The two main steps of Apriori algorithm are:

1. Candidate generation
2. Pruning

Candidate generation is basically a form of networking. Grid Computing is a form of data grids that is used for distributed and large scale cluster computing and it is a form of network distributed parallel processing. With the help of data grids in Grid Computing it is easy to discover or generate frequent itemsets [2].

The premise of this paper is to discover frequent patterns by the use of data grids in WEKA 3.8 environment. Workload imbalance occurs due to the dynamic nature of the grid computing hence data grids are used for the creation and validation of data. Association rules are used to extract the useful information from the large database. In this paper the researcher generate the best rules by using WEKA 3.8 for better performance. WEKA 3.8 is used to accomplish best rules and implementation of various algorithms.

KEYWORDS: Grid Computing, Association Rule Mining, Apriori Algorithm, WEKA 3.8, Visualization tools.

1. INTRODUCTION

1.1 Grid Computing

Grid Computing is considered as one of the promising platform for data and computation intensive applications like data mining. A grid is basically defined as data grids that are useful for hardware as well as software that provides dependable, consistent and transparent access for large scale distributed resources and shared by various multiple domain organizations in order to provide support for wide range of applications with the quality of service [1].

Grid Computing is a collection of machine learning and visualization tools for easy access with graphical user interface (GUI) for data mining algorithms. The version of Weka works with the modeling method and implemented in other programming languages [7].

Association Rule Mining is one of the important technique among the data mining techniques that is used for finding the interesting correlation, frequent patterns, associations or structures among the voluminous and transactional database. Usually Association Rule Mining is widely used in areas like telecommunication networks, marketing and inventory control etc. By using Association Rule Mining minimum support and confidence can easily be defined from a database [3].

Association Rule Mining has basically two problems:

Firstly find those item sets whose occurrence exceeds a already predefined threshold in database. These item sets are called frequent item sets. This problem can further divided into two subproblems candidate large itemsets generation process and frequent itemsets generation process.

Second problem is to generate Association Rules related with those frequent itemsets with the constraints of minimal confidence [4].

2. APRIORI ALGORITHM

Apriori Algorithm is used to operate on large database that contains transactions that is collection of items. Each transaction is a set of items or itemset. A threshold is already defined in apriori algorithm. Itemsets in the transaction must be subset of that threshold value [5].

Let D be a dataset. If p and q are itemset such that q is a subset of p then support of q is greater or equal to the support of p. All for the need of itemsets to be frequent all its subsets must be in frequent manner. Algorithm makes multiple passes over the data. In the every pass item set should be frequent and discovers frequent itemset of next bigger size. In the first pass it generates all the frequent-1 itemsets. After this by using a second pass with the combination of first pass it generates frequent-2 itemset by determining their supports. Similarly second pass is used to generate frequent-3 item set with their support. The candidate generation in the apriori algorithm reached to the mth passes or an m itemset is considered as candidate only if all (m-1) itemsets contained in it [6].

The two main steps of Apriori algorithm are:

The Join step: To find Lk a candidate k-itemsets is generated by joining Lk-1 with itself. This set of candidate is denoted by Ck.

• The prune step: In the prune step, delete all the itemsets related to the Ck, where (k-1) subset of C is not in Lk-1.
• Pseudo code of Apriori Algorithm:
In this research paper data is obtained from “Open Flight Airport Database”. In this data includes (airports, train stations and ferry terminals). It contains 5888 airlines. Each entry in the dataset includes information of (Airport ID, Name, City, Country, Time zone, DST, Departure Time and Arrival Time).

The attributes are retrieved from Airports- extended.dat and are implemented in Weka 3.8 by obtaining the best results.

5. EXPERIMENTAL RESULTS

Experimental results are based on the performance in Weka 3.8. As Airline dataset is used for obtaining the best rules. In the below figure1 researcher get results by using a visualization tool for obtaining best rules. Various attributes are used for a particular dataset entry that will give results based on the Apriori algorithm.

In this paper researcher is using Apriori algorithm that is used to calculate Association rules with minimum support and minimum confidence. By using the Apriori algorithm in Weka 3.8 researcher will get the effective results with the increasing performance in appropriate manner.

In the figure1 given below 25 attributes are used in the Airline dataset entry. Thus it will show the results which concludes process time decreases and threshold value increases and obtain better results. In the given results it proved that association have a minimum support as well as minimum confidence and rules produced by the association will also generate better performance results.

For the frequent patterns Association rules are created by analyzing the data using minimum support and confidence for a particular relationship. For the creation of association rules Apriori algorithm is used to perform the operations in Weka.

Implementation in Weka is basically done in arff format or dataset that is used for the best Association rules. First of all open the file in the preprocess segment after that it will show two phases. In the left phase it shows all the attributes with their names. In the right phase it will show the items of the attributes in the upper portion and in the lower portion it will show the graphical representation of the attributes.

Next step is to open the associate tab. In the associate tab selection of algorithm is done. In this research paper the researcher is using Apriori algorithm for finding the frequent patterns with the help of Association Rules. In the associate tab the value of the car must be true because it will mine class Association rules instead of (general Association Rules). Then after pressing the start button it will show the best Association Rules. In the Associate tab figure2 the results are shown which gave the information of 10 best Association rules by applying Apriori algorithm.

In the figure3 it shows the Minimum support and Minimum confidence with the number of cycles performed in Associate tab by using Apriori Algorithm. After this it will generate the best rules by using this Associate tab and increase the performance of the data grids or frequent patterns that is created by the Association Rules.

By the Scheduling method it will increase the performance and shows best rules for the interpretation of the results with minimum support and confidence. The next step is to generate the best rules create by the Associate tab in the Weka 3.8. Ten best rules found for each class with the different values of Confidence, Lift leverage and Conviction. After performing the algorithm the results are the following:

6. ANALYSIS AND INTERPRETATION

In the above figure finally best rules are found with the help of visualization tool for the minimum support and confidence by generating frequent patterns.

7. CONCLUSION

In this research paper the researcher used Weka 3.8 with the help of Apriori algorithm on a large scale dataset of Airlines in terms of the departure time and arrival time for increasing the performance in an effective manner.

In this research paper performance in both the cases by using Apriori algorithm as well as by using Weka 3.8 so Weka 3.8 produce the best Association rules for that dataset after performing the operations on it. Implementation of the Apriori algorithm is most compatible and the observed results show the effective use of the dataset and produce best rules after performing the operations on it.

REFERENCES


