International Journal of Scientific & Engineering Research Volume 8, Issue 5, May-2017
57
ISSN 2229-5518

INFREQUENT WEIGHTED ITEM SET MINING USING FREQUENT PATTERN GROWTH

K.Siva Ramakrishnan, Assistant Professor,
K.Brindha, N.R.Deepika, P.Gokila, UG Scholar
Department of Computer Science and Engineering
KPRIET, Coimbatore.
deepikanr29@gmail.com,brindhakannaksb@gmail.com,gokilapv3@gmail.com

Abstract

Frequent weighted item sets represent correlations frequently holding in data in which items may weight differently. However, in some contexts, e.g., when the need is to minimize a certain cost function, discovering rare data correlations is more interesting than mining frequent ones. This paper tackles the issue of discovering rare and weighted item sets, i.e., the infrequent weighted item set (IWI) mining problem. Two novel quality measures are proposed to drive the IWI mining process. Furthermore, two algorithms that perform IWI and Minimal IWI mining efficiently, driven by the proposed measures, are presented. Experimental results show efficiency and effectiveness of the proposed approach.

EXISTING SYSTEM

1. In traditional association rule mining used to identify frequently occurring patterns item set.
2. ARM model treats all the items in the database equally by only considering ,if an item is present in transaction or not.
3. The frequent item set mining approach may not satisfy sales managers goal.
4. The supports measures and reflects the statistical correlation of items. But it does not reflect their semantic significance .In other words statistical correlation may not measure how useful an item set is an accordance with users preferences(i.e. profit).

Dis-Advantages

The practical usefulness of the frequent item set mining is limited by the significance of the discovered item set. Their 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 two many uninteresting frequent item sets are found the user is forced to additional work to select the item sets that are indeed interesting.

PROPOSED SYSTEMS

1. Two novel qualities measures proposed to drive the IWI mining process. Infrequent item sets that do not contain any infrequent subsets have been proposed.
2. Experiment's performed on both synthetic and real life data sets ,show efficiency and effectiveness of the proposed approach .
3. In particular they show the characteristics and usefulness of the item sets discovered from data coming from benchmarking and real. To reduce the computational time the authors introduce the residual trees.
4. The item sets that are both high frequent and high utility can be obtained using the method.
5. The customers relationship management is incorporated into the system by tracking the customers who are frequent buyer's of the different kinds of item set.

Advantages

In proposed customers relationship management is incorporated into the system by tracking the customers who are frequent buyers of the different kinds of item set. So we are spitted the frequent utility frequent mining.

PROJECT DESCRIPTION

This paper faces the issue of discovering infrequent item sets by using weights for differentiating between relevant items

Aim & Objective

- To find **infrequent weighted items** in transactions.
- To improve the efficiency identification infrequent weighted items in **transactions Contributions**

The proposed work contributes the Unique that may concentrates on decision-making system that supports domain expert's targeted actions based to the characteristics of the discovered IWIs.

Implementation Modules:

1. Item Set Mining
2. Weighted Transaction Equivalence
3. The Infrequent Weighted Item set Miner Algorithm
4. The Minimal Infrequent Weighted Item set Miner Algorithm

Two FPGrowth- like algorithms that accomplish IWI (Infrequent Weighted Itemset Miner) and MIWI(Minimal Infrequent Weighted Itemset Miner Algorithm) mining efficiently are also proposed.

Item Set Mining

ITEMSET mining is an exploratory data mining technique widely used for discovering valuable correlations among data. The first attempt to perform item set mining was focused on discovering frequent item sets, i.e., patterns whose observed frequency of occurrence in the source data (the support) is above a given threshold. Frequent Item sets find application in a number of real-life contexts (e.g., market basket analysis , medical image processing , biological data analysis . 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 item set mining process, the notion of weighted item set has also been introduced . A weight is associated with each data item and characterizes its local significance within each transaction.

Weighted Transaction Equivalence

The weighted transaction equivalence establishes an association between a weighted transaction data set T, composed of transactions with arbitrarily weighted items within each transaction , and an equivalent data set TE in which each transaction is exclusively composed of equally weighted items. To this aim, each weighted transaction tq 2 T corresponds to an equivalent weighted transaction set, which is a subset of TE's transactions. Item weights in tq are spread, based on the irrelative significance, among their equivalent transactions in TEq. The proposed transformation is particularly suitable for compactly representing the original data.

The Infrequent Weighted Item set Miner Algorithm

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 whose IWI-support satisfies . Since the IWI Miner mining steps are the same by enforcing either IWI-support-min or IWI-support-max thresholds, we 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 (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.

The Minimal Infrequent Weighted Item set Miner Algorithm:

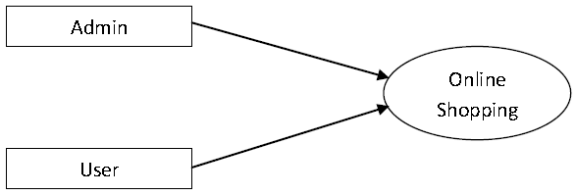
A weighted transactional data set and a maximum IWI-support (IWI-support-min or IWI-support-max) threshold , the Minimal Infrequent Weighted Item set Miner algorithm extracts all the MIWIs that satisfy. The pseudo code of the MIWI Miner algorithm is similar to the one of IWI Miner, constraints, the pseudo code is not reported. However, in the following, the main reported in Algorithm 1.Hence, due to space differences with respect to IWI Miner are outlined. At line 10 of Algorithm 1, 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 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.

FLOW DIAGRAM

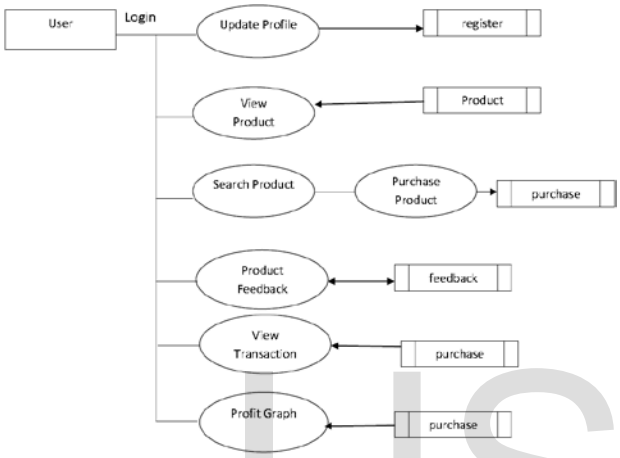
CONCLUSION AND FUTURE WORK

We experimentally demonstrate that our identifying local influencing regions technique is effective and the proposed method is mostly at least an order of magnitude faster than any other system. 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 plan 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. Next, we will apply our proposed System to the linear threshold model.

Data Flow Diagram:
Level 0:



Level 1:



REFERENCES

[1] Y. Wang, G. Cong, G. Song, and K. **Xie, "Community**-based greedy algorithm for mining top-k influential **nodes in mobile social networks," in** Proc. 16th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining, 2010, pp. 1039–1048

[2] W. Chen, C. Wang, and Y. Wang, **"Scalable influence maximization for** prevalent viral marketing in large-scale **social networks," in Proc. 16th ACM** SIGKDD Int. Conf. Knowl. Discovery Data Mining, 2010, pp. 1029–1038.

[3] Q. Jiang, G. Song, C. Gao, Y. Wang, W. Si, and K. Xie**, "Simulated annealing** based influence maximization in social **networks," in Proc. 25th AAAI Conf.** Artif. Intell., 2011, pp. 127–132.

[4] Goyal, F. Bonchi, and L. V. S. 4.**Lakshmanan, "A data**-based approach **to** social influence maximization," Proc. VLDB Endowment,vol. 5, no. 1, pp. 73– 84, Sep. 2011.

Web References

- www.jspin.com
- www.javaworld.com
- www.sun.java.com
- www.wrox.com