

# A Combined Approach of Frequent Pattern Growth and Decision Tree for Infrequent Weighted Itemset Mining

Mr. Amol. S. Dange, Miss. Snehal J. Patil

Department of Computer Science & Engineering  
Annasaheb Dange College of Engineering & Technology, Ashta  
Shivaji University, Kolhapur  
Email - amol.dange@gmail.com  
Email - spatilsnehal26@gmail.com

\*\*\*

**Abstract** - The frequent item set mining is one of the popular data mining techniques and it can be used in many data mining fields for finding highly correlated item sets. An infrequent item set mining finds rarely occurring item sets in the database. The proposed system uses clustering or logical grouping concepts for finding infrequent weighted item sets. The algorithm which is used in the proposed system, works well with real-time databases and is highly scalable which is suited for real-time applications. Incessant weighted item sets to associate frequently holding an information in which item sets may weight distinctively. The paper handles the issue of running across an extraordinary and weighted itemsets, i.e., An Infrequent Weighted Item set (IWI) mining. In this paper two novel quality measures are proposed to test the IWI (Infrequent Weighted Itemset) mining procedure. The two calculations that perform IWI and Neglectable IWI mining efficiently, which is determined by the proposed measures, are displayed. Test outcomes show the efficiency and adequacy of the proposed methodology.

**Key Words:** Infrequent weighted itemset, Frequent pattern growth, Data Mining, Frequent pattern Mining, Weighted mining, Decision tree.

## 1.INTRODUCTION

In the recent years, the majority of research society has been focused on the problem of infrequent item set mining, i.e. generating an item sets whose frequency of occurrence in the analyzed data is less than or equal to a maximum threshold [2].

The frequent weighted item sets represent correlations frequently holding in data in which items may weight differently. However, in some contexts, for e.g., when the need is to minimize a certain cost function, discovering rare data correlations is more interesting than the mining frequent ones. The paper overcome the issue of discovering rare and weighted item sets, i.e., an infrequent weighted item set (IWI) mining problem. In this paper two novel quality

measures are proposed to test the IWI mining process. Furthermore, The two calculations that perform IWI and Minimal IWI mining efficiently, and driven by the proposed measures, are presented.

Item set mining is an exploratory information mining system generally utilized for uncovering profitable connections among information. The main endeavor to perform item set mining was concentrated on uncovering successive item sets, i.e., the recurrence of event in the source information is over a given edge. Incessant item sets discover provision in various real connections (for e.g., market wicker container dissection, medicinal picture handling, and biotic information investigation). In any case, many customary methodologies overlook the impact/enthusiasm of everything/ transaction inside an investigated information. To permit the treating item sets/transactions diversely focused around their significance in the incessant item set mining process, the thought of weighted item set has additionally been presented. A weight is connected with every item set and portrays its neighborhood essentialness inside every transaction. The consideration of an Exploration group has an additionally been centered on the occasional item set Mining issue, i.e., uncovering item sets whose recurrence of event in the investigated information is short of what or equivalent to a most extreme edge. For example, in the calculations for the finding insignificant occasional item sets, i.e., rare item sets that don't hold any occasional subset have been proposed. An occasional item set finding is material to information hailing from the distinctive genuine provision connections, for example:

(I) Measurable divulgence hazard an evaluation from registration information and

(II) Misrepresentation recognition. In any case, customary sanguinary item set mining calculations still experience the ill effects of their powerlessness to consider nearby the things interestingness throughout the mining stage. Indeed, from one point of view, item set quality measures which are utilized to derive the regular weighted item set mining methodology are not specifically appropriate to fulfill the occasional weighted item set mining. The state-

of-the-craftsmanship rare item set mine workers are unable to adapt the weights information. Event weights are inferred from the weights connected to every transaction by applying a given charge capacity. Specifically, according to our consideration there are two diverse IWI help measures:

(a) The IWI-help min measure, which depends on a base expense capacity, i.e., the event of an item set in a given transaction is weighted by the weight of its slightest interesting thing.

(b) The IWI-help max Measure, which depends on a greatest expense capacity, i.e., an event of an item set in a given transaction is weighted by the weight of the most intriguing thing.

Data mining is known for discovering earlier, suitable, original, useful and reasonable patterns in large databases. Due to the availability of vast amount of data and the need to transform such data into useful information and knowledge, data mining has grown to be the most generally used technique in the society as a whole. Thus data mining can be used for an applications ranging from the market analysis, fraud detection and customer retention, to production control and science exploration [8]. One of the data mining technique termed as an item set mining is an exploratory data mining strategy is broadly used for the generating precious correlation between the data. The first effort for performing item set mining was focused on generating the frequent item sets i.e. patterns whose observed frequency of incidence in the source data is more than the given threshold.

## 2. LITERATURE SURVEY

The major motivation to this work is done by Vaidya Seema Bhagwan, A. B. Bagwan. The authors suggest the connection of frequently holding in data in which an items may weight differently represented frequented weighted itemsets. Though, in some situation, for example when there is necessitating diminishing a certain cost function, generating rare data correlation is more motivating than the mining frequent one. Here in this paper address the topic of generating sanguinary and weighted itemsets, i.e. an infrequent weighted itemset mining problem. The two new excellence measures are proposed for solving an infrequent weighted itemset mining problem [2].

The major motivation to this work is done by D.J. Haglin and A.M. Manning. The authors suggest a new algorithm for minimal infrequent itemset mining. Potential applications of finding infrequent itemsets include statistical disclosure the risk assessment, bioinformatics, and fraud detection. This is the first algorithm designed specifically for finding these sanguinary itemsets. Many itemset properties used implicitly in the algorithm are proved. The problem is shown to be NP-complete [3].

The major motivation to this work is done by F. Tao, F. Murtagh, and M. Farid. The authors suggest the issues of discovering significant binary relationships in the transaction datasets in a weighted setting. Traditional model of association rule mining is suitable to handle weighted association rule mining problems where each item is allowed to have a weight. The goal is to the mining focus to those significant relationships involving items with significant weights rather than being overfill in the combinatorial explosion of insignificant relationships. Author identify the challenge of using weights in an iterative process of generating large itemsets. The problem of invalidation of the downward closure property in the weighted setting is solved by using the improved model of weighted support measurements and exploiting a weighted downward closure property. The new algorithm called WARM (Weighted Association Rule Mining) is developed based on the improved model [4].

The major motivation to this work is done by G. Cong, A.K.H. Tung, X. Xu, F. Pan, and J. Yang. The authors suggest microarray datasets commonly contain vast number of columns but small number of rows. Association rules have been proved to be useful in an analyzing such datasets. However, most existing association rule mining algorithms are unable to efficiently handle the datasets with vast number of the columns. Moreover, the number of association rules generated from such datasets is enormous due to the large number of the possible column combinations. In this paper, author describe a new algorithm called FARMER that is specially designed to the discover association rules from microarray datasets. Instead of finding individual association rules, FARMER finds an interesting rule groups which are essentially a set of rules that are generated from the same set of rows. Unlike the conventional rule mining algorithms, FARMER searches for interesting rules in the row enumeration space and an exploits all user-specified constraints including minimum support, confidence and chi-square to support efficient pruning [5].

The major motivation to this work is done by A. Gupta, A. Mittal, and A. Bhattacharya. The authors suggest itemset mining has been an active area of the research due to its successful application in various data mining scenarios including finding an association rules. Though most of the past work has been on finding the frequent itemsets, infrequent itemset mining has demonstrated its utility in the web mining, bioinformatics and other fields. In this paper, author propose a new algorithm based on pattern-growth paradigm to find the minimally infrequent itemsets. A minimally infrequent itemset has no subset which is also an infrequent. also introduce the novel concept of residual trees. further utilize the residual trees to mine the multiple level minimum support itemsets where different thresholds are used for finding frequent itemsets for different lengths of the itemset [6].

The major motivation to this work is done by M. Hamsathvani, D. Rajeswari, R. Kalaiselvi. The authors suggest a new algorithm for an infrequent item set mining, finding infrequent weighted item set in transaction database. Frequent weighted item sets represent the correlation regularly holding in data in which items may weight differently. The research society has focused on an infrequent weighted item set mining problem. Infrequent weighted item set discover item sets whose frequency of an occurrence in the analyzed data is less than or equal to a maximum threshold. To discover infrequent weighted item set, two algorithms are discovered an Infrequent weighted item set IWI and Minimal infrequent item set MIWI. In this survey is focused on an infrequent weighted item sets, from transactional weighted data sets to address IWI support measure is defined as a weighted frequency of an occurrence of an item set in the analyzed data [8].

### 3. NEED OF WORK

- In the traditional association rule mining used to identify frequently occurring patterns item set.
- The frequent item set mining approach may not satisfy sales manager’s goal.
- The supports, measures and reflects the statistical correlation of an item. But it does not reflect their semantic significance. In other words, statistical correlation may not measure how useful an item set is in accordance with user’s preferences (i.e. profit).
- The practical usefulness of the frequent item set mining is limited by significance of the discovered item set. There are two principal limitations. A huge number of frequent item set that are not interesting to user are often generated when the minimum support is low. For example, thousands of combinations of the products that occur in 1% truncations. If too many uninteresting frequent item sets are found, the user is forced to an additional work to select the item sets that are indeed interesting.

### 4. OBJECTIVE OF THIS WORK

The objective of our work is to provide a solution to the issue of discovering rare and weighted itemsets, i.e., an infrequent weighted itemset mining problem. There are two novel quality measures are proposed to drive the IWI mining process. Two algorithms that perform IWI and the Minimal IWI mining efficiently, driven by the proposed measures, are presented. We address the support value is calculated with max-min normalization, which makes the proposed methodology is more independent of any parameter value. The results show efficiency and effectiveness of the proposed approach.

### 5. MODEL DESCRIPTION

Figure shows the system architecture of our proposed system. As shown in system architecture, 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 finding 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 these common types are taken as an output.

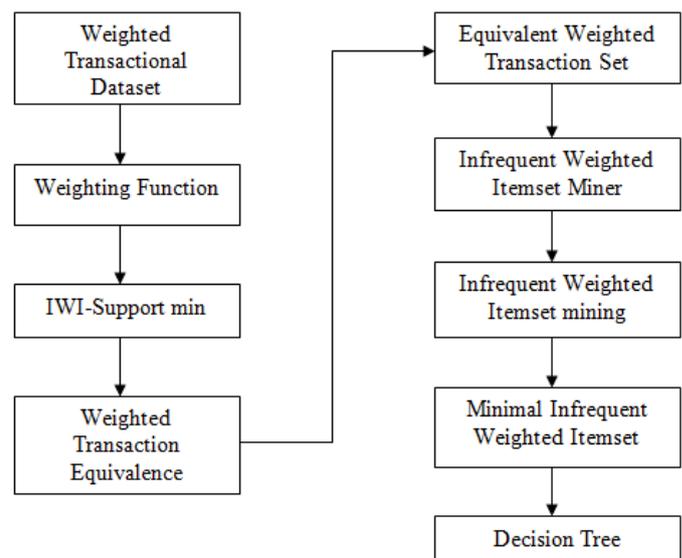


Fig: Flowchart of Infrequent Weighted Itemset Mining Using Frequent Pattern Growth and Decision Tree

## Uploading Dataset

In this approach we have to find infrequent weighted items in transactions. and improve the efficiency identification infrequent weighted items in transactions.

## Item Set Mining

An itemset mining is an exploratory data mining technique widely used for discovering valuable correlations among data. The first attempt to perform an item set mining was focused on discovering frequent item sets, i.e. patterns whose observed frequency of an occurrence in the source data (the support) is above a given threshold. Frequent Item sets find application in a number of the real-life contexts (e.g. market basket analysis, medical image processing, biological data analysis).

However, the many traditional approaches ignore an influence/interest of each item/transaction within the analyzed data. To allow treating an items/transactions differently based on their relevance in the frequent item set mining process, the notion of the weighted item set has also been introduced. A weight is an associated with each data item and the characterizes its local significance within each transaction.

## Weighted Transaction Equivalence

The weighted transaction equivalence is establishes 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 exclusively composed of an equally weighted items. To this aim, each weighted transaction  $t_q$  is belongs to  $T$  corresponds to an equivalent weighted transaction set  $TE_q$  is subset of  $TE$ , which is a subset of  $TE$ 's transactions. Item weights in  $t_q$  are spread, based on an irrelative significance, among their equivalent transactions in  $TE_q$ . The proposed transformation is particularly suitable for the compactly representing the original dataset.

## Mining Infrequent Weighted Itemset

A weighted transactional data set and the maximum IWI-support (IWI-support-min or IWI-support-max) threshold, the Infrequent Weighted Item set Miner algorithm extracts all the IWIs whose IWI support value satisfies threshold. Since the IWI Miner mining steps are same by enforcing either IWI-support-min or IWI-support-max thresholds. IWI Miner is a FP-growth-like mining algorithm that performs the projection-based item set mining. Hence, it performs the main FP-growth mining steps:

- FP-tree creation.
- recursive item set mining from the FP tree index.

Unlike FP-Growth, IWI Miner discovers infrequent weighted item sets instead of the frequent (unweighted) ones. To

accomplish this task, the following main modifications with respect to the FP-growth have been introduced:

- A novel pruning strategy for pruning part of the search space early.
- A slightly modified FP tree structure, which is allows storing the IWI-support value associated with each node.

## Algorithm

Input- weighted transaction dataset and support value)

IWI ( $T, E$ )

- 1)  $F=0$
- 2) Count item IWI ( $T$ )
- 3) Construct FP tree
- 4) For all weighted transaction
- 5) Calculate Equivalent transaction
- 6) For all transaction create and insert into FP tree

Output:

Set of satisfying  $E$

## Mining in Minimal Infrequent Weighted Itemset

A weighted transactional data set and the maximum IWI-support (IWI-support-min or IWI-support-max) threshold, the Minimal Infrequent Weighted Item set Miner algorithm extracted all the MIWIs that satisfy threshold. the pseudo code of the MIWI Miner algorithm is identical to the one of IWI Miner. Hence, due to space constraints, the pseudo code is not reported. However, in following procedure, the main differences with respect to the IWI Miner are outlined.

The MIWI Mining procedure is identical to IWI Mining. However, since MIWI Miner focuses on generating only minimal infrequent patterns, the recursive extraction in the MIWI (minimal infrequent weighted itemset) Mining procedure is stopped as soon as the infrequent item set occurs. In fact, whenever an infrequent item set  $I$  is exposed, all its extensions are not minimal. It finds both an infrequent item sets and minimal infrequent item set mining.

**Algorithm**

IWI mining (T, E, P)

- 1) F=0 initialization
- 2) Create header table holds for all items i in tree
- 3) Generate a new item set I with prefix and support of item i
- 4) I: Infrequent item
- 5) Construct I as conditional pattern and FP tree
- 6) Select the infrequent items from the set
- 7) Remove from Tree and finally apply recursive mining

**Decision Tree**

It contributes the techniques that may concentrate on decision tree, which show the combination of different products, and also show the minimum support transactions.

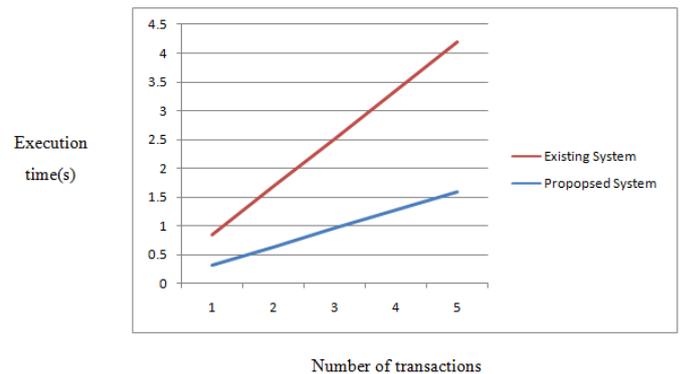
**Steps**

- 1) Suggest the combined solution.
- 2) Combination of different products.
- 3) Minimum support transaction.

**6. ANALYSIS OF THE RESULT**

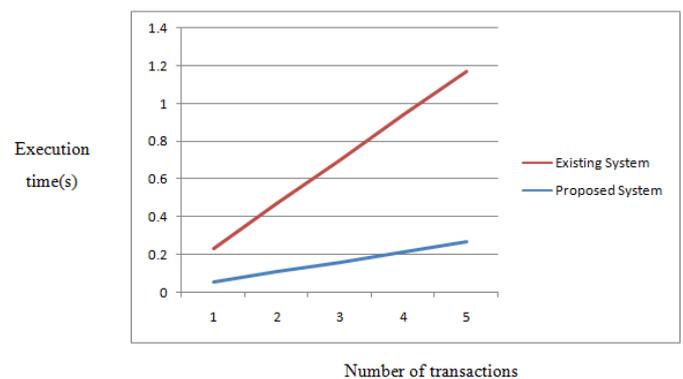
The program is implemented in JAVA. The experiment is carried out using NetBeans IDE 8.0 on a single machine with windows 7 operating system. The infrequent items drawn from the dataset is dependent of the support values. As the frequent item sets negates, the infrequent items, resulting in increase of support value also increase the infrequent transaction count. We analyzed the algorithm scalability, in terms of execution time on the dataset. To test the algorithm scalability with the number of dataset transactions (i.e. the dataset cardinality). We generated dataset of size ranging from 0 to 10000 transactions.

Figure a and b respectively report IWI miner and MIWI miner execution times by varying the dataset cardinality. In general, IWI miner execution time significantly increases for higher IWI-support-min threshold values due to the combinatorial growth of the number of extracted patterns.



**Chart -1:** MIWI miner with IWI-support-max

Figure b shows the result of MIWI miner execution times by varying the dataset cardinality. MIWI miner execution times by varying the transaction length. Both IWI miner and MIWI miner execution time scale roughly linearly with the dataset size for all the tested settings.



**Chart -2:** MIWI miner with IWI-support-min

**7. 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.

**8. 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 infrequent weighted itemset mining using frequent pattern growth that supports at the time variable datasets instead of one dataset. Furthermore, the application of different an aggregation function besides minimum and maximum will be studied.

## ACKNOWLEDGMENT

I would like to express my deep sense of gratitude towards my guide Prof. A. S. Tamboli and Prof. A. S. Dange, Assistant Professor, Annasaheb Dange College of Engineering & Technology, Ashta, for his invaluable help and guidance for the project.

## REFERENCES

- [1] Luca Cagliero and Paolo Garza, "Infrequent Weighted Itemset Mining Using Frequent Pattern Growth," IEEE TRANSACTION ON KNOWLEDGE AND DATA ENGINEERING, VOL 26,NO.4,APRIL 2014.
- [2] Vaidya Seema Bhagwan, A. B. Bagwan "A Survey: By Using Frequent Pattern Growth Infrequent Weighted Item set Mining," International Journal of Science and Research (IJSR) Volume 3 Issue 10, October 2014, ISSN (Online): 2319-7064pp 621-623.
- [3] D.J. Haglin and A.M. Manning, "On Minimal Infrequent Item set Mining," Proc. Intl Conf. Data Mining (DMIN 07), Pp. 141-147, 2007.
- [4] F. Tao, F. Murtagh, and M. Farid, "Weighted Association Rule Mining Using Weighted Support and Significance Framework," Proc. ninth ACM SIGKDD Intl Conf. Knowledge Discovery and Data Mining (KDD 03), Pp. 661-666, 2003.
- [5] G. Cong, A.K.H. Tung, X. Xu, F. Pan, and J. Yang, "Farmer: Finding Interesting Rule Groups in Microarray Data sets," Proc. ACM SIGMOD Intl Conf. Management of Data (SIGMOD 04), 2004.
- [6] A. Gupta, A. Mittal, and A. Bhattacharya, "Minimally Infrequent Itemset Mining Using Pattern-Growth Paradigm and Residual Trees," Proc. Intl Conf. Management of Data (COMAD), pp. 57-68, 2011.
- [7] R. Agrawal, T. Imielinski, and Swami, "Mining Association Rules between Sets of Items in Large Databases," Proc. ACM SIGMOD Intl Conf. Management of Data (SIGMOD 93), pp. 207-216, 1993.
- [8] M. Hamsathvani, D. Rajeswari, R. Kalaiselvi "Survey on Infrequent Weighted Item set Mining Using FP Growth," International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering, Vol. 3, Issue 11, November 2014, ISSN (Print): 2320 3765, ISSN (Online): 2278 8875, Pp: 13128-13131.