Utility Co-occurrence Identification) to reduce the number of joins operations when mining low-utility item sets using the SM (split and Merge) data structure.

Keywords: Clustering, classification, and association rules, data mining

I. INTRODUCTION

Frequent itemsets mining is a core component of data mining and variations of association analysis, like association-rule mining and sequential-pattern mining. In frequent itemsets are produced from very big or huge data sets by applying some rules or association rule mining algorithms like Partition method, Apriori technique, Incremental, Border algorithm Pincer-Search, and numerous other techniques that take larger computing time to compute all the frequent itemsets. Extraction of frequent itemsets is a core step in many association analysis techniques. An itemset is known as frequent if it presents in a large-enough portion of the dataset. This frequent occurrence of item is expressed in terms of the support count. Therefore, it needs complicated techniques for hiding or reforming users’ private information during a data gathering process. Moreover, these techniques should not surrender the correctness of mining results. For example some common words or information that repeated frequently in a data set can be treated as frequent itemset for that data set. For example, buying a digital camera followed by Akash tablet and then a memory card, if it occurs regularly in a shopping database. It is known as (frequent) sequential pattern. Similarly substructure is referring to dissimilar structural forms, like sub-trees, sub-graphs or sub-lattices.

Infrequent patterns in Data Mining can be used in many applications

In text mining, indirect associations can be used to find synonyms, antonym or words that are used in different contexts. For example, the word data might be indirectly associated with the word gold, using the mediator mining. In the market basket domain, indirect associations can be used to find competing items, such as desktop computers and laptops, which states that people whom buys desktop computers won’t buy laptops.

Infrequent patterns can be used to detect errors. For example, if [Fire = Yes] is frequent, but [Fire = Yes, Alarm = On] is infrequent, then the alarm system probably is faulting.

Frequent weighted itemsets represent correlations frequently holding in data in which items may weight differently. However, in some contexts, e.g., when the need is to minimize a certain cost function, discovering rare data correlations is more interesting than mining frequent ones. This paper tackles the issue of discovering rare and weighted itemsets, i.e., the infrequent weighted itemset (IWI) mining problem. Two novel quality measures are proposed to drive the IWI mining process. Furthermore, two algorithms that perform IWI and Minimal IWI mining efficiently, driven by the proposed measures, are presented. Experimental results show efficiency and effectiveness of the proposed approach.

II. RELATED WORK

A. Uniform Distribution of items

R. Agarwal introduces Frequent itemset mining which is widely used data mining technique. Here, the rules are framed based on the itemset mined which is said to be frequent. Those itemset satisfying minimum support and confidence are taken as frequent and is used for framing association rules. Most approaches to association rule mining assume that all items within a dataset have a uniform distribution with respect to support. The main problem with this is items in a transaction are treated equally.

B. Significance of item

W. Wang introduces the concept of weight to be assigned for item in each transaction which reflects the intensity or the importance of the item within the transaction. The main problem with this is weights are introduced only during the rule generation step not used for the mining purposes.

C. Weighted Association Rule Mining

Feng Tao et.al presents Weighted Association Rule Mining for frequent itemset mining. In this work the limitation of the conventional Association Rule Mining model is avoided specifically its inability for treating units differently. The presented method uses weights which can be incorporated in the mining process to resolve this difficulty. Then the challenge is solved when doing enhancement towards using weight, especially the invalidation of downward closure property. In order to adapt weighting in the new setting, a set of new concepts are used. With this weighted downward
closure term is used as a substitute of the unique downward closure property. At last this method is confirmed as suitable and gives reason for the efficient mining scheme in the new construction of weighted support. By learning the simulation of the lattice building, solution is suggested that weight can be utilized to guide the mining focus to those significant itemsets with high degree of consequence.

Transaction weight is a type of itemset weight. It is a value attached to each of the transactions. Usually the higher a transaction weight, the more it contributes to the mining result. However weights are to be priorly assigned which is difficult in real life cases.

D. Data trimming framework

Data trimming framework is presented for mining frequent itemsets from uncertain data under a probabilistic framework. This method uses the U-Apriori algorithm, which is a customized part of the Apriori algorithm, to process on various datasets. Then the computational problem of U-Apriori is identified by using a data mining technique. Then LGS-Trimming method is used under the framework and confirmed, by widespread experiments, that it attains very high performance gain by means of Input/output cost and computational cost. In contrast to U-Apriori, LGS-Trimming process well on datasets with increased percentage of low probability items.

E. W-support mechanism

Ke Sun and Fengshan Bai presented novel framework of w-support mechanism in association rule mining. Initially, the HITS model and algorithm are utilized to obtain the weights of transactions from a database record with simply binary attributes. By derived from these weights, a novel assessment of w-support is described to provide the consequence of item sets. However the presented method differs from the conventional support in taking the quality of transactions into account. Then, the w-confidence and w support of association rules are described in similarity to the description of confidence and support. Then an Apriori-like algorithm is presented to extract association rules whereas

In the traditional itemset mining problem, items belonging to transactional data are treated equally. To allow differentiating items based on their interest or intensity within each transaction. In existing system the authors focus on discovering more informative association rules, i.e., the weighted association rules (WAR), which include weights denoting item significance. The main drawback of the WAR is weights are introduced only during the rule generation step after performing the traditional frequent itemset mining process.

To overcome the above issue the Weighted Support and Significance Framework proposed. This framework attempted to push item weights into the item set mining process. It proposes to exploit the anti-monotonicity of the proposed weighted support constraint to drive the Apriori-based itemset mining phase. But the drawback of the technique is weights have to be pre-assigned, while, in many real-life cases, this might not be the case. To address this issue, some existing technique that analyzed transactional data set is represented and evaluated by means of a well-known indexing strategy which is named as HITS. The HITS helps to automate item weight assignment.

Weighted item support and confidence quality indexes are defined accordingly and used for driving the itemset and rule mining phases.

F. Advantages

- To advance the profit of rarely originated datasets in the transactions.
- The efficiency of performance has been improved when the large databases has been accounted.

III. OBJECTIVES & OVERVIEW OF THE PROPOSED MECHANISM

A. Objectives

In this paper, we propose the method to find infrequent items by the temporal partition of the database rather than scanning the entire database by the method of Split and Merge Algorithm(SMA). The main objective is to reduce the time of search and quick retrieval of result. The algorithm also used to handle the database at the time when the load of data increases beyond a certain level. It is used to keep the database stable without crashing when the load increases.

B. Overview of the proposed Mechanism

The Proposed system focuses on the issue of discovering infrequent itemsets by using weights for differentiating between relevant items and not within each transaction. To reduce the complexity of the mining process in high dimensional databases, the temporal infrequent weighted itemset TIWI Miner adopts an FP-tree node pruning strategy to early discard items (nodes) that could never belong to any itemset satisfying the TIWI-support threshold.

The proposed system performs the infrequent itemset mining using weighted value calculation. The system also performs the temporal partition of the database for timely analysis. Using the temporal scheme the user can identify the frequent and infrequent item weights for the specific time interval. Using the above the user can take decision according to the item transaction. The proposed system implements the aggregation techniques in the transactional database using FP-Growth algorithm.

IV. THE ALGORITHMS

A. Split And Merge Algorithm (SMA)- low utility Mining

The steps are illustrated in Figure 1 for a simple example transaction database.

Step 1: Shows the transaction database in its original form.
Step 2: The frequencies of individual items are determined from this input in order to be able to discard frequent items
support of three-old. An first divide the database into five either-M operates is built transaction table, where each distinct item is encoded as a

Initially, the original database is converted into the C.

Output: F values which is a set of TIWI extending prefix

Input:

prefix process.

maximum TIWI support threshold. And finally perform the tree from the TIWI support algorithm. Another one is B.

contained items).

counter and a pointer to the sorted transaction (array of which each element consists of two fields: an occurrence counter and a pointer to the sorted transaction (array of contained items).

**B. TIWI support mining algorithm**

The TIWI Mining algorithm takes three parameters one is the tree from the TIWI support algorithm. Another one is maximum TIWI support threshold. And finally perform prefix process.

**Input:** An FP tree

Maximum TIWI support threshold (St)

The set of items with patterns (prefix)

Output: F values which is a set of TIWI extending prefix

**Steps:**

1. Initially assign 0 for F.

   F=0

2. For each item I in the header tree table HTree.

3. =prefix U[1]-generate a new itemset I by joining prefix and I with TIWI support set to the TIWI support item i

4. If I is infrequent

   a. Store I.

5. End if

6. If TIWI-support(I) <= St then

   F=F\cup\{I\}

7. End if

8. Conditional_pattern (P)=generate(Htree, I)

9. HTree=\text{createFP-tree(Conditional_pattern)}

10. Perform pruning

    Prune=identify(Htree I, St)

    HTree=remove(Htree I, prune)

11. If HTree I #0 then

    F=F \cup \text{TIWIMining(HTree, St, I)}

12. End

13. Return the output F.

**C. Dataset**

Initially, the original database is converted into the transaction table, where each distinct item is encoded as a unique transaction id. The In-INFREQUENT weighted item set and in In-INFREQUENT transactions are separately identified by tables and the rule information. The transaction-rule index is also constructed using the concept of inverted lists to correlate the tables for efficient retrieval.

**D. Temporal Segmentation**

This module presents the proposed algorithm, TIWI, for maxing general temporal association rules in this section. A novel algorithm, TIWI, is proposed in this section to discover general temporal association rules efficiently. The basic idea behind TIWI is to first divide the database into partitions according to the time granularity imposed. Then, in light of the exhibition period of each item, TIWI employs procedure to segment the database into sub-databases in such a way that items in each sub-database will have either the common starting time or the common ending time.

**E. TIWI support mining algorithm**

The third module represents the infrequent item generation using the TIWI algorithm. This has the following procedure. Find max_sup. It is the maximum support threshold. An item set satisfies maximum support if the occurrence frequency of the item set is greater than or equal to max_sup. If an item set satisfies maximum support, then it is an infrequent item set.

**F. Strong Rules**

Rules that satisfy both a maximum support threshold and a maximum confidence threshold are called strong.

**G. Association Rule Maxing**

Find all infrequent item sets. Generate strong association rules from the infrequent item sets.

**H. TOP k result Generation**

An infrequent item set is an item set whose support is greater than some user-specified maximum support. A candidate item set is a potentially infrequent item set. This can produce the top k values. Here k represents the threshold of the user specified limit.

**I. Aggregation**

Finally the system finds the infrequent items and its aggregated results.

**V. PERFORMANCE EVALUATION**

**A. Execution Time Analysis**

Figure 1 shows the results of algorithm efficiency for the different threshold level eg. 5, 10 etc. Clearly, our TIWI algorithm achieves more efficiency even when the threshold level increases.

Figure 2 compares the results of previous algorithm with the TIWI. It shows the reduced execution time as compared to the MIWI and MINIT.
A Novel Algorithm for Temporal Infrequent Weighted Item Set using Frequent Pattern Growth (IJSRD/Vol. 3/Issue 03/2015/396)

Fig. 1

**B. Comparison Chart**

We evaluate mainly the performance according to the following metrics.

Fig. 2

VI. CONCLUSION

In this research work, the issue of discovering infrequent item sets by using weights for differentiating between relevant items and not within each transaction. Two FP Growth-like algorithms that accomplish Infrequent Weighted Itemset (IWI) and Minimal Infrequent Weighted Itemset (MIWI) mining efficiently are also proposed which improves the efficiency and consumes time. IWI Miner and MIWI Miner, which perform IWI and MIWI mining driven by IWI-support thresholds. It is a two step process. First, the weight of an itemset associated with a weighted transaction is defined as an aggregation of its item weights. Second, the significance with respect to the whole data set is estimated by combining the itemset significance weights associated with each transaction. The usefulness of the discovered patterns has been validated on data coming from a real-life context with the help of a domain expert.

REFERENCES


