IJCSE International Journal of Computer Sciences and Engineering Open Access
Research Paper Volume-3, Issue-7 E-ISSN: 2347-2693

A Novel Low Utility Based Infrequent Weighted Itemset Mining Approach Using Frequent Pattern

Akilandeswari. S[1*] and A.V.Senthil Kumar[2]
[1*,2] *Department of Computer Application*, Bharathiyar University,Coimbatore, India
www.ijcseonline.org

Received: Jun/23/2015 Revised: July/06/2015 Accepted: July/21/2015 Published: July/30/ 2015

***Abstract*—** Item set mining is one of the famous data mining methods in which frequent and infrequent items can be mined. Now a days, the research society has focused on the problem of infrequent itemset mining, i.e., find out the item sets which has frequency of occurrence in transactional data base is less than or equal to a maximum threshold. Discovering rare item set is more interesting than mining frequent ones. The existing system deal with the issue of discovering infrequent weighted item sets. Infrequent Weighted Itemset Miner (IWI Miner) and Minimal Infrequent Weighted Itemset Miner (MIWI Miner) algorithms are introduced for efficient IWI and Minimal IWI mining. In many real world situations, utility of item sets depends on user's perspective such as cost, profit or revenue which are major significance. The existing Infrequent weighted item set mining algorithms are used to find out infrequent item sets from weighted transactional database, it does not compute utility of items. So in the proposed system introduced low utility based Infrequent Weighted Itemset mining (LUIWIM) algorithm. The proposed system is used for effectively mine the low utility infrequent weighted item set according to the profit, sale, etc. of items and it can improve the performance of the system compared to the existing system.

***Keywords*—** Data mining, infrequent item set, utility item.

I. INTRODUCTION

Data mining is the technique of investigating data from various viewpoints and summarizing it into helpful information - information that can be used to either enhance profits, cuts costs, or both. Thus data mining can be used for the purpose ranging from market investigation, fraud recognition and customer maintenance, to fabrication control and science exploration. Itemset discovering is an exploratory data mining scheme which is used for discovering valuable correlations among data [1].

Frequent Pattern Mining is a technique of predicting frequently occurred itemset in the database. This itemset mining method has interest to deal with frequent itemset. Frequent item sets mining scheme is used in the number of real-life beneficial application [2].

Often when considering data mining, the focal point is on frequent itemset. Although the greater part of the majority interesting patterns will lie within the regular ones, there are important patterns that will be disregarded with this process. This is called as infrequent patterns. Get for example the sale of VHS:s and DVD:s . Here the small occurrences of public buying both of them. In data mining, the item set {VHS, VD} will be rare and therefore it is ignored. However, the most of the people buy a DVDs and they does not consider VHS:s and vice versa. These item sets will be challenging and mining infrequent itemset is a demanding one because there are a huge number of such item sets that can be derived from a given data set or data base [3].

The trouble in mining infrequent patterns is: (1) how to determine the interesting rare patterns, and (2) how to professionally find out them in huge data sets. To get a different perspective on different scheme of mining interesting rare patterns, the couple of related concepts are negative patterns and negatively correlated patterns are considered. Here negative item set and negative association rule are introduced.

Now a day's more attention has been given to the issues of item sets discovering process that appears regularly in a dataset or data base. The discovering task (items) that focuses on find out frequent patterns from the databases which is called as frequent pattern extraction. In the frequent pattern extraction, only regular patterns are returned while rare patterns are simply discarded without further concern. This is because of well expensive Informations are carried by the regular patterns and the infrequent patterns cannot acceptably reflect the typical individuality from the data because of their rare occurrence [4].

Association Rule Mining (ARM) is a most investigating area in data mining. The hidden relationships among items are mined by ARM. This association rules find out the infrequent or frequent item set by utilizing a given a user-specified threshold which is called as minimum support[5] [6]. It mostly used for evaluate the itemset after the customer purchasing in the shopping malls. Association

rules also used in different areas such as telecommunication networks, market, risk management and inventory control etc.

II. RELATED WORK

Apriori [7] introduced an algorithm in association rule mining, which is used to categorize the repeated item sets in the huge transactional database. Apriori mining process considered in two phases. In that first phase, it produces all probable Item set arrangements. These arrangements will take action as probable candidates. The candidate will be used in subsequent phase. In Apriori algorithm method, first the smallest amount of support is applied to discover all frequent item sets in a database. In Second phase, these frequent item sets and the smallest amount of confidence constraint are used to form a particular rule. The main weakness of Apriori is production of huge number of applicant sets.

David et.al [8] introduced a new algorithm of MINIT method which is used for discovering minimal τ -infrequent or minimal τ -concurrent item sets. First, a ranking of objects is controlled by approximation need of every item and then produced a record of items in increasing order of support. Minimal τ –infrequent item sets are resolute by using every item in rank order and iteratively calling MINIT on the dataset with regard to items using only those items with superior rank than present items. After that checking every applicant of minimal infrequent items (MII) against the unique dataset is performed. A system can be making a use of judge only superior-ranking items in the iteration which is used to preserve a "liveness" vector indication.

Gao Cong et.al [9] proposed a novel approach for find out an interesting Rule Groups in Microarray datasets. Instead of discovering an individual association rules, the proposed FARMER algorithm find outs an interesting rule groups which are created from similar set of rows. These algorithms searches for attractive rules in the row details and develop all user-specified restriction including minimum support, confidence and chi-square to support well-organized pruning. The FARMER algorithm is faster than the association rule mining algorithm.

Luca Cagliero et.al [10] proposed a novel algorithm for find out the Infrequent Weighted Itemset from the transactional data base. One of the famous data mining approaches is item set mining. Infrequent Weighted Itemset (IWI) and minimal Infrequent Weighted Itemset (MIWI) mining algorithms are used for find out the rare item set in the data base. IWI-support-max measure and IWI-support-min measures are used for discovering infrequent weighted item set. This system efficiently mine the rare item set from transaction data base but it does not consider the utility of item set.

III. PROPOSED SYSTEM

Weighted Transactional Data Set construction

A weighted transaction database (D) includes a set of transactions T and each transaction having number of item set. In weighted transaction database, weight value is associated with each item in I.

T= {t1, t2 …tm}
I = {i1, i2. . . in}
W= {w1, w2,…. . wn}

Where,

T - Set of transactions
I- Set of items
W- Set of weight value

Item sets mined from weighted transactional data sets are called weighted item sets. Their appearance is similar to the one used for conventional item sets, i.e., a weighted itemset is a subset of the data items occurring in a weighted transactional data set. The problem of mining item sets by considering weights associated with each item is called as the weighted itemset mining problem.The significance of a weighted transaction, i.e., a set of weighted items, is generally computed in terms of the equivalent item weights.

Infrequent Weighted Itemset Miner Algorithm

The infrequent weighted item set whose weighted frequency of occurrence is less than a specified weighted threshold. Here two support measures are used for discover infrequent weighted item sets. There are IWI-support-min measure and IWI-support-max measure. The IWI-support-min depends on a minimum cost function, i.e., the occurrence of an item set in transaction is weighted by the weight of its minimum interesting item. The IWI-support-max depends on a maximum cost function, i.e., the occurrence of an item set in transaction is weighted by the weight of the huge interesting item.

By using a maximum IWI-support threshold, all IWIs and MIWIs are discovered. Infrequent Weighted Itemset Miner and Minimal Infrequent Weighted Itemset Miner algorithm are used to tackle the Infrequent Weighted Itemset (IWI) and minimal Infrequent Weighted Itemset (MIWI) mining problems. These algorithms are FP-Growth-like mining algorithms.

The system carry out the main FP-growth mining in following steps: First one is FP-tree creation and another one is recursive itemset mining from the FP tree. Here Pruning is applied to remove frequent itemset which does

not belong to any itemset satisfying the IWI-support threshold ξ.

Algorithm 1: IWI-Miner (T, ξ)

Input: weighted transactional dataset, ξ.
Output: F (Frequent item)

Step 1: initialize FP-Tree construction
Step 2: for each T
// T - weighted transaction
Step 3: FP tree ← T
Step 4: FP tree ← IWI Mining
Step 5: Return F

Given a weighted transactional data set and a maximum IWI-support (IWI-support-min or IWI-support-max) threshold , the Minimal Infrequent Weighted Itemset Miner algorithm extracts all the MIWIs that satisfy . The MIWI Miner algorithm perform operation same as the IWI Miner. However the minimal infrequent weighted item set miner only extract minimal infrequent Patterns.

Algorithm 2 IWI Mining (Tree, ξ)

Input: FP –tree, maximum IWI-support threshold
Output: F, the set of IWIs extending prefix

Step 1: Initialize HT
// HT –Header table
Step 2: HT ← I in tree
Step 3: compute new I with support
Step 4 : I compared with threshold
Step 5: Select IF
// IF –I infrequent weighted item set
Step 6 : compute recursive mining

Minimum Utility based Infrequent Weighted Itemset mining

The projected Low Utility based Infrequent Weighted Itemset mining (LUIWIM) is used to compute minimum utility infrequent weighted itemset based on minimum threshold a value which depends on customer preferences. The utility of items is determined by taking into account factors such as profit, deal, temporal characteristics, etc. of items.

The utility of weighted items is less than the minimum utility threshold which is known as low Utility based Infrequent Weighted Itemset.

$(u(i_m) \leq \text{minutil})$

Where
$u(i_m)$ – Utility of item i_m

Algorithm 3:

Input: IWI
Output: low utility IWI

Step 1: Initialize IWI
Step 2: To compute Utility u (i_m, IWI)
$u(i_m) = l(i_m) \times p(i_m)$
// profit value of item i_m
// number of occurances

Step 3: Compute utility of an itemset I in IWI

$$U(I, IWI) = \sum_{i_m \in I} u(i_m, IWI)$$

Step 4: compute minimum utility
$\text{Minutil} = \delta \times \sum_{IWI} u(i_m, IWI)$
Step 5: if ($u(i_m) \leq$ minutil)
Step 6: low utility value
Step 7: else
Step 8: High value
Step 9: Return low utility IWI

If the utility of items is lower than the minimum utility that have a low utility function ($u(i_m) \leq$minutil). Otherwise that has high utility functions. Minimum utility value is computed from product of minimum utility threshold and utility of particular items in the tree.

Performance evaluation

Data Set Description

Two data sets are taken for infrequent itemset mining. There are Real-life data sets and Synthetic data sets. Real-life data sets to validate the usefulness of the proposed algorithms technique we analyzed 10 collections, each one composed of 31 real-lives weighted data sets. In Synthetic data sets we also exploited a synthetic data set generator to estimate algorithm performance and scalability. The data generator is based on the IBM data generator.

Result analysis

In this experimental analysis, the performance of the existing and the proposed system is compared. Data access speed is how match of time taken for mining the infrequent itemset from the data base.

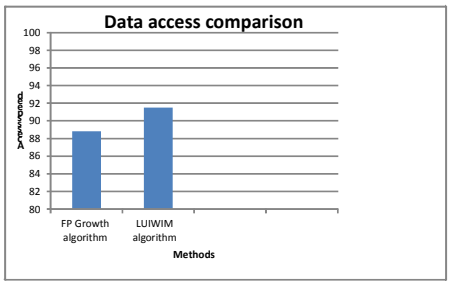

Figure. 1. Data access comparison

Figure .1.shows that access speed comparison. FP Growth access speed of discovering infrequent itemset is compare with the proposed Low Utility based Infrequent Weighted Itemset mining (LUIWIM).Here the access speed of proposed system is high compared with existing one.

Accuracy comparison

Accuracy is defined as the degree of generating the experimental output that is matches with the expected output.

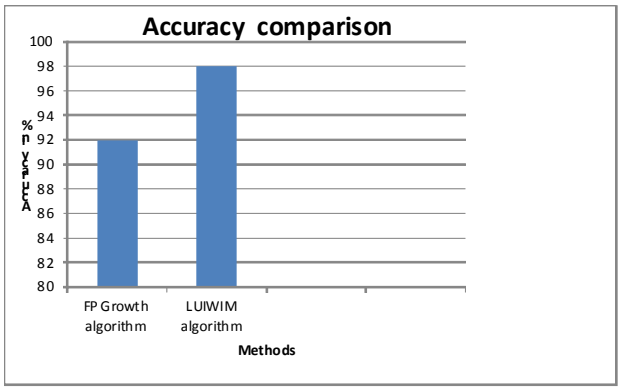

Figure. 1.2. Accuracy Comparison

We analyze and compare the performance offered by existing and proposed methods. In this graph FP Growth and LUIWIM algorithms are predicted in the x axis and the accuracy value is predicted in the y axis. The following accuracy graph shows that proposed system is having higher accuracy compared to other system.

Conclusion

One of the most recent data mining research areas is Utility Mining which prominence on all kind of utility factors and incorporates the utility idea in data mining tasks. In this proposed system, we are evaluating the utility parameters of the rare item sets. The system considers a both the individual profit of each infrequent item and quantity of each one in the data base simultaneously. The utility-based expressive data mining which aims at mining item sets having low utility value which is termed as low utility itemset mining. The utility of items is determined by considering factors such as profit, sale and etc. of infrequent items. The proposed system is used for successfully mine low utility infrequent weighted itemset. From the experimentation result, the proposed system can improve the performance of the system compared to the existing system.

References

[1]. A Review on Infrequent Weighted Itemset Mining using Frequent Pattern Growth Shipra Khare , Prof. Vivek Jain, Shipra Khare et al, / (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 5 (2) , **2014**, 1642-1647, ISSN:0975-9646.

[2]. J.Jenifa, Dr.V. Sampath Kumar," Study on Predicting Various Mining Techniques Using Weighted Item sets ", IOSR Journal of Pharmacy and Biological Sciences (IOSR-JPBS) e-ISSN: 2278-3008, p-ISSN:2319-7676. Volume 9, Issue 2 Ver. VI (Mar-Apr. **2014**), PP 30-39.

[3]. Kamepalli,Sujatha and K. Raja Sekhara Rao," a survey on infrequent pattern mining" International Journal of Advances in Engineering & Technology, ISSN: 22311963 ,1728 Vol. 6, Issue 4, pp. 1728-1732 ,Sept. **2013**.

[4]. Sujatha Kamepalli1 , Raja Sekhara Rao Kurra and Sundara Krishna.Y.K, " Apriori Based: Mining Infrequent and Non-Present Item Sets from Transactional Data Bases", International Journal of Electrical & Computer Science IJECS-IJENS Vol:**14** No:**03**.

[5]. Kumudbala Saxena, Dr. C.S. Satsangi, "A Non Candidate Subset-Superset Dynamic Minimum Support Approach for sequential pattern Mining", International Journal of Advanced Computer Research (IJACR),Volume-2 Number-4 Issue-6 December-**2012**.

[6]. Manish Shrivastava, Mr. Kapil Sharma, MR. Angad Singh, "Web Log Mining using Improved Version of Proposed Algorithm", International Journal of Advanced Computer Research (IJACR),Volume 1 Number 2 December **2011**.

[7]. Sakthi Nathiarasan, Kalaiyarasi, Manikandan, "Literature Review on Infrequent Itemset Mining Algorithms", International Journal of Advanced Research in Computer and Communication

Engineering Vol. 3, Issue 8, August **2014** Copyright to IJARCCE www.ijarcce.com 7670 , ISSN (Online) : 2278-1021,ISSN (Print) : 2319-5940.

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

[9]. Gao Cong, Anthony K. H. Tung, Jiong Yang, "FARMER: Finding Interesting Rule Groups in Microarray Datasets

[10]. Luca Cagliero and Paolo Garza," Infrequent Weighted Itemset Mining Using Frequent Pattern Growth"IEEE transactions on knowledge and data engineering, vol. 26, no. 4, april **2014**.