Journal of Advanced Engineering Research
ISSN: 2393-8447
Volume 2, Issue 2, 2015, pp.42-45

Implementation of miner algorithm for finding infrequent item set from frequent pattern

V. Kumararaja*, N. Shunmugakarpakam

Department of Computer Science and Engineering, Er. Perumal Manimekalai College of Engineering, Hosur, Tamil Nadu, India,

*Corresponding author email: mail2kumarcs@gmail.com

ABSTRACT

Pattern mining is important task in data mining process. Pattern mining focuses on identifying rules that define specific patterns within the data. Data mining is a process of discovering interesting knowledge from large amounts of data. Frequent item set mean a set of items that appear frequently together in a transaction data set is a frequent itemset. In this frequent item set mining is widely used in data mining technique. Frequently item set in which items may weight differently. The proposed work has been finding frequent item set retrieve the frequent data from database. The resultant data is the infrequent data retrieved all infrequently data from database. In this Paper provide two algorithms to get the infrequent data. Infrequent weighted item set miner (IWI) and Minimal infrequent weighted item set miner (MIWI). By applying these algorithms for mining infrequent patterns which are basis for future research in the field of pattern mining.

Keywords - Frequent Item sets, Infrequent Item sets, Patterns, Data Mining, Miner Algorithm and Pattern Mining.

1. INTRODUCTION

Frequent item set mining is a widely used data mining technique that has been introduced In the traditional item set mining problem items belonging to transactional data are treated equally. This approaches focuses on mining infrequent item sets from weighted data instead of frequent ones. Hence, different pruning techniques are exploited. To address this issue, probabilistic models have been constructed and integrated in priori-based or projection-based algorithms.

Discovery of infrequent and weighted item sets, i.e., the infrequent weighted item sets, from transactional weighted data sets. To address this issue, the IWI-support measure is defined as a weighted frequency of occurrence of an item set in the analyzed data. Occurrence weights are derived from the weights associated with items in each transaction by applying a given cost function.

Two different IWI-support measures are defined. The IWI-support- min measure, which relies on a minimum cost function, i.e., the occurrence of an item set in a given transaction is weighted by the weight of its least interesting item, The IWI-support-max measure, which relies on a maximum cost function, i.e., the occurrence of an item set in a given transaction is weighted by the weight of the most interesting item. 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.

To accomplish tasks two novel algorithms, namely Infrequent Weighted Item set Miner (IWI Miner) and Minimal Infrequent Weighted Item set Miner (MIWI Miner), which performs IWI and MIWI mining driven by IWI-support thresholds.

IWI Miner and MIWI Miner are FP-Growth-like mining algorithms, whose main features may be summarized as follows:

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 cost function-independence, i.e., they work in the same way regardless of which constraint (either IWI-support-min or IWI-support-max) is applied, early stopping of the recursive FP-tree search in MIWI Miner to avoid extracting non-minimal IWIs. 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.

2. PREVIOUS WORKS

Frequent item set mining is a widely used data mining technique that has been introduced in. In the traditional item set mining problem items belonging to transactional data are treated equally. To allow differentiating items based on their interest or intensity within each transaction, in the authors focus on discovering more informative association rules, i.e., the weighted association rules (WAR), which include weights denoting item significance.

It addresses the discovery of infrequent and weighted item sets, i.e., the infrequent weighted item sets, from transactional weighted data sets. To address this issue, the IWI-support measure is defined as a weighted frequency of occurrence of an item set in the analyzed data. Occurrence weights are derived from the weights associated with items in each transaction by applying a given cost function.

Two different IWI-support measures are illustrated. The IWI-support- min measure, which relies on a minimum cost function, i.e., the occurrence of an item set in a given transaction is weighted by the weight of its least interesting item, The IWI-support-max measure, which relies on a maximum cost function, i.e., the occurrence of an item set 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. 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:

Two novel algorithms, namely Infrequent Weighted Item set Miner (IWI Miner) and Minimal Infrequent Weighted Item set Miner (MIWI Miner), which performs IWI and MIWI mining driven by IWI-support thresholds. IWI Miner and MIWI Miner are FP-Growth-like mining algorithms, whose main features may be summarized as follows: Early FP-tree node pruning driven by the maximum IWI-support constraint.

The early discarding of part of the search space thanks to a novel item pruning strategy, and cost function-independence, i.e., they work in the same way regardless of which constraint (either IWI-support-min or IWI-support-max) is applied, early stopping of the recursive FP-tree search in MIWI Miner to avoid extracting non-minimal IWIs.

3. SYSTEM ARCHITECTURE

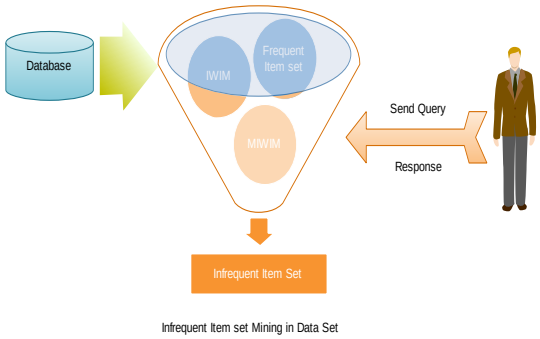

Fig 1. Retrieval of Infrequent Data

4. PROBLEM STATEMENTS

Discovery of infrequent and weighted item sets, i.e., the infrequent weighted item sets, from transactional weighted data sets. To address this issue, the IWI-support measure is defined as a weighted frequency of occurrence of an item set in the analyzed data. Occurrence weights are derived from the weights associated with items in each transaction by applying a given cost function.

The IWI-support- min measure, which relies on a minimum cost function, i.e., the occurrence of an item set in a given transaction is weighted by the weight of its least interesting item, The IWI-support-max measure, which relies on a maximum cost function, i.e., the occurrence of an item set 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.

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: To accomplish tasks two novel algorithms, namely Infrequent Weighted Item set Miner (IWI Miner) and Minimal Infrequent Weighted Item set Miner (MIWI Miner), which performs IWI and MIWI mining driven by IWI-support thresholds.

IWI Miner and MIWI Miner are FP-Growth-like mining algorithms, whose main features may be summarized as follows: 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 cost function-independence, i.e., work in the same way regardless of which constraint (either IWI-support-min or IWI-

support-max) is applied, early stopping of the recursive FP-tree search in MIWI Miner to avoid extracting non-minimal IWIs.

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. Two algorithms to get the infrequent data they are infrequent weighted item set miner and minimal infrequent weighted item set miner. By applying these algorithms the infrequent data is retrieved. The issue of discovering infrequent item sets by using weights for differentiating between relevant items and not within each transaction. As proposed it is integrated in advanced decision making systems that supports domain expert's targeted action based on the characteristics of discovered IWI's. The application of different aggregation functions besides maximum and minimum will be studied.

5. ALGORITHMS FOR RETRIEVING THE INFREQUENT DATA FROM THE DATABASE

5.1 The 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 Infrequent Weighted Item set Miner algorithm extracts all IWIs who's IWI-support satisfies. Since the IWI Miner mining steps are the same by enforcing either IWI-support-min or IWI-support-max thresholds, it will not distinguish between the two IWI support measure types in the rest of this section.

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 (unweighted) 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.

5.2 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 pseudo code of the MIWI Miner algorithm is similar to the one of IWI Miner, reported in Algorithm.

The main differences with respect to IWI Miner are outlined. , the MIWI Mining procedure is invoked instead of IWI Mining. The MIWI Mining procedure is similar to IWI Mining. However, since MIWI Miner focuses on generating only minimal infrequent patterns, the recursive extraction in the MIWI Mining procedure is stopped.

6. IMPLEMENTATION AND RESULT

The work addresses the discovery of infrequent and weighted item sets, i.e., the infrequent weighted item sets, from transactional weighted data sets. To address this issue, the IWI-support measure is defined as a weighted frequency of occurrence of an item set in the analyzed data. Occurrence weights are derived from the weights associated with items in each transaction by applying a given cost function.

The traditional techniques were unable to mine the infrequent item set. Rather it retrieves the frequent items from the data base. The first attempt to pushing item weights into the item set mining process has been done in. It proposes to exploit the anti-mono city of the proposed weighted support constraint to drive the A priori-based item set mining phase. However, in weights have to be pre assigned, while, in many real-life cases, this might not be the case might not useful to real world scenario.

5. CONCLUSION

The work 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, the plan is 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.

REFERENCE

[1] Gupta, A. Mittal, and A. Bhattacharya, Minimally Infrequent Itemset Mining Using Pattern-Growth Paradigm and Residual Trees, *Proc. International Conf. Management of Data (COMAD)*, 2011, 57-68.

[2] IBM Quest Synthetic Data Generation Code, http:// www.almaden.ibm.com/, 2009.

[3] T. Bernecker, H.-P. Kriegel, M. Renz, F. Verhein, and A. Zuefle, Probabilistic Frequent Itemset Mining in Uncertain Databases, *Proc. 15th ACM SIGKDD Int'l Conf. Knowledge Discovery and Data Mining (KDD '09)*, 2009, 119-128.

[4] A.M. Manning, D.J. Haglin, and J.A. Keane, A Recursive Search Algorithm for Statistical Disclosure Assessment, *Data Mining and Knowledge Discovery*, 16 (2), 2008, 165-196.

[5] K. Sun and F. Bai, Mining Weighted Association Rules Without Preassigned Weights, *IEEE Trans. Knowledge and Data Eng*., 20 (4), 2008, 489-495.

[6] D.J. Haglin and A.M. Manning, On Minimal Infrequent Itemset Mining, *Proc. International Conf. Data Mining (DMIN '07)*, 2007, 141-147.

[7] C.K.S. Leung, C.L. Carmichael, and B. Hao, Efficient Mining of Frequent Patterns from Uncertain Data, *Proc. Seventh IEEE International Conf. Data Mining Workshops (ICDMW '07),* 2007, 489-494.

[8] D.J. Haglin and A.M. Manning, On Minimal Infrequent Itemset Mining, *Proc. International Conf. Data Mining (DMIN '07),* 2007, 141-147.

[9] Manning and D. Haglin, A New Algorithm for Finding Mini-mal Sample Uniques for Use in Statistical Disclosure Asses-sment, *Proc. IEEE Fifth International Conf. Data Mining (ICDM '05)*, 2005, 290-297.