ISSN:0975-9646

Shipra Khare et al, / (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 5 (2) , 2014, 1642-1647

A Review on Infrequent Weighted Itemset Mining using Frequent Pattern Growth

Shipra Khare , Prof. Vivek Jain

Dept of Computer Science
SRCEM, Gwalior ,India

Abstract— **In data mining and knowledge discovery technique areas, frequent pattern mining plays an important role but it does not consider different weight value of the items. The frequent itemsets are patterns or items like itemsets, substructures, or subsequences that come out in a data set frequently or rapidly. The frequent itemsets are patterns or items like itemsets, substructures, or subsequences that come out in a data set frequently or rapidly. In this paper we are presenting review of various frequent pattern mining and weighted itemset mining.**

Index Terms— **Data Mining, frequent pattern Mining, itemset mining, infrequent weighted itemset.**

INTRODUCTION

Data mining is the process of analyzing data from different perspectives and summarizing it into useful information - information that can be used to either enhance profits, cuts costs, or both. Data mining has concerned about a great deal of attention in the information industry and in society as entire in recent years, due to the wide preventability of large amounts of data and the forthcoming need for forming such data into useful information and knowledge. Itemset mining is an exploratory data mining technique widely used for discovering valuable correlations among data [1].

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 [2]. 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, which may be jointed with itemsets or subsequences. If a substructure occurs recurrently, it is called a (frequent) structured pattern. Discovery such frequent pattern plays an important role in mining relations, correlations, and many other appealing relationships along with data. Additionally, it helps in data clustering, classification and other data mining tasks as well [3].

However, significantly less attention has been paid to mining of infrequent itemsets, although it has acquired significant usage in (i) mining of negative association rules from infrequent itemsets [4], (ii) statistical disclosure risk assessment where rare patterns in anonymous census data can lead to statistical disclosure [5], (iii) fraud detection where rare patterns in financial or tax data may suggest unusual activity associated with fraudulent behavior [5], and (iv) bioinformatics where rare patterns in microarray data may suggest genetic disorders [5]. The large body of frequent itemset mining algorithms can be broadly classified into two categories: first is candidate generation-and-test paradigm and second is pattern-growth paradigm. As per previous studies, it has been exposed experimentally that pattern-growth based algorithms are computationally faster on dense datasets [6].

Patterns that are rarely found in database are often considered to be uninteresting and are eliminated using the support measure. Such patterns are known as infrequent patterns. An infrequent pattern is an itemset or a rule whose support is less than the minsup threshold. Although a vast majority of infrequent patterns are uninteresting, some of them might be useful to the analysis, particularly those that correspond to negative correlations in data. Some infrequent patterns may also suggest the occurrence of interesting rare events or exceptional situations in the data. For example, if {Fire = Yes} is frequent but {Fire = Yes, Alarm = On} is infrequent, then the latter is an interesting infrequent pattern because it may indicate faulty alarm systems. To detect such unusual situations, the expected support of a pattern must be determined, so that, if a pattern turns out to have a considerably lower support than expected, it is declared as an interesting infrequent pattern.

Mining infrequent patterns is a challenging endeavor because there are an enormous number of such patterns that can be derived from a known data set. More exclusively, the key issues in mining infrequent patterns are: (1) how to identify interesting infrequent patterns, and (2) how to efficiently discover them in large data sets. To get a different perspective on various types of interesting infrequent patterns, two connected conceptions are negative patterns and negatively correlated patterns. The negative itemsets and negative association rules are collectively known as negative patterns. Infrequent patterns, negative patterns, and negatively correlated patterns are three closely related concepts. Although infrequent patterns and negatively correlated patterns refer only to itemsets or rules that contain positive items, while negative patterns refer to itemsets or rules that contain both positive and negative items.

MINING TECHNIQUES FOR INTERESTING INFREQUENT PATTERNS

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

Mining Negative Patterns

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

Support Expectation

Another class of techniques considers an infrequent pattern to be interesting only if its actual support is considerably smaller than its expected support. For negatively correlated patterns, the expected support is computed based on the statistical independence assumption. Two alternative approaches for determining the expected support of a pattern using (1) a concept hierarchy and (2) a neighborhood-based approach known as indirect association.

Support Expectation Based on Concept Hierarchy

Objective measures alone may not be sufficient to eliminate uninteresting infrequent patterns. For example, support bread and laptop computer are frequent items. Even though the itemset {bread, laptop computer} is infrequent and perhaps negatively correlated, it is not interesting because their lack of support seems obvious to domain experts. Therefore, a subjective approach for determining expected support is needed to avoid generating such infrequent patterns. In the preceding 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 items, (a, b), that are rarely bought together by customers. If a and b are unrelated items such as bread and DVD player, then their support is expected to be low. On other hand, if a and b are related items, then their support is expected to be high. The expected support was previously computed using a concept hierarchy. Here, an approach for determining the expected support between a pair of items by looking at other items commonly purchased together with these two items. Indirect association has many potential applications. In the market basket domain, a and b may refer to computing items such as desktop and laptop computers. In text mining, indirect association can be used to identify synonyms, antonyms, or words that are used in different contexts. For example, given a collections of documents, the word data may be indirectly associated with gold via the mediator mining. This pattern suggests that the word mining can be used in two different contexts − data mining versus gold mining.

WEIGHTED FREQUENT ITEMSETS MINING

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

BACKGROUND

The main mining algorithm based on association rule, Apriori not only predisposed the association rule mining community, but it pretentious other data mining fields as well. In recent years, the attention of the research community has also been focused on the infrequent itemset mining problem, i.e., discovering itemsets whose frequency of occurrence in the analyzed data is less than or equal to a maximum threshold. Traditional infrequent itemset mining algorithms still suffer from their inability to take local item interestingness into account during the mining phase. 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, the authors focus on discovering more informative association rules, i.e., the Weighted Association Rules (WAR), which includes weights denoting item significance. However, weights are introduced only during the rule generation step after performing the traditional frequent itemset mining process.

RELATED WORK

In recently 2013, Luca Cagliero and Paolo Garza suggested Infrequent Weighted Itemset Mining using Frequent Pattern Growth. They addresses the discovery of infrequent and weighted itemsets, i.e., the Infrequent

Weighted Itemsets (IWIs), from transactional weighted datasets. To address this issue, the IWI-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. They mainly focuses on following: (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 [1].

It is important 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. Specifically, the following problems have been addressed:

1) IWI and Minimal IWI mining driven by a maximum IWI-support-min threshold, and
2) IWI and Minimal IWI mining driven by a maximum IWI-support-max threshold.

Task (1) entails discovering IWIs and Minimal IWIs (MIWIs) which include the item(s) with the least local interest within each transaction. Task (2) entails discovering IWIs and MIWIs which include item(s) having maximal local interest within each transaction by exploiting the IWI-support-max measure. To accomplish tasks (1) and (2), they 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 [1].

In year 2012 Xin Li et al proposed Frequent Itemsets Mining in Network Traffic Data. They think about the problem of frequent itemset mining problem in network traffic data, and propose an algorithm for mining frequent itemsets. They try to minimize the size of results and only maximal frequent itemsets are considered. To protect the privacy, intermediate mining results are encrypted using hashing method by different servers. The proposed algorithm is evaluated from the perspectives of accuracy and efficiency [2]. In same year, Soumadip Ghosh et al presented Mining Frequent Itemsets Using Genetic Algorithm. This work carried out with logic of GA to improve the scenario of frequents itemsets data mining using association rule mining. The main benefit of using GA in frequent itemsets mining is to perform global search with less time complexity. This scheme gives better results in huge or larger data set. It is also simple and efficient. They had dealt with a challenging association rule mining problem of finding frequent itemsets using their recommended GA based method. This method is very simple and efficient one. This is successfully tested for different large data sets. The results obtained are correct and appropriate [3].

Efficient mining of both positive and negative association rules [4]. They focus on identifying the associations among frequent itemsets. They need to minimize the harmful impacts as well as maximize possible benefits. They designed a new method for efficiently mining both positive and negative association rules in databases. This approach is novel and different from existing research efforts on association analysis. Some infrequent itemsets are of interest in this method but not in existing research efforts. They had also designed constraints for reducing the search space, and had used the increasing degree of the conditional probability relative to the prior probability to estimate the confidence of positive and negative association rules. Their experimental results have demonstrated that the proposed approach is effective, efficient and promising [4].

On minimal infrequent itemset mining was proposed by D. J. Haglin and A. M. Manning. They present a new algorithm for finding minimal infrequent patterns. This is the first algorithm designed specifically for finding minimal infrequent itemsets. The easiest way to describe the differences in dataset properties is to consider the matrix form. For traditional itemset mining, the matrix consists of binary entries. They can transform a SUDA2-type matrix into a binary matrix by enumerating all of the <column, value> pairs. For each of such pairs, a column is formed in the transformed binary matrix. For every value in a column in the SUDA2-type input matrix, the corresponding <column, value> location in the transformed binary matrix is given a one [5].

Ashish Gupta, Akshay Mittal and Arnab Bhattacharya presented Minimally Infrequent Itemset Mining using Pattern-Growth Paradigm and Residual Trees. they recommend a new algorithm based on the pattern-growth paradigm to find minimally infrequent itemsets. It has no subset which is also infrequent. They also introduce the novel concept of residual trees. Later on utilized the residual trees to mine multiple level minimum support itemsets where different thresholds are used for finding frequent itemsets for different lengths of the itemset. Finally, they analyze the behavior of our algorithm with respect to different parameters and show through experiments [6].

They made a fair contribution like; they proposed a new algorithm IFP min for mining minimally infrequent itemsets and an optimization on the Apriori algorithm to mine minimally infrequent itemsets. They introduce the concept of residual trees using a variant of the FP-tree structure termed as inverse FPtree. They also present a detailed study to quantify the impact of variation in the density of datasets on the computation time of Apriori, MINIT and our algorithm. They extend the proposed algorithm to mine frequent itemsets in the MLMS framework [6].

Weighted frequent itemset mining with a weight range and a minimum weight also known as WFIM is proposed by Yun, Unil, and John J. Leggett in year 2005. A weight range and a minimum weight constraint are defined and items are given different weights within the weight variety.

The weight and sustain of each item are considered separately for pruning the search domain. The numerous weighted frequent itemsets can be reduced by setting a weight range and a minimum weight, permitting the user to equilibrium support and weight of itemsets. WFIM produces more summarizing and significant weighted frequent itemsets in huge databases, predominantly intense databases with low least support, by correcting a minimum weight and a weight range. They used the term, weighted itemset to represent a set of weighted items [7].

A simple way to achieve a weighted itemset is to calculate the average value of the weights of the items in the itemset. WFIM uses FP-trees as a compression technique. FP-trees are mainly used in pattern growth algorithms. WFIM computes local frequent items of a prefix by scanning its projected database. The FP-trees in our algorithm are made as follows. Scan the transaction database one time and count the support of each item and check the weight of each item. After this, sort the items in weight ascending order. Although supports of items may be lower than the minimum support and infrequent, the items cannot be deleted since infrequent items may become weighted itemsets in the next step. The extensive performance analysis shows that WFIM is efficient and scalable in weighted frequent itemset mining. Many improved algorithms using divide and conquer methods have been also suggested [7].

In recent year 2013, Diti Gupta and Abhishek Singh Chauhan presented a survey on Mining Association Rules from Infrequent Itemsets. They survey about the finding of negative and positive association rules form the infrequent itemset. They come with several advantages and the problem formulation which can be implemented in future. Based on this study they suggested that to extend the support-confidence framework in a dynamic fashion. In addition to finding confident positive rules that have a strong correspondence, the algorithm determines negative association rules with strong negative correlation between the antecedents and consequents. So they can formulate a proficient system which generates both positive and negative association rules. They also generate all types of restrained rules, therefore permitting to be used in diverse applications where all these types of rules could be needed or just a subset of them. As a result they obtain better frequency result set for the entire item set in both positive and negative associations [8].

Association rule mining is to find out association rules that satisfy the predefined minimum support and confidence from a specified database. In the real time situation they can subdivide the problem in two parts. First is to find the set exceed a predefined threshold in the database; those item sets are called frequent or large item sets. The second stage is the occurrence generation from association rules. Therefore if applying dynamic minimum support then level wise decomposition is easy. If we think of the most appropriate and efficient data mining algorithm then we always think about Apriori algorithm. However there are two bottlenecks of the Apriori algorithm. The process of candidate generation is first which can increase the time as well as the space. Therefore the second thing is produced from the first that it needed multiple scan when it in the iteration procedure. Founded on Apriori algorithm, various new algorithms were designed with some modifications or improvements. The computational cost of association rules mining can be reduced by reducing the passes, using sampling and through adding extra constrains as per demand [8].

In 2012, YihuaZhong et al. [9] suggest that association rule is an important model in data mining. Still, conventional association rules are mostly based on the confidence metrics and support, and most algorithms and researches unspecified that each attribute in the database is equivalent. In fact, since the user inclination to the item is dissimilar, the mining rules using the offered algorithms are not always suitable to users. By introducing the idea of weighted dual confidence, a new algorithm that can mine efficient weighted rules is suggested by the authors. This study demonstrate that the algorithm can reduce the large number of meaningless association rules and mine interesting negative association rules in real life. By introducing the concept of weighted dual confidence, another new algorithm can mine effective weighted rules that are on the basis of the dual confidence association rules used in algorithm.

In 2012, He Jiang et al. [10] support the technique that allows the users to specify multiple minimum supports to reflect the natures of the itemsets and their varied frequencies in the database. This scheme is very efficient for huge databases to use algorithm of association rules based on numerous supports. The presented algorithms are typically mining positive and negative association rules from frequent itemsets. Excluding the negative association rules from infrequent itemsets are ignored. Furthermore, they set different weighted values for items according to the importance of each item. According to three factors mentioned above, an algorithm for mining weighted negative association rules from infrequent itemsets based on multiple supports (WNAIIMS) is proposed by the author. They set different minimum support for itemsets. During association rule mining, if the specified minimum support is excessively high, then the items with low occurrence of emergence couldn't be mined. Otherwise, if the specified minimum support is excessively low, then grouping explosion may arise.

In 2010 Younghee Kim et al [11] presented Mining Frequent Itemsets with Normalized Weight in Continuous Data Streams. They consider the problem of mining with weighted support over a data stream sliding window using limited memory space, called WSFI-Mine. WSFI-Mine stands for Weighted Support Frequent Itemsets Mine. This algorithm allows the user to specify the weight for each item. It can discover useful recent knowledge from a data stream by using a single scan. Based on the weighted support, we propose a new algorithm, to efficiently discover all the frequent itemsets from streams. This method is driven by an external weight table or weight function. The proposed WSFI-Mine method is designed to mine all frequent itemsets from one scan in the data streams. They propose a WSFI-Mine that can mine dynamically maintained usage patterns using information from a previous sliding time that can be updated in real time. The WSFI-Mine algorithm has three phases: the normalization of weight support and dividing patterns into three categories, the construction of the WSFP-Tree, and a frequent itemset discovery scheme. Construction of a WSFP-Tree ensures that frequent pattern mining can be performed efficiently. A WSFP-Tree is a data structure based on an extended FP-tree. It serves to store compressed crucial information about frequent patterns [11].

A Survey on Algorithms for Mining Frequent Itemsets over Data Streams was presented by James Cheng et al. [12] in year 2008. survey a number of representative state-of-the-art algorithms on mining frequent itemsets, frequent maximal itemsets, or frequent closed itemsets over data streams. They structured the algorithms into two categories based on the window model that they adopt: the landmark window or the sliding window. Every window representation is then categorized as time-based or count-based. According to the number of transactions that are updated each time, the algorithms are further classifying into update per transaction or update per batch. Then, classify the mining algorithms into two categories: exact or estimated. They also categorized the estimated algorithms according to the results they return: the false-positive approach or the false-negative approach. The false-positive approach returns a set of itemsets that includes all frequent itemsets but also some infrequent itemsets, while the false-negative approach returns a set of itemsets that does not include any infrequent itemsets but misses some frequent itemsets. They also discussed the different issues raised from the different window models and the nature of the algorithms. They also make clear the fundamental principle of the ten algorithms and analyze their merits and limitations [12].

They [12] also focused on frequent itemset mining and have tried to cover both early and recent literature related to mining frequent itemsets (FIs) or frequent closed itemsets (FCIs). In particular, they have discussed in detail a number of the state-of-the-art algorithms on mining FIs, FMIs or FCIs over data streams. Moreover, we have addressed the merits and the limitations and presented an overall analysis of the algorithms, which can provide insights for end-users in applying or developing an appropriate algorithm for different streaming environments and various applications [12].

In 2012, Idheba Mohamad Ali O. Swesi et al. [13] study is to develop a new model for mining interesting negative and positive association rules out of a transactional data set. Their proposed model is integration between two algorithms, the Positive Negative Association Rule (PNAR) algorithm and the Interesting Multiple Level Minimum Supports (IMLMS) algorithm, to propose a new approach (PNAR_IMLMS) for mining both negative and positive association rules from the interesting frequent and infrequent item sets mined by the IMLMS representation. The investigational results demonstrate that the PNAR_IMLMS model provides significantly better results than the previous model.

In 2009, Yuanyuan Zhao et al. [14] suggest that the Negative association rules become a focus in the data mining field. Negative association rules are functional in market-basket analysis to identify products that conflict with each other or products that accompaniment each other. The negative association rules frequently consist in the infrequent items. The experiment proves that the number of the negative association rules from the infrequent items is larger than those from the frequent.

Asha Rajkumar and G.Sophia Reena [15] presened a study on Frequent Item set Mining Using Global Profit Weight Algorithm in year 2010. This study focused to implement the Global Profit Weight (GPW) for the frequent item set. Generally the profit can be measured in a traditional way. In this study they proposed multi criteria based profit calculation. They also discussed the major problem is to mine the association rules with weighted items, based on the different types of the association rules, which are binary association rules and quantitative association rules. New algorithms are required to solve such problems since the available algorithms cannot be solved. The objective of study is to implement the global profit weight measure and test the performance of the algorithm with traditional weighted association rule mining. The implementation is expressed as step wise visual presentation and its performance is measured with weighted ARM. The accuracy is classified using classification algorithm such as Navie Bayes, VFI, BF Tree and IB1and the results are compared using WEKA. It can be concluded from the study result that GPW is required high computation power to generate the weight. The result is compromised with its quality. According to the research problem, the calculated weight can be reused it for many times and as required [15].

This work employs global profit weight algorithm is implemented using visual basic to find the profit of the item set in the transaction. Classification algorithm is used to evaluate the profit measure such as high (H), medium (M) and low (L). Widely used supervised machine learning techniques namely Naïve Bayes Decision tree classifier, VFI and IB1 Classifier are utilized for learning the representation. The outcomes of the models are evaluated and observed that Naïve Bayes performs well. The project generalizes previous work on profit measure. The profit measure of the items in the transaction is defined by using the characteristic of the item. Considering the profit of an item, there are a number of important factors to consider as well. They focus on the mining of weighted association rules for which the weight of an item set is normalized by the size of the item set. The choice of using unorganized or normalized weight was depended on the individual need of each application [15].

PROBLEM STATEMENT

The technique implemented here for the mining of infrequent weighted item sets provides less execution time and contains less storage and the number of nodes created are also very less on the basis of support and confidence. But future enhancements can be done as to integrate the proposed approach in an advanced decision-making system that supports 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.

PROPOSED PLAN

1. Input dataset.
2. Pass Support and confidence on the basis of which minimum support is calculated.
3. Apply Association rule mining algorithm for the generation of frequent sets and association rules.
4. Classify frequent and infrequent item sets using Naive Bayes classification.

Naïve Bayes Algorithm

Abstractly, the probability model for a classifier is a conditional model.

$$p(C|F_1,\dots,F_n)$$

over a dependent class variable C with a small number of outcomes or *classes*, conditional on several feature variables F_1 through F_n. The problem is that if the number of features n is large or when a feature can take on a large number of values, then basing such a model on probability tables is infeasible. We therefore reformulate the model to make it more tractable

. EXPECTED OUTCOME

The result analysis will be on the basis of following factors:

a. Error rate
   It is defined as the difference of the original data and the changed data after performing any algorithm on the dataset.

b. Frequent and Infrequent Item Sets generated.
   The sets that are generated using any classification algorithm that are dependant and are independent of each other.

c. Rules Generated.
   The number of rules generated using classification algorithm.

d. Time Computation.

REFERENCES

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

[2] Xin Li, Xuefeng Zheng, Jingchun Li, Shaojie Wang "Frequent Itemsets Mining in Network Traffic Data", 2012 Fifth International Conference on Intelligent Computation Technology and Automation, pp. 394- 397, 2012.

[3] Soumadip Ghosh, Sushanta Biswas, Debasree Sarkar, Partha Pratim Sarkar "Mining Frequent Itemsets Using Genetic Algorithm", International Journal of Artificial Intelligence & Applications (IJAIA), Vol.1, No.4, pp. 133 – 143, October 2010.

[4] X. Wu, C. Zhang, and S. Zhang "Efficient mining of both positive and negative association rules", ACM Transaction Information System, vol. 22, issue 3, pp. 381–405, 2004.

[5] D. J. Haglin and A. M. Manning "On minimal infrequent itemset mining", In DMIN, pp. 141–147, 2007.

[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] Yun, Unil, and John J. Leggett. "WFIM: weighted frequent itemset mining with a weight range and a minimum weight", In Proceedings of the Fifth SIAM International Conference on Data Mining, pp. 636 – 640, 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] Yihua Zhong; Yuxin Liao, "Research of Mining Effective and Weighted Association Rules Based on Dual Confidence," Fourth International Conference on Computational and Information Sciences (ICCIS), pp.1228 - 1231, Aug. 2012.

[10] He Jiang, Xiumei Luan and Xiangjun Dong," Mining Weighted Negative Association Rules from Infrequent Itemsets Based on Multiple Supports" , International Conference on Industrial Control and Electronics Engineering, pp. 89 – 92, 2012.

[11] Younghee Kim, Wonyoung Kim and Ungmo Kim "Mining Frequent Itemsets with Normalized Weight in Continuous Data Streams", Journal of Information Processing Systems, Vol.6, No.1, March 2010.

[12] James Cheng, Yiping Ke, and Wilfred Ng "A Survey on Algorithms for Mining Frequent Itemsets over Data Streams", Knowledge and Information Systems, Volume 16, Issue 1, pp. 1 - 27, July 2008.

[13] IdhebaMohamad Ali O. Swesi, Azuraliza Abu Bakar, AnisSuhailis Abdul Kadir, "Mining Positive and Negative Association Rules from Interesting Frequent and Infrequent Itemsets" , 9th International Conference on Fuzzy Systems and Knowledge Discovery, pp. 650 – 655, 2012.

[14] Yuanyuan Zhao,He Jiang; RunianGeng; Xiangjun Dong, "Mining Weighted Negative Association Rules Based on Correlation from Infrequent Items," Advanced Computer Control,ICACC '09. International Conference on, vol., no., pp.270 - 273, 22-24 Jan. 2009.

[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.