

Enhanced Data Processing Using Positive Negative Association Mining on AJAX Data

Sumayya Khan, Shrikant Lade

Information Technology R.K.D.F. College Bhopal, India

Information Technology R.K.D.F. College Bhopal, India

Abstract: Knowledge discovery is the process of analyzing data from different perspectives and summarizing it into useful information. [1] Association rule mining is a data mining process used widely in traditional databases to find the positive association rules. Association rules are created by analyzing data for frequent patterns and 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. Organizations want to concentrate on their own business and outsource the rest of their work. This approach is named "Software as a Service Concept" and provides lots of benefits to data owner, but, at the same time, brings out some security problems. In this research, a rule mining approach has been proposed that provides efficient and secure solution to positive and negative association rule computation on AJAX data streams in software as a service (SaaS) concept. It is important to note that SaaS typically encapsulates enterprise as opposed to consumer-oriented web-hosted software. Software as a service (SaaS) is a way of delivering applications over the Internet—as a service. Instead of installing and maintaining software, you simply access it via the Internet, freeing yourself from complex software and hardware management. SaaS applications are sometimes called Web-based software, on-demand software, or hosted software. By using [3] AJAX, we get the search result in the form of semantic.

Index Terms: Association rules mining, Ajax data stream, horizontal tree approach, Apriori algorithm

I. Introduction

Due to the development of database technology and systems in recent years, the importance of the data mining has been increased rapidly for business domains like marketing, financing and telecommunication. [4] Data processing is done by using data mining techniques. Data processing means any operation performed on the data such as collection, use, management or disclosure. 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. Association rule mining is a data mining task that discovers relationships among items in a transactional database. Data Processing Outsource Services ranging from: Data conversion, Data entry, Word processing, Image processing, Forms processing, Survey processing, Database management, and Script processing.

The most popular association rule mining algorithm is called [5] Apriori algorithm. This algorithm is easy to implement but slow due to the lots of passes over the data set. Therefore, another and fast rule mining algorithm, FP (Frequent Pattern)-Growth, has been proposed. There are two main 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. Association rule mining is understood as positive association rule mining. Positive association rule is stated as "if A occurs in a transaction, then B is likely to occur in the same transaction". Researchers have recently focused on finding alternative patterns like negative associations. The following example illustrates negative association rules: "birds can fly is a well-known fact, but penguins cannot fly although they are birds". Data streams arrive continuously with high speed and contains large amount of data so that fast processing of the data is very important. Also, because of fast data flow, algorithms may scan the data set only once. There are few data stream mining algorithms. 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. Since managing the large database systems is an expensive and complicated task, organizations want to concentrate on their own professions and they prefer to outsource their data centric works. The outsourcing of database processes like administration, backup and optimization to a service provider is named as "software as a service". This concept provides lots of benefits to data owner. On the other hand, these arise some problems about the security of data. Therefore, providing security has a vital importance. 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.

II. Basic Concepts

2.1 Association Rule mining:

The problem of association rule mining is defined as: Let $I = \{i_1, i_2, \dots, i_n\}$ be a set of n binary attributes called items. Let $D = \{t_1, t_2, \dots, t_m\}$ be a set of transactions called the database. Each transaction in D has a unique transaction ID and contains a subset of the items in I . A rule is defined as an implication of the form $X \Rightarrow Y$ where $X, Y \subseteq I$ and $X \cap Y = \emptyset$. The sets of items (for short itemsets) X and Y are called antecedent (left-hand-side or LHS) and consequent (right-hand-side or RHS) of the rule respectively

2.2 Negative Association Rule Mining:

The traditional definition of itemset is maintained (so $X, Y \subseteq I$), and to each positive rule $X \Rightarrow Y$ correspond three negative ones, $X \Rightarrow \neg Y$, $\neg X \Rightarrow Y$ and $\neg X \Rightarrow \neg Y$. A transaction t supports $X \Rightarrow \neg Y$ if $X \subseteq t$ and $Y \not\subseteq t$. Hence, the meaning of a rule like $\{i_1\} \Rightarrow \neg \{i_2, i_3\}$ is that "the appearance of i_1 in a transaction t induces that i_2 and i_3 are unlikely to appear simultaneously in t "; hence a record containing i_1 and i_2 , but not i_3 , supports this rule. It can be verified that $\text{supp}(X \Rightarrow \neg Y) = \text{supp}(X \neg Y) = \text{supp}(X) - \text{supp}(XY)$ for $X, Y \subseteq I$, and similarly support and confidence of the other kinds of negative association rules can be straightforwardly deduced from the corresponding positive itemset supports.

2.3 Apriori Algorithm:

Apriori is a classic algorithm for frequent item set mining and association rule learning over transactional databases. It proceeds by identifying the frequent individual items in the database and extending them to larger and larger item sets as long as those item sets appear sufficiently often in the database. The frequent item sets determined by Apriori can be used to determine association rules which highlight general trends in the database: this has applications in domains such as market basket analysis.

2.4 Ajax:

Ajax (an acronym for Asynchronous JavaScript and XML) is a group of interrelated web development techniques used on the client-side to create asynchronous web applications. With Ajax, web applications can send data to, and retrieve data from, a server asynchronously (in the background) without interfering with the display and behavior of the existing page. Data can be retrieved using the XMLHttpRequest object. Despite the name, the use of XML is not required and the requests do not need to be asynchronous.

Ajax is not a single technology, but a group of technologies. HTML and CSS can be used in combination to mark up and style information. The DOM is accessed with JavaScript to dynamically display, and allow the user to interact with, the information presented. JavaScript and the XMLHttpRequest object provide a method for exchanging data asynchronously between browser and server to avoid full page reloads.

III. Related Work

The most well-known, example 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, financing and telecommunication. An association rule is an implication of the form $X \Rightarrow Y$, 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 $X \Rightarrow \neg Y$.

Another well-known association rule mining algorithm is Apriori. 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. The advantage of the Apriori algorithm is that it's easy implementation. However, disadvantage of Apriori algorithm are, requires too many scans over the database to find rules and it leads to high CPU usage, more search space is needed and I/O cost. Because of the iterative scans, it is not suitable for data stream mining in which data should be scanned only once.

Another popular mining algorithm is FP-Growth. The Mining Process of the FP-Growth algorithm was divided into two steps is as follows. First the FP-Tree is constructed, 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. A data mining algorithm Positive Negative Rule Mining on AJAX Set Growth is divided into three steps as follows:

2.1 DATA TRANSFORMATION: - Which 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 is that rule mining algorithms can be applied with 100% accuracy.

2.2 MINING PROCESS: - It is applied on transformed XML data streams and used 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 transaction. Positive Negative Rule Mining on AJAX Set Growth mines both positive and negative association rules at the same time.

2.3 DATA RE-TRANSFORMATION:- In Mining Process, the algorithm finds the positive and negative association rule transformed items and sends them back to the data owner. In this step re-transform the items using the mapping table generated.

Another popular approach is the Horizontal Tree Approach, which makes database searching faster than the number of times a words repeated then the associative words provided with extra weight. This research focuses on horizontal Tree Approach by using AJAX where AJAX is a programming language which has new ways of using standards.

IV. PROPOSED SOLUTION

The aim of this work is to develop a system which provides faster data searching from data base. There are different data mining algorithm like Apriori Algorithm, K-Means Algorithm, and Horizontal Tree Approach. There are some disadvantages in Apriori, K-Means and FP-Growth as mentioned in introduction. To remove the disadvantage and faster access to the data stream from the data base propose to use Horizontal Tree Approach. AJAX is new technology on which is Horizontal Tree Approach has not been applied. AJAX also provides fast access to data base.

A centralized database can be made, for professions like doctors, pharmacist, or engineers, researchers, using Software as a Service. By using this approach faster data processing will be done, due to reduce code length. The structure of the process is given in the following flowchart.

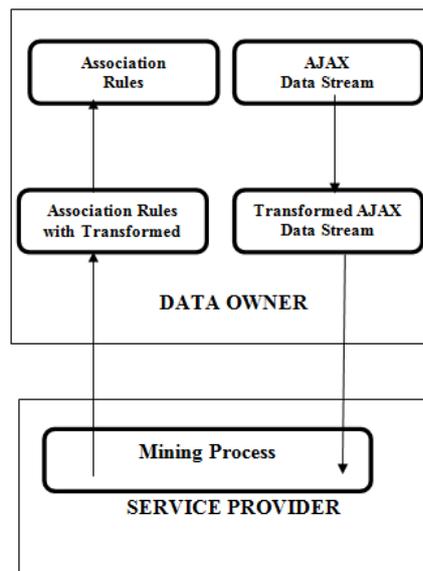


FIGURE 1: Flowchart of the Positive Negative Rule Mining on AJAX Set

V. CONCLUSION

It reduces the time and gives the next probable word. So, by using this the retrieval time will be reduced. This technique will also check the probability of the error and confidence of value of transaction. Also the probability of word detection will be checked. The performance parameters include execution time to retrieve data which checks on different AJAX data streams Acknowledgment

References

- [1]. Agrawal, R.; Imieliński, T.; Swami, A. (1993). "Mining association rules between sets of items in large databases". Proceedings of the 1993 ACM SIGMOD international conference on Management of data - SIGMOD '93. p. 207. doi:10.1145/170035.170072. ISBN 0897915925.
- [2]. Jesse James Garrett (18 February 2005). "Ajax: A New Approach to Web Applications". AdaptivePath.com. Archived from the original on 2 July 2008. Retrieved 19 June 2008.
- [3]. Samet Cokpinar, Taflan Imre Gunden, "Positive and negative rule mining on XML data streams in database as a service concept", Expert Systems with Applications, Vol.39, pp.7503-7511, 2012.
- [4]. "Data Processing", Available: http://wiki.answers.com/Q/What_are_the_importance_of_data_processing_in_the_society
- [5]. Rakesh Agrawal and Ramakrishnan Srikant Fast algorithms for mining association rules in large databases. Proceedings of the 20th International Conference on Very Large Data Bases, VLDB, pages 487-499, Santiago, Chile, September 1994
- [6]. Xu, Z. M. & Zhang, R., "Financial revenue analysis based on association rules Mining", In Asia-Pacific Conference on Computational Intelligence and Industrial Applications, Vol. 1, pp. 220-223, 2009.
- [7]. "Horizontal Tree Algorithm", Available: <http://sydney.edu.au/engineering/it/~comp5048/Lect2-trees.pdf>.
- [8]. PENG Hui-ling, SHU Yun-xing, "A New FP-tree-based Algorithm MMFI for Mining the Maximal Frequent Itemsets", Page 61-64, 2012 IEEE.
- [9]. 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 (FSKD 2012), 2012, IEEE.
- [10]. Goethals, B., Zaki, M., eds.: FIMI'03: Workshop on Frequent Itemset Mining Implementations. Volume 90 of CEUR Workshop Proceedings series.
- [11]. Honglei Zhu, Zhigang Xu: *An Effective Algorithm for Mining Positive and Negative Association Rules*. International Conference on Computer Science and Software Engineering 2008.