EFFICIENT PROCESSING OF AJAX DATA USING MINING ALGORITHMS

Mudra Doshi¹, Bidisha Roy²

¹(Student, M.E (Computer Engineering), Faculty of Computer Engineering, St. Francis Institute of Technology, Mumbai University, India)
²(Associate Professor, Department of Computer Engineering, St. Francis Institute of Technology, Mumbai University, India)

ABSTRACT

Knowledge discovery is an important process in data mining wherein data can be analyzed from different perspectives and summarized for future use. One of the most widely used data mining process is association rule mining. Association rules are created by analyzing data for frequent patterns and by using the criteria support and confidence to identify the most important relationships. However, there are some other challenging rule mining topics like negative association rule mining. In this research, a rule mining approach has been proposed that provides efficient solution using positive and negative association rule computation on Asynchronous JavaScript and XML (AJAX) data. In association rule mining rules are created and important relationships identified by analyzing data for frequent pattern using criteria support and confidence. A Horizontal Tree Approach is proposed for efficient data processing. The performance of this approach is compared with Apriori and FP-Growth algorithm.

Keywords: Association rule mining, Database, AJAX, Apriori Algorithm, Frequent Pattern-Growth (FP-Growth) Algorithm, Horizontal Tree Approach.

I. INTRODUCTION

Data mining has become extremely important for business domains like marketing, financing and telecommunication. This has become possible because of the development of database technology and systems in recent years [1]. Data mining techniques is used for data processing. Operations performed on the data such as collection, use or management is called data processing. A few examples are, a shop keeper asking customers to fill in an answer slip for data process; a hotel offering the possibility of online reservation also processes data if it requires guest names, the dates
of their stay and their credit card number. Association rule mining is a data mining technique that finds frequent patterns or associations in large data sets. In order for information to be valuable, it must have the following characteristics: Accurate, Complete, Flexible, Reliable, Relevant, Simple, Timely Retrievable, and Verifiable [2]. An important application area of mining is in the field of Data processing Outsourcing which is helpful in a various divisions of BPO industry for instance, services providers as well as BPO professionals. Data Processing Outsource Services range from: Data conversion, Data entry, Word processing, Image processing, Forms processing, Survey processing, Database management, and Script processing [3].

The most widely used association rule mining algorithm is called Apriori algorithm [4]. This algorithm is easy to implement but slow due to the many number of passes are required to run over the data set. Therefore, another fast rule mining algorithm, FP (Frequent Pattern)-Growth, has been proposed [5]. There are two most important improvements in FP-Growth. First, FP-Growth algorithm uses FP-Tree data structure. FP-Tree is the compressed form of the database, providing memory savings. Also there is no candidate set generation in FP-Growth which makes the overall algorithm fast [1]. Association rule mining is understood as positive association rule mining. Positive association rule is stated as ‘‘if person buy the product like bread and milk, then that person is likely to buy butter at the same time’’ [5] [6]. Researchers have recently focused on finding alternative patterns like negative association rule mining. The following example illustrates negative association rules: ‘‘birds can fly is a well-known fact, but penguins cannot fly although they are birds’’ [6].

AJAX data stream is another challenging research area for data mining. AJAX is widely used to transmit and store the data. AJAX is an important approach for improving rich interactivity between Web server and end users. The structured data in AJAX Web pages cannot be extracted easily due to its asynchronous loading [7]. This research proposes a rule mining methodology that mines positive and negative association rules on AJAX data streams as a service, efficiently and securely. The proposed system is based on the Horizontal Tree approach with positive negative association rule mining.

The remanent part of this paper is arranged as follows. Section II will give the detailed information of association rule mining and data mining algorithms. In section III, the proposed work and overview of entire system is discussed. Section IV gives the information about implementation detail and experimental results, and finally the paper concludes in section V.

II. REVIEW OF LITERATURE

This section will give the background knowledge and related works about the association rule mining and data mining algorithms.

The well-known, examples of association rule mining is market basket analysis. It has emerged as the next step in the evolution of retail merchandising and promotion. Besides market basket analysis, association rule mining is also applicable to other domains like marketing [8], financing [9] and telecommunication [10]. An association rule is an implication of the form X ⇔ Y, where X and Y are item sets where X is called as the antecedent and Y is called as the consequent of the rule. The strength of an association rule can be measured with its support and confidence value. The support value of an item set is the proportion of transactions in the data set which contain the item set. The confidence value of a rule indicates its reliability. The other type of association rule is negative association rule. A negative association rule can be illustrated by the following example: customers who buy product X, but not product Y. The search space in negative rule mining is much bigger than that in positive rule mining. The definition of a negative association rule is similar to that of positive one. The only difference is that in a negative association rule, the antecedent or the consequent part of the rule is negated. The support and confidence values of a negative rule are of the form X ⇔ ¬ Y [1].
Another most well-known association rule mining algorithm is Apriori algorithm [4]. This model, basically, divides the rule Mining Process into two basic steps. In the first step, the algorithm generates the 1–k large item sets where k is the count of separate items in the transactions. After the candidate item sets are generated, the algorithm finds the frequent large item sets that have support values greater than the predefined minimum support value. The next basic step generates the association rules from the frequent large item sets which satisfy the minimum confidence constraint [1]. The advantage of the Apriori algorithm is it’s easy to implementation. However, disadvantage of Apriori algorithm are that it requires too many scans over the database to find rules and it leads to high CPU usage as more search space is needed and by using that increasing the I/O cost. Because of the iterative scans, it is not suitable for data stream mining in which data should be scanned only once.

FP-Growth is another most popular mining algorithm [11]. The Mining Process of the FP-Growth algorithm was divided into two steps as follows. First the FP-Tree is constructed as explained below. In order to find the support value of each item, the data set is first scanned. The infrequent items are eliminated, because they do not have importance in the Mining Process. The frequent items are sorted in decreasing support value. After that, the data set is scanned once more to construct FP-Tree. The first transaction is read and the nodes are created. Also, the support value of each item is set to 1. If the transactions do not contain common prefixes, the process is continued with creating new nodes. Otherwise, if there are transactions with common items, their paths are overlapped partially or fully. Because of overlapping paths, support values of common items are increased by 1 and support values of the others are set to 1. The process is continued until all transactions have been mapped. Lastly, frequent item sets which have higher support value than the user-specified support threshold are mined without candidate set generation [1].

In [1] authors propose a data mining algorithm PNRMXS_Growth (Positive Negative Rule Mining on XML Set Growth). This algorithm is divided into three steps are shown in Fig. 1:

![Flowchart of the PNRMXS_Growth](image)
A. Data Transformation

This is used to hide the data from the service provider by using encryption technique i.e. one-to-one mapping technique. By using this technique, mining algorithms can be applied with 100% accuracy [1].

B. Mining Process

It is applied on transformed XML data streams and uses landmark windows data processing model which is based on FP-Growth algorithm approach. In this approach the data stream is processed block by block and each block contains the same number of transactions. PNRMXS mines both positive and negative association rules at the same time [1].

C. Data Re-Transformation

In Mining Process, the algorithm finds the positive and negative association rules with transformed items and sends them back to the data owner. In this step re-transform the items using the mapping table generated [1].

Another popular approach is the Horizontal Tree Approach [12] [13], which makes database searching much faster than other algorithms. Previously implemented algorithms had the drawback that each time an alphabet repetition occurs the weight for the corresponding alphabet is incremented accordingly. So, all the other algorithms took more time than the Horizontal Tree Approach. This research focuses on Horizontal Tree Approach using by AJAX where AJAX is not a new programming language, but a new way to use existing standards.

The Horizontal Tree Approach creates a tree like structure which is horizontally aligned and to the right of previous node. It will retrieve the data from the database horizontally i.e. row-by-row. The example of Horizontal Tree Algorithm is shown in Fig. 2. If user want to search from database like word “RANDOM”. So first user type “AN” it will give next possible word like “ANT” and “AND”. Then user is fixed the word “AND”. So database will give next possible word i.e. “RAND” and “NAND” likewise it will search from the database. AJAX dataset gives the next probable word from database.

III. PROPOSED WORK

The aim of this work is to develop a system which provides faster data searching from database. There are different data mining algorithm like Apriori Algorithm [4], K-Means Algorithm [8], and Horizontal Tree Approach [12] [14] [15]. There are some disadvantages in Apriori, K-Means and FP-Growth as mentioned in introduction and literature survey. To remove the disadvantage and faster access to the data stream from the database the use of Horizontal Tree Approach is proposed. Also Horizontal Tree Approach has not been used in this area of research. There are two types of approach i.e. Horizontal and Vertical Tree Approach. Vertical Tree Approach will retrieve the data
column wise, which is not useful as data in a database is stored row wise. So, Horizontal Tree Approach is used for the process. It will retrieve the data row wise as explained in the example in Fig. 2. Till now authors have proposed algorithms to be used on XML data streams but not on AJAX data streams. So, AJAX is new technology on which is Horizontal Tree Approach has not yet been applied, is being used. AJAX also provides fast access to database. A centralized database can be made, for different categories of data for e.g. doctors, pharmacist, engineers, researchers etc. By using this approach faster data processing will be done, due to reduced code length.

Fig. 3: Flowchart of the PNRMAS[16]

The proposed data mining algorithm PNRMAS (Positive Negative Rule Mining on AJAX Set) is shown in Fig. 3. This process works on AJAX Dataset and applies the Horizontal Tree Approach for the mining process. Mining is applied on AJAX data streams using Horizontal Tree Approach. In this approach the data stream is processed block by block and each block contains the same number of transactions. PNRMAS mines both positive and negative association rules at the same time. Here, transformation technique is not used because it increases the retrieval time due to the encryption and decryption technique to retrieve the data from database. Apriori Algorithm takes the data from the static database i.e. the content that is stored in the database that will be shown as in the cookie. Whereas FP-Growth and Horizontal Tree Algorithm will take the data from dynamic database i.e. whatever the user wants to search can be searched from the database dynamically.

IV. EXPERIMENTAL RESULTS

The proposed approach has many modules. The first and the most important step is preprocessing on AJAX [17] Data. Next module is based on mining the data using three algorithms. Eclipse tool, WAMP server, Java 1.7 and Apache Tomcat Server have been used for implementing the proposed approach. The proposed system provides faster retrieval as compared to those given by previous authors.

Front end of the project which is used to give the link to connect all the three types of algorithms. Dummy database of product camera from where data is retrieved and stored in database. There are 150 entries in the database from where data is retrieved and this data is then processed and retrieval time of all three algorithms is determined. Implementation of three algorithms is shown below,
Fig. 4 shows the front end of the project which is used to search the product from the database using Apriori Algorithm. It takes 646ms to search the word “Camera Cannon EOS 600D SLR” from 150 entries. Same way FP-Growth and Horizontal Tree is implemented. FP-Growth will take 470ms time to search the same word and Horizontal Tree takes 180 ms to run the algorithm and search the word.

Comparison of three algorithms is based on retrieval time of the algorithm. Table I given below shows the retrieval time of different data mining algorithm i.e. Apriori, FP-Growth and Horizontal Tree Algorithm. Table I shows the results of number of entries used on different data mining algorithms. Here, product “Camera Cannon EOS 600D SLR” is searched from the database using three algorithms and retrieval time is determined in ms.

<table>
<thead>
<tr>
<th>No. of Entries</th>
<th>Apriori Algorithm (ms)</th>
<th>FP-Growth Algorithm (ms)</th>
<th>Horizontal Tree Algorithm (ms)</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>1198</td>
<td>779</td>
<td>338</td>
</tr>
<tr>
<td>30</td>
<td>727</td>
<td>234</td>
<td>141</td>
</tr>
<tr>
<td>50</td>
<td>582</td>
<td>229</td>
<td>129</td>
</tr>
<tr>
<td>100</td>
<td>1168</td>
<td>331</td>
<td>186</td>
</tr>
<tr>
<td>150</td>
<td>646</td>
<td>470</td>
<td>180</td>
</tr>
</tbody>
</table>

V. CONCLUSION

The proposed approach has many modules like preprocessing of AJAX Data, application of association rule on AJAX Data set. The performance parameter is primarily based on execution time (ms) to retrieve data which checks on same AJAX data streams. The proposed work aims at reducing the time and giving the next probable word. Hence, by using this, the retrieval time will be reduced. Performance analysis is done on the basis of retrieval time of three algorithms i.e. Horizontal Tree Approach, FP-Growth and Apriori algorithm. The proposed system is beneficial for faster retrieval of data from the database using Horizontal Tree Approach. Data processing using data mining on AJAX data stream as a project has a potential impact on data retrieval. Various issues related to data processing using data mining are addressed. Using Horizontal Tree Approach on AJAX provides
useful information that the client requires and has been proved to work faster than FP-Growth algorithm and Apriori algorithm.

VI. REFERENCES


