

# Web Mining for Efficient Web Page Navigation Prediction Model

Harsha Dubey<sup>1</sup> Megha Mishra<sup>2</sup> Dr.V.K.Mishra<sup>3</sup>

<sup>1</sup>Research Scholar <sup>2</sup>Sr.Assistant Professor <sup>3</sup>Associate Professor

<sup>1,2,3</sup>Department of Computer Science & Engineering

<sup>1,2</sup>Shri Shankaracharya Technical Campus <sup>3</sup>Bharti College of Engineering & Technology

**Abstract**— Web mining is the utilization of information mining systems to discover patterns from the World Wide Web. Web mining can be broadly classified in to three types-Web usage Mining, Web content Mining and Web structure Mining. Web mining is a new developing research discipline and also it is subdivision of data mining, it has attracted a great deal of attention in the information technology and society, due to the availability of huge amount of heterogeneous data. The Process of web server log files in order to extract patterns to web page prediction with the help of Apriori algorithm. The prediction of web page and user navigation behavior and this technique the web link prediction this the process to predict the Website page to be gone by a client in view of the pages visited by the user. The target of these web mining is to foresee the web item to be prefetched with high precision and serves to great prescient exactness with various log record. In the web mining, pre-fetching and prediction is done by the user by preprocessing to access the log files. The approaches result in prediction of better web page and navigation behavior and it also help to decrease the accuracy time. So that the web-prefetching method lessens the dormancy of web & predict the object to pre-fetched with high accuracy and achieve the better accuracy of different log file.

**Key words:** Web Mining, Apriori Algorithm, Web log file

## I. INTRODUCTION

Web mining is the application of data mining having large web data repositories and better serves the needs of web based application usage. World Wide Web is very popular and interactive. It has become an important source of information and services. Web is immense, various and dynamic and thus raises the Scalability, multimedia data and temporal issue respectively. Web Mining consists of three categories – Web Content Mining, Web Structure Mining and Web Usage Mining.

Web Content Mining - web content mining is the mining, extraction and reconciliation of helpful information, data and learning from site page content. In this Kind of Web mining process is to discover all links of hyperlinks in a document to generate the structural report on a web page. Web Structure Mining - web structure mining is the method of using graph theory to separate the center point and association structure of a site.

Web Usage Mining: Web use mining is the use of information mining system to find the fascinating use designs from web data remembering the finished objective to grasp and better serve the necessities of online applications.

Web usage Mining are three tasks to performed – Pre-processing, Pattern Discovery and pattern Analysis. It involves finding the frequency of page access by the client. The data is collected from the Web server log file.

## A. Data Pre-Processing

This paper presents the overview of various steps of preprocessing stage.

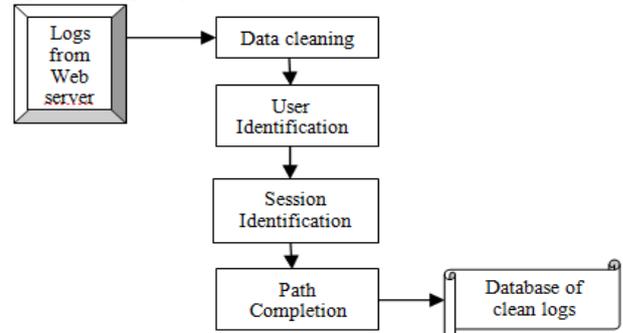


Fig. 1: Data Preprocessing

The input for web usage is a user session file, to preprocess the file and information accessed by the web site and which page and how long the user can access the web page. In the web log files, the unnecessary information is reduced or minimized by data to preprocessing the web log data.

## B. Pattern Discovery

Pattern Discovery is the implement techniques from data mining and information theory on the web data collected. Design Discovery systems discover all examples for which there are adequately visit illustrations in the specimen information.

## C. Pattern Analysis

The examination of the pre-handled information is exceptionally gainful to all associations business over the web. The sites are keen on successive use design like "How the client are utilizing site?" "Which pages are oftentimes gotten to by the user's?"

## II. RELATED WORK

This paper proposed that the Web page prediction model to give the best answer of web user's to access the next web page to better prediction with the apriori algorithm. Web Navigation is present to improve the web services by extracting useful information and knowledge from vast amount of web data. Data Used for web mining, the data can be collected at the server side, Client Side and Proxy side for organization's database.[1]

The web mining process of web server log files in order to extract usage patterns to web link prediction with the help of Apriori Algorithm. The approaches result in prediction of popular web page or stage and user navigation behaviour. It proposed that the cluster user navigation based on their similarity measure combined with the concept of apriori algorithm is used for web link to predict the web pages to be visited by the user. Web Navigation Patterns

should be updated when the new data are inserted into original web navigation sequence database. [2]

In the web page the prediction model is give the best answer of web user to access the next web page and better prediction of web page. Web prefetching techniques reduces the web latency & they predict the web object to be pre-fetched with high accuracy and it also help to achieve better accuracy among different log file. The evolutionary approach helps to train the model to make predictions to current web browsing patterns. [3]

In the web log file the preprocessing is the first step for web usage mining. Quality of pattern mining and pattern analysis is fully depending on the preprocessing. We can use some new technique to provide the user to analyze the log file at different level such as user sessions. There are some advantages and disadvantages in web log preprocessing to improve the efficiency and its quality. [4]

We know that web site construction is always changed, many Prediction proficiency need not consider the behavior of frequency but the web site structure to mine web navigation patterns for navigation prediction, dynamic mining approach is based on the previous mining results and formed new patterns of the web data.

### III. PROBLEM DEFINITION

Due to their browsing patterns have to be extracted from web server log files. Web page prediction is the web usage mining by performing preprocessing of the data from a web site. The need for predicting the user's needs in order to improve the usability and user maintenance of a web site is marked now a day's lacking proper guidance, a visitors often purpose without visit the significant pages and leaves the site and effectiveness of the discovered knowledge not much improved.

### IV. METHODOLOGY

In the mining, prefetching and prediction is done by the user by preprocessing to get to the log. We incorporate the affiliation standard with Apriori calculation and grouping to get to the better site page. This system used to group the comparative move to enhance the productivity of expectation. These sources can incorporate web server access logs, intermediary server logs, program logs or some other information that is created by clients communicating with a site. Preprocessing so as to prefetching and forecast is finished the client access logs and coordinating the methods i.e. clustering model and apriori calculation to accomplishes better website page access expectation precision. We can take care of this issue to discover low proficiency and precise result.

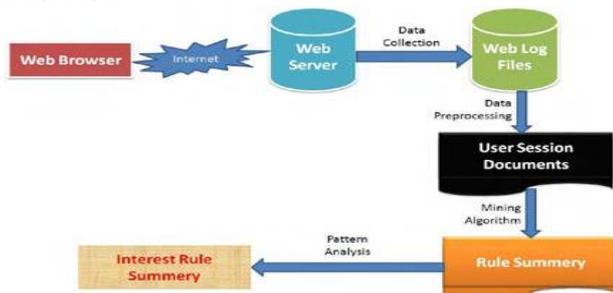


Fig.2. Work flow for web mining for efficient web page navigation prediction model

This proposed procedure used to bunch comparative move practices for productive used to facilitate enhance the proficiency of prediction. Prediction methods were connected utilizing every group and utilizing the entire data. Results show that the grouping markov model calculation on the information set enhances the exactness of the expectation model.

Site page expectation is the web utilization performing so as to mine preprocessing of the information from a web site. The requirement for anticipating the client's needs to enhance the convenience and client support of a site is more than checked now a day's missing appropriate guidance, a guest frequently meanders capriciously without going by noteworthy pages, loses consideration and leaves the website sooner than anticipated.

### V. RESULT

The web mining, every client has distinctive number of time got to by the client in site page. It betters expectation and it decreases the client access time and enhances the web execution. Every site page has diverse number of times got to by the user. Since every page is not of the same hobby and clients dependably make the comparable transaction. Thus, the client's entrance history can be utilized to use for mining client access designs.

### VI. CONCLUSION & FUTURE WORK

In this work, an effective web prefetching mechanism that makes predictions for the group of users who have similar requirements. The web user access can predict accurately and minimize user perceived latency. The Apriori algorithm presented analyzed and evaluate to predict web access which providing high accuracy.

The future extension is that the page rank of a site page is registered utilizing recurrence of the watchwords which is advancements to sites by including the catchphrases rehash at ordinarily.

### REFERENCES

- [1] Priyanka S.Panchal, Urmi D.Agravat "HYBRID TECHNIQUE FOR USER'S WEB PAGE ACCESS PREDICTION BASED ON MARKOV MODEL" IEEE, 4th ICCNT 2013, Tiruchengode.
- [2] Smriti Pandya, Rajesh Nigam "REVIEW PAPER ON WEB PAGE PREDICTION USING DATA MINING" IISTE, Vol.6, 2015.
- [3] Prasad J.Koyande, Kavita P.Shirsat "COMPARISON OF DIFFERENT NAVIGATION PREDICTION TECHNIQUES" IJCSIT, Vol.6 (2), 2015
- [4] Arshi Shamsi, Rahul Nayak, Pankaj Pratap Singh, Mahesh Kumar Tiwari, "WEB USAGE MIMNING BY DATA PREPROCESSING". IJCSIT Vol.3, Issue 1, Jan.- March 2012.
- [5] C.P.Sumathi, R.Padmaja Valli, T.Santhanam "AN OVERVIEW OF PREPROCESSING OF WEB LOG FILES FOR WEB USAGE MINING" Journal of Theoretical and Applied Information Technology 31st December 2011. Vol.34 No.2 IC.P.
- [6] Deepti Razdan "THE NEXT PAGE ACCESS PREDICTION USING MARKOV MODEL", IJECCT Volume 1 Issue 1, September 2011.

- [7] L.K.Joshila Gracel, V.Maheswari, Dhinakaran Nagamalai, "ANALYSIS OF WEB LOGS AND WEB USER IN WEB MINING", International journal of Network Security & its Applications, Vol.3, No.1, January 2011.
- [8] Navin Kumar Tyagi, A.K.Solanki and Manoj Wadhwa, "ANALYSIS OF SERVER LOG BY WEB USAGE MINING FOR WEBSITE IMPROVEMENT", IJCSI International journal of computer science, Vol.7, Issue 4, No 8, July 2010.
- [9] Srivasta,J., Cooley,R., Deshpande,M., and Tan P.N.(2000). "WEB USAGE MINING: DISCOVERY AND APPLICATION OF WEB USAGE PATTERN FROM WEB DATA". Department of Computer Science and Engineering, University of Minnesota.
- [10] Kosala, R., Blockeel, H. (2000) "WEB MINING RESEARCH: A SURVEY." ACM SIGKDD (special Interest Group on Knowledge Discovery and Data Mining) Explorations. June, (2:1). Pp 1-10.

