INFREQUENT WEIGHTED ITEM SET MINING USING FREQUENT PATTERN GROWTH

R. Lakshmi Prasanna*1, Dr. G.V.S.N.R.V. Prasad2

1M.Tech Scholar, Dept. of CSE, Gudlavalleru Engineering College, Gudlavalleru, Krishna (A.P), India.
2Prof. & Dean, Dept. of CSE, Gudlavalleru Engineering College, Gudlavalleru, Krishna (A.P), India.

ABSTRACT

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. Infrequent item set mining finds rarely occurring item sets in the database. Most of the existing infrequent item set mining techniques find infrequent weighted item sets with high computing time and are less scalable when the database size increases. The proposed system uses clustering or logical grouping concepts for finding infrequent weighted item sets. The proposed algorithm works well with real-time databases and is highly scalable which suits for real-time applications. Continuous weighted item sets to associate frequently holding information in which item sets may weight distinctively. This paper handles the issue of running across extraordinary and weighted item sets, i.e., Infrequent Weighted Item set (IWI) mining. Two novel quality measures are proposed to test the IWI mining procedure. The two calculations that perform IWI and Negligible IWI mining efficiently, determined by the proposed measures, are displayed. Test outcomes show efficiency and adequacy of the proposed methodology.

Keywords: Infrequent weighted itemset, Frequent pattern growth, Data Mining, Frequent pattern Mining, Weighted mining.

1. INTRODUCTION

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 genuine connections (e.g., market wicker container dissection, medicinal picture handling, and biotic information investigation). In any case, numerous customary methodologies overlook the impact/ enthusiasm of everything / transaction inside the investigated information. To permit 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 the Exploration group has 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 calculations for finding insignificant occasional item sets, i.e., rare item sets that don't hold any occasional subset have been proposed. Occasional item set finding is material to information hailing from distinctive genuine provision connections, for example:

(I) Measurable divulgence hazard evaluation from registration information and (II) misrepresentation recognition. In any case, customary rare item set mining calculations still experience the ill effects of their powerlessness to consider nearby thing 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 expense 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 fascinating thing.

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

2. PROBLEM DEFINITION

In association rule mining, each item set has certain occurrence frequency which could be termed as the weight of the item set. The weight could be positive, negative or null. Mining of such weighted transactional datasets for finding frequent patterns are called weighted item sets. In state of art of the infrequent item set mining algorithms, the ability of taking the small frequent item set into consideration is negligible. Finding the frequent item set can be negated for the entire transactional dataset which would result in finding the infrequent patterns. The support measure in most of the algorithms treats the item set equally even though they don’t have the same relevance in the dataset.

3. LITERATURE REVIEW

- In (2011), Weimin Ouyang et al. [1] illustrated three drawbacks of conventional algorithms such as quantitative databases are not considered in some cases. The detection of the frequent item sets is based on the minimum support that falsely accumulates the frequency of items to be same. Finally some algorithms perform mining where simple rules are obtained. To overcome these limitations, rule mining is done by assigning fuzzy value to the data items that have multiple support values.

- In (2012), Yihua Zhong et al. [2] introduced an algorithm based on the concept of weighted dual confidence for extracting efficient weighted rules in the database because the traditional association rule approaches are based on the support and confidence metrics with attributes considering an equal weight, resulting in ineffective rules which are not suitable for taking decision making. This method extracts interesting negative association rules from the database.

- In (2013), Johannes K. Chiang et al. [3] underlined some drawbacks in conventional mining techniques such as these can perform the mining process based on a predefined schemata, therefore scanning is required for addition of new attributes. Since the rule mining can be properly decided based on certain level, they are designed for extracting either frequent or infrequent rules. The author overcomes these limitations by suggesting the concept which is used as a data structure for representing associations of the data in the database.

- In (2014), Shipra Khare and Prof. Vivek Jain [4] proposed a mining of infrequent weighted item set which provides less amount of computational time. This approach plays a significant role in decision-making. To discard the infrequent rules, the support and confidence threshold values are necessary parameters that are required to the mining algorithm to avoid generating misleading association rules. Therefore, this algorithm is useful for discovering minimum number of items based on support and confidence values obtains interesting weighted negative association rules from the database. This approach extracts minimum number of frequent item sets for analyzing large amount of data. However the resultant negative association rules from the original transactional data cannot be recovered.

4. EXISTING SYSTEM

Item set mining is an exploratory information digging procedure generally utilized for uncovering profitable connections among information. The primary endeavor to perform item set mining was centered on finding continuous thing sets. That is examples whose watched recurrence of event in the source information is over a given limit. To permit treating transactions contrastingly focused around their pertinence in the continuous item set mining process, the thought of weighted item set has likewise been presented. A weight is connected with every information thing and describes its nearby noteworthiness inside every transaction. The weaknesses has an incessant weighted thing set mining is not specifically material to achieve the rare weighted thing set. Time for doing the methodology is likewise high in view of the successive item sets.
Disadvantages: The practical usefulness of the frequent item set mining is limited by the significance of the discovered item set. There are two principle limitations. A huge number of frequent item set that are not interesting to the user are often generated when the minimum support is low. For example thousands of combinations of products that occur in 1% truncations. If too many uninteresting frequent item sets are found the user is forced to additional work to select the item sets that are indeed interesting.

5. PROPOSED SYSTEM

The finding of occasional and weighted item sets, i.e., the Infrequent Weighted Item sets from transactional weighted datasets. The IWI-help measure is characterized as a weighted recurrence of event of an item set in the broke down Information. The IWI-help min measure, which depends on a base expense Capacity. That is the event of an item set in a given transaction is weighted by the weight of its slightest fascinating thing. The IWI-help max measure, which depends on a greatest expense capacity. That is the event of an item set in a given transaction is weighted by the weight of the most intriguing thing. The favorable circumstances has an occasional thing set finding is pertinent to information originating from diverse genuine requisition settings, for example, Measurable divulgence hazard appraisal from enumeration information and misrepresentation location. Suitable for deriving the choice of an advantageous subset of rare weighted information relationships. Which means this proposed system is mining of rarely occurred item sets of weighted transactional dataset by using IWI miner and MIWI Miner.

5.1 Advantages of Proposed Methodology

In the existing approach the weights are assigned

With local interestingness whereas in proposed approach the weights are assigned by a fuzzy membership value having a sigmoid distribution.

- In existing method, the support value is a user-defined value and in proposed method the support is calculated with max-min normalization, which makes the proposed methodology more independent of any parameter value.
- In existing approach the FP tree mining is done which would take large space and time complexity. The tree compression is done to reduce the space and time complexity by merging the similar path for a data item.

5.2 System Architecture

![System Architecture Diagram](image-url)
6. ALGORITHMS

This section presents two algorithms, namely Infrequent Weighted Item set Miner and Minimal Infrequent Weighted Item set Miner and also use FP-Growth Algorithm and Apriori Algorithm.

6.1 The Infrequent Weighted Item set Miner Algorithm

IWI Miner is a FP-growth-like mining algorithm that performs projection-based item set mining. Hence, it performs the main FP-growth mining steps: (a) FP-tree creation and (b) recursive item set mining from the FP-tree index. Unlike FP-Growth, IWI Miner discovers infrequent weighted item sets instead of frequent (un weighted) ones. To accomplish this task, the following main modifications with respect to FP-growth have been introduced:

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

Algorithm (IWI Miner (T, E))

Input:
- T, a weighted transactional dataset
- E, a maximum IWI-support threshold
Output:
- F, the set of IWI satisfying E

Step1: Count the infrequent weighted item sets with the support value.
Step2: Create header table which is a data structure which holds information about total weight values.
Step3: For each transaction, create equivalent transaction.
Step4: Create an FP-Tree, for each transaction.
Step5: Iterate the process until all transactions are traced.
Step6: Create conditional pattern base calculate weight value.
Step7: Obtain the infrequent item sets.

To reduce the complexity of the mining process, IWI Miner adopts an FP-tree node pruning strategy to early discard items (nodes) that could never belong to any item set satisfying the IWI-support threshold. Hence, an item (i.e., its associated nodes) is pruned if it appears only in tree paths from the root to a leaf node characterized by IWI-support value greater than E.

6.2 The Minimal Infrequent Weighted Item set Miner Algorithm

Given a weighted transactional data set and a maximum IWI-support (IWI-support-min or IWI-support-max) threshold, the Minimal Infrequent Weighted Item set Miner algorithm extracts all the MIWIs that satisfy the IWI-support threshold. The pseudo code of the MIWI Miner algorithm is similar 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. At line 10 of Algorithm 1, the MIWI Mining procedure is invoked.

The MIWI Mining procedure is almost similar to IWI Mining. In IWI Mining infrequent patterns will be generated, where as in MIWI Mining Minimal Infrequent patterns will be generated. However, since MIWI Miner focuses on generating only minimal infrequent patterns, the recursive extraction in the MIWI Mining procedure is stopped as soon as an infrequent item set occurs (i.e., immediately after line 5 of Algorithm 2). In fact, whenever an infrequent item set I is discovered, all its extensions are not minimal.
7. IMPLEMENTATION

7.1 Weighted Transactional Dataset

The weighted transaction information set holds the transaction of the everything. The weight is then computed for everything in the transactional dataset. The weight implies the use of framework is considered as the weight. Utilizing this mining of the occasional thing is figure out. Let $I = \{i_1, i_2, \ldots, i_m\}$ be a set of things. A weighted transactional dataset $Tw$ is a situated of weighted Transactions, (where each one weighted transaction $tw_q$ is a situated of weighted things such that $i_k \in I$ and $wq$.)

7.2 Weighted Transaction Equivalence

Weighted transaction equivalence is utilized make acquaintanceship between weighted transaction dataset $T$ and an identical dataset $TE$. Each one weighted transaction $tq \in T$ compares to a proportional weighted transaction set. $T$ be a weighted transactional dataset and $TE$ its comparing identical dataset. $TE$ of a weighted transactional dataset $T$ is the union of all comparable transactional set.

8. PERFORMANCE ANALYSIS

All the experiments were performed on 2.65 GHz Intel core processor with 2 GB RAM running Windows 7. The algorithms were implemented in C programming. The Connect database having 67557 instances and 42 attributes in UCI Repository is considered. The infrequent item sets 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 increases the infrequent transaction count. Figure 2 shows the graphical representation of the support values and the infrequent transaction count. It shows the experimented count ratio performance for respective support value, proposed algorithm is showing efficiency in terms of Transaction Count when compared to existing. Figure 4 plots the execution time of the algorithm with the variation of the support values. It shows the Computation Time ratio performance for corresponding support value, proposed algorithm is showing efficiency in terms of Computation Time by taking less when compared to existing.

![Support vs Infrequent Transactions](image)

**Fig. 2: Infrequent Transactions**

We analyzed IW1 Miner and MIWI Miner performance on standard synthetic and real data sets. In particular, we analyzed:

(i) The impact of the equivalence procedure on the data set size
(ii) The impact of the IW1-support thresholds on both the number of mined patterns and the algorithm execution time and
(iii) The comparison in terms of execution time, between IW1 Miner and MIWI, a state-of-the-art of minimal infrequent (Unweighted) item set miner.

We also analyzed the algorithm scalability, in terms of execution time, on synthetic data sets. To test the algorithms capability with the number of data set transactions (i.e., the data set cardinality), we generated data
sets of size ranging from 0 to 1,000,000 transactions by using the proposed technique. This technique is taking less computation time and generating more Transaction count for the given dataset even if the size of data sets is having huge number of transactions compared to existing, which is represented as graphs in figure 4 and 5 for sample data set.

Fig. 3: Execution Time

9. CONCLUSION AND FUTUREWORK

This technique faces 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 IWI and MIWI mining efficiently are also proposed. The usefulness of the discovered patterns has been validated on data coming from a real-life context with the help of a domain expert. As future work, we would like 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 IWLs. Furthermore, the application of different aggregation functions besides minimum and maximum will be studied.

REFERENCES


