Improving Association Rule Mining with Apriori Algorithm and Charm

K. R. Badole¹ P. Kulurkar²
1M.Tech. Scholar ²Assistant Professor
1,2Department of Computer Science & Engineering
1,2V.I.T, Nagpur, Maharashtra, India

Abstract— Many other algorithm proposed after the introduction of apriori which retain the same general structure ,added several technique to optimize certain steps within algorithm, this paper improve the performance of apriori to improve high speed and better recommendation for any application by using to find support and confidence. Association rule mining originally proposed with its apriori algorithm to solve frequent itemset, mining problem as efficient as possible, which used to find out interesting pattern from database with the help of CHARM. Several business application used this technology for transaction databases. This paper include comprehensive survey and study between the performance of the describe approaches.

Key words: Apriori, CHARM, Association Rule, Frequent Itemset

I. INTRODUCTION

Data mining is a collection of techniques for effective automated discovery of formerly unknown, novel, valid, useful and understandable pattern in large databases. The pattern must be usable so that they can be used in the enterprise’s decision making process. Data mining is also an important step in the overall process of knowledge discovery which consist of various segments.

Association rule mining is a kind of data mining process. It is done to extract interesting correlations among items in the transaction database or other data repositories. For example an association rule: fruit => milk generated from the transaction database of a grocery store can help in formulating marketing strategy around the rule. Mining association rules from databases is a time-consuming process. Apriori like algorithms performs more number of scans and generates huge number of candidate keys. Apriori algorithm needs to scan the database multiple times. When mining a huge database, multiple database scans are costly. One feasible strategy to improve the efficiency of Apriori algorithm is to reduce the number of database scan .Apriori algorithm actually solves the complete association rule mining problem, of which mining all frequent sets was only the first, but most difficult phase.

The problem of mining association rules is to find all association rules that have support and confidence exceeding the user-specified threshold of minimum support (called MinSup) and threshold of minimum confidence (called MinConf ) respectively. Finding frequent itemsets has gained popularity because it has more number of applications viz., market basket analysis, catalog design, add-on sales, and store layout and customer segmentation. To find the frequent K-itemset, it is necessary to start the algorithm from frequent 1-itemset, 2-itemset……… to frequent K-itemset. The proposed method generates frequent K-itemset with a minimum support directly from the database.

Charm is an algorithm for generating closed frequent itemsets for association rules from transactional data. A closed frequent itemset is a subset of the corresponding frequent itemset. This subset is necessary and sufficient to capture all of the information about the frequent itemset. The closed frequent itemsets are the smallest representative subset of frequent itemsets without loss of information. For the formal definition of a closed frequent itemset, see. All possible association rules can be generated from the association rules created from only the closed frequent itemsets.

The first algorithm to generate all frequent itemsets and confident association rules was the AIS algorithm by Agrawal et al., which was given together with the introduction of this mining problem. Shortly after that, the algorithm was improved and renamed Apriori by Agrawal et al., by exploiting the monotonicity property of the support of itemsets and the confidence of association rule. The same technique was independently proposed by Mannila et al. Both works were cumulated afterwards.

This paper is organized as follows: In section II, we describe the related work. In section III present the Problem occurs in literature review. In section IV, we proposed the Genetic algorithm. Finally in section V presents the conclusion of this work.

II. RELATED WORK

This section presents a comprehensive survey, mainly focused on the study of research methods for mining the frequent itemsets and association rules with utility considerations. Most of the existing works paid attention to performance and memory perceptions.

In March-April 2013, Vimal Ghorecha [1] has proposed the comparative evaluation of Association Rule Mining algorithms with frequent item sets . In this paper he studied MaxClosure algorithm, FP-Growth, Closure, and he conclude that MaxClosure algorithm is better than all algorithms. FP-Growth was better than Apriori algorithm , Closure was better than FP-Growth and MaxClosure is better than Closure algorithm.

In Feb 2014, Thabet Slimani and Amor Laazez, [2] have proposed the efficient analysis of pattern and association rule mining approaches.In this paper he studied different kinds of frequent pattern itemset mining i.e. Closed, Maximal, Constrained, Approximate, Near-match and Top-k frequent itemset. He also studied kinds of Association Rule Mining and Pattern Mining Approaches. In this paper, he had presented a brief overview of the current status and future directions of frequent mining pattern. They conduct a deep research that based on several critical issues so that domain had factual existence and deep impact in data mining applications.

In June 2012, Saravanan Suba and Chistopher.T [3] have proposed a study on milestones of association rule mining algorithms in large databases . In this paper he observe that existing association rule mining algorithm generate lot of rules for future predictions and some of them
had not actionable. Those rules should be filtered to get meaningful rules that are immediately actionable. So they integrate user domain expert knowledge, ontology concept and related interesting measures in the post processing step effectively than existing methods in order to reduce the number of rules. So the founded rules would applied to improve any business.

In November 2012, K.Rajeswari, V.Vaithiyanathan, Swati.Tonge & Rashmi Phalnikar [4] have proposed mining association rules using hash table. In this paper, they observe that the use of hashing technique improves retrieval of item sets, thereby improving overall efficiency of that proposed algorithm. The application of frequent pattern mining varies from bioinformatics, web mining, software bug detection and analysis and improves the performance of XML management systems.

In 2009, S. Shankar and T. Purusothaman [5] have presented numerous efficient algorithms used in the literature for mining frequent itemsets and association rules and that all about to improve performance and memory uses. T.Y. Lin, Xiaohua Hu and Eric Louie [6] have presented a fast association rule algorithm which was based on bitmap and granular computing. In this paper they present that Bit-Association Rule doesn’t follow the generation and test strategy of Apriori algorithm and adopts the divide and conquer strategy, therefore avoids the time consuming table scan to find and prune the itemsets, also all the operations of finding large itemsets from the datasets performed the fast bit operations based on corresponding granular and it helped to improve performance of association rule algorithm and were very promising for data mining applications. They also presented parallelism that was another crucial aspect of DSS and data mining performance.

In June 2011, Bay Vol and Bac Le2 [7] have presented a frequent closed itemsets lattice-based approach for mining minimal non-redundant association rules. In this paper they have proposed that to reduce a lot of time for generating rules. First they used CHARM.L for building FCIL and then based on FCIL, an algorithm for fast generating MNAR had proposed. After that Experimental results show that the proposed algorithm is much faster and more efficient than frequent itemsets lattice based algorithm in the mining time.

Chris Cornelis, Peng Yan, Xing Zhang, Guoqing Chen [8] have proposed the mining positive and negative association rules from large databases. In this paper, they have studied emerging topic by cataloguing, and identifying problems with, existing negative AR definitions and mining approaches, and proposed a new Apriori based algorithm (PNAR) that exploits the upward closure property of negative association rules. They conclude that PNAR was very efficient in finding all positive and negative ARs within a reasonable time framework.

III. PROBLEM DEFINITION

After doing comparative study of various papers, it observed that the apriori algorithm is failed for various concept. Due to failing of apriori algorithm without using CARM, its need to be improve. So in future it need to be improve. And after improving this apriori algorithm automatically its performance will be improve with high speed which will beneficial for time reducing method which will also saving memory without scanning large dataset with accuracy to apply accurate rules and better recommendation for any application.

IV. OBJECTIVES OF THE STUDY

- Collection of data for processing.
- Reducing data using convert and process
- Applying apriori algorithm on reduced data.
- Improvement in apriori algorithm using charm algorithm.

V. PROPOSED WORK

This paper improving Apriori algorithm is applied over the rules fetched from Apriori association rule mining for web application. The proposed Algorithm is Reducing the data using convert and process then find out the support and confidence values to apply Apriori algorithm to find the frequent itemsets with the minimum support. Improving apriori algorithm using charm performance analysis between apriori and improved apriori algorithm.

VI. CONCLUSIONS

This paper proposed the method of finding frequent dataset from large databases by improving apriori algorithm with the help of CHARM. Experimental result show that proposed algorithm used this approach for any application and perform more efficiently. The proposed algorithm is more efficient, improving high speed and better recommendation for any application for dealing with frequent items due to its improving apriori algorithm.

REFERENCES

[8] Chris Cornelis, Peng Yan, Xing Zhang, Guoqing Chen,” Mining Positive and Negative Association Rules from Large Databases”. All rights reserved by www.tjsrd.com