

A Survey Paper on Efficient Processing of AJAX Data using Mining Algorithms

Pooja P. Taral¹, Prof. Pravin D. Soni²

^{1,2}CSE Department, PRPCOET, Sant Gadge Baba Amravati University, Maharashtra, India

Abstract— Association rule mining deals with discovering associations between n itemsets in transactions. It is one of the most important data mining task. Association rules are created by analysing data for frequent patterns. Typical association rules consider only items involved in transactions. Such rules are referred to as positive association rules while negative association rules consider same items but also consider negated items which are absent from transactions. Negative association rules are very useful in some spatial problems and are capable of extracting some useful and previously unknown hidden information. In this research, a rule mining system has been proposed that provides efficient solution to positive and negative association rule computation on XML data using AJAX with the application of Top-K to get the top-k association rules.

Keywords— “association rule mining”, “positive rules”, “negative rules”, “AJAX”, “top-k association rules”.

I. INTRODUCTION

Association rule mining is one of the most popular data mining techniques to find associations among items in a set by mining necessary patterns in a large database. Typical association rules consider only items enumerated in transactions. Such rules are referred to as positive association rules. Negative association rules also consider the same items, but in addition consider negated items (i.e. absent from transactions). Negative association rules are useful in market-basket analysis to identify products that conflict with each other or products that complement each other. They are also very useful for constructing associative classifiers [7]. In this paper, we propose a methodology that mines positive and negative association rules with the application of Top-k association rules. In the knowledge management process, the techniques that are applied to discover significant knowledge from a large amount of data is called data mining techniques[10].

It was first introduced by Agrawal, Imielinski, and Swami (1993) in order to mine association rules on large transactional databases. Agrawal and Srikant (1994) have developed the most popular association rule mining algorithm called Apriori. 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-Growth, is proposed by Han, Pei, and Yin (2000). 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 [3].

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 [9]. Data mining on AJAX data stream is one of challenging research area. 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 [1]. This research proposes a rule mining methodology that mines positive and negative association rules on AJAX data streams as a service, efficiently with the application of Top-K algorithm.

With the enormous amount of data stored in databases, files, and other repositories, it is very important to develop powerful software or tool for analysis and interpretation of such data and for the extraction of interesting knowledge that could help in decision-making. The only answer to all above is ‘Data Mining’[12]. Data mining methodology consists of ten processes; translate a business problem into a data mining problem, select appropriate data, get to know the data, create a model set, fix the problems with the data, transform data, build models, assess models, deploy models and assess results. However, these steps can be further summarized into five main stages: the initial exploration incorporates with business and data understanding, data preparation, model building, evaluation and deployment. These are the entire steps taken to generate predictions[11].

II. BASIC CONCEPTS

A. 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 $X, Y \subset I$ 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 [4].

B. Positive Rules

Positive association rule is stated as “if A occurs in a transaction, then B will likely also occur in the same transaction” [3].

C. Negative Rules

Negative Associations illustrates negative association rules: “birds can fly is a well-known fact, but penguins cannot fly although they are birds”. Negative association rules provide valuable information to data owners, but there are few algorithms that are proposed in the literature for negative rule mining [3].

The traditional definition of itemset is maintained (so $X, Y \subset I$), 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 $\{i1\} \Rightarrow \neg \{i2, i3\}$ is that “the appearance of $i1$ in a transaction t induces that $i2$ and $i3$ are unlikely to appear simultaneously in t ”; hence a record containing $i1$ and $i2$, but not $i3$, supports this rule. It can be verify that $\text{supp}(X \Rightarrow \neg Y) = \text{supp}(X \neg Y) = \text{supp}(X) - \text{supp}(XY)$ for $X, Y \subset I$, and similarly support and confidence of the other kinds of negative association rules can be straightforwardly deduced from the corresponding positive itemset supports[4].

D. Top-K Association Rules

Several algorithms have been proposed for top-k association rule mining. However, most of them do not use the standard definition of an association rule. For instance, KORD finds rules with a single item in the consequent, whereas the algorithm of You et al. mines association rules from a stream instead of a transaction database. To the best of our knowledge, only TopKRules discovers top-k association rules based on the standard definition of an association rule (with multiple items, in a transaction database). TopKRules takes as parameters k and minconf , and it returns the k rules with the highest support that meet the minconf threshold. The reason why this algorithm defines the task of mining the top-k rules on the support instead of the confidence is that minsup is much more difficult to set than minconf because minsup depends on database characteristics that are unknown to most users, whereas minconf represents the minimal confidence that users want in rules and is therefore generally easy to determine [6].

E. 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 [1].

III. LITERATURE REVIEW

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”. Researchers have recently focused on finding alternative patterns like negative association rule mining [1]. Unexpected patterns and exceptional patterns are referred to as exceptions of rules, also known as surprising patterns. An exception is defined as a deviational pattern to a well-known fact, and exhibits unexpectedness. For example, while ‘bird(x) => flies(x)’ is a well-known fact, an exceptional rule is ‘bird(x), penguin(x) => ¬flies(x)’. This exception indicates that unexpected patterns and exceptional patterns can involve negative terms and therefore can be treated as a special case of negative rules [8].

To find Top-K association rules, a question is how to combine the concept of top-k pattern mining with association rules? For association rule mining, two thresholds are used. But, in practice minsup is much more difficult to set than minconf because minsup depends on database characteristics that are unknown to most users, whereas minconf represents the minimal confidence that users want in rules and is generally easy to determine. For this reason, we define “top-k” on the support rather than the confidence [5].

IV. PROPOSED WORK

Mining association rules is a fundamental data mining task. However, depending on the choice of the parameters (the minimum confidence and minimum support), current algorithms 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 analyzing 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. 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 [5].

To solve previous problems, proposed system include implementation of Top-k algorithm to find top-k association rules with the positive negative rules finding approach using AJAX and some advanced technology framework.

V. CONCLUSION

The proposed approach has many modules like collection of XML data stream for processing, development of XML parser for parsing the data stream, rule mining approach on the data stream to get all the rules from data stream, positive and negative rule mining approach, modification of existing algorithm and implementation of top-k rules finding approach such that performance of the system should get improve.

REFERENCES

- [1] Mudra Doshi¹, Bidisha Roy² “Efficient Processing Of Ajax Data Using Mining Algorithms”, *International Journal on Computer Science and Engineering*, Vol.5, Issue 8, August (2014), pp. 48-54.
- [2] “DataProcessing”, http://wiki.com/Q/What_are_the_importance_of_data_processing_in_the_society, 11th August 2013.
- [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] Sumayya Khan¹, Shrikant Lade² “Enhanced Data Processing Using Positive Negative Association Mining on AJAX Data”, *IOSR Journal of Computer Engineering*, p- ISSN: 2278-8727Volume 16, Issue 2, Ver. II (Mar-Apr. 2014), PP 15-18.
- [5] Philippe Fournier-Viger¹, Cheng-Wei Wu² and Vincent S. Tseng², “Mining Top-K Association Rules”, *National Cheng Kung University*.
- [6] Philippe Fournier-Viger¹ and Vincent S. Tseng², “Mining Top-K Non-Redundant Association Rules”, National Cheng Kung University, Taiwan.
- [7] R.Sumalatha¹, B. Ramasubbareddy², “Mining Positive and Negative Association Rules”, *International Journal on Computer Science and Engineering*, Vol. 02, No. 09, 2010, 2916-2920.
- [8] Xindong Wu¹, Chengqi Zhang² and Shichao Zhang³, “Efficient Mining of Both Positive and Negative Association Rules”, *ACM Transactions on Information Systems*, Vol. 22, No. 3, July 2004.
- [9] Yongmei Liu and Yong Guan, “FP-Growth Algorithm for Application in Research of Market Basket Analysis”, *Capital Normal University*, Beijing, China.
- [10] N.Kamalraj¹, Dr.A.Malathi², “Applying Data Mining Techniques in Telecom Churn Prediction”, *International Journal of Advanced Research in Computer Science and Software Engineering*, Volume 3, Issue 10, October 2013.
- [11] Khalida binti Oseman¹, Sunarti binti Mohd Shukor¹, Norazrina Abu Haris¹, Faizin bin Abu Bakar¹, “Data Mining in Churn Analysis Model for Telecommunication Industry” *Journal of Statistical Modeling and Analytics*, Vol. 1 No. 19-27, 2010.
- [12] Aruna J. Chamatkar¹ and P.K. Butey², “Comparison on Different Data Mining Algorithms”, *International Journal of Computer Sciences and Engineering*, E-ISSN: 2347-2693, Volume-2, Issue-10, 2014.