

# International Journal of Advance Research in Computer Science and Management Studies

Research Article / Survey Paper / Case Study

Available online at: [www.ijarcsms.com](http://www.ijarcsms.com)

## *A Review on Different Frequent Pattern Mining Techniques*

**Dhaval S Patel<sup>1</sup>**

Computer Science Engineering  
Parul Institute of Technology  
Vadodara, India

**Jayraj M Desai<sup>2</sup>**

Computer Science Engineering  
Parul Institute of Technology  
Vadodara, India

**Swapnil Andhariya<sup>3</sup>**

Assistant Professor  
Computer Science Engineering  
Parul Institute of Technology  
Vadodara, India

*Abstract: Data mining is the process of identifying useful pattern from large amount of data. Web mining is the process of finding web pattern from web data. Web link mining uses data from log file. Web link mining finding useful things from log file. Log file contains all the user actions. In existing system all the data are mine by apriory algorithm. It is only used for finding sequence from log file but it does not classified log data according to our need. Support vector machine is used to classify all the data of web log file that analyze data and identify patterns. Support vector machine classify all the data into two classes. It differentiates two classes by hyper plane. After applying SVM, We discover all the data into one pattern. This pattern is useful for business analyst to take a useful decision. Finally we get classify data from web navigation data.*

*Keywords: Data Mining, Web Link Mining, Log File, SVM classifier, Pattern*

### I. INTRODUCTION

Data mining is the proces used for provide the useful knowledge and extracting the interesting pattern from the huge amount of data set[1] .

Frequent patterns are item sets, subsequences, or substructures that appear in a data set with frequency no less than a user-specified threshold. Frequent pattern mining is a first step in association rule mining.[Comparative Analysis of Various Approaches Used in Frequent Pattern Mining] One of the most frequent users of frequent mining is to analysis of large amount of data. It is also called as web data. There are following type of web data in web mining [1].

**a) Web Server Data:**

User logs are gathered by the web server and regularly incorporate IP location, page reference and access time.

**b) Application Server Data:**

Commercial provision servers, for example, Weblogic, Storyserver, have huge characteristics to empower E-business requisitions to be based on top of them with little exertion. A key characteristic is the capability to track different sorts of business occasions and log them in requisition server logs.

**c) Application Level Data:**

New sorts of occasions could be characterized in a provision, and logging might be turned on for them — producing histories of these occasions. It must be noted, in any case, that numerous end provisions oblige a combo of one or a greater amount of the systems connected in the over the classification.

## II. RELATED WORK

The existing research works is divided into three main categories a) Association Rule Mining (ARM), b) Clustering, c) Classification.

In Association Rule Mining, Ming-Syan et al., [3] proposed a new data mining algorithm that involves mining path traversal patterns in a distributed information-providing environment where documents or objects are linked together to facilitate interactive access. Jianhan Zhu et al., [4] applied the Markov chains to model user navigational behavior. They proposed a method for constructing a Markov model of a web site based on past visitor behavior. Then the Markov model is used to make link predictions that assist new users to navigate the Web site.

In Clustering, Paola Britos et al., [5] described the capacity of use of Self Organized Maps, kind of artificial neural network, in the process of Web Usage Mining to detect user's patterns. Mehrdad Jalali et al., [6] presented an approach which is based on the graph partitioning for modeling user navigation patterns. In order to mining user navigation patterns, they establish an undirected graph based on connectivity between each pair of the web pages and also proposed novel formula for assigning weights to edges of the graph.

In Classification, Mahdi Khosravi et al., [7] proposed a novel approach for dynamic mining of users' interest navigation patterns, using naïve Bayesian method.

## III. VARIOUS TECHNIQUES OF FREQUENT PATTERN MINING

### 1. *Apriori Based Algorithms:*

It is a very first algorithm for frequent pattern mining. By using item set's downward closure property apriori algorithm decrease the size of search space. If an item set of length  $k$  is not frequent, none of its superset patterns can be frequent. Candidate frequent item sets,  $C_k$  where  $k$  is the length of the item set, are generated before each data scan. The supports of candidate frequent item sets are counted. Candidate  $k$  item sets,  $C_k$  are generated with frequent  $(k - 1)$  item sets. [8]. Performance of Apriori is increase by reducing the candidate set. At the end of each scan, transactions that are potentially useful are used for the next iteration. A technique called scan reduction uses candidate 2 item sets to generate subsequent candidate item sets [9].

#### **Advantage:**

It is very easy to understand and implimentant.

#### **Disadvantage:**

It needs  $k$  scans to find all frequent  $k$ -item sets.

### 2. *Partition-based Algorithms:*

Partition –Based algorithm is to overcome the problem of high number of database scans. In this Algorithm whole data set is scan in just two scans. In this algorithm dataset is dividing into many number of small dataset. In First data scan, algorithm create local frequent item set in each partition. Because whole partition can be stored into main memory mining process done without any disk I/O operation. The local frequent item sets are added into the global candidate's frequent item set [8]. In second step, false candidates are discarded from the global candidates frequent item set. By applying previous knowledge during the candidate item set extracting process. Is will help in finding the false global candidate frequent item set. False item set then removed from list. In worst case, number of scan for partition-base algorithm is  $(2b-1)/b$  where  $b$  is number of partition [8].

**Advantage:**

Less number of database scan.

**3. Pattern-growth Algorithm:**

There are two main drawback of apriori algorithm one is generate candidate frequent item set and other is cost related to I/O operation. Second issue can be solve but for first issue Han et al. [10] proposed a data structure called frequent pattern tree or FP Tree. FP-growth mines frequent item sets from FP-Tree without generating candidate frequent item sets. It is a extension of prefix tree structure [8]. Each node of tree contain item label and it's frequency. Support value defined the root of item in tree form root to leave. All scanning process is complete in two steps. In first step, this calculate the support value and in second scan, this count is used to sort the item in descending order. If two transactions share a common prefix, the shared portion is merged and the frequencies of the nodes are incremented accordingly. Nodes with the same label are connected with an item link. The item link is used to facilitate frequent pattern mining [8]. In this technique we ensure that all pattern set are generate properly.

**IV. COMPARISON BETWEEN THE ABOVE MANTION TECHNIQUE**

Comparison between above mentioned techniques is shown in the following table.<sup>[8]</sup>

**TABLE I**

Comparison between methods

Techniques	Apriori Based	Partition Based	FP Tree	SQL Based
Number of Database scan(Best Case)	2	1	2	1
Number of Database scan(Worst Case)	A+1	(2B-1)/B	2	1
Candidate Generation Needed or Not?	Yes	Yes	No	No
Incremental Mining Possible	No	No	No	No
Sensitive to Change in User Parameter	Yes	Yes	Yes	Yes

**V. CONCLUSION**

In this paper, various techniques for frequent pattern mining are explained. The techniques are selected based upon its application. Output gives us frequent pattern in particular session. There are so many researches going on pattern mining but some more research is required, because in today's world behavior of user is very much important for web site. Finding frequent item set is very useful in various tasks such as select time of site maintenance, find most visited page of site, finding place for post advertisement etc.

**ACKNOWLEDGMENT**

With the cooperation of my guide, I am highly indebted to Asst. Prof. SWAPNIL ANDHARIYA, for his valuable guidance and supervision regarding my topic as well as for providing necessary information regarding review paper.

**References**

1. Pawel Weichbroth, Mieczyslaw Owoc," Web User Navigation Patterns Discovery from WWW Server Log Files", 978-83-60810-48-4, 2012 IEEE.
2. Mr. Akshay A. Adsod, Prof. Nitin R. Chopde, "A Review on: Web Mining Techniques", ISSN: 2231-5381 IJETT – Volume 10 Number 3 - Apr 2014.
3. Ming-Syan Chen, Jong Soo Park, Philip S. Yu, "Efficient Data Mining for Path Traversal Patterns", IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 10, NO. 2, MARCH/APRIL 1998.

4. Jianhan Zhu, Jun Hong, John G. Hughes, "Using Markov Chains for Link Prediction in Adaptive Web Sites", Soft-Ware 2002, LNCS 2311, pp.60–73, 2002
5. Paola Britos, Damián Martinelli, Hernán Merlino, Ramón García-Martínez, "Web Usage Mining Using Self Organized Maps", International Journal of Computer Science and Network Security, VOL.7 No.6, June 2007
6. Mehrdad Jalali, Norwati Mustapha, Ali Mamat, Md. Nasir B Sulaiman, "WEB USER NAVIGATION PATTERN MINING APPROACH BASED ON GRAPH PARTITIONING ALGORITHM", Journal of Theoretical and Applied Information Technology
7. Mahdi Khosravi, Mohammad J. Tarokh, "Dynamic Mining of Users Interest Navigation Patterns Using Naive Bayesian Method", 978-1-4244-8230-6/10/\$26.00 ©2010 IEEE
8. Deepak Garg, Hemant Sharma, 'Comparative Analysis of Various Approaches Used in Frequent Pattern Mining', International Journal of Advanced Computer Science and Applications, Special Issue on Artificial Intelligence
9. Lee Chang Hung, Lin Cheng Ru, and Chen Ming Syan. "Sliding Window Filtering: An Efficient Method for incremental Mining on a Time-Variant Database". Proceedings of 10th International Conference on Information and Knowledge Management, 263-270, November 2001
10. Han Jiawei, Pei Jian, Mortazavi-Asl Behzad, Chen Qiming, Dayal Umeshwar, and Hsu Mei-Chun. "FreeSpan: Frequent pattern-projected sequential pattern mining". Boston, Ma, August 2000
11. X. Shang, K.-U. Sattler, and I. Geist "SQL based frequent pattern mining with fp-growth. In INAP/WLP", pages 32–46, 2004

#### AUTHOR(S) PROFILE



**Dhaval Patel**, received the B.E degree in Computer Engineering from Vidhyabharti trust institute and research center of Gujarat technological University in 2013, and pursuing M.E degrees in computer Science & Engineering from Parul Institute of technology from Gujarat technological University.



**Jayraj Desai**, received the B.E degree in Computer Engineering from Vidhyabharti trust institute and research center of Gujarat technological University in 2013, and pursuing M.E degrees in computer Science & Engineering from Parul Institute of technology from Gujarat technological University.