An Efficient Approach for Mining Association Rules from Web Log Data

Jabed Al Faysal[1], Md. Anisur Rahman[2], Rokebul Anam[3]

Computer Science and Engineering Discipline, Khulna University
Khulna-9208, Bangladesh

DOI: 10.29322/IJSRP.10.12.2020.p10876
http://dx.doi.org/10.29322/IJSRP.10.12.2020.p10876

***Abstract*—** Mining of association rules from frequent patterns has recently been a large field of interest in data mining studies. In addition to that, the demand of mining association rules from large web log data is increasing rapidly. When we discover hidden information from large web log data is known as web data mining. The main objective behind this mining is obtaining information regarding navigational behavior of the web users that can be used for system improvement, advertising purpose, e-commerce or business application as well as understanding user reaction, network communication etc. In this paper, we have tried to introduce a new algorithm which mines association rules from a web log dataset in which access of the users to different pages are given in some sequence of page visits. The analysis is performed over the server log dataset to generate the required association rules. Experiments have been done with our algorithm using large web log data and considerable improvement has been found.

***Keywords*—** data mining, apriori algorithm, association rules.

I. Introduction

Extraction of meaningful and hidden knowledge from a large collection of data is the main goal of data mining. And the extraction of that meaningful knowledge from large web data has become a big part of data mining in the recent years, which is known as web data mining. Being a massive unstructured data repository, web provides us an incredible amount of data information. Web mining is the integration of information gathered by traditional data mining techniques and strategies with information gathered over the World Wide Web.

We have been observing a growing trend among various organizations and individuals to gather information through the web data mining in order to utilize that information of their best interest. Large web log data increases the complexity when dealing with the information from different types of users, business analysts and service providers. However, web mining doesn't mean only to apply data mining strategies on data stored in the web. The algorithms need to be modified in such ways that they'll better suit the demands for web datasets.

Different areas of web data mining includes web content mining (WCM), Web structure mining (WSM), and web usage mining (WUM). However, our research part is about web usage or log mining and extracting interesting rules from the log data which consists of textual data and is represented in a sequential format which describes the page visits of users. There are many important application areas of WUM such as E-Commerce/Business Applications, Site Reorganization, System improvement, Web Personalization etc. The aim of this kind of mining is to model and examine the web log data in such a way that it can be able to determine the usage behavior of the site users. As an example, suppose an Internet Service Provider (ISP) wants to get single level knowledge on the websites their clients frequently visit. Now, if they have a large dataset from the core repository (if available) in which each row represents a user and each column is for a website, then they need to use mining approach for determining user behavior. Association rules play a big role in understanding those user behavior. These rules usually used in finding the relationship between attributes from an itemset. In case of web usage mining, item set is a set of pages. Association rules are generated from frequent items with respect to minimum support and minimum confidence, which is an important parameter of any apriori algorithm based approach. Rules are applied to understand pages which often looked together in order to disclose associations between groups of users with specific interests.

II. Literature Review

A. Data Mining

Data mining is the process by which we analyze data from different perspectives and summarize those data into meaningful information. It is an interdisciplinary research field which has drawn from areas like database systems, data warehousing, high-performance computing, information retrieval, statistics etc. Data mining is the task where interesting patterns are discovered from a large amount of data which can be stored in databases, data warehouses or other information repositories. That knowledge is applied in decision making, process control and query processing. Also, the discovered knowledge can be used for various applications ranging from market analysis, fraud detection to customer retention and production control.

B. Web Mining

Another area referred as web mining which involves the integration of information that is gathered by conventional mining techniques using available data over the web. According to analysis, various types of web mining includes web usage mining, web content mining and web structure mining.
Content mining discovers productive information available on the web. Different types of web content provides useful information to users which includes unstructured data (plain text), semi-structured (html document), structured (xml document) and multimedia data. The main objective of such mining is to give an efficient mechanism that will assist the system users to discover required information. There have been a rapid development in approaches of web content mining in the past few years.

Another type of mining, named as web structure mining**,** means the process of finding the hyperlink structure throughout the web. In reality, this kind of mining emphasis on internal document information. Also, web structure mining finds the link structures at the internal document level. The focus is to know the authoritative and a given subject for the hub pages. A page containing a large amount of referencing hyperlink means that the information of the page is important, and also trustworthy. Hubs are web pages having many links to authoritative pages, so they always help to cluster the authorities. Structure mining can be gained only in a single portal and sometimes on the whole web too. The task of web content mining is supported by mining the structure of the web. The document retrieval process becomes more efficient by using the information about the structure of the web, while the reliability of these documents can be greater. Web structure mining exploits the graph structure of the web in order to develop the performance of the knowledge retrieval and to develop document classification.

Web usage mining uses three types of log files. The client side, the server side and the proxy servers store Log files. Navigational patterns of the users, storing the information more than one place make the mining process harder. Only if one has data from all these three types of log file, then reliability could be achieved in results. Generally, server side doesn't contain or save all the web page accesses which are cached on the proxy servers or in client side. The stored page requests in the client side are removed. Yet, it becomes difficult to get all the information of the client side. Thus, most algorithms are based on the server side data. Some commonly used algorithms of data mining for web usage mining are sequence mining, clustering and association rule mining.

There are three stages of web use mining, including pre-processing, pattern discovery and pattern analysis. Again, pre-processing has three steps. The data collected should be cleaned up initially, which ensures that graphical and multimedia entries are removed. Then it is important to recognize various users having different sessions.

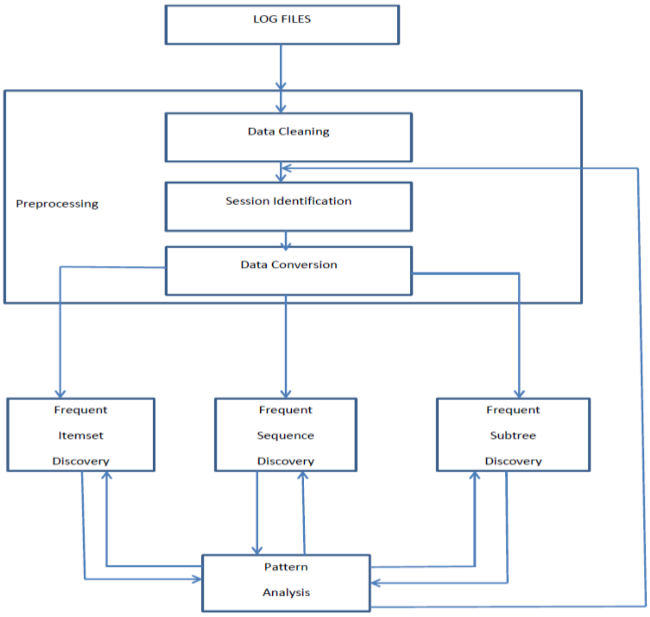

Fig 1: Process of web usage mining

A session is considered as a group of activities that when a user navigates via a given site, he/she conducts. It is not an easy move to understand the sessions from raw data, since the server logs do not always have all the required details. In this work, pattern discovery implies that the frequent sequential pattern discovery procedures implemented have been applied to the desired log data. That is why in the pre-processing stage, the data must be transformed so that the transformation output can be used as the algorithm input. Analysis of patterns involves knowing and drawing a conclusion on the outcomes obtained by the algorithms. The motive behind this study is to figure out the rules in the pattern discovery stage from the collection discovered. The methodology of research is generally controlled by the application for which this web mining is performed. However, web usage mining is our working part. In more detail, extracted interesting rules from the web usage or log data which consists of textual data and is represented in a sequential format which describes the page visits of users.

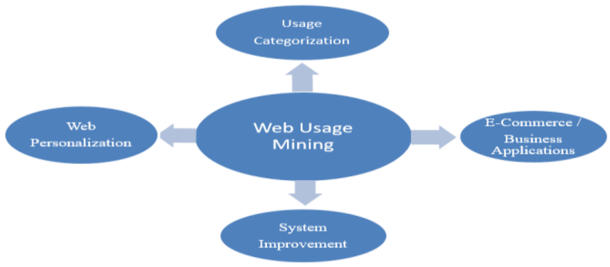

Fig 2: Major Applications of Web Usage Mining

C. Itemset

An Itemset $X = \{x_1, x_2, x_3 \ldots x_n\}$ is a set of one or more items where $X \subseteq I$. Here, I is the set of all possible items in the database. An item-set with k items is called a k-itemset. For instance, {google; facebook; yahoo} is a 3-itemset with three items.

D. Frequent Itemset

An Item-set is said to be frequent if all the items in the set is available in the transaction database, D at least more than or equal to the minimum support threshold, σ. For example, an item-set $I = \{I_1, I_2, I_3 \ldots I_n\}$ is frequent if frequency(I) $\geq \sigma$, minimum support. If a database holds a minimum support threshold of 30% and number of transactions in the database is 200, any itemset $\{i_1, i_2, i_3 \ldots i_n\}$ will be frequent if it satisfies the minimum support threshold i,e. the frequency of the itemset must be at least 60 out of the 200 transactions.

E. Support

Fraction of transactions involves a collection of objects. The sup(X) support of an itemset X is defined as the proportion of transactions that the itemset contains in a data set. If a dataset has 5 transactions, and within those a 3-itemset {local, news, sports} is present together in 2 transactions, the 2/5 is the support of that itemset. since it occurs in 40% of all transactions (2 out of 5 transactions). Again, the support of a sequential pattern includes the number of sequences where the pattern takes place, divided by the total number of sequences in the database.

F. Confidence

Confidence is an interesting measure of an association rule that refers to the state of likelihood given the precedent of the rule resulting from the rule. If n% of sessions in D that have X also have Y, the rule $X \rightarrow Y$ holds in a set of sessions D with confidence n. The following formula is used to measure the confidence of a rule, where Supp represents candidate support.

$$Conf(X \rightarrow Y) = Supp(X \cap Y) / Supp(X)$$

G. Association Rules

Let $X = \{x_1, x_2 \ldots x_p\}$ and $Y = \{y_1, y_2 \ldots y_p\}$ be two Item sets where $X,Y \subseteq I$ and $X \cap Y = \phi$. Here I $\{I_1, I_2 \ldots I_n\}$ be a finite set of n items. Thus an association rule is defined as representation of the form $X \rightarrow Y$. The set of items X is here called the antecedent and the set of items Y is called the consequent.

For example, let I = {facebook, google, yahoo, youtube} and a set of frequent items be {facebook, google, yahoo}. Now an association rule could be {facebook, yahoo} → {google} which means that if people visit facebook and yahoo, then they visit google as well. One of the well-researched data mining approaches is association rule mining. It focuses on extracting interesting similarities and regular trends in the transaction databases or other data repositories between sets of objects. In fields such as risk management, business research, inventory control, these rules are extensively used.

H. Related Works

There have been advancements in the sector of web data mining. Different techniques of web data mining have been proposed recently. Some of them were frequent sequential pattern mining from web logs, applying OLAP and data mining technology on web log etc. Pattern mining are important and useful for knowledge extraction. This extracted knowledge can be used in improvement of web site design, system performance analysis, network communication, user response and motivation learning and building adaptive website [11,12]. Different data mining techniques can be used in web usage mining. Association rules are used to find out the pages that are visited together, which shows the interest of specific group of people [2,6]. Using this knowledge of the prediction of visiting of next page of user can be determined. In sequence mining WAP-tree is used for efficient pattern storing [10]. Tree-like tropology pattern and path traversal is used in searching in data mining [8, 9, 10]. For mining closed sequential pattern NCSP used MSNBC dataset [7]. NCSP Algorithm first scans the whole database and removes the infrequent items. Then it builds vertical bitmap. From the bit map it finds 1-sequence frequent item set. Then it generate the candidates.

Agrawal and Srikant introduced sequential pattern mining problem [4]. A sequence set is given here, and each sequence has a list of elements while each element consists of a set of objects, and sequential pattern mining finds all frequent subsequences given a user-specified minimum support threshold, i.e., the subsequence that occurs frequently within the sequence set is no less than minimum support. GSP was introduced by Agrawal and Srikant, which generalized the sequential pattern concept that surpasses their Apriori All algorithm [3,4]. GSP works efficiently in the case where transactions are not large and the sequences are not long. However, when the length of sequences increases or transactions are large, the generated candidate sequence number can grow in exponential way, and GSP will fall in difficulties. All of the above studies that relates with (either sequential or periodic) frequent pattern mining takes the form of an Apriori like paradigm, advancing a generate-and- test methodology (generate a set of candidate patterns, then test when each candidate have enough support in the database). To find frequent set of pages, ItemsetCode algorithm [3,16] was introduced in web data mining which is a level-wise “candidate generate and test” method. It also follows apriori algorithm, but not fully. The algorithm enhances the Apriori on lower level. Here, small frequent itemsets are discovered in an improved way and it is considered as a better and faster way to find greater itemsets.

An improved apriori algorithm for association rules was introduced by Al-Maolegi and Bassam[17]. Firstly, all transactions are scanned by the algorithm to get frequent itemset of one item that contains the items, their support count and the transactions ids. Then eliminate the candidates which are infrequent that means their support are less than the minimum support. Then it stores all the transaction ids against the individual frequent items. After that, candidate 2-itemset is generated from the first frequent itemsets. It generally perform join operation in frequent 1-itemset. After that, it checks whether support of an itemset is minimum or not. Then those transactions are scanned to count the support of newly created set or join set. This is done for all join sets and terminate those itemsets whose support value is less than the given support value. Then it performs the same operation of storing. After that it performs the 3-itemset operation and so on until the candidate set is empty.

III. Problem Statement

Being a core algorithm of data mining, Apriori plays a vital role in web data mining and traditional mining procedure as well. Here the target is to generate frequent itemsets and producing association rules based on those itemsets. In web data mining this algorithm is used for mining frequent pattern from a transaction data set which is a repetitive process. This algorithm searches level wise. Apriori organizes a "subset frequency based pruning optimization" which means, it only process a itemset whose subsets are also frequent. The major disadvantage here includes: i) It performs n number of passes through the database, where n stands for the distance of largest frequent itemset. The count of candidate itemsets having a length k can be obtained through k^{th} pass, ii) Requires much runtime for large real datasets. Apriori algorithm divides the complexity of mining patterns into two different steps: Firstly, find all combinations of items which have transaction support beyond the minimum support, and then generate frequent patterns. Secondly, a pass k (being a subsequent pass), consists of two phases. First, the frequent pattern L_k-1 (the set of all frequent (k-1) - patterns) found in the (k-1)th pass are used to generate the candidate set C_k. The algorithm ends when L_k turns out to be blank.

Here for n itemset, it explores n+1 itemset. First it generates frequent itemset of 1-item comparing with support value. After that it produces candidate set. Then it generates the 2-itemset frequent value using support and confidence. It produces 2-itemset using combination of first frequent itemset. It generates more frequent itemsets until n_itemset are found which satisfies both the support and confidence. It scans the whole database for each and every time of iteration. It is very costly to generate candidate set. In web mostly the data sets are huge so it is not cost effective at all. Apriori takes more time to scan the large data set and produce frequent itemset. For instance if we take 104 1-item sets, apriori will produce 107 length-2 candidates and accumulate and test if their frequency occur there. Moreover, to find out a frequent pattern of size 100 such as {a1.a2……a100}, it should generate 2100-2 ˜ 1030 candidates totally. This is the candidate generation cost. So the number of database scan is always increasing thus the candidate generation increase and so the computational cost. As the search space is increased and so the I/O cost will increase. Because of lot of iteration it takes more time to execute and memory consuming as well.

IV. Proposed System

A. Proposed Method

In this section, we discuss our strategies to achieve the goal of our algorithm which is mining association rules from web log data. We are given with a transaction database which is an anonymous web log data, where each transaction represents an user's web page hits in a sequential form. Besides that, the required input parameters are given which consists of minimum support, confidence thresholds, web categories, sort flag.

Here is a flowchart of our working procedure. The sequential steps include reading log data file, cleaning the data, data conversion (that refers to making the matrix format), frequent itemset generation, pattern analysis and at last association rule mining.

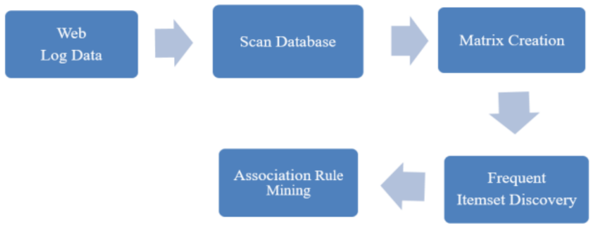

Fig. 3: Working Procedure

B. Algorithm

In this part we discuss about our proposed algorithm and its input and desired output. Our algorithm is based on the apriori algorithm, a widely used data mining algorithm. Association rules can be generated from frequent items with respect to the minimum support as well as minimum confidence, which is an important parameter of any apriori algorithm based approach. However, the algorithm have modified in a way so that they better suit the required demands for web datasets having sequence of page visits of users.
The algorithm can be described as below:

Input: Dataset D,Minimum Support δ,Minimum Confidence γ

Output: Association rules from frequent itemsets

```
begin
    Scan D
        build a zero matrix(R,C);
        for each item Ii in Transaction Ti in D do
                matrix(Ri, Ci) = 1;
        end
    end
    call ruleFinder (matrix, δ, γ);
    display result;
end
```

Now, in the *ruleFinder* procedure (which have the matrix, minimum support δ and minimum confidence γ) includes the following steps:

Step 1. Find frequent itemsets of size 1 (one) (list of all items with δ)
Step 2. Find frequent item sets of size >=2 and from those itemsets, identify rules with γ
Step 3. Generate all the possible combinations of items that are in frequent itemset
Step 4. Store the freqent itemsets
Step 5. Sort the rules in descending order based on the support or confidence level
Step 6. Save the rule in a text file.

Here the function that performs association analysis is *ruleFinder*. Given a set of transactions,
It is able to find rules that'll predict the occurrences of different items based on the occurrences of other items in the same transaction. Generally, rules are of the form A-> B (e.g. {news, misc} -> {local}). Here, Other parameters are sortFlag, categories, nRules.

V. Experimental Results

A. Dataset

The MSNBC is an Anonymous Web Data Set. We used this big dataset for the experiment of our algorithm as it is suitable for our proposed methodology. The data type is of discrete sequence and a very dense dataset. This set provides information about the page visits of the users who already visited msnbc.com during twenty-four hour period on a specific day.

This dataset has taken from Internet Information Server (IIS) logs for msnbc.com, for the whole day of September 28, 2011. A page view of the user over a span of twenty-four hours reflects every single sequence in the dataset. Again each event in a series represents the request for a page from a user. Here, requests have not been reported in depth at the very best level (not at the URL level) but have been recorded at the level of the page category (as decided by a site administrator). The types available are as below:
"frontpage", "news", "tech", "local", "opinion", "on-air", "misc", "weather", "health", "living", "business", "sports", "summary", "bbs" (bulletin board service), "travel", "msn-news" and "msn-sports".

Also, any page request that is served via a caching mechanism did not record in the server logs. Hence, they were not present in the data. David Heckerman (heckerma@microsoft.com) is the Dataset provider. Total number of users inside the dataset is 989818.

Here is the table of page category number with the code that is given inside the dataset:

Page Category	Code	Page Category	Code	Page Category	Code
frontpage	1	misc	7	summary	13
news	2	weather	8	bbs	14
tech	3	health	9	travel	15
local	4	living	10	msn-news	16
opinion	5	business	11	msn-sports	17
on-air	6	sports	12		

Table 1: msnbc.com page category codes

B. Result Analysis (Tabular)

After implementing our proposed algorithm we got our required association rules as output. Several sample rules (1st 25 rules) which were found by the implementation of the algorithm are given in the following table. Here, the given minimum support is 0.1% and minimum confidence is 50%. Note that, the table is sorted in decreasing order of support.

Rule	Support	Confidence
news,misc -> local	1.04870%	52.1137%
tech,health -> news	0.54939%	57.4174%
local,health -> news	0.52596%	55.7626%
tech,living -> news	0.51484%	52.9290%
health,business -> news	0.49211%	54.4611%
misc,weather -> local	0.48120%	55.7924%
misc,summary -> on-air	0.47524%	79.4192%
misc,health -> on-air	0.45089%	57.7510%
health,living -> news	0.41614%	57.4798%
misc,health -> news	0.40573%	51.9669%
tech,misc -> news	0.38765%	50.9021%
misc,living -> news	0.35017%	53.6948%
misc,living -> local	0.34218%	52.4710%
misc,living -> on-air	0.33309%	51.0767%
health,sports -> news	0.31602%	57.5847%
opinion,living -> news	0.31309%	57.6880%
health,summary -> on-air	0.31117%	64.8148%
tech,opinion -> news	0.31087%	59.3214%
local,opinion -> news	0.28581%	55.3512%
opinion,business -> news	0.25894%	62.7110%
opinion,summary -> on-air	0.25005%	64.1026%
opinion,misc -> on-air	0.24328%	62.5780%
opinion,health -> news	0.23873%	66.9595%
living,summary -> on-air	0.23651%	56.6965%
opinion,sports -> news	0.22317%	60.0761%

Table 2: Rules with support and confidence (Our Approach)

Now, we will show the results that we get after using msnbc dataset in an improved apriori algorithm [19]. Note that we implemented the algorithm in the same environment and the results are shown below:

Rule	Support	Confidence
misc -> on-air	3.3625%	41.3382%
summary -> on-air	1.4178%	48.0616%
news,misc -> local	1.0487%	52.1137%
tech,business -> news	0.69669%	46.9084%
on-air,health -> news	0.62062%	43.2819%
local,business -> news	0.58728%	41.4977%
tech,local -> news	0.58182%	41.2477%
local,living -> news	0.55394%	48.3424%
tech,health -> news	0.54939%	57.4174%
local,health -> news	0.52596%	55.7626%
on-air,living -> news	0.51999%	44.6943%
tech,living -> news	0.51484%	52.929%
travel -> news	0.46908%	42.1861%
travel -> living	0.45241%	40.6869%
tech,sports -> news	0.43553%	44.9718%
tech,misc -> news	0.38765%	50.9021%
news,summary -> on-air	0.37866%	44.0164%
opinion,on-air -> news	0.36027%	48.656%
misc,weather -> news	0.35299%	40.9277%
opinion,living -> news	0.31309%	57.688%
tech,opinion -> news	0.31087%	59.3214%
local,opinion -> on-air	0.28581%	55.3512%
opinion,business -> living	0.25894%	62.711%
opinion,health -> news	0.23873%	66.9595%

Table 3.Rules with support and confidence (Improved Apriori)

C. *Result Analysis (Graphical)*

Here is a graphical view of time required to build up the binary matrix which helps us to count the itemset frequency in an efficient way. The execution time of the algorithm after building the matrix is found less than the traditional apriori approach. The required time for building the matrix for upto 800k transactions is given below:

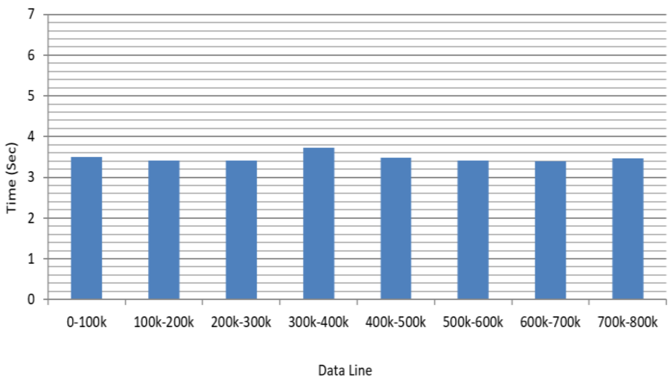

Fig. 4: Time required to build data matrix

The performance of our proposed algorithm has been showed that it takes less runtime than an existing algorithm improved apriori. We used 32000 transactions for runtime evaluation. So, we used the same number of transaction and compared with their performance and got runtime less than improved apriori.

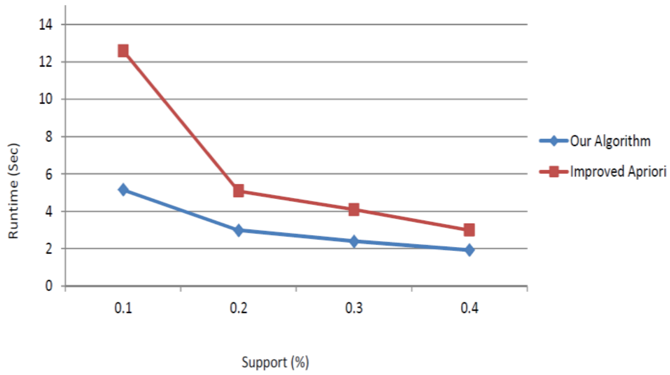

Fig. 5: Performance comparison using MSNBC dataset

VI. CONCLUSION

There have been a increasing trend to explore required information by mining web data among individuals and organizations in order to utilize the collected information of their best possible interest for increasing personal or company profit. Web data mining does not mean only to apply data mining methods to the data that stored in the web, rather the algorithm need to modify in a way so that they better meets the demands for web datasets. And we have done this work. Our main contribution is that we have modified the traditional apriori approach is such a way so that a large web log data (that contains sequential user access patterns) can be processed in an efficient way without scanning many times and then generate association rules. Result proved that the transaction database matrix has reduced from the very first scan where new matrix is created that contains only frequent itemsets. It also reduced the computational cost and we found our proposed algorithm faster. Thus we can easily conclude that, we have found a remarkable improvement in our proposed work for large web log data.

REFERENCES

[1] Prof Dr. Saba Hilal, Neetu Anand, "Identifying the User Access Pattern in Web Log Data", Neetu Anand et al, / (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 3 (2) , 2012, 3536-3539.
[2] Asst. Prof. Rupali Bhartiya, Pooja Sharma, "An Efficient Algorithm for Improved Web Usage Mining", Int.J.Computer Technology & Applications,Vol 3 (2), 766-769, 2011.
[3] S.Vijayalakshmi, V.Mohan, S.Suresh Raja, "Mining of Users Access Behaviour for Frequent Sequential Pattern From Web Logs", International Journal of Database Management Systems (IJDMS) Vol.2, No.3, August 2010.
[4] Rakesh Agrawal, Tomasz Imielinski, "Mining Association Rules Between Sets of Items in Large Database". In Proc. 1995 Int. Conf. Data (ICDE'95), p.3–14, March 1995.
[5] J. Borges and M. Levene, "Data Mining of User Navigation Patterns" in WEBKDD, 1999, pp. 92-111
[6] M. Eirinaki and M. Vazirgiannis,"Web Mining For Web Personalization", ACM Trans. Inter. Tech., Vol. 3, No. 1, pp. 1-27, 2003
[7] J. Pei, J. Han, B. Mortazavi-Asl, and H. Zhu, "A Novel Algorithm For Mining Closed Sequential Patterns" International Journal of Data Mining & Knowledge Management Process (IJDKP) Vol.5, No.1, January 2015, pp. 41-50

ISSN 2250-3153
[8] M. S. Chen, J. S. Park, and P. S. Yu, "Data Mining for Path Traversal Patterns in a Web Environment" in Sixteenth International Conference on Distributed Computing Systems, 1996, pp. 385-392

[9] X. Lin, C. Liu, Y. Zhang, and X. Zhou, "Efficiently Computing Frequent Tree-like Topology Patterns in a Web Environment" in TOOLS '99: Proceedings of the 31st International Conference on Technology of Object-Oriented Language and Systems. Washington, DC, USA: IEEE Computer Society, 1999, p. 440

[10] A. Nanopoulos and Y. Manolopoulos, "Finding Generalized Path Patterns For Web Log Data Mining", in ADBIS-DASFAA'00: Proceedings of the East-European Conference on Advances in Databases and Information Systems, a joint International Conference on Database Systems for Advanced Applications. London, UK: Springer-Verlag, 2000, pp. 215- 228

[11] M. Perkowitz and O. Etzioni, "Adaptive Sites: Automatically Learning from User Access Patterns". In Proc. 6th Int'l World Wide Web Conf., Santa Clara, California, April 1997.

[12] R. Srikant and R. Agrawal, "Mining Quantitative Association Rules in Large Relational Tables". In Proc. 1996 ACM-SIGMOD Int. Conf. Management of Data, pages 1→12, Montreal, Canada, June 1996

[13] O. Zaiane, M. Xin, and J. Han, "Discovering Web Access Patterns and Trends by Applying OLAP and Data Mining Technology on Web Logs". In Proc. Advances in Digital Libraries Conf. (ADL'98), Melbourne, Australia, pages 144-158, April 1998

[14] A. Anitha, "A New Web Usage Mining Approach for Next Page Access Prediction", International Journal of Computer Applications (0975 – 8887) Volume 8– No.11, October 2010.

[15] R. Iváncsy and I. Vajk, "Time and Memory Efficient Frequent Itemset Discovering Algorithm for Association Rule Mining" International Journal of Computer Applications in Techology, Special Issue on Data Mining Applications.

[16] Mohammed Al-Maolegi and Bassam Arkok, "An Improved Apriori Algorithm for Association Rules", International Journal on Natural Language Computing (IJNLC) Vol. 3, No.1, February 2014.