Scientific Journal Impact Factor (SJIF): 1.711
International Journal of Modern Trends in Engineering and Research
ISSN (ONLINE) : 2349 - 9745
ISSN (PRINT) : 2393 - 8161

TREE BASED PATTERN AND EVENT DISCOVERY SCHEME

Ms. N.M. Indumathi[1], Ms. C. Nithya[2]
ME, *Assistance Professor, Department of Computer Science[1,2],*
Nandha Engineering College (Autonomous), Erode, Tamilnadu, India

Abstract - Frequent and infrequent itemsets are discovered using the association rule mining methods. Weighted rule mining methods are used to identify the rules with attribute weight values. Rare patterns are discovered using the Infrequent Weighted Itemset (IWI) mining mechanism. Weight and frequency values are analyzed in the Infrequent Weighted Itemset mining process. Two IWI support measures are used in the system. Minimum cost function is used in the IWI-support-min. Maximum cost function is used in the IWI-support-max measure estimation process. Infrequent weighted itemsets are discovered using two algorithms. They are Infrequent Weighted Itemset Miner (IWI Miner) and Minimal Infrequent Weighted Itemset Miner (MIWI Miner). IWI miner uses the IWI support with Max measure value. IWI support Min measure is used in MIWI miner algorithm. IWI Miner and MIWI Miner are Frequent Pattern (FP) Growth-like mining algorithms. Frequent Pattern (FP) tree construction and regressive itemset mining in the FP-tree index tasks are carried out in the IWI mining process. Event discovery is carried out using the IWI mining scheme. Accuracy is improved with Aggregation functions. Optimal cost functions are applied in the mining process. Data usefulness analysis is integrated with the infrequent itemset mining operations.

I. INTRODUCTION

Association rule analysis is one of the most important fields in data mining. It is commonly applied to market-basket databases for analysis of consumer purchasing behavior. Such databases consist of a set of transactions, each containing the items a customer purchased. The most important and computationally intensive step in the mining process is the extraction of frequent itemsets - sets of items that occur in at least minSup transactions.

It is generally assumed that the items occurring in a transaction are known for certain. However, this is not always the case. In many applications the data is inherently noisy, such as data collected by sensors or in satellite images. In privacy protection applications, artificial noise can be added deliberately [2]. Finding patterns despite this noise is a challenging problem. By aggregating transactions by customer, we can mine patterns across customers instead of transactions. This produces estimated purchase probabilities per item per customer rather than certain items per transaction.

II. RELATED WORK

The problem of mining frequent itemsets was first introduced by Agrawal et al. [1], who proposed the Apriori algorithm. Apriori is a bottom-up, breadth-first search algorithm that exploits the downward closure property “all subsets of frequent itemsets are frequent”. Only candidate frequent itemsets whose subsets are all frequent are generated in each database scan. Apriori needs l database scans if the size of the largest frequent itemset is l. In this paper, we propose a variation of the Apriori algorithm for mining minimally infrequent itemsets (MIIs).

Volume 03, Issue 01, [January – 2016] ISSN (Online):2349–9745 ; ISSN (Print):2393-8161

In [8], Han et al. introduced a novel algorithm known as the FP-growth method for mining frequent itemsets. The FP-growth method is a depth-first search algorithm. A data structure called the FP-tree is used for storing the frequency information of itemsets in the original transaction database in a compressed form. Only two database scans are needed for the algorithm and no candidate generation is required. This makes the FP-growth method much faster than Apriori. In [6], Grahne et al. introduced a novel FP-array technique that greatly reduces the need to traverse the FP-trees. We use a variation of the FP-tree for mining the MIIs. To the best of our knowledge there has been only one other work that discusses the mining of MIIs. In [7], Haglin et al. proposed the algorithm MINIT which is based upon the SUDA2 algorithm developed for finding unique itemsets [9]. The authors also showed that the minimal infrequent itemset problem is NP-complete.

Dong et al. proposed the MLMS model for constraining the number of frequent and infrequent itemsets generated. A candidate generation-and-test based algorithm Apriori MLMS was proposed in [5]. The downward closure property is absent in the MLMS model, and thus, the Apriori MLMS algorithm checks the supports of all possible k-itemsets occurring at least once in the transaction database, for finding the frequent itemsets. Generally, the support thresholds are chosen randomly for different length itemsets with the constraint $\sigma_i \geq {}^{\gamma}{}_i < {}_j$. In [4], Dong et al. extended their proposed algorithm from [5] to include an interestingness parameter while mining frequent and infrequent itemsets.

III. INFREQUENT WEIGHTED ITEM SET MINING

Itemset mining is an exploratory data mining technique widely used for discovering valuable correlations among data. The first attempt to perform itemset mining as focused on discovering frequent item sets, i.e., patterns whose observed frequency of occurrence in the source data is above a given threshold. Frequent item sets find application in a number of real-life contexts. 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, the notion of weighted itemset has also been introduced. A weight is associated with each data item and characterizes its local significance within each transaction.

This paper addresses the discovery of infrequent and weighted itemsets, i.e., the infrequent weighted itemsets, from transactional weighted data sets. To address this issue, the IWI-support measure is defined as a weighted frequency of occurrence of an itemset in the analyzed data [3]. Occurrence weights are derived from the weights associated with items in each transaction by applying a given cost function. In particular, we focus our attention on two different IWI-support measures: (i) The IWI-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, (ii) The IWI-support-max 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. They are deemed suitable for driving the selection of a worthwhile subset of infrequent weighted data correlations. Specifically, the following problems have been addressed:

A. IWI and Minimal IWI mining driven by a maximum IWI-support-min threshold and

B. IWI and Minimal IWI mining driven by a maximum IWI-support-max threshold.

To accomplish tasks (A) and (B), we present two novel algorithms, namely Infrequent Weighted Itemset Miner (IWI Miner) and Minimal Infrequent Weighted Itemset Miner (MIWI Miner), which perform IWI and MIWI mining driven by IWI-support thresholds [10]. IWI Miner and MIWI Miner are FP-Growth-like mining algorithms, whose main features may be summarized as follows: (i) Early FP-tree node pruning driven by the maximum IWI-support constraint, i.e., early discarding of part of the search space thanks to a novel item pruning strategy 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. Property (ii) makes tasks (A) and (B) equivalent, from an algorithmic point of view, as long as a preliminary data transformation step, which adapts data weights according to the selected aggregation function, is applied before accomplishing the mining task.

IV. PROBLEM STATEMENT

Infrequent Weighted Itemset (IWI) mining mechanism is used to fetch rare patterns with weight values. IWI support threshold is used to mine infrequent weighted itemsets. IWI-support measure is defined as a weighted frequency of occurrence of an itemset in the analyzed data. The system uses two different IWI-support measures. The IWI-support-min measure relies on a minimum cost function. The IWI-support-max measure relies on a maximum cost function. Two different algorithms are used to mine infrequent weighted itemsets. They are Infrequent Weighted Itemset Miner (IWI Miner) and Minimal Infrequent Weighted Itemset Miner (MIWI Miner). IWI miner uses the IWI support with Max measure value. IWI support Min measure is used in MIWI miner algorithm. IWI Miner and MIWI Miner are Frequent Pattern (FP) Growth-like mining algorithms. Frequent Pattern (FP) tree construction and regressive itemset mining in the FP-tree index tasks are carried out in the IWI mining process. The following problems are identified from the IWI mining methods.

- Decision making process is not included
- Event detection process is not supported
- Aggregation functions based analysis is not performed
- Data usefulness measure is not estimated

V. PATTERN AND EVENT DISCOVERY PROCESS

The weighted itemset mining mechanism is used to fetch infrequent rules with frequency and weight values. Event detection and decision making process are integrated in the system. Scheduling process is performed in the decision making process. The system is divided into six major modules. They are Data Preprocess, Support Estimation, FP Tree Construction, Weighted Itemset Miner, Enhanced Weighted Itemset Miner and Event Detection. Data preprocess module is used to perform data populate and cleaning process. Probability based support estimation is performed in the support estimation module. FP tree construction module is used to construct the itemset tree values. Weighted itemset mining module is designed to fetch infrequent patterns. Aggregation function based weighted itemset mining process is carried out under the enhanced weighted itemset mining module. Targeted event detection and decision making are performed in event detection process.

Computational resource data is used for the pattern mining process. Resources and weight values are assigned with energy consumption levels. Data populate process is used to transfer the textual data values into database. Missing values are assigned in the data cleaning process. Candidate set and itemsets are prepared in the support estimation process. Attributes and their values are used in the candidate set estimation process. Itemsets are prepared using candidate sets. Support values are estimated with frequency and weight values.

Frequent Pattern (FP) tree is constructed with itemset values. Tree nodes are updated with frequency information. Itemsets are compared in the tree update process. Weight and support values are used in the FP tree construction process. The weighted itemset mining is performed to fetch infrequent patterns. Infrequent Weighted Itemset (IWI) miner is used for the mining process. Maximal cost function is used in the IWI miner process. Minimal IWI miner is used to fetch the rules with minimum cost functions.

Aggregation functions are used to improve the IWI miner process. Adaptive weight assignment model is performed with aggregation functions. Optimal cost function identification process is used for the mining process. Cost functions are integrated with rule boundary levels. Event detection process is applied on the infrequent rule collections. Event threshold values are collected from the user. Decision making is performed with detected events. Resource scheduling is carried out with reference to the events.

VI. PERFORMANCE ANALYSIS

The weighted rule mining methods are used to fetch the association rules using the items weight values. The Infrequent Weighted Itemset (IWI) miner algorithm is used to discover rules with weight values. The IWI miner algorithm uses the minimum and maximum cost functions to fetch the rules. The Enhanced Infrequent Weighted Itemset (EIWI) miner algorithm uses the weighted support with optimal weight values. The system supports event detection process with the rule mining tasks. The weighted rule mining system is tested with three parameters. They are rule retrieval rate, time complexity and rule relevancy rate. The rule retrieval rate analysis between the Infrequent Weighted Itemset (IWI) miner and the Enhanced Infrequent Weighted Itemset (EIWI) miner techniques are analyzed with different data values. The system analysis is carried out with the CPU resource allocation traces. Synthetic data generation mechanism is used to prepare the data set.

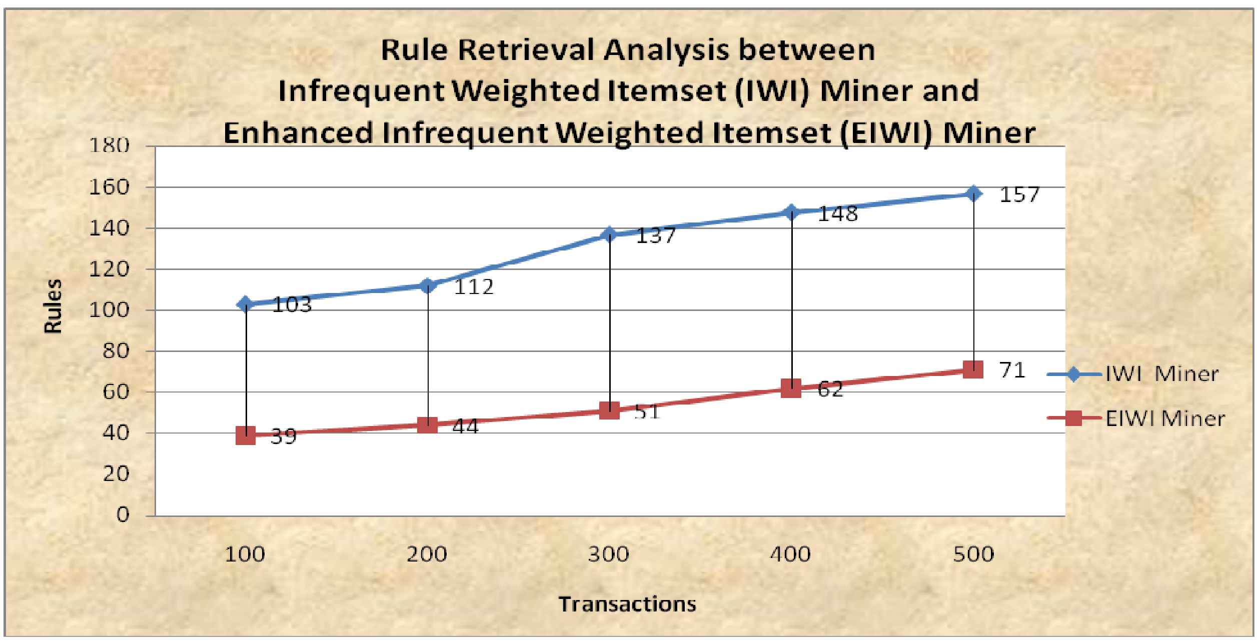

Figure No. 6.1. Rule Retrieval Analysis Infrequent Weighted Itemset (IWI) Miner and Enhanced Infrequent Weighted Itemset (EIWI) Miner

The Infrequent Weighted Itemset (IWI) miner and Enhanced Infrequent Weighted Itemset (EIWI) miner algorithm rule retrieval rate is compared in figure 6.1. The analysis result shows that the (EIWI) miner technique reduces the irrelevant rule selection 55% than the Enhanced Infrequent Weighted Itemset (EIWI) miner technique.

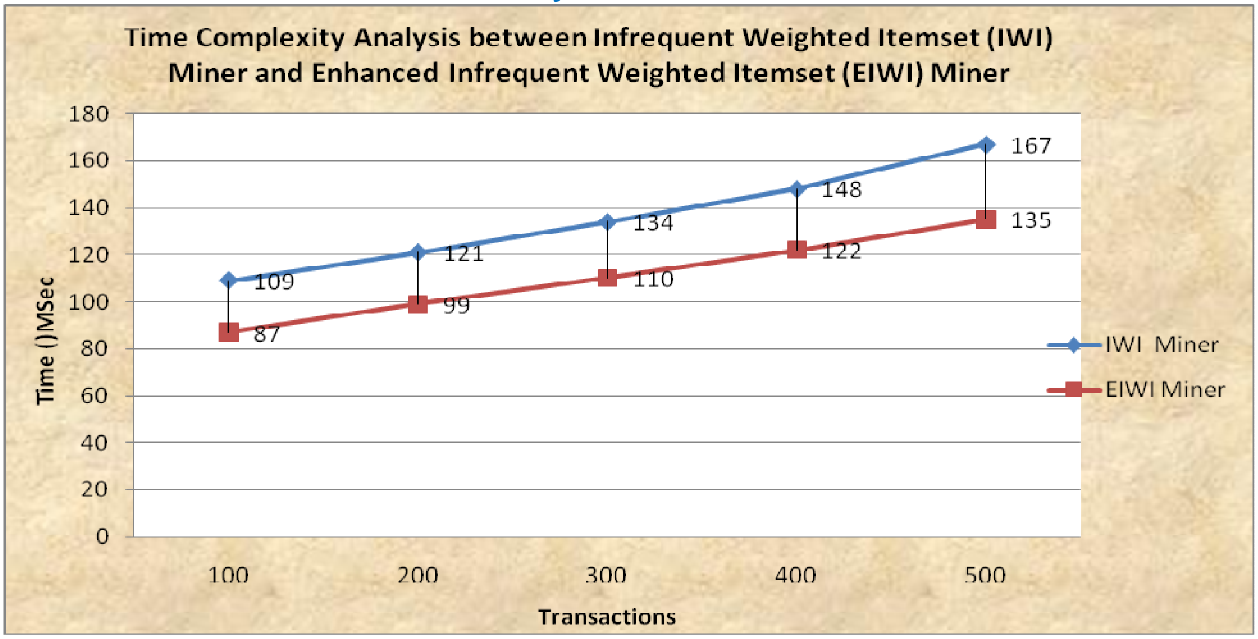

Figure No. 6.2. Time Complexity Analysis between Infrequent Weighted Itemset (IWI) Miner and Enhanced Infrequent Weighted Itemset (EIWI) Miner

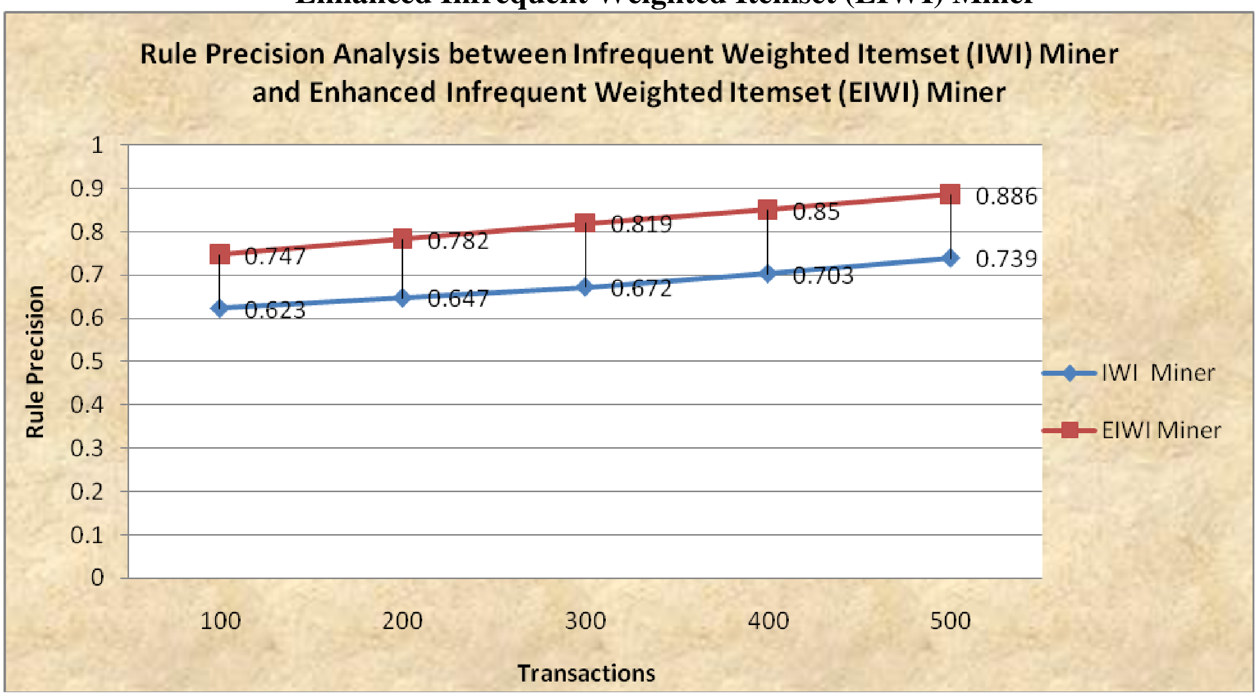

Figure No. 6.3. Rule Precision Analysis between Infrequent Weighted Itemset (IWI) Miner and Enhanced Infrequent Weighted Itemset (EIWI) Miner

The time complexity analysis between the Infrequent Weighted Itemset (IWI) miner and the Enhanced Infrequent Weighted Itemset (EIWI) miner techniques are compared in figure 6.2. The analysis result shows that the Enhanced Infrequent Weighted Itemset (EIWI) miner technique reduces the process time 20% the (IWI) miner technique.

The rule relevancy rate analysis between the Infrequent Weighted Itemset (IWI) miner technique and Enhanced Infrequent Weighted Itemset (EIWI) miner technique are compared in figure 6.3. The analysis result shows that the Enhanced Infrequent Weighted Itemset (EIWI) miner technique increases the rule relevancy rate 15% than the Infrequent Weighted Itemset (IWI) miner technique.

VII. CONCLUSION AND FUTURE WORK

The association mining techniques are used to identify the frequent patterns or attribute relationships. Rare and weighted itemsets are extracted using Infrequent Weighted Itemset (IWI) mining mechanism. Infrequent Weighted Itemset Miner (IWI Miner) and Minimal Infrequent Weighted Itemset Miner (MIWI Miner) are used extract infrequent patterns. The IWI miner is enhanced to support targeted event detection based decision support system. The system improves the rule retrieval accuracy level. The system support high scalability in weighted itemset mining process. Targeted action detection is performed. Computational overhead is reduced by the system. The weighted support based infrequent rule mining mechanism is designed with minimum and maximum cost functions. The Infrequent Weighted Miner is enhanced to detect events based on the rules. The rule mining model is applied on the CPU resource analysis environment.

The system can be enhanced with the following features. The weight based rule mining mechanism can be integrated with the support probability based mechanism. The system can be enhanced to support rule mining under distributed environment. The event detection process can be improved with event reaction mechanism. The weighted rule mining scheme can be adapted to discover rule under dynamic weight environment.

REFERENCES

[1] R. Agrawal and R. Srikant. Fast Algorithms For Mining Association Rules In Large Databases. In VLDB, pages 487–499, 1994.

[2] Lukasz Golab and Divesh Srivastava, "Discovering Conservation Rules", IEEE Transactions On Knowledge And Data Engineering, Vol. 26, No. 6, June 2014

[3] Ran Wolff, "Local Thresholding in General Network Graphs", IEEE Transactions On Knowledge And Data Engineering, April 2014

[4] X. Dong, Z. Niu, D. Zhu, Z. Zheng, and Q. Jia. Mining Interesting Infrequent And Frequent Itemsets Based On Mlms Model. In ADMA, pages 444–451, 2008.

[5] X. Dong, Z. Zheng and Z. Niu. Mining Infrequent Itemsets Based On Multiple Level Minimum Supports. In ICICIC, page 528, 2007.

[6] G. Grahne and J. Zhu. Fast Algorithms For Frequent Itemset Mining Using FP-Trees. Trans. Know. Data Engg., 17(10):1347–1362, 2005.

[7] D. J. Haglin and A. M. Manning. On Minimal Infrequent Itemset Mining. In DMIN, pages 141–147, 2007.

[8] J. Han, J. Pei, and Y. Yin. Mining Frequent Patterns Without Candidate Generation. In SIGMOD, pages 1–12, 2000.

[9] A. M. Manning and D. J. Haglin. A New Algorithm For Finding Minimal Sample Uniques For Use In Statistical Disclosure Assessment. In ICDM, pages 290–297, 2005.

[10] Zhou Zhao and Wilfred Ng, "Mining Probabilistically Frequent Sequential Patterns in Large Uncertain Databases", IEEE Transactions On Knowledge And Data Engineering,, May 2014