

Efficient Processing of AJAX Data using Mining Algorithms

Miss Pooja P. Taral¹, Prof. Pravin D. Soni²

¹Department of Computer Science and Engineering, P.R.Patil COET, Maharashtra, India

²Professor Department of Computer Science and Engineering, P.R.Patil COET, Maharashtra, India

Abstract - One of the fundamental data mining tasks is mining association rules. Positive association rules referred to as typical association rules which consider only items enumerated in transactions. Negative association rules consider the same items but negated items which are absent from transactions. To identify products that conflict with each other or products that complement each other in market-basket analysis, negative association rules are used. Mining for such rules requires the examination of an exponentially large search space. Depending on minimum confidence and minimum support, algorithms can become very slow and generate large amount of results or generate too few results, regardless of valuable information. As users have limited resources for analyzing the results in practice, an algorithm to mine the top-k association rules has been proposed using Sequential Pattern Mining. In addition to improve performance of an algorithm AJAX is used, which is an important approach for improving rich interactivity between Web server and end users. If we adapt client-server communication based on AJAX feature in communicating client and servers, decrease traffic and increase faster transmission. It delivers requests and responses are better, smarter and richer.

Key Words: Association rule mining, Positive rules, Negative rules, AJAX, Top-k association rules, Sequential Pattern Mining.

1. INTRODUCTION

An association rule is an implication of the form $X \Rightarrow Y$, where X and Y are frequent itemsets in a transaction database and $X \cap Y = \emptyset$. In practical applications, the rule $X \Rightarrow Y$ used to predict that 'if X occurs in a transaction, then Y will likely also occur in the same transaction', and we can apply this association rule to place 'Y close to X' in the items placement of supermarket management [1]. Nowadays, data mining is extremely important for business domains like marketing, financing, trading and telecommunication [2]. Users have limited resources for analyzing the results and thus are mostly interested in discovering a certain amount of results. As it is time-consuming to get whole result, Top-k mining rules are useful [3]. Support counting is one of the main advantages of the vertical bitmap representation of the data in Sequential Pattern Mining [4].

2. LITERATURE REVIEW

Association rule mining is interpreted as positive association rule mining. Positive association rule is stated as "if a person buy the product like bread and butter, then that person is likely to buy milk at the same time". Whereas negative association rules stated as "birds can fly is a well-known fact, but penguins cannot fly although they are birds" [2]. 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 $X \Rightarrow Y$ [4]. AJAX (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 without interfering with the of the existing page. HttpRequest object provide a method for exchanging data asynchronously between browser and server to avoid full page reloads [5].

Ayres et al. [4] proposed SPAM algorithm based on the idea of SPADE. The difference is that SPAM utilizes a bitmap representation of the database regardless of sid, tid pairs used in the SPADE algorithm. SPAM is to be one of best depth-first search strategy for mining sequential patterns. An additional salient feature of SPAM is its property of online outputting sequential patterns of different length. Vertical bitmap data layout is used in the implementation of SPAM allowing for simple, efficient counting.

The problem of top-k association rule utilizes a new approach for generating association rules named "rule expansions" and several optimizations. An evaluation of the algorithm with datasets used in the literature shows that TopKRules has excellent performance and scalability. Moreover, results show that Top K rules is an advantageous option to association rule mining algorithms for users who want to control over the number of association rules generated [3].

TopKRules finds 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 minimum confidence value, and returns the k rules with the highest support that meet the minimum confidence threshold. The reason why this algorithm defines

the task of mining the top-k rules on the support instead of the confidence is that minimum support is much more difficult to set than minimum confidence because minimum support depends on database characteristics [6].

To verify the independence between two variables, statistical test has been used whereas for positive or negative relationship, a correlation metric was used. To find negative itemsets involve following steps:

(1) First, find all the generalized large itemsets in the data (i.e., itemsets at all levels in the taxonomy whose support is greater than the user specified minimum support).

(2) Next, identify the candidate negative itemsets based on the large itemsets and the taxonomy and assign them expected support.

(3) In the last step, count the actual support for the candidate itemsets and retain only the negative itemsets. AJAX is a technology which can be used to reduce client server traffic [8].

3. PROPOSED WORK

In the proposed work, Sequential Pattern Mining is combined with Top-k rule mining to get better mining results as well analysis is done based on mining without AJAX and mining with AJAX to correlate performance of an algorithm. Proposed work focuses on retrieving efficient results with AJAX using various mining techniques as Sequential Patten Mining, Top-k mining, pruning with depth first search strategy. Proposed algorithm follows steps as:

- Step 1. Input itemset, pattern length and support threshold.
- Step 2. Scan database for various sequences.
- Step 3. Perform classification and mining on transactions.
- Step 4. Find relationship between data using Top-k rule.
- Step 5. Separate Positive and Negative rules.
- Step 6. Show mining result with AJAX and without AJAX.

Proposed mining algorithm works as:

Step 1. Scan the database to store first bit position of each sequence and calculate the total no of bit for each bitmap.

- Read a file for each sequence until the end.
- Record length of a current sequence
- Split sequence according to spaces into tokens.
- For each token if it is not an itemset separator, record last bit position for the bitmaps.

Step 2. Scan the database to create vertical database representation.

- sid used to know which sequence is going to scan and tid to know which itemset is going to scan.
- Split sequences with spaces and if token is -1 then increment tid and if token is -2 then increment, otherwise get a bitmap for the item.
- Register the bit in the bitmap for current item with sid, tid and sequence size.

Step 3. Remove infrequent items as they will not appear in any sequential frequent patterns.

- Iterate over items to get bitmap for the item.
- If cardinality of the bitmap is lower than minimum support then remove the item.
- Otherwise add item to the list of frequent items.

Step 4. Perform recursive depth first search to find longer sequential patterns recursively.

- For each frequent item, create a prefix with that item.
- Call DFS method with a particular prefix.
- Perform DFS pruning.

In proposed work, depending on the choice of parameters, association rule mining algorithms can generate an large number of association rules which leads to long execution time and huge memory consumption. With Top-k rules on positive as well as negated items, an algorithm found the top-k rules having the highest support, where k value as constant is set by the user in an algorithm.

4. RESULT ANALYSIS

As per result analysis of proposed algorithm, input support value 0.1, 0.2, 0.3 and 0.4 with the pattern length 3, 5, 10 provides good set of positive and negative mining rules. In case of support value 0.5, we got resultset of negative rules only. As 0.5 thresholds value has been set for class value to separate positive and negative rules.

Results of proposed algorithm are as:

Table -1: Time analysis with support constant 0.1

Time Analysis			
Support	Pattern Length	Time (without AJAX)	Time(With AJAX)
0.1	3	65 ms	16 ms
0.1	5	16 ms	15 ms
0.1	10	20 ms	15 ms

Table -2: Time analysis with support constant 0.2

Time Analysis			
Support	Pattern Length	Time (without AJAX)	Time(With AJAX)
0.2	3	65 ms	16 ms
0.2	5	16 ms	15 ms
0.2	10	20 ms	15 ms

Table -3: Time analysis with support constant 0.3

Time Analysis			
Support	Pattern Length	Time (without AJAX)	Time(With AJAX)
0.3	3	65 ms	16 ms
0.3	5	16 ms	15 ms
0.3	10	20 ms	15 ms

Pattern length 3, 5, 7 and 10 works good for all support values 0.1, 0.2, 0.3 and 0.4 whereas pattern length 2 works good for support values 0.1, 0.2 and 0.3.

Table -4: Time analysis with pattern length constant 0.3

Time Analysis			
Pattern Length	Support	Time(Without AJAX)	Time(With AJAX)
2	0.1	2 ms	0 ms
2	0.2	1 ms	0 ms
2	0.3	2 ms	0 ms
2	0.4	0 ms	0 ms

Table -5: Time analysis with pattern length constant 3

Time Analysis			
Pattern Length	Support	Time(Without AJAX)	Time(With AJAX)
3	0.1	5 ms	3 ms
3	0.2	5 ms	0 ms
3	0.3	2 ms	0 ms
3	0.4	1 ms	0 ms

Table -6: Time analysis with pattern length constant 5

Time Analysis			
Pattern Length	Support	Time(Without AJAX)	Time(With AJAX)
5	0.1	57 ms	16 ms
5	0.2	21 ms	16 ms
5	0.3	35 ms	32 ms
5	0.4	3 ms	0 ms

In the consideration of above all analysis results, it is proved that Association Rule mining using Top k rules with Sequential Pattern Mining provides efficient results than without AJAX.

6. CONCLUSIONS

The retrieval method of rule mining proposed in this paper is using classification as well mining algorithm to improve results in an efficient manner. The proposed method is trying to show results in an effective way by separating positive and negative rules which can be used for product placement in supermarket management.

REFERENCES

- [1] Xindong Wu, Chengqi Zhang and Shichao Zhang, "Efficient Mining of Both Positive and Negative Association Rules", ACM Transactions on Information Systems, Vol. 22, No. 3, July 2004.
- [2] Mudra Doshi, Bidisha Roy, "Efficient Processing Of Ajax Data Using Mining Algorithms" International Journal of Computer Engineering and Technology (IJCET), Volume 5, Issue 8, August (2014), pp. 48-54.
- [3] Philippe Fournier-Viger¹, Cheng-Wei Wu² and Vincent S. Tseng² "Mining Top-K Association Rules", National Cheng Kung University.
- [4] Jay Ayres, Johannes Gehrke, Tomi Yiu, and Jason Flannick, "Sequential Pattern Mining using A Bitmap Representation", Dept. of Computer Science Cornell University.
- [5] Sumayya Khan, Shrikant Lade , "Enhanced Data Processing Using Positive Negative Association Mining on AJAX Data", IOSR Journal of Computer Engineering, Volume 16, Issue 2, Ver. II (Mar-Apr. 2014), PP 15-18.
- [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(IJCSE), Vol. 02, No. 09, 2010, 2916-2920.
- [8] S.Vikram Phaneendra, "Minimizing Client-Server Traffic Based On Ajax", International Journal Of Computer Engineering & Technology (IJCET), Volume 3, Issue 1, January- June (2012), pp. 10-16.