Infrequent Weighted Itemset Mining Using Frequent Pattern Growth

Snehal J. Patil¹, A. S. Dange²
Department of Computer Science & Engineering
AnnasahebDange College of Engineering &Technology, Ashta, Shivaji University, Kolhapur, India

Abstract:
Itemset mining is an effective area of the research due to its successful application in various data mining scenarios like finding association rules. There are two types of itemset mining namely, Frequent Itemset Mining and Infrequent Itemset Mining. The frequent items are patterns or items like itemsets, substructures, or subsequences that come out in a data set frequently or rapidly. Frequent weighted items represent correlation regularly holding in data in which an item may weight differently. The research society has focused on the Infrequent Weighted Itemset Mining problem. Infrequent weighted item set discover an item sets whose frequency of occurrence in the analyzed data is less than or equal to a maximum threshold value. To discover an infrequent weighted item set, two algorithms are discovered an Infrequent weighted item set IWI and the Minimal infrequent item set MIWI.

Keywords: clustering, classification, FP tree, FP growth and data mining

I. INTRODUCTION

The itemset mining is an exploratory data mining technique widely used for discovering valuable correlations among data. The first attempt to perform an itemset mining was focused on discovering frequent items, i.e., patterns whose observed frequency of an occurrence in the source data (the support) is above a given threshold. Frequent items find application in a number of real-life contexts (e.g., market basket analysis, medical image processing, biological data analysis). Nevertheless, many traditional approaches ignore an influence/interest of each item/transaction within the analyzed data. To allow treating an item/transactions differently based on their relevance in the frequent itemset mining process, the concept of weighted itemset has also been introduced. A weight is associated with each data item and the characterizes its local significance within each transaction.

Here in this paper tackles the innovation of an infrequent weighted itemsets. To overcome this subject, the IWI support measures are defined as a weighted frequency of an instance of the itemsets in the analyzed data. After applying the given cost function to the weights associated with an item in each transaction an occurrence weights are derived. Particularly we focus on the two distinct IWI support measures: (i) The IWI support min measure, which relies on minimum cost function. (ii) The IWI support max measures, which relies on the maximum cost function. In this paper we overcome the following problems:

a. IWI and minimal IWI mining driven by a maximum IWI support min threshold,
b. IWI and minimal IWI mining driven by a maximum IWI support max threshold.

To accomplish above tasks, we represent two algorithm which are, infrequent weighted itemset miner and minimal infrequent weighted itemset miner. An algorithm performs the IWI and MIWI mining driven by the IWI support thresholds. Both these algorithms are FP growth like algorithm which is satisfied to the following features: (i) FP tree node pruning determined by the maximum IWI support constrained and (ii) cost function independence i.e., they work in the same way regardless of which constraint is applied, (iii) early stopping of the recursive FP tree search in MIWI miner to avoid extracting non minimal IWIs.

II. LITERATURE SURVEY

[3] Introduced the frequent itemset mining technique which is broadly used in data mining technique. The items which are belonging to the transactional data are treated uniformly in the existing approached.

In [4], the author focuses on generating more revealing an association rules i.e., the weighted associations rules for allowing the problem of differentiating an items based on their interest on intensity within each transaction. Though, weights are initiated only during a rule generation step after performing the traditional frequent itemset mining process.

In [5], author first time try to push an items weights into the itemset mining process. For deriving an apriori based itemset mining, process the work produces an antimonotonicity of the proposed weighted support constraint.

III. MINING TECHNIQUES FOR INTERESTING INFREQUENT PATTERNS

In principle, an infrequent itemsets are given by all itemsets that are not extracted by standard frequent itemset generations algorithms like these Apriori and FP-growth. Since the number of infrequent patterns exponentially large, especially for sparse, high dimensional data, the techniques developed for mining infrequent patterns focus on finding only interesting infrequent patterns.

Mining Negative Patterns
Transaction data can be binarized by increasing it with negative items. By applying existing frequent itemset generation algorithm such as Apriori on the increases transactions, all the negative itemsets can be derived. Such an approach is probable only if a few variables are treated as symmetric binary.

Support Expectation
Another class of techniques considers the infrequent pattern to be interesting only if its actual support is considerably smaller
than its anticipated support. For negatively correlated patterns, the expected support is computed based on the statistical an independence assumption. Two alternative approaches for determining the expected support of a pattern using a concept hierarchy and a neighborhood-based approach known as indirect association.

Support Expectation Based on Concept Hierarchy
An objective measures alone may not be sufficient to eliminate uninteresting infrequent patterns. For example, support bread and laptop computer are the frequent items. Even though the itemset \{bread, laptop computer\} is infrequent and perhaps negatively correlated, it is not absorbing because their lack of support seems obvious to domain experts. Therefore, a subjective approach for determining anticipated support is needed to avoid generating such infrequent patterns. In the example, bread and laptop computers belong to two completely different product categories, which is why it is not surprising to find that their support is low.

Support Expectation Based on Indirect Association
Consider a pair of an items, \((a, b)\), that are rarely bought together by customers. If \(a\) and \(b\) are unrelated an items such as bread and DVD player, then their support is an expected to be low. On other hand, if \(a\) and \(b\) are related items, then their support is anticipated to be high. The expected support was previously computed using a concept hierarchy. Here, an approach for determining the anticipated support between a pair of items by looking at other items commonly purchased together with two items. Indirect association has many potential applications. In the market basket domain, \(a\) and \(b\) may refer to computing an items such as desk top and laptop computers. In text mining, an indirect association can be used to an identify synonyms, an antonyms, or words that are used in different contexts. For example, given a collections of documents, the word data may be an indirectly associated with gold via the mediator mining. This pattern suggests that the word mining can be used in the two different contexts data mining versus gold mining.

IV. WEIGHTED FREQUENT ITEMSETS MINING
Researchers have proposed weighted frequent an itemset mining algorithms that reflect the significance of items. The foremost focus of weighted frequent an itemset mining is concerns satisfying the downward closure belongings. Every weighted association rules mining algorithms implied so far have been based on the Apriori algorithm. Nevertheless, pattern growth algorithms are much more effectiveness than Apriori based algorithms. An efficient weighted frequent itemset mining algorithm is main approach used to push weight constraints into the pattern growth algorithm and provide ways to keep the downward closure assets. The WFIM accepts an rising weight ordered prefix tree. The tree is crossed bottom-up because the previous matching are not maintain the downward closure property. A support of each itemset is usually decreased as the length of an itemset is an enlarged, but the weight has a unusual characteristic. An itemset which has a low weight sometimes it can get a higher weight after adding another item with a higher weight, so it is not assured to keep the downward closure property [7].

V. MODEL DESCRIPTION
The weighted transaction dataset is input for the project. This dataset is created at every time if we run the project, this dataset is created by 5 system, system A, system B, system C, system D and system E. These are the 5 systems using in this transaction dataset. The utilization of this system is take it as input in the weighted transaction dataset, and then next we find out the weighting function of this weighted transaction dataset. The weighting function is useful to the infrequent support minimum function.

The infrequent weighted itemset support minimum function is calculated for all these transaction in the dataset. The weighted transaction equivalence is find out by using the weighted transaction dataset. The weighted transaction equivalence is based on the original dataset. It is just related to the original dataset, it have some calculations to generate the equivalence weighted transaction set, using these equivalent transaction set we calculate the infrequent weighted itemset support. Then the threshold is calculated for all these data.

The system is below the threshold weight then it is taken for the consideration otherwise it is not consider, using infrequent weighted itemset miner algorithm we can find the common systems in between the weighted transaction dataset and the equivalent weighted transactions. The combination of this two are getting the infrequent weighted itemset miner and then the transactions included in this common types are taken as a output.

The flowchart of proposed module is shown in Figure 1.

**VI. THE ALGORITHMS**

**Weighted Transaction Equivalence**
The weighted transaction equivalence establish an association between a weighted transaction data set \(T\), composed of transactions with an arbitrarily weighted items within each transaction , and an equivalent data set \(TE\) in which each transaction is an exclusively composed of equally weighted items. To this aim, each weighted transaction \(tq \in T\) corresponds to an equivalent weighted transaction set, which is the subset of \(TE\)’s transactions. Item weights in \(tq\) are spread, based on an irrelative significance, among their
equivalent transactions in TEq. The proposed transformation is particularly suitable for compactly representing the original data.

**The Infrequent Weighted Itemset Miner Algorithm**

IWI miner is the FP growth like mining algorithm which is carry out projection based itemset mining. The algorithm carry out the following FP growth mining steps: (i) creation of FP tree and (ii) recursive itemset mining from the FP tree index. Distinct from FP tree IWI miner generates infrequent weighted itemsets instead of frequent ones. For achieving the task, there are few changes in the FP tree growth which are as follows: (i) A novel pruning approach for pruning part of the search space an early and (ii) a slightly modified tree structure, which allows storing the IWI support value associated with each node.

**The Minimal Infrequent Weighted Itemset Miner Algorithm**

A weighted transactional dataset and the maximum IWI-support (IWI-support-min or IWI-support-max) threshold, the Minimal Infrequent Weighted Itemset Miner algorithm extracted all the MIWIs that satisfy . The pseudo code of the MIWI Miner algorithm is identical to the one of IWI Miner, reported in Algorithm 1. Hence, due to space constraints, the pseudo code is not reported. However, in the following, the main differences with respect to IWI Miner are outlined. The MIWI Mining procedure is identical to IWI Mining. However, since MIWI Miner attention on generating only minimal infrequent patterns, the recursive extraction in MIWI Mining procedure is canceled as soon as an infrequent item set occurs. In fact, whenever an infrequent item set I is exposed, all its extensions are not minimal.

**VII. CONCLUSION**

This paper faces the problem of discovering infrequent itemsets by using weights for differentiating between relevant items and not within each and every transaction. Two FP Growth- like algorithms that accomplish IWI and MIWI mining efficiently are proposed. The benefit of the discovered patterns has been validated on data coming from a real-life context with the help of a domain expert.

**VIII. FUTURE WORK**

The future work will be exposing maximum and minimum infrequent item-set by integrating existing algorithm with residual trees. As future work, we plan to integrate the proposed approach in an advanced decision-making system that supports the domain expert’s targeted actions based on the characteristics of the discovered IWIs. Furthermore, the application of different aggregation functions besides minimum and maximum will be studied.

**XI. REFERENCES**


