International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 4 **Issue:** I **Month of publication:** January 2016

DOI:

International Journal for Research in Applied Science & Engineering Technology (IJRASET)

Minimize the Utilization of household items using FP-Tree

S.P. Siddique Ibrahim[1], Dr. M. Sivabalakrishnan[2]

[1]S.P. Siddique Ibrahim is Assistant Professor and with Department of Computer Science and Engineering, Kumaraguru college of Technology, Coimbatore-641049.
[2]Dr. M. Sivabalakrishnan is Associate Professor with School of Computing Science and Engineering, VIT University, Chennai..

Abstract*— Association Rule Mining (ARM) is the promising popular data mining technique. All Previous work was based on frequent itemset mining. Frequent itemset mining finding application in number of real-life contexts e.g., market basket analysis, telemarketing analysis, inventory control, medical image processing, biological data analysis, fraud detection etc. In recent years , researchers has been point of infrequent itemset mining to propose a rage of new algorithms to gain the accuracy of output prediction which is useful while the user point out the rare occurence. This paper address the issues of produce the rare and weighted itemsets in large utilization based datasets. The infrequent itemset mining problem is focus to bring the itemsets whose support count is is less than or equal to maximum threshold fixed by customer. This paper will address the various method of mining infrequent itemset. Finally, discover the rate itemset which is generated by novel algorithm is presented.

***Keywords*—Association Rule Mining, Weighted Itemsets, Rare Itemsets, Infrequent itemset.**

I. INTRODUCTION

Data is unprocessed and set of values of qualitative variables like a varity of forms or raw fact. It does not give any meaning and use. Information can be considered as an aggregation of data and it gives some anwer to the question for the particular context. Information can be converted into knowledge for decision making for future trends. Data Mining is defined as "Analysis process of discovering fascinating pattern huge databases from different sources.". Data mining is the process of discovering data from different viewpoints and finally summarizing into useful information. Discovering of usual patterns noramally hidden in a database, which is emergent research area to extract such pattern constraints in several data mining task like supermarket analysis, automobile purchase in developing countries. There are two predict models in data mining concept. One is predictive model which uses data with known result like classification and develop a model that can use clearly to predict values. Another is descriptive model, which describes the real world events and relationship of them.

Classification is most important task in data mining to build accurate and efficient classifier for datasets. classifier is constructed to predict class label based on some metrics. It collected of two steps: supervised learning of a training set of data to create a model based on testing dataset derived from original dataset, and then constructing classifying the data according to the model. Regression analysis is a another data mining technique that often used for numerical prediction. It is used between dependent variable and one or predictors. Many real world data mining applications can be applied to predicting future data states based on historical and current data. Summarization, Association rule, Clustering, and sequence discovery are numaric in nature. Clustering involves identifying a finite set of meaningful sub classes from unpartioned large set of data. Each member in a cluster should be very similar to other member and dissimilar value is high with to other cluster. Sequential discovery is used to determine statically relevant patents in data where the values are delivered in sequence manner. These patterns are closely related time sequence mining. String mining is popular and deals with characters for items that appears in sequence.

FP Growth algorithm [13] used to generate frequent pattern without candidate generation based on tree based construction and conditional pattern which increase the mining process by left of candidate rule. The main disadvantage of FP-growth is depends on searching of paths in FP tree and it need unnecessary memory requirement still for reasonably sized datasets, The patterns are generated by concatenating the suffix pattern with the frequent pattern generated for conditional FP tree. The process continues until either the tree contains only the root or the tree has a single path .If the tree contains only a single path , then all the combinations of that path are generated and considered to be frequent patterns.

The large body of frequent itemset mining algorithms can be broadly classified into two categories: (i) candidate generation-and-test paradigm and (ii) pattern-growth paradigm. In earlier studies, it has been shown experimentally that pattern-growth based algorithms are computationally faster on dense datasets.

A. Applications of Data mining

Cross Market Prediction
Fraud Detection
Theft Alert
Prediction analysis
Information Retrival
Text Analysis
Utilization Finding
Biomedical...etc

B. Measures of Association Rule Mining:

Support: The rule X=>Y holds with support

Sup(X)= sup X U Y / N

Mining high utility itemsets from databases is not an easy task because the itemsets must satisfy the downward closure property in frequent itemset. In FP-Tree item pruning for high utility will satisfy the minimum threshold set by the user. In some cases this problem more difficult to pruning like large search space with databases contains large transactions or low utility itemsets. Existing algorithm often generates a huge set of Potential High Utility Itemsets (PHUIs) [13] then identifies the utility of item set. This may degrade the mining performance in particularly when the database contains much long transaction. Consider the database in table 1. There are five items in the profit table and five transaction in the transaction table in the database.

Minimal and not minimal infrequent weighted itemset (IWI) address the discovery of infrequent weighted itemsets, i.e., mining infrequent weighted form supermarket datasets. To obtain useful infrequent itemsets IWI-support count measure is proposed. Weights are derived from the occurrence of an item in each supermarket by applying a given expenditure function. Such as

1) The IWI-support-Min measure, which depends on a min cost function, i.e., the rate of an itemset in a given transaction is weighted by the weight of its least interesting item weight.
2) The IWI-support-Max measure, which depends on a max cost function, i.e.,the rate of an itemset in a given transaction is weighted by the weight of its largely interesting items weight. When dealing with optimization technique, the minimum and maximum are commonly used cost function to evaluate the efficient of computer program. Minimal infrequent itemset mining is the problem of discover minimal weight of the itemset among database, whose weight always less than neighbor item weight. Not minimal infrequent itemset mining is the problem of discover maximal weight of the itemset among database, whose weight should higher than neighbor item weight.

For example, in Table 1. It consists of eight Utilization of Airconditionar (which is identified by corresponding ids) each transaction composed of four AC Units.Each transactions may represent consumer frequently Utilize the AC Unit.

Table1: Weighted Transactional Itemset

TID	AC1	AC2	AC3	AC4
1	100	0	44	40
2	94	0	45	99
3	10	72	0	43
4	0	100	60	70
5	0	42	30	10
6	60	70	0	70
7	0	28	7	0
8	23	0	0	0

A. Itemset Mining

Itemset ming is an exploratory data mining technique widely used for discovering valuable correlation among data. Itemset mining of two types they are

1) *Frequent itemset mining:* Itemset mining focused on discovering frequent itemset. Itemset is frequent if its support satisfies given minimum support threshold. Frequent itemset find application in many real life contexts. For example buying a PC first, then a digital camera and then a memory card, if it occurs frequently in shopping history, then it is called frequent pattern. Market basket analysis is one of the frequent itemset mining applications.
2) *Infrequent itemset mining:* Patterns that are rarely found in database are considered to be unexciting and are eliminated using the support measure. Item set is infrequent if its support is less than or equal to predefined support threshold. This method has a great interest as they deal with rare but crucial cases. Applications in infrequent itemset mining include identifying rare diseases, predicting equipment failure, and finding association between infrequently purchased items.

II. LITERATURE SURVEY

Author in [1] has been proposed a new algorithm called MINIT (MINimal Infrequent Itemset), which is used for mining minimal infrequent itemset (MIIs) [1]. This is the first algorithm for finding rare itemset. The itemset that satisfy a maximum support threshold and does not contain any infrequent subset, from transactional data set. It is based on SUDA2 algorithm. A Dataset property is to consider the matrix form is a main difference between MINIT and SUDA2. Matrix consists of binary entries for traditional itemset mining. But for SUDA2, the matrix entries can contain any integer. Minimal infrequent itemset problem is NP-complete.

Author in [2] has proposed a measure called w-support, which is used to find weight of itemset and weight of transaction does not require preassigned weight. These weights are completely based on internal structure of the database. HITS model and algorithm are used to derive the weights of transactions from a database with only binary attributes. A new measure w-support is defined to give significance of itemset and it differs from the traditional support in taking the quality of transactions into consideration based on these weights. An apriori-like algorithm is proposed to extract association rules whose w-support and w-confidence are above some give thresholds.

Probabilistic frequent itemset mining in uncertain transactional database. Based on world semantics Thomas Bernecker in [3] introduces new probabilistic formulations of frequent itemsets. In a probabilistic model, an itemset is said to be frequent if the probability that itemset happens in at least min support is higher than that of given threshold. In addition to probabilistic model, framework is presented which has a capacity to solve the Probabilistic Frequent Itemset Mining (PFIM) problem powerfully.

Based on interest/intensity of the item within the transaction Wei Wang in [4] has proposed by allowing weight to be associated with each item within the transaction. In turn, to associate a weight parameter with each item in a resulting association rules. Then it is called as weighted association rule WAR). For example, bread[4,6] and jam[3,5] is a weighted association rule indicating that if a customer purchase a bread quantity between 4 and 6 pack, he is likely to purchase 3 and 5 pack of jam. This method produces a higher quality results than previous known method on quantitative association rules.

Author in [5] present SUDA2 a recursive algorithm for finding Minimal Sample Uniques (MSUs). It uses a new method for demonstrating the search space for MSUs and observe about the properties of MSUs to prune and traverse this space. It has a ability to identify the boundaries of the search space with an execution time which is numerous orders of magnitude faster than that of SUDA. SUDA2 is a good candidate for parallelism as a search can be divided up according to rank, it will produce efficient load balancing.

In weighted settings author in [6] deal with the problem of finding significant binary relationship in transactional dataset. In weighted association rule mining problem each item is allowed to have a weight. The main focus is to mining significant relationship relating items with significant weights rather than insignificant relationship. A new algorithm called WARM (Weighted Association Rule Mining) is developed. This algorithm proposed a “weighted downward closure property” as a substitution of original “downward closure property”. Weighted downward closure is a idea of replacing support with significance is proposed. WARM algorithm is both scalable and efficient in discovering significant relationship in weighted settings.

Ashish Gupta, Akshay Mittal, Arnab Bhattacharya [7] propose a new algorithm based on pattern-growth paradigm for finding minimally infrequent itemset. They introduce a new concept called residual tree. To mine a multiple level minimum support itemsets ,where different threshold are used for finding frequent itemsets for different lengtha of the itemset by using residual tree. For mining minimally infrequent itemsets (MIIs) author [7] introduce a new algorithm called IFP min. Here Apriori algorithm is proposed to find MIIs. Extension of the algorithm is designed for finding frequent itemset in the multi level minimum support (MLMS) model.

Luca Cagliero and Paolo Garza [8] deal with the issues of discovering rare and weighted itemsets, i.e., the infrequent itemset mining problem. Finding rare data correlations is more interesting than mining frequent ones. [8] is like FP-growth algorithm. The IWI-support measure is defined as a weighted frequency of occurrence of an itemset in the analyzed data. Occurrence weights are derived from the weights associated with items in each transaction by applying a given cost function. They mainly focuses on IWI-support-min measure and IWIsupport-max measure. Author proposes two IWI mining algorithm that carry out IWI and Minimal IWI mining efficiency. A result shows the efficiency and effectiveness of the proposed approach.

In [4] author focus on finding association among frequent itemsets. X. Wu, C.Zhang and S.zhang designed a new method for mining both positive and negative association rules efficiently. This approach is new and different from existing approach i.e., association analysis. This method reduces the search space and had used the increasing degree of the conditional probability to assess the confidence of positive and negative association rules. This result shows that the proposed approach is effective, efficient and promising [4].

Negative association rule (NAR) is used for discovering of interesting pattern during mining process. Apriori algorithm is used for mining negative association rule for frequent absence and presence (FAP) itemset. Association rule mining only explores positive relationship in the beginning. Positive relationship implies the purchase one item or itemset with the purchase of another item or itemset. Negative relationship implies the presence of items by the absence of other item in the same transaction. FAP itemsets effectively generate optimum number of rules including NAR compared to others. The search space can be significantly reduced in this approach.

III. PROPOSED APPROACH

The proposed work follows two phrases on datasets. In the first phase equivalent weight can be calculated for the item sets. Evaluation of the equivalence is checked with FP-Tree of the weighted system load datasets. The second phase rules in FP-Tree is pruned using enhanced pruning method.

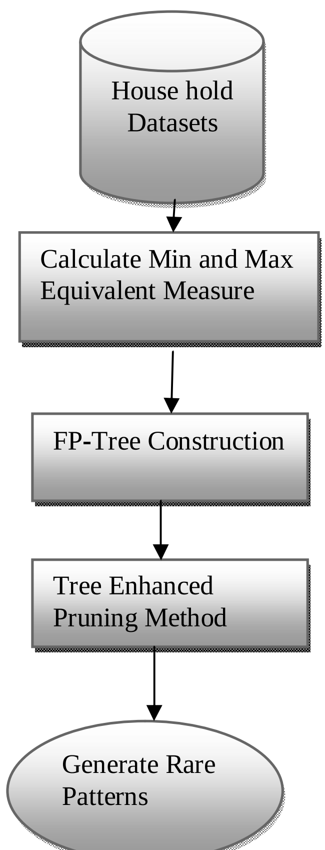

Figure 1. Proposed method chart

A. Weighted Dataset

Each item in the dataset associated with weight that will useful to analysis and monitoring of multi core system usage is commonly associated to detecting system malfunctiong, optimizing and regulate the load balancing and proper resource sharing among systems and also the expert will perform system resizing with available resources. IWIs that satisfy a IWI-support value represent combinations of underutilized ot idle cores. When IWI-suppot min threshold, represent the smallest core combinations that contain at least one underutilized core. Moreover, when the IWI-support max threshold represent the smallest core combinations that contain only underutilized cores.

B. FP-Tree Pruning Method:

In this procedure tree pruning can be done in two steps

1) FP tree construction and
2) Mining infrequent items sets recursively from enhaced FP tree

the proposed FP growth mining infrequent itemsets intead of frequent ones. To achieve this task, the following modifications with respect to FP tree have been followed

a) a new pruning policy for pruning part fo the search space premature and b) a little modified FP tree construction which allow each node storing the corresponding infrequent support value independently.

III. SAMPLE CODE

```
 import java.util.Collections;
 import java.util.Comparator;
 import java.util.LinkedList;
 import java.util.List;
 import java.util.Scanner;
 import java.util.Stack;
 import java.util.Vector;
 import java.io.*;
import java.util.HashSet;
import java.util.Set;
public class Infreq
{
public static void main(String[] arg)
  {
    try
    {
       Vector vc=new Vector();
       Vector vc1=new Vector();
       BufferedWriter bw=new BufferedWriter(new FileWriter("e:/max.txt"));
       BufferedWriter bw1=new BufferedWriter(new FileWriter("e:/min.txt"));
       BufferedWriter bw2=new BufferedWriter(new FileWriter("e:/max1.txt"));
       BufferedWriter bw3=new BufferedWriter(new FileWriter("e:/min1.txt"));
       Main.ta.append("\n ORIGINAL DATASET \n");
      Main.ta.append("------------------------------------------------------------------ \n");
       String ghg="",tempk="";
       int row=0,col=0;
       int temp=0;
       int temp1=0;
       String[] atrr={"bread","jelly","oil","milk"};
       int[] atr={1,2,3,4,5};
       int[] aa=new int[100000];
       BufferedReader brs=new BufferedReader(new FileReader("e:/inputdata.txt"));
       while((ghg=brs.readLine())!=null)
       {
```

```
        Main.ta.append(""+ghg+"\n");
        String gh[]=ghg.split(",");
        row++;
    }
    int[][] dd=new int[row][5];
    int[][] ff=new int[row][5];
     brs=new BufferedReader(new FileReader("e:/inputdata.txt"));
     row=0;
    while((ghg=brs.readLine())!=null)
    {
        String gh[]=ghg.split(",");
        for(int i=2;i<gh.length;i++)
        {
            String[] hh=gh[i].split(":");
            dd[row][i-2]=Integer.parseInt(hh[1]);
            ff[row][i-2]=Integer.parseInt(hh[1]);
        }
        row++;
    }
    int[][] ee=dd;

   Main.ta.append("------------------------------------------------------------------------ \n");
   Main.ta.append("Equivalent Maximum Weighted Transaction \n");
   Main.ta.append("------------------------------------------------------------------------ \n");


for(int i=0;i<ff.length;i++)
    {
    while(true)
    {
     int max=-10000;
     for(int j=0;j<ff[i].length;j++)
     {
      if(ff[i][j]!=0)
      {
      if(max<ff[i][j])
      {
       max=ff[i][j];
      }
  }
     }
     int yyy=0;
     int g1=0;
     int[] cc=new int[5];
     for(int j=0;j<ff[i].length;j++)
     {
       if(ff[i][j]<=max)
       {
           ff[i][j]-=max;
```

Volume 4 Issue I, January 2016
ISSN: 2321-9653
International Journal for Research in Applied Science & Engineering Technology (IJRASET)

```
Main.ta.append(" "+atrr[j]+"("+max+")"+"  ");
if(max<0)
{
   bw.write(" "+ -max);

 if((atr[j]!=0) && (atr[j]!=temp1))
   {
        //ab=" "+atr[j];
       cc[g1++]=atr[j];
       temp1=atr[j];
   }
```

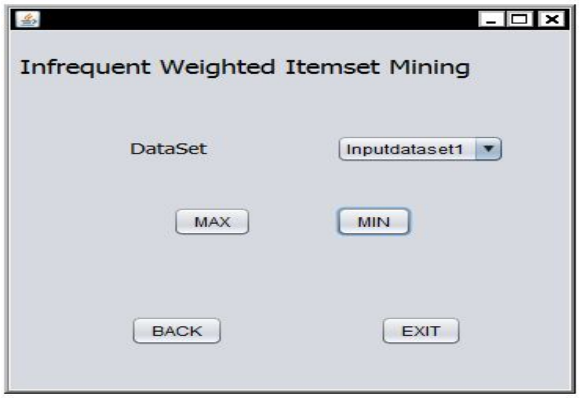

Figure 2: Selection of Dataset

Figure 3: IWI selection

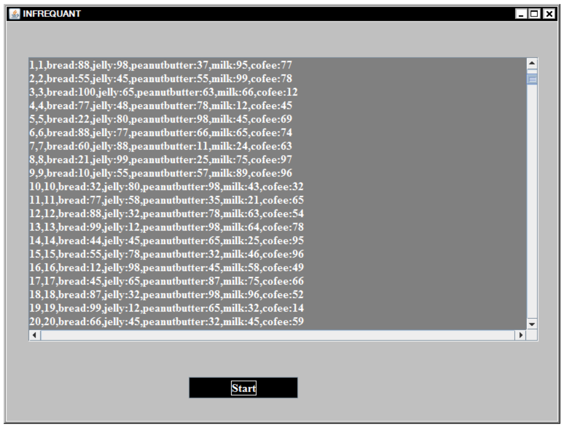

Figure 4: IWI generation

International Journal for Research in Applied Science & Engineering Technology (IJRASET)

IV. CONCLUSION

Frequent Itemset Mining has attracted plenty of attention but much less attention has been given to mining Infrequent Itemsets. This survey is focused on infrequent weighted itemset. Occurrence weights derived from the weights associated with each items in transaction and applying a given cost function. The related concepts of positive and negative correlated pattern and its association rules are mined. The major advantage for mining infrequent itemset was to advance the profit of rarely originated datasets in the transactions. Merits and demerits of each method are described in comparative table to efficiently differentiate the each methods functionality.

REFRENCES

[1] D. J. Haglin and A.M. Manning, "On Minimal Infrequent ItemsetMining," Proc. Int"l Conf. Data Mining (DMIN "07), pp. 141-147, 2007.
[2] K. Sun and F. Bai, "Mining Weighted Association Rules Without Preassigned Weights," IEEE Trans. Knowledge and Data Eng.,vol. 20, no. 4, pp. 489-495, Apr. 2008.
[3] C.-K. Chui, B. Kao, and E. Hung, "Mining Frequent Itemsets from Uncertain Data," Proc. 11th Pacific-Asia Conf. Advances in Knowledge Discovery and Data Mining (PAKDD "07),pp.47-58,2007.
[4] W. Wang, J. Yang, and P.S. Yu, "Efficient Mining of Weighted Association Rules (WAR)," Proc. Sixth ACM SIGKDD Int"l Conf. Knowledge Discovery and data Mining (KDD "00), pp. 270-274, 2000.
[5] A. Manning and D. Haglin, "A New Algorithm for Finding Minimal Sample Uniques for Use in Statistical Disclosure Assessment," Proc. IEEE Fifth Int. Conference in Data Mining (ICDM "05), pp. 290-297, 2005.
[6] 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.
[7] A. Gupta, A. Mittal, and A. Bhattacharya, "Minimally Infrequen Itemset Mining Using Pattern-Growth Paradigm and Residual Trees," Proc. Int"l Conf. Management of Data
[8] Luca Cagliero and Paolo Garza," Infrequent Weighted Item set Mining Using Frequent Pattern Growth" IEEE Transactions On Knowledge And Data Engineering, Volume 26, No.4, April 2014
[9] 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.
[10] Anis Suhailis Abdul Kadir, Azuraliza Abu Bakar, Abdul Razak Hamdan, "Frequent Absence and Presence Itemset for Negative Association Rule Mining," 11th International Conference On Intelligent System Design and Applications, 2011.
[11] 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), pp. 119-128, 2009.
[12] 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
[13] J. Han, J. Pei, and Y. Yin, "Mining Frequent Patterns without Candidate Generation," Proc. ACM SIGMOD Int"l Conf. Management of Data, pp. 1-12, 2000.
[14] Siddique Ibrahim S P, Priyanka R, "A Survey on Infrequent Weighted Itemset Mining Approaches" , 2015, IJARCET, Vol.4, pp. 199-203.
[15] G. Cong, A.K.H. Tung, X. Xu, F. Pan, and J. Yang, "Farmer: Finding Interesting Rule Groups in Microarray (COMAD), pp. 57-68, 2011.Datasets," Proc. ACM SIGMOD Int"l Conf. Management of Data (SIGMOD "04), 2004.
[16] Mehdi Adda, Lei Wu, Sharon White(2012), Yi Feng "Pattedetection with rare item-set mining" International Journal on Soft Computing, Artificial Intelligence and Applications (IJSCAI), Vol.1, No.1, August 2012.