Innovative Science and Technology Publications

International Journal of Future Innovative Science and Technology
ISSN: 2454- 194X Volume - 2, Issue - 2

Manuscript Title

Mining The Association Rules Using F-CARM

[1]R.Archana, [2]M.Shanthamani., M.Tech.,

[1]PG Scholar, [2]Assistant Professor,

Department of Computer Science and Engineering,

Velalar College of Engineering and Technology, Thindal, Erode, India.

E-Mail: ***archana93.me@gmail.com, shanthalaya@gmail.com***

May – 2016

www.istpublications.com

International Journal of Future Innovative Science and Technology, ISSN: 2454- 194X
Volume-2, Issue-2, May - 2016 editor@istpublications.com

Mining The Association Rules Using F-CARM

[1]R.Archana, [2]M.Shanthamani., M.Tech.,

[1]PG Scholar, [2]Assistant Professor,

Department of Computer Science and Engineering,

Velalar College of Engineering and Technology, Thindal, Erode, India

E-Mail: ***archana93.me@gmail.com, shanthalaya@gmail.com***

ABSTRACT

Data mining is the process of obtaining useful information from large databases. Association rule mining is the main task in the field of data mining whose aim is to extract asoociation rules in set of transaction item database. CARM algorithm on the fuzzy itemset database measures inspired by cogency also with support thershold. Cogency values are based pairwise item conditional probability. so the proposed algorithm mines association rules by only one pass through the file and also more efficient for dealing with infrequent items. This paper evaluates CARM over data sets. the proposed algorithm is faster and consumes less memory space The problem of associative classification is used here for evaluating the proposed algorithm. In addition research work has been emerging in Location-Based Service and due to a wide range of potential data mining applications. One of the active topics is the mining and prediction of movable actions and associated transactions. The proposed work on mining and prediction of sub sequences rare item sets with considerations of user relations and temporal property simultaneously. Through experimental evaluation under various simulated circumstances, the proposed schemes are shown to deliver excellent performance using Fuzzy Association Rule Mining (FARM) with location based services.

***Keywords*—** Data mining, Association rule mining, confabulation, cogency, frequent patterns item.

I. INTRODUCTION

Mining frequent itemset in transaction databases, time-series databases, and many other kinds of databases has been studied popularly in data mining research. Most of the previous studies adopt an Apriori-like candidate set generation-and-test approach. However, candidate set generation is still costly, especially when exist a large number of patterns and/or long patterns. Further, dynamic itemset counting algorithm [1], an extension to Apriori algorithm used to reduce number of scans on the dataset. It was alternative to Apriori Itemset Generation .In this, itemsets are dynamically added and deleted as transactions are read .It relies on the fact that for an itemset to be frequent, all of its subsets must also be frequent, so we only examine those itemsets whose subsets are all frequent. Both Apriori and DIC are based on candidate generation. Previous algorithm is costly to handle a number of candidate's sets. Considering the example, If 1000 frequent 1-items set, the previous algorithms will need to generation more than 1000X 1000 length candidate and accumulated and test their occurrence frequent itemset.

This is the inherent cost of candidate generation, no matter what implementation techniques is applied. It is tedious to repeatedly scan the database and check large set of candidates by itemset matching, which is especially true for mining long patterns. Previous algorithm scans the database too many times, when the database storing a large number of data servies, the limit memory capacity, the system I/O load, considerable scanning the database will be a very long time, so efficiency is very low.

A new ARM algorithm is called CARM that is inspired by the mechanism of thought in the human brain, and specifically the theory of confabulation, to mine association rules. The proposed algorithm contains two steps of knowledge acquisition and rule extraction. Knowledge acquisition consists of two modules where the axonal communication links between these two modules are used to store all domain knowledge. The second step of rule extraction is then performed based on the strength of these communication links. For performance evaluation of CARM, we use associative classification.

The aim of association rule mining is to detect interesting associations between items in a database[2]. It was initially proposed in the context of market basket analysis in transaction databases, and has been extended to solve many other problems such as the classification problem. Association rules for the purpose of classification are often referred to as predictive association rules. Usually, predictive association

rules are based on relational databases and the consequences of rules are a pre-specified column, called the class attribute.

This paper addresses the problem of finding interesting predictive association rules in datasets with unbalanced class distributions[3]. To address these issues, a single phase incremental association mining technique has been reported in this paper, which can extract reduced set of interesting rules from real-life datasets.

II. RELATED WORKS

Associative classification is a rule-based approach to classify data relying on association rule mining by discovering associations between a set of features and a class label. Support and confidence are the de-facto "interestingness measures" used for discovering relevant association rules[6]. The support confidence framework has also been used in most, if not all, associative classifiers. Although support and confidence are appropriate measures for building a strong model in many cases, they are still not the ideal measures and other measures could be better suited[7][8].

Rare association rule mining has received a great deal of attention in the recent past. In this research they use transaction clustering as a pre-processing mechanism to generate rare association rules. The basic concept underlying transaction clustering stems from the concept of large items as defined by traditional association rule mining algorithms. They make use of an approach proposed by Koh & Pears (2008) to cluster transactions prior to mining for association rules [9] [10] [11]. They show that pre-processing the dataset by clustering will enable each cluster to express their own associations without interference or contamination from other sub groupings that have different patterns of relationships. Their results show that the rare rules produced by each cluster are more informative than rules found from direct association rule mining on the un partitioned dataset[12 [13].

Frequent patterns are an important class of regularities that exist in a transaction database. Certain frequent patterns with low minimum support (minsup) value can provide useful information in many real-world applications. However, extraction of these frequent patterns with single minsup based frequent pattern mining algorithms such as Apriori and FP-growth leads to "rare item problem[1][16][17] [18]." That is, at high minsup value, the frequent patterns with low minsup are missed, and at low minsup value, the number of frequent patterns explodes. In the literature, "multiple minsup frameworks" was proposed to discover frequent patterns. Furthermore, frequent pattern mining techniques such as Multiple Support Apriori and Conditional Frequent Pattern-growth (CFP-growth) algorithms have been proposed. As the frequent patterns mined with this framework do not satisfy downward closure property, the algorithms follow different types of pruning techniques to reduce the search space. In this paper, they proposed an efficient[14] [15] CFP-growth algorithm by proposing new pruning techniques. Experimental results show that the proposed pruning techniques are effective [19][20].

III. PROPOSED METHODOLOGY

Fuzzy Association Mining Rule Temporal Sequential Patterns (FAMRTSPs) a prediction strategy is proposed to predict the subsequent mobile behaviors, in FAMRTSPs - Mine, user clusters are constructed by a novel algorithm named Cluster Affinity Search Technique (CAST) and similarities between users are evaluated by the proposed measure, Location-Based Service Alignment (LBS-Alignment).

At the same time, a time segmentation approach is presented to find segmenting time intervals where similar mobile characteristics exist. The proposed system considers mining and prediction of mobile behaviors with considerations of user relations and temporal property simultaneously. The domain knowledge links are building and rules are estimated using the confabulation theory by means of single search or scan make over the database. In the proposed algorithm (FAMRTSP), only one-item consequent association rules are generated, where there can be multiple predecessor items.

In addition, incremental information extraction is applied from the data sets. The new items from new transactions are found out and L Matrix size is incremented. Old L Matrix values are updated based on old items found in new transactions. MinCog threshold is set based on the average link strength between items.

IV. PROPOSED METHODOLOGY PROCEDURE

Association rule mining is one of the most significant tasks of the Data Mining. Association rule mining finds the interesting or correspondence relationship among a huge set of data items. The distinctive example of association rule mining is the market basket analysis. Frequent item set mining directs to the finding of associations and association among items in huge transactional or relational datasets. In short, an association rule is an expression X =>Y, where X and Y are large item sets. The association rule mining can be viewed as a two step process.

- Discovery of all frequent item sets: The items which are frequently purchased together in one transaction are called the recurrent itemsets.
- Create tough association rules from the recurrent itemsets: the rules which are generated must satisfy the minimum confidence and minimum support.
- Let I = $\{i_1, i_2..., i_m\}$ be a set of literals, called items. Let D be a set of transactions, where each transaction T is a set of items such that T_i. and it quantities the

items bought in a transaction are considered, which means that each item is a binary variable representing if an item was bought. Each transaction is allied with an identifier, called T_{id}.

- Let X is a set of items. A transaction T is supposed to contain X if and only if X(T). An association rule is a proposition of the form X→ Y, where X Y and Y I, and X ∩ Y = ¢. The rule X→ Y holds in the transaction set D with confidence c if among those transactions that contain X c% of them and also contain Y. The rule X→ Y has support s in the transaction set D if s% of transactions in D contains X Y.
- Rule support and confidence are two measures of rule interestingness. Certainty and utility are the most important measures of rule interestingness, which describes confidence and support accordingly.

Confabulation Theory

Confabulation theory offers a comprehensive detailed explanation of the mechanism of thought (i.e., "cognition": vision, hearing, reasoning, language, planning, origination of movement and thought processes, etc.) in humans and other vertebrates (and possibly in invertebrates, such as octopi and bees). For expositional simplicity, only the human case is considered here.

Confabulation (verb: confabulate) is a memory disturbance, defined as the production of fabricated, distorted or misinterpreted memories about oneself or the world, without the conscious intention to deceive. Confabulation is distinguished from lying as there is no intent to deceive and the person is unaware the information is false.

Although individuals can present blatantly false information, confabulation can also seem to be coherent, internally consistent, and relatively normal. Individuals who confabulate present incorrect memories ranging from "subtle alterations to bizarre fabrications", and are generally very confident about their recollections, despite contradictory evidence.

Rare Item Mining

The discovery of new and interesting patterns in large datasets, known as data mining, draws more and more interest as the quantities of available data are exploding. Data mining techniques may be applied to different domains and fields such as computer science, health sector, insurances, homeland security, banking and finance, etc. In this project they are interested by the discovery of a specific category of patterns, known as rare and non-present patterns. They present a novel approach towards the discovery of non-present patterns using rare item-set mining.

CARM Algorithm

In CARM, only one-item consequent association rules are generated, where there can be multiple antecedent items. The proposed CARM approach using a cogency inspired measure for generating rules. Cogency inspiration can lead us to more intuitive rules. Moreover, cogency-related computations only need pair-wise item co-occurrences, hence, They can find rules only by one file scan. Rule mining is performed in two main phases: knowledge acquisition and structure construction and rule generation by confabulation and cogency measure. In this algorithm, only one item con-sequent association rules are generated, which means that the consequents of these rules only contain one item.

Below is the pseudo code for the CARM algorithm:

Algorithm 1: CARM

```
1-   i=1
2-   while not EOF
3-   read transaction t_i
4-   for each x ∈ t_i
5-     for each y ∈ t_i
6-       L xy = L xy + 1
7-     end for
8-   end for
9-   i=i+1
10-  end while
11-  end
```

Algorithm 2: FARM

/* FARM Algorithm */

Input: Two mobile transaction itemset sequences s and s'

Output: The similarity between s and s'

```
01 LBS-Alignment (s,S')
02 p ← 0.5 / (s.length + s'.length)   /*p is the location penalty */
03 M0,0 ← 0.5
04 Mi,0 ← Mi-1,0 - p  i = {1,2,....,s.length)
05 M0,j ←M0,j-1 - p  j = {1,2,....,s'.length)
06 For i ←1 to s.length
07 For j ←1 to s'.length
08 For j ←1 to s'.length
09If si.location = sj'.location
10 TP ←p * | si.time = sj'.time| / len          /* time penalty */
11 SR ←p * ( si.service ∩ sj.'service/ si.service U sj.'service)
                        /* service reward */
12 Mi,j ←Max(Mi-1,j-1 – TP + SR, Mi+1,j – p, Mi,j-1 – p)
13 Else
14Mi,j ←Max(Mi-1,j– p, Mi,j-1 – p)
```

V. THEORY OF F-CARM ALGORITHM

Input:
transaction_items: transaction items.
count_items: limit the items scan and finding the frequent items.
temp_items : maintains the temporary frequent or non frequent items.
classify_items: classify the transaction items.
fuzzy_items: If select already set the frequent items or non frequent items.
dynamic_items: increment frequent items.
Output:
association_items: result in frequent item sets (finding minimum support items & update database)
Notations:
ts: transaction items
cu: count the limit items (cuT, cuF, cuD)
temp: temporary frequent or non frequent items.
cs: classify_itemset schema (Classify transaction items)
fs: fuzzy_itemset schema (Classify already set the frequent or non frequent items)
ds: dynamic_itemset schema (finding the increment frequent itemset between **fs** and **cs**)
aso: association_itemset schema (finding the frequent itemset in **ds**)
Method:
Initially state (cs, fs and ds is create and empty the schema)
Each step finding the frequent items and frequent count (cs, fs, and ds is drop the schema)
Step1: Set the **cuT** and partition of **cuT** items in **ts**.
Step2: Scan and partition the **ts** and store the **cs**.
Step3: Set the **cuT** and partition of **cuF** items and store in **fs**.
Step4: Finding the frequent in itemset until **ts** items between **fs** and **cs**.

If **fs** is set frequent items means:
If (items is frequent) means store **aso**
Else maintains non frequent items in **temp**
If **fs** is set non frequent items means:
If (items is non frequent) means store **temp**
Else maintains frequent items in **aso**
Return frequent items and frequent count

Step5: The **temp** item partitions of **cuD** set and store the **ds.**
Step 6: Find the increment frequent itemset between **cuD** to **ds.**

If **ds** is set frequent items means:
If (items is frequent) means store **aso**
Else maintains non frequent items in **temp**
If **ds** is set non frequent items means:
If (items is non frequent) means store **temp**
Else maintains frequent items in **aso**
Return frequent items and frequent count

Step 7: Repeat the process finding frequent items and its counts.
Step 8: **Drop** the schema **fs, cs** and **ds** is each finding frequent items.
Step 9: Finding the association rule mining in minimum support, and confidence following state.

Support = No. of. Count in Frequent Items / No. of. Total Transaction >= **min_support**

***[i.e. s' Support(A →B) = P(AUB)* >= min_support]**

Confidences = No. of. Count in Frequent Items / No. of. Transaction Items >= **min_conf**

***[i.e. s' confidences(A →B) = P(B/A)* >= min_conf]**

Where **min_support:** The minimum support threshold**.**
min_conf **:** The minimum confidences threshold.
Step 10: Update the original database is new frequent min_support itemset.
Step 11: Maintains non frequent items future refer the next frequent items.

VI. EXPERIMENTAL RESULTS

The calculated the association mining rule the frequent items F is 40%, A, B, G and H contains 30%, C, E, I, J is 20% and D is 10%. Finally already select the fuzzy items frequent is A, B, C, F, G , H and extra items updates the original database is E, I and J. D is store the temporary list because Dis minimum count is 1 and minimum support 10%.
So market basket dataset is frequently items encountered easily and maintains current items in the market dataset, the followed the chart minimum support and maintains frequent items.

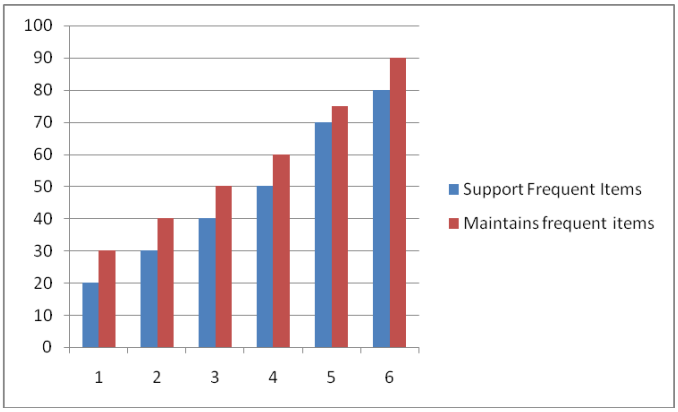

Figure 1: Performances for Frequent Items

The chart is refer maintains the frequent items in original database. It is assume approximately First frequent items support is 20% contains 30% of frequent items is maintain. The second frequent items support is 30% contains 40% of frequent items is maintains for original dataset and 40% contains 50%, 50% contains 60%, 60% contains 70%, 70% contains 75%, 80% contains 90% frequent items maintains the original dataset.

Experimental results in this section show that the proposed research outperforms the state-of-the-art algorithms almost in all cases on medicine transaction data sets.

This thesis is used to process using eliminate time complexity rate while finding high utility item sets in a transaction database. In this proposed paper, tree construction two This thesis is used to process using eliminate time complexity rate while finding high utility item sets in a transaction database. In this proposed paper, tree construction two strategies, namely CARM (Confabulation Association Mining Rule) and F-CARM (Fuzzy Confabulation Association Mining Rule)). It also used to reduce number scans to the database.

Transaction No	Data item set (N)	CARM [%]	F-CARM [%]
1	100	55.33	61.33
2	200	62.30	64.33
3	300	67.33	69.22
4	400	72.11	72.45
5	500	76.57	78.01
6	600	80.08	81.09
7	700	81.44	81.98
8	800	82.55	83.08
9	900	83.55	84.03
10	1000	84.67	85.04

Table 1- Performances for CARM and F-CARM Algorithm

In this research work , present the use of an association rule mining driven application is to manage market basket dataset that provide items with report regarding prediction of product purchase or sales trends and customer behavior. Our goal of the research is to find a new schema based rare frequent items for finding the rule of the market basket transactional dataset, which outperforms in terms of running time, number of database scan, memory consumption and the interestingness of the rules over the classical F-CARM algorithms.

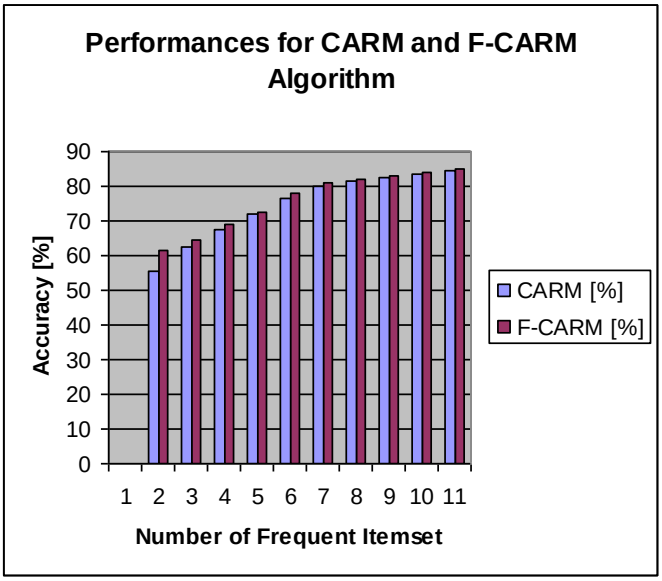

Figure 2: Performances for CARM and F-CARM Algorithm

VII. CONCLUSION

Market basket dataset is one of most important part of research process. The main goal of super market industry sales of frequent items maintains and increasing profit. So there are strong association rule finding frequent items must the data mining works. In this algorithm proposed is some techniques added the future analysis of mining frequent items. Because select the fuzzy items frequent optimizing select and compare transactions items, so strong fuzzy items created techniques applied our proposed algorithm and update the original database increasing times, access the speed of processors implementation of our proposed algorithms.

The new system become useful if the below enhancements are made.

- ✓ In future work, the method can be applied to real data sets. In addition, the CTMSP-Mine can be applied to other applications, such as GPS navigations, with the aim to enhance precision for predicting user behaviors.
- ✓ If the application is developed as web based application, then it can be used from anywhere.

The new system is designed such that those enhancements can be integrated with current modules easily with less integration work.

VIII. REFERENCES

[1] Agrawal R. and Srikant R., "Fast Algorithms for Mining Association Rules," Proc. 20th Int'l Conf. Very Large Data Bases (VLDB), pp. 487-499, 1994.

[2] Cai, C.H. Fu A.W.C., Cheng C.H., and Kwong W.W., "Mining Association Rules with Weighted Items," Proc. Int'l Database Eng.and Applications Symp. (IDEAS '98), pp. 68-77, 1998.

[3] Chen M.-S., Park J.-S., and Yu P.S., "Efficient Data Mining for Path Traversal Patterns," IEEE Trans. Knowledge and Data Eng., vol. 10, no. 2, pp. 209-221, Mar. 1998

[4] Creighton C. and Hanash S., "Mining Gene Expression Databases for Association Rules," Bioinformatics, vol. 19, no. 1, pp. 79-86, 2003.

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

[6] Georgii E., Richter L., Ru ¨ ckert U., and Kramer S., "Analyzing Microarray Data Using Quantitative Association Rules," Bioinformatics, vol. 21, pp. 123-129, 2005.

[7] Han J., Dong G., and Yin Y., "Efficient Mining of Partial Periodic Patterns in Time Series Database," Proc. Int'l Conf. on Data Eng., pp. 106-115, 1999.

[8] Han J. and Fu Y., "Discovery of Multiple-Level Association Rules from Large Databases," Proc. 21th Int'l Conf. Very Large Data Bases, pp. 420-431, Sept. 1995.

[9] Han J., Pei J., and Yin Y., "Mining Frequent Patterns without Candidate Generation," Proc. ACM-SIGMOD Int'l Conf. Management of Data, pp. 1-12, 2000.

[10] Lee S.C., Paik J., Ok J., Song I., and Kim U.M., "Efficient Mining of User Behaviors by Temporal Mobile Access Patterns," Int'l J. Computer Science Security, vol. 7, no. 2, pp. 285-291, 2007.

[11] Li H.F., Huang H.Y., Chen Y.C., Liu Y.J., and Lee S.Y., "Fast and Memory Efficient Mining of High Utility Itemsets in Data Streams," Proc. IEEE Eighth Int'l Conf. on Data Mining, pp. 881- 886, 2008.

[12] Li Y.-C., Yeh J.-S., and Chang C.-C., "Isolated Items Discarding Strategy for Discovering High Utility Itemsets," Data and Knowledge Eng., vol. 64, no. 1, pp. 198-217, Jan. 2008.

[13] Lin C.H., Chiu D.Y., Wu Y.H., and Chen A.L.P., "Mining Frequent Itemsets from Data Streams with a Time-Sensitive Sliding Window," Proc. SIAM Int'l Conf. Data Mining (SDM '05), 2005.

[14] Y. Liu, W. Liao, and A. Choudhary, "A Fast High Utility Itemsets Mining Algorithm," Proc. Utility-Based Data Mining Workshop, 2005.

[15] Martinez R., Pasquier N., and Pasquier C., "GenMiner: Mining nonredundant Association Rules from Integrated Gene Expression Data and Annotations," Bioinformatics, vol. 24, pp. 2643-2644, 2008.

[16] Pei J., Han J., Lu H., Nishio S., Tang S., and Yang D., "H-Mine: Fast and Space-Preserving Frequent Pattern

[17] Mining in Large Databases," IIE Trans. Inst. of Industrial Engineers, vol. 39, no. 6, pp. 593-605, June 2007.

[18] Pei, J. Han J., Mortazavi-Asl, H. Pinto H., Chen Q., Moal U., and M.C. Hsu, "Mining Sequential Patterns by Pattern-Growth: The Prefixspan Approach," IEEE Trans. Knowledge and Data Eng., vol.16, no.10, pp. 1424-1440, Oct. 2004.

[19] Pisharath J., Y. Liu, Ozisikyilmaz B., Narayanan R., Liao W.K., Choudhary A., and Memik G. MineBench NU- Version 2.0 Data Set and Technical Report, http://cucis.ece.northwestern.edu/projects/DMS/MineBench.html, 2012

[20] Shie B.-E., Hsiao H.-F., Tseng V., S., and Yu P.S., "Mining High Utility Mobile Sequential Patterns in Mobile Commerce Environments," Proc. 16th Int'l Conf. DAtabase Systems for Advanced Applications (DASFAA '11), vol. 6587/2011, pp. 224-238, 2011

[21] Shie B.-E., Hsiao H.-F., Tseng V., S., and Yu P.S.,, "Online Mining of Temporal Maximal Utility Itemsets from Data Streams," Proc. 25th Ann. ACM Symp. Applied Computing, Mar. 2010.