Available Online at www.ijcsmc.com

International Journal of Computer Science and Mobile Computing

A Monthly Journal of Computer Science and Information Technology

ISSN 2320–088X
IMPACT FACTOR: 5.258

IJCSMC, Vol. 5, Issue. 4, April 2016, pg.391 – 396

Positive and Negative Association Rule Mining using Improved MOPNAR

Ms. S.G.Hiwase[1], Prof. R.R.Keole[2]
[1]Department of Computer Science and Engineering S.G.B.A.U, India
[2]Department of Information Technology S.G.B.A.U, India
[1] hiwase.suwarna@gmail.com; [2] ranjitkeole@gmail.com

Abstract— Data mining is one of the significant topics of research in recent years. Association rule is a method for discovering interesting relations between variables in large databases. Support and Confidence are the two basic parameters used to study the threshold values for each database. For the mining of positive and negative rules, a variety of algorithms are used such as Apriori algorithm and tree based algorithm. A number of algorithms have a good performance but produce large number of rules which were difficult to make decision and also suffered from multi-scan problem. In this paper, we propose an algorithm that mines positive and negative association rules without adding an additional measure and extra database scans. We propose a system to evaluate positive and negative quantitative association rules from E-Commerce Dataset, in order to accomplish this we are using MAPNAR and Improved MAPNAR association rule mining algorithm on the datasets. MAPNAR is a multi-objective algorithm used for mining which may produce large number of rules which puts overhead on time and space resources. To improve performance we propose an improved MAPNAR which is a combination of MAPNAR and Top K Association Mining. These are the rules with minimum confidence. We are trying to obtain rules that are easy to understand, provide good coverage of the dataset, and more efficient in time and space. Once these rules are generated they will be classified for analysis purpose. Most of the algorithms used for mining quantitative association rule generally focus on positive quantitative association rule without paying particular attention to negative quantitative association rule. The propose system gives study for generating negative and positive rule generation as demand of modern data mining techniques requirements. We are trying to develop the algorithm that is more efficient and flexible than the previous algorithms which are used for obtaining reduced set of positive and negative quantitative association rules.

Keywords— "Data mining", "positive and negative association rules"," quantitative association rules", "Firefly Algorithm", "support", "confidence"

I. INTRODUCTION

In the 1990s, the evolution of information technologies and especially the networks like the Internet enabled companies to easily record data from their customers. Since then, huge amounts of data have been collected and stored in the databases of many enterprises. Due to the fact that a lot of business intelligence is hidden in the data, the companies need tools to find out patterns and regularities. As many of the databases are very large containing

a huge number of tuples and attributes, efficient automated tools are necessary for acquiring useful information. Association rules are used to discover the relationships, and potential associations, of items or attributes among huge data. These rules can be effective in uncovering unknown relationships, providing results that can be the basis of forecast and decision. Therefore, the application and development of association rules is a popular area of data mining research [1]. Mining association rules is a very important issue in data mining, the general transaction database are stored in a huge number of transaction data, transaction records are records related to the project (item), contains user information and project or the time of the transaction; data mining techniques can from these large data mining association rules between the various projects. For example: 80% of students are opting for "C language" but also elective "data structures"[2]. Association rules have been extensively studied in the literature for their usefulness in many application domains such as recommender systems, diagnosis decisions support, telecommunication, intrusion detection, etc. Efficient discovery of such rules has been a major focus in the data mining research. From the celebrated Apriori algorithm there have been a remarkable number of variants and improvements of association rule mining algorithms Apriori is the first algorithm proposed in the association rule mining field and many other algorithms were derived from it [3]. It is very well known that mining algorithms can discover a prohibitive amount of association rules; Starting from a database, it proposes to extract all association rules satisfying minimum thresholds of support and confidence. For instance, thousands of rules are extracted from a database of several dozens of attributes and several hundreds of transactions [4].The generation of frequent itemsets requires more effort and the rule generations are discussed. Association rule mining on high dimensional data, now a day a high topic of research interest in many fields of data mining tasks. Mining association rules on these data bases is a challenging issue. The emergence of various new application domains, such as bioinformatics and e-commerce, emphasize the need for analysing high dimensional data. Many organizations have enormous amounts of data containing valuable information for running and building a decision making system. Extracting the value of that data is a big challenge [5]. A typical example of association rule mining application is the market basket analysis in which the behaviour of the customers is studied with reference to buying different products in a shopping store. The discovery of interesting patterns leads to important marketing and management strategic decisions. All the traditional association rule mining algorithms were developed to find positive associations between items. By positive associations, we refer to associations between items exist in transactions containing the items bought together. What about associations of the type: "customers that buy Coke do not buy Pepsi" or "customers that buy juice *do not* buy bottled water"? In addition to the positive associations, the negative associations can provide valuable information, in devising marketing strategies [6]. In this paper an algorithm for mining positive and negative association rules is proposed. The process of discovering association rules is decomposed into two steps: in step one, generate all the item combinations i.e. frequent itemsets whose support is greater than the user specified minimum support. In step two secondly, use the identified frequent itemsets to generate the rules that satisfy a user specified confidence. The generation of frequent itemsets requires more effort and the rule generations are discussed. Each item can be seen as a Boolean variable presenting the absence or presence of that item in the transaction or in the row [7].

II. RELATED WORK

Diana Mart´ın, Alejandro Rosete, Jes´us Alcal´a-Fdez propose MOPNAR, a new multiobjective evolutionary algorithm, in order to mine a reduced set of positive and negative quantitative association rules with low computational cost. To accomplish this, the proposal extends a recent multiobjective evolutionary algorithm based on decomposition to perform an

evolutionary learning of the intervals of the attributes and a condition selection for each rule, while introducing an external population and a restarting process to store all the nondominated rules found and to improve the diversity of the rule set obtained. Moreover, this proposal maximizes three objectives—comprehensibility, interestingness, and performance—in order to obtain rules that are interesting, easy to understand, and provide good coverage of the dataset. The effectiveness of the proposed approach is validated over several real-world datasets. It might produce large set of association rules giving less weightage to the important rules from user perspectives [8]. Mining To k association rule is the unique technique of mining limited number of rules, where *k* is the number of association rules to be found and is set by the user. The algorithm utilizes a new approach for generating association rules named rule expansions and includes several optimizations. Experimental results show that the algorithm has excellent performance and scalability, and that it is an advantageous alternative to classical association rule mining algorithms when the user want to control the number of rules generated [9]. Firefly Algorithm (FA) is among the most powerful algorithms for optimization. The Firefly Algorithm was developed by Yang is inspired by biochemical and social aspects of real fireflies and it was based on the idealialized behavior of the flashing characteristics of fireflies. Most fireflies produce short and rhythmic flashes are to attract mating partners (communication) and to attract potential prey. In addition, flashing may also serve as a protective warning mechanism [10].

III. PROPOSED WORK

Generally association rule mining algorithm generates a large number of rules which were different to make decisions and also suffered from multiscan problems. To limit the number of rules generated we developed a system which an extension to MAPNAR to mine only Top K rules. MAPNAR generates too many rules ignoring valuable information. Due to limited resources it becomes more expensive to handle such kind of problems. Our proposed system is designed in such a way that it will generate only Top K rules. K is the number of rules to be generated, the value of K specified by the user. This is system also uses several optimization to generate Top K rules. For analysis purpose we use the Firefly algorithm which classifies the rules in to two categories normal rules and outlier rules. It improves time and space efficiency. Normal rules are interesting and non-repeated rules. Outlier rules are the unwanted and repeated rules from the datasets. Finally normal rules are used for decision.

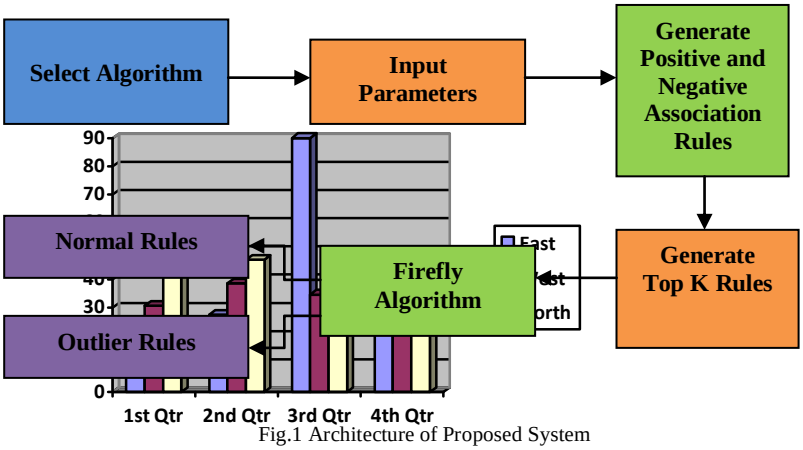

Fig.1 Architecture of Proposed System

Figure 1 shows details of system architecture. It shows different phases of the system. As a input we provide dataset and additional parameters like minimum confidence and value of K the number indicating the best top most rules to be calculated. The Improved MAPNAR initially considers minimum confidence and generates the rules containing each item on each side. The item on the left hand side is called antecedent and item on the right hand side is

called consequent. The list of rules along with number of rules with efficiency parameters is generated. Finally Firefly algorithm is used for classification. These rules will be saved in database and used for decision purpose.

IV. EXERIMENTAL RESULTS

Figure 2 shows the GUI of the proposed work. We need to specify the algorithm which is used for association rule mining, dataset to be considered, the files in which output to be stored, the minimum confidence value and value of K. Table I shows the time needed to calculate the number of rules generated by MAPNAR and Improved MAPNAR. We keep the value of minimum confidence. Time needed by the algorithm varies with minimum confidence. Figure 3 shows piechart for MAPNAR and Figure 4 shows piechart for Improved MAPNAR. The piechart shows the number of rules that are produced by the system after the classification algorithm is applied.

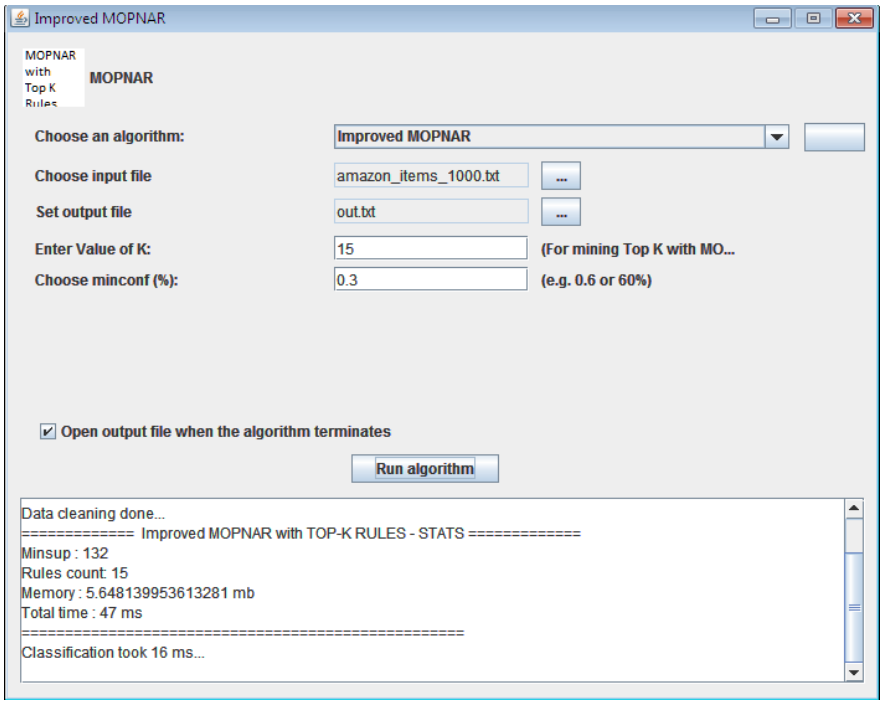

Fig. 2 Snapshot

TABLE I
COMPARING OUTPUTS

Input	Confidence	MAPNAR	Improved MAPNAR	
		Time(ms)	K	Time(ms)
1000	.2	453	15	93
1000	.3	407	15	109
2000	.2	407	15	109
2000	.3	94	13	79
2000	.5	109	18	73
5000	.2	172	17	125
5000	.3	141	15	125
5000	.4	141	15	125
10000	.2	297	18	156
10000	.3	844	22	156

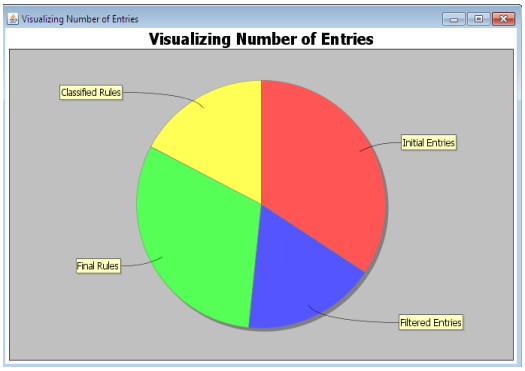

Fig. 3 Piechart for MAPNAR

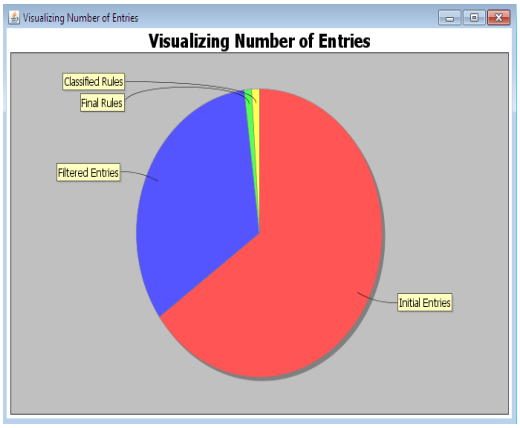

Fig. 4 Piechart for Improved MAPNAR

V. CONCLUSIONS

Generally algorithms for mining quantitative association rules can become very slow and generate an extremely large amount of results or generate too few results, omitting valuable information. This is a serious problem because in practice users have limited resources for analysing the results and thus are often only interested in discovering a certain amount of results, and fine tuning the parameters is time-consuming. To address this problem, we propose an algorithm to mine the top-*k* association rules, where *k* is the number of association rules to be found and is set by the user. Finally Firefly algorithm is used for classification purpose. Our aim is to improve time and space efficiency.

REFERENCES

[1] Data Mining Lecture Notes Available: http://wwwdb.stanford.edu/~ullman/mining/mining.html.
[2] Zhang Chunying, Liu Baoxiang, Wang Jing, "λ -Association Rules Extracted on Fuzzy Concept Lattice and Parameters Optimized", JCIT, Vol. 7, No. 15, pp. 294 ~ 302, 2012.
[3] B.Ramasubbareddy, A.Govardhan, and A.Ramamohanreddy. Mining Positive and Negative Association Rules, IEEE ICSE 2010, Hefeai, China, August 2010.
[4] Feng Yucai, "Association Rules Incremental Updating Algorithm", Journal of Software, Sept., 1998.
[5] Agrawal R., Srikant R., "Fast algorithm for mining association rules in large databases", Proceedings of 1994 International Conference on VLDB, 1994 pp.487-499.
[6] Data Mining: Concepts and Techniques Jiawei Han, Micheline Kamber.
[7] International Journal of Engineering Research and Applications (IJERA) ISSN: 2248-9622 www.ijera.com Vol. 3, Issue 4, Jul-Aug 2013, pp.1442-1448" General Survey on Multidimensional And Quantitative Association Rule Mining Algorithms".
[8] IEEE TRANSACTIONS ON EVOLUTIONARY COMPUTATION, VOL. 18, NO. 1, FEBRUARY 2014 "A New Multiobjective Evolutionary Algorithm for Mining a Reduced Set of Interesting Positive and Negative Quantitative Association Rules" Diana Mart´ın, Alejandro Rosete, Jes´us Alcal´a-Fdez, Member, IEEE, and Francisco Herrera, Member, IEEE.

[9] Philippe Fournier-Viger1, Cheng-Wei Wu2 and Vincent S. Tseng2 1Dept. of Computer Science, University of Moncton, Canada 2Dept. of Computer Science and Information Engineering, National Cheng Kung University" Mining Top-K Association Rules".

[10] X.S. Yang, "Firefly algorithms for multimodal optimization", SAGA 2009, Lecture Notes in Computer Science, 5792, 2009, 169-178.

