ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 5, Issue 4, April 2018

Infrequent Pattern Mining Techniques: A Review

[1] G. Sophana, [2] Dr. V. Joseph Peter
[1] Research Scholar, [2] Associate Professor
[1][2] Department of Computer Science, Kamaraj College, Thoothukudi.

Abstract: Data Mining is the process of extracting useful information or patterns from the data in large relational databases. In data mining, frequent pattern mining plays an important role for finding correlations among data. Similarly infrequent patterns are finding the uninteresting patterns that are rarely found in the database provide useful information. The aim of Infrequent Pattern Mining is to mine patterns which are not frequent that is the support value less than the threshold. Infrequent pattern mining (IPM) also plays a major role in the field of research and it has wide application domains such as medical, banking, biology, market basket analysis, telecommunication etc. Mining infrequent patterns from large dataset is challenging. In this paper we focus on study of various existing infrequent pattern mining techniques like Positive and Negative Association Rule (PNAR), Minimal Infrequent Itemset Mining (MIIM), Rare Association Rule (RAR), Pattern-Growth Paradigm and Residual Trees, Optimization Rule Based and Confabulation Inspired Association Rule Mining.

Keyword: - Data Mining, Infrequent, MINIT, PandNAR, RAR, CARM

I. INTRODUCTION

In recent years, the data mining technology has been developed rapidly. Data Mining extracts useful and hidden knowledge from large databases to discover existing and newer patterns. Data mining has gained advantage for decision making and behavioral analysis. Data mining is the technique of automatic finding of hidden, valuable and interesting patterns and relationships from huge volume of data stored in databases in order to help make better decisions. Discovering useful patterns hidden in a database plays an essential role in several data mining tasks. The knowledge obtained with help of data mining tools can be utilized for solving complex problems such as detection of fraud identification to enhance customer buying behaviour. As most of the users are not professionally trained to analyze the patterns of the data, data mining tool in such cases resolve the problem to identify frequent and infrequent patterns for better decision making.

A. Frequent Pattern Mining

Itemset mining was made on discovering frequent itemsets. Frequent itemset mining is a central part of data mining and distinctions of association. From large amount of data, Frequent Itemset is constructed by concerning some rules or association rule mining algorithms to calculate all the frequent itemsets. Association rule mining is of two phases finding frequent itemset and constructing rules.

B.Infrequent Pattern Mining

Frequent itemset mining characterizes data through frequent correlations among data, but in some situation where certain functions are to be minimized for determining rare correlations. Even though infrequent itemsets has obtained major usage in mining of negative association rules from infrequent itemsets, it is noticed that less consideration is for mining infrequent itemsets. Patterns that are rarely found in the database are considered to be uninteresting and are eliminated using the support measure. These patterns are known as infrequent patterns. An itemset or a rule whose support is less than the minimum support of threshold is an infrequent pattern. The infrequent patterns that are relate to rare cases are likely to be of great interest. The applications of infrequent pattern mining are identifying rare diseases, predicting failure of an equipment, and finding associations between the items purchased infrequently.

In the field of medical, by analyzing clinical dataset a user can discover rare patterns or trends that will assist doctors to make decisions about the clinical care.

Infrequent patterns can be used to detect unusual errors. For example, if {Fire = yes} is frequent, but {Fire = yes, Alarm = on} is infrequent, then the alarm system probably is faulting [5]. To detect such error the expected support of a pattern must be determined, so that if a pattern turns out to have a considerably lower support than expected, it is considered as an interesting infrequent pattern.

In the domain of market basket analysis, the indirect associations of buying products can be analyzed. It can be used to find competing items, such as desktop computers and laptops, which states that people whom buys desktop computers won't buy laptops [5].

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 5, Issue 4, April 2018

In data mining extracting specific pattern is an important task. Frequent itemsets find application in a number of real-life contexts. Most of the work has been done on finding frequent itemsets, but also infrequent itemset mining has demonstrated its own importance in web mining, bioinformatics, fraud detection, etc. The infrequent pattern mining is to discover itemsets whose frequency of occurrence in the analyzed data is less than or equal to a maximum threshold. In this paper we provide an overview of the different techniques related to infrequent patterns and gives the knowledge on the algorithms for mining infrequent patterns which are the basis for future research in the field of infrequent pattern mining.

II.TECHNIQUES

1. Positive and Negative Association Rule (PandNAR)

PandNAR is a method for efficiently mining of both positive and negative association rules in databases, and it focuses on identifying the associations among frequent itemsets. This approach is novel and different from existing research efforts on association analysis. Some infrequent itemsets are of very interest in this method. In [3] they had also designed constraints for reducing the search space, and had used the increasing degree of the conditional probability relative to the prior probability to estimate the confidence of positive and negative association rules. PandNARs may cause some problems, such as how to find a moderate degree of infrequent itemsets, how to discover PNARs properly, how to deal with the problem caused by single minimum support [3]. In [9] NARs of forms $A \Rightarrow \neg B$, $\neg A \Rightarrow B$, and $\neg A \Rightarrow \neg B$ are referred to as negative associations between itemsets when their support and confidence are lower than the user specified minimum support and minimum confidence thresholds, respectively. The support and confidence of such rules are difficult to calculate in a direct way. However, it can be computed using positive association rules. Despite these factors, the support confidence framework is the most popular approach used in association rules mining for pruning associations among items in a database. Many uninteresting rules may be generated, especially when mining PNARs instantaneously [11][13].

B. Minimal Infrequent Itemset Mining (MINIT)

The algorithm MINIT is for finding minimal infrequent or minimal concurrent item sets. Firstly, a ranking of items is organized by estimating the need of each of the items and then generating a record of items in rising order of support. In [3][10] the minimal infrequent itemsets are determined by using each item in rank order and iteratively calling MINIT on the maintained set of the dataset with regard to items using only those items which are superior rank than current items, after checking that each candidate of minimal infrequent items (MII) against the original dataset is performed. A system that can be utilized for superior-ranking items in the iteration is to preserve a "likeness" vector representing which items stay feasible at each level of the iteration[6][7].

The Minimally Infrequent Weighted Itemset is for creating only minimal infrequent patterns and the recursive mining in the MIWI Mining process is stopped once an infrequent itemset occurs, and it finds both the Infrequent Weighted Itemset and Minimally Infrequent Weighted Itemset. MIWI algorithm is, decreasing the computational time, better the efficiency of algorithm performance related to FPGrowth algorithm and reduces the creation of the Candidate sets.

C. Rare Association Rule (RAR)

The main objective of this method is to generate the rare rules which might give valuable information by extracting rare itemsets. The rare association rules are generated for mining of infrequent itemsets. This method is to take out rare association rules that stay hidden for traditional frequent itemset mining algorithms. When compared with other method this method finds strong but rare associations that are local regularities in the data are found. These rules are said to be "minimal Rare Itemset (mRI) rules". Apriori computes the support of minimal rare itemsets i.e. rare itemsets such that all proper subsets are frequent, mRIs are retained instead of pruning [4][14]. The mRIs form a generator set of rare itemsets that is all rare itemsets can be restored from the set of mRIs which have two merits. Firstly, they are highly informative in the case that they have an ancestor which is a producer itemset while adding up the resultant to give ways for a closed itemset. Secondly, the amount of these rules is minimal, that is the mRG rules comprise a dense illustration of all largely confident associations that can be taken from the least rare itemsets. The main advantage of this algorithm is that it restores all the minimal rare itemsets, but it fails to find all the rare itemsets.

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 5, Issue 4, April 2018

D. Pattern-Growth Paradigm and Residual Trees

The Pattern-Growth Paradigm is to discover minimally infrequent itemsets and it is to find minimally infrequent itemsets and it has no subset which is also infrequent. The Infrequent Pattern mining algorithm is for mining minimally infrequent itemsets [7].

The residual tree concept has been incorporated by using a variant of the FP-Tree structure which is known as inverse FP-tree [3]. The optimization of Apriori algorithm is performed in order to mine the minimally infrequent itemsets. Finally the presented tree is used for mining of frequent itemset as well as infrequent itemset [8][10].

E. Optimization Rule based algorithm

The algorithm for optimization of association rule mining determines the negative rule generation and as well as optimized the method of rule generation. This method used a multi-level multiple support (MLMS) of data table as binary values of 0 and 1 [12]. The divided process minimizes the examining time of database and this method works in the combination of genetic algorithm and MLMS. An algorithm has been given for mining interesting negative and positive rule from infrequent and frequent pattern sets[14]. The algorithm is divided into three stages. First it extracts frequent and infrequent pattern sets by incorporating apriori approach. Secondly positive and negative rule are generated. And finally prune redundant rule has been applied for interest measurements.

F.Confabulation-Inspired Association Rule Mining (CARM)

In [15] Confabulation-Inspired Association Rule Mining (CARM) uses a cogency inspired measure for producing rules. In this CARM algorithm rules are detected only by one file scan. Rule mining is performed in two steps. In the first step Knowledge acquisition and structure construction is done and in the second step Rule is generated by confabulation and cogency measure. Only one item consequent association rules are generated by this algorithm. Items are considered as symbols in this algorithm. After finding all frequent itemsets the algorithm generates all rules using their support and confidence. Knowledge acquisition consists of two modules, the axonal communication links between the two modules are used to store all domain knowledge, and the rule extraction is completed based on the strength of the communication links [16][17].

III. CONCLUSION

The presented work surveys various methods for extracting infrequent item sets of data. The major advantage for finding infrequent itemset was to advance the profit of rarely originated datasets in the transactions. Merits and demerits of each method are discussed to effectively differentiate the each methods functionality.

Table 1: Merits and Demerits of Infrequent Pattern Mining Techniques

Techniques	Merits	Demerits
PandNAR	Efficient mining	Patterns not reliable
MINIT	Better performance	Complexity increased
RAR	Mine frequent itemsets involving rare item	Non-interesting pattern can be found
PGP and Residual Trees	Improved performance	Better scalability is not achieved (maximal)
Optimization Rule Based	Space complexity reduced	Over Parameterized
Confabulation Inspired ARM	More efficient	Manually configured

REFERENCES

[1] Nikky Rai, Susheel Jain and Anurag Jain, "Mining Positive and Negative Association Rule from Frequent and Infrequent Pattern Based on IMLMS_GA," International Journal of Computer Applications, vol 77, September 2013.

[2] Ling Zhou, and Stephen Yau, "Efficient Association Rule Mining among Both Frequent and Infrequent Items," An International Journal of Computers and Mathematics with Applications," pp. 737-749, 2007.

[3] J. Jaya, and S.V.Hemalatha, "A Survey of Frequent and Infrequent Weighted Itemset Mining Approaches," International Journal of Innovative Research in Computer and Communication Engineering, vol. 2, Issue 10, pp. 6086-6090, October 2014.

[4] A.Varsur Jalpa, A., P. Desai Sonali, and B. Hathi Karishma, " Performance Analysis of Rare Itemset Mining Algorithms," Journal of Emerging Technologies and Innovative Research, vol. 2, Issue 2, February 2015.

[5] KalyaniTukaramBhandwalkar, and MansiBhonsle, "Study of Infrequent Itemset Mining Techniques," International Journal of Engineering Research and General Science," vol. 2, Issue 6, pp. 676-679, October 2014.

[6] Sonia Jadhav and G. M. Bhandari, "A Review on Efficient Mining Approach of Infrequent Weighted Itemset," International Journal of Advance Research in Computer Science and Management Studies, vol 2, Issue 11, November 2014.

[7] K.Hemanthakumar, and J.Nagamuneiah, "Study on Mining Weighted Infrequent Itemsets Using FP Growth," International Journal of Engineering and Computer Science, vol 4, Issue 6, pp.12719-12723, June 2015.

[8] Luca Cagliero and Paolo Garza, "Infrequent Weighted Itemset Mining Using Frequent Pattern Growth," IEEE Trans. on Knowledge and Data Engineering, vol. 26, no. 4, pp. 903-915, April 2014.

[9] Chris Cornelis, Peng Yan, Xing Zhang, Guoqing Chen, "Mining Positive and Negative Association Rules from Large Databases," IEEE, 2006.

[10] Shipra Khare, and Vivek Jain, " A Review on Infrequent Weighted Itemset Mining using Frequent Pattern Growth," International Journal of Computer Science and Information Technologies, vol. 5 (2) , pp. 1642-1647, 2014.

[11] Idheba Mohamad Ali O. Swesi, Azuraliza Abu Bakar, Anis Suhailis Abdul Kadir, "Mining Positive and Negative Association Rules from Interesting Frequent and Infrequent Itemsets," 9th International Conference on Fuzzy Systems and Knowledge Discovery, 2012.

[12] Rakesh Duggirala, P.Narayana, "Mining Positive and Negative Association Rules Using Coherent Approach," International Journal of Computer Trends and Technology, vol 4, Issue 1, 2013.

[13] R.Sumalatha and B. Ramasubbareddy, "Mining Positive And Negative Association Rules," International Journal on Computer Science and Engineering, vol. 02, no. 09, pp. 2916-2920, 2010.

[14] B.Kavitha Rani, K.Srinivas, B.Ramasubba Reddy, and Dr.A.Govardhan, "Mining Negative Association Rules," International Journal of Engineering and Technology, vol. 3 (2), pp. 100-105, 2011.

[15] T. Malathi, T. Senthil Prakash, P.V. Jothi Kantham, "Discovering Frequent Item Set Using Confabulation-inspired Association Rule Mining," International Journal on Engineering Technology And Sciences, vol. 3, Issue 5, pp. 156-160, May 2016.

[16] P. Kavya, and A. KalaiSelvi, "Mining Infrequent Itemset Using Association Rule," International Journal and Magazine of Engineering, Technology, Management and Research, vol. 4, Issue 7, July 2017.

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 5, Issue 4, April 2018

[17] Neha Srivastava, and Jyoti Gautam, "Prognosis of Disease that may Occur with Growing Age using Confabulation Based Algorithm," Defence Life Science Journal, Vol. 2, No. 4, pp. 399-405, October 2017.