RESEARCH ARTICLE OPEN ACCESS

An Ontological Approach for Mining Association Rules from Transactional Dataset

Sivanthiya.T[1], G.Sumathi[2]
[1]ME-CSE (Final Year), Muthayammal Engineering College, Rasipuram
[2] Assistant professor, Muthayammal Engineering College, Rasipuram

Abstract
Infrequent item sets are mined in order to reduce the cost function and to make the sale of a rare data correlated item set. In the past research, algorithms like Infrequent Weighted Item Set Miner and Minimal Infrequent Weighted Item Set Miner were used. Since, mining of infrequent item set is done by satisfying support count less than or equal to the maximum support count many number of rules were generated and the mined result do not guarantee that only interesting rules were extracted, as the interestingness is strongly depends on the user knowledge and goals. Hence, an Ontology Relational Weights Measure using Weighted Association Rule Mining approach is introduced to integrate user's knowledge, minimize number of rules and mine the interesting infrequent item sets.

Index Terms- Ontology Relational Weights, Weighted Association Rule Mining, Infrequent Weighted Item set, Minimal Infrequent Weighted Item set.

I. INTRODUCTION

Association rule mining [1], is considered as an important task in Knowledge Discovery among Databases to retrieve either frequent item sets or infrequent item sets. It aims at discovering valuable information for the decision-maker. An association rule is described as X→Y where X and Y are the sets of items. To find the frequent item set Apriori [1] was the first algorithm proposed in the association rule mining field and following this algorithm many other algorithms were derived. However many traditional approaches ignore the interest of each item within the analyzed data. To allow treating items differently based on their importance in the frequent item set mining process, the thought of weighted item set has been introduced in [2], [3], [4]. A weight is associated with each data item and the importance of each transaction is characterized. The importance of a weighted transaction is evaluated in terms of the corresponding item weights. The item set quality measures have also been tailored to weighted data and are used for driving the frequent weighted item set mining process. In [2], [3], [4] different approaches were proposed to incorporate item weights, but they are all tailored for retrieving frequent item sets. Recently, the attention of the research community has turned to infrequent item set mining problem, i.e., discovery of item sets whose frequency of occurrence in the analyzed data is less than or equal to the maximum threshold. The quality measure used in [2], [3], [4] to drive the frequent weighted item set mining process are not directly applicable to mine infrequent weighted item set efficiently. So the discovery of infrequent weighted item set is done by two effective Miners like Infrequent Weighted Item Set Miner and Minimal Infrequent Weighted Item Set Miner algorithms. Though these algorithms retrieve the infrequent item sets and improve the quality of items, the integration among user knowledge and interestingness of the patterns are not considered. Moreover huge numbers of rules were produced. To overcome this drawback, along with the effective miners, Ontology Relational Weights concepts are introduced. It uses a Weighted Association Rule approach to prune and filter the discovered infrequent item sets. The classical model of association rule mining employs the support measure, which treats every transaction equally. In contrast, different transactions have different weights in real-life data set. For example, consider a transactional dataset with items and their corresponding weights. If item a has weight 0, item b has weight 100, item c has weight 57, item d has weight 71 in the same transaction then each item should be treated differently. The work introduces the measure called w-support with only binary attributes and the ranking of discovered infrequent item sets are obtained by HITS [5] model.

Table 1
Example of Weighted Transactional Data set

Tid	CPU usage readings
1	(a,0) (b,100) (c,57) (d,71)
2	(a.0) (b,43) (c,29) (d,71)
3	(a,43) (b,0) (c,43) (d,43)
4	(a,100) (b,0) (c,43) (d,100)
5	(a,86) (b,71) (c,0) (d,71)
6	(a,57) (b,71) (c,0) (d,71)

II. RELATED WORK

In the traditional item set mining approach the items belonging to transactional data are treated equally. To differentiate items based on their interest or intensity in [2] the authors focused on discovering more informative association rules. However, weights are introduced only during the rule generation step and were not tailored for infrequent item sets. The pushing of item weights into the item set mining process has been done in [3]. In [2], [3] weights have to be preassigned, in many real-life cases this is not possible. This issue is addressed in [4], where the analyzed transactional data set is represented as a bipartite hub-authority graph and evaluated by means of indexing strategy called HITS [5]. The survey differs from the above mentioned approaches because it mainly focuses on mining infrequent items instead of frequent ones using ORWM, i.e., Ontology Relational Weights Measure algorithm. The authors in [6] addressed the issue of discovering minimal infrequent item sets from the transactional data sets. Recently in [7] FP-Growth like algorithm for mining minimal infrequent item sets has also been proposed. The authors used the concept of residual tree to reduce the computational time. Similarly the proposed work mentioned in the survey also uses Frequent Pattern tree based approach with the modification in pruning strategy. This approach is used at the initial stage to mine the general infrequent item sets. The attempt of mining positive and negative association rules using infrequent item sets has made in [8], [9].

III. ONTOLOGIES IN DATA MINING

Ontology is next to taxonomy. In the early 1990s, Ontology was defined by Gruber as, A formal, explicit specification of a shared conceptualization [10]. By conceptualization the abstract model of some phenomenon is understood. The formal describes the idea that machine should be able to interpret an ontology. Explicit refers to the transparent definition ontology elements. Finally shared brings the outline from some knowledge common to a certain group, but not individual knowledge. In 2001, Ontology was viewed as a logical theory accounting for the intended meaning of a formal vocabulary. The main definition of Ontology is described as data schemas, providing a controlled vocabulary of concepts each with an explicitly defined and machine processable semantics. Depending upon the granularity there are four types of Ontologies (i) Upper Ontologies (ii) Domain Ontologies (iii) Task Ontologies (iv) Application Ontologies. Upper Ontologies are the top level ontologies that deal with general concepts and the rest of the ontologies deal with the domain specific concepts.

IV. ASSOCIATION RULE MINING FROM INFREQUENT ITEM SETS USING ONTOLOGY RELATIONAL WEIGHTS

Association rule mining is done after the finding of frequent item sets or infrequent item sets. The work is based on infrequent item sets hence, the infrequent item sets are mined based on Frequent Pattern Growth like algorithm, where the process follows projection based item set mining. The FP-growth like algorithm turns into a infrequent weighted item set mining algorithm when the modification with respect to Frequent pattern growth is made. The modification is done in the pruning strategy and allows to store the IWI support value associated with each other. An infrequent item set is said to be minimal if none of its subsets is infrequent [11]. Hence a minimal infrequent weighted item set mining algorithm is introduced. The pseudocode of the MIWI algorithm is similar to the one of IWI miner. The main difference with respect to IWI miner is, procedure invoked here is MIWI mining instead of IWI mining. An item set is infrequent if its support is less than or equal to a predefined maximum support threshold. The problem produced here is the volume of rules becomes high and the knowledge analysis is found to be difficult because the rules interestingness is strongly depends on user knowledge and goals. Hence the Ontology Relational Weights using Weighted Association Rule is introduced. The measure called w-support is used to measure the item sets with only binary attributes. The basic idea behind w-support is the infrequent item set derived may not be as important as it appears, because the weights of transactions are different. It works on the assumption that good transaction consist of good item sets. This assumption is based on HITS model. The main advantage of using w-support is, it can be worked out without much overhead and interesting infrequent patterns can be discovered. According to the traditional model, weights are described in a different manner. Ram et al introduced weighted support of association rules based on the costs assigned to items as well as transaction. Cai et al took only item weights into account, but downward closure property is broked. Tan et al provided another definition to retain the weighted downward closure property. Using this methodology the weights are assigned to items and new measures are introduced using these items. The weights introduced in the survey are based on the occurrences of each item in each transaction. In ORWM method the directed graph is created, where the nodes represent items and links denote association rules. A HITS model is applied to the graph to rank the items, all the nodes and links have weights associated with it. Under HITS model the Authority and Hub values are measured. Authority is the number of items within the transactions and the

authority values are computed as the sum of the scaled hub values. Hubs are the items relevant to the process of finding the authoritative items and the hub values are computed as the sum of scaled authority values.

V. THE ALGORITHMS

This section presents three algorithms, namely Infrequent Weighted Item Set Miner, Minimal Infrequent Weighted Item Set Miner, Ontology Relational Weights Measure. The proposed Algorithm is ORWM, i,e., Ontology Relational Weights Measure whose main characteristics is to integrate user knowledge, minimize number of rules and produce interesting infrequent item sets using the measure called w-support.

5.1 THE INFREQUENT WEIGHTED ITEM SET MINER ALGORITHM

In Item set mining considerably less attention has been remunerated to mine infrequent item sets, but it has acquired major usage in mining unconstructive association rules from infrequent item set, fraud detection, statistical disclosure risk assessment from census data, market basket analysis and bioinformatics. IWI Miner is a FP-Growth-like mining algorithm that performs projection-based item set mining. The processing steps are similar to FP-Growth algorithms such as (a) FP-tree creation and (b) recursive item set mining from FP-tree index. Unlike FP-Growth, IWI Miner discovers infrequent weighted item sets instead of frequent ones. So the modification with respect to FP-growth has been introduced. A novel pruning strategy and a slightly modified FP-tree structure allows to store the IWI-support value associated to each node.

5.2 THE MINIMAL INFREQUENT WEIGHTED ITEM SET MINER

The Minimal Infrequent Weighted Item set Miner performs IWI and MIWI mining driven by IWI-support threshold. These miners are FP-Growth like mining algorithm whose main feature is to prune maximum IWI-support constraint and to avoid extracting non-minimal IWIs. The mining procedure is stopped as soon as the infrequent item sets occurs.

5.3THE ONTOLOGY RELATIONAL WEIGHTS MEASURE

The first step of the Ontology Relational Weight algorithm is to retrieve the set of results to the searched query. There are two types of updates used, Authority Update Rule and Hub Update Rule. In order to calculate the hub/authority scores of each node, repeated iterations of the Authority Update Rule and Hub Update Rule are applied.

i Authority Update Rule, $\sum_{i=1}^{n} \text{hub(i)}$,

Where n is the total number of items. It updates each node's authority score to be equal to the sum of hub scores of each node.

ii Hub Update Rule, $\sum_{i=1}^{n} \text{auth(i)}$,

Where n is the total number of items. It updates each node's hub score to be equal to the sum of authority scores of each node.

VI. CONCLUSION

The survey discusses the problem of obtaining interesting infrequent item sets, integrating user knowledge and minimizing the number of rules produced. The ORWM helps to assign weights to the attributes and perform the process based on Weighted Association Rule Mining Approach and encounter the above mentioned problems efficiently.

REFERENCES

[1] R. Agrawal, T. Imielinski, and A. Swami, "*Mining Association Rules between Sets of Items in Large Datasets,"* Proc. ACM SIGMOD '93, pp. 207-216, 1993.

[2] Erwin.A, Gopalan.R.P, and Achuthan.N.R, 2008 "*Efficient Mining of High Utility Item sets from Large Data Sets*," Proc.12th Pacific-Asia Conf. Advances in Knowledge Discovery and Data Mining (PAKDD), pp. 554-561.

[3] F. Tao, F. Murtagh, and M. Farid, "*Weighted Association Rule Mining Using Weighted Support and Significance Framework*,"Proc. nineth ACM SIGKDD Int'l Conf. Knowledge Discovery and Data Mining (KDD '03), pp. 661-666, 2003.

[4] Sun.K and Bai.F, Apr. 2008 "Mining Weighted Association Rules W*ithout Preassigned Weights,"* IEEE Trans. Knowledge and Data Eng., vol. 20, no. 4, pp. 489-495.

[5] J.M. Kleinberg, "*Authoritative Sources in a Hyperlinked Environment*,"J. ACM, vol. 46, no. 5, pp. 604-632, 1999.

[6] Haglin.D.J and Manning.A.M, 2007 "*On Minimal Infrequent Item set Mining,"* Proc. Int'l Conf. Data Mining (DMIN '07), pp. 141-147.

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

[8] X. Wu, C. Zhang, and S. Zhang, "*Efficient Mining of Both Positive and Negative association Rules,"* ACM Trans. Information Systems, vol. 22, no. 3, pp. 381-405, 2004.

[9] X. Dong, Z. Zheng, Z. Niu, and Q. Jia, "*Mining Infrequent Itemsets Based on Multiple Level Minimum Supports,"* Proc. Second Int'l Conf. Innovative Computing, Information and Control (ICICIC'07), pp. 528-531, 2007.

[10] T.R. Gruber, *" A Translation Approach to Portable Ontology Specifications,"* Knowledge Acquistion,Vol.5, pp.199-220, 1993.

[11] Manning.A and Haglin.D, 2005 "*A New Algorithm for Finding Minimal Sample Uniques for Use in Statistical Disclosure Assessme*nt," Proc. IEEE Fifth Int'l Conf. Data Mining (ICDM '05), pp. 290-297.

[12] Peng Zhu and Fei Jia, Nov (2012)" *A new ontology based association rules mining algorithm*", journal of theoretical and applied information technology ISSN:1992-86452012. Vol. 45 No.1.