

A Survey of Web Usage Mining: Concepts With Applications and Its Future Scope

Jitendra N. Shrivastava^[1], Shivendra Pratap Singh^[2]

M.Tech Research Scholar^[2]

Department of Computer Science and Engineering

Invertis University Bareilly, 243123

Uttar Pradesh – India

ABSTRACT

In this paper we have to focus on data mining concept and its tools and technology which help us for user behaviour analysis to take a proper decision and get a proper result. Web usage mining supports for the creation of web site designing; provide personalization web server and other business decision, etc. Web usage mining consists of three stages, namely pre-processing, pattern discovery, and pattern analysis. Web Usage Mining is an important type of Web Mining, which deals with extraction of interesting knowledge from the web log files. The lots of research has done in this field but basically this paper emphasize on user future next request prediction using web log record, click streams record and user information. The aim of this paper is to provide past, current evaluation and update in web usage mining- future request prediction. This paper also reports the comparisons and summary of various methods of future request prediction with application, which gives the overview of development in research.

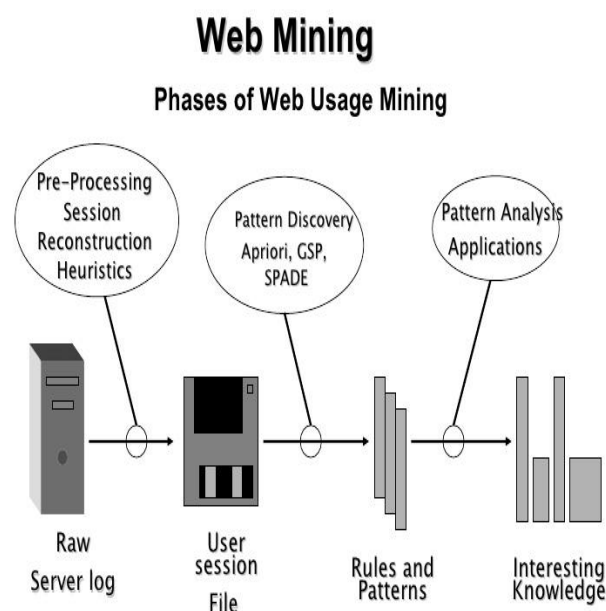
Keywords:- Web Usage Mining, Data Mining, Web Server, Pattern Discovery, Pattern Analysis, web Log Files.

I. INTRODUCTION

Many companies are adopting internet search practices for their intranets. While the underlying search process is the same for both the Internet and the intranet, the search needs of the respective users and their environments are very different. The Internet consists of users who have individualized information needs and share no understanding with the information providers. Internet users have access to an unbounded document set that may include advertisements and spam [1]. The web is a huge and hard to estimate the growth of web data every day, the web provides different kind of services such as government, electronic commerce, news, etc. Mining of the web is potentially useful patterns of implicit information from World Wide Web. The work has various parts. First part describes the various steps of web usage mining. Second part elaborates about various techniques. web usage mining also consist of various categories involved in it. Then various applications are discussed in the work done here. web usage mining generally have three categories like web content mining, web structure mining and finally web usage mining. These various categories are discussed in details later.

II. STAGES OF WEB USAGE MINING

It supports for the creation of web site designing; provide personalization web server and other business decision, etc. Web usage mining consists of three stages, namely pre-processing, pattern discovery, and pattern analysis.



The first step is data collection of resources using the second step is pre-processing, the third step is pattern discovery and the final step is pattern analysis. Web

usage mining is the application of data mining. : The information that formulates the hyper-link structure of a website i.e., various HTML tags used to link one page to another page and one website to another.

III. WEB USAGE

The information that reflects the usages of Web resources, i.e., entries in Web browser's history and Internet temporary files, proxy server and Web server logs. There are two very distinct environments when it comes to web search [6]

- A) The internet and
- B) The intranet.

The way these environments are viewed from both users and researchers are very different. There are only a handful of published studies regarding intranet search, but internet search reports are published nearly every three months. The most recent internet statistics published was in February (Nielson, 2010), which reported that Google is the most preferred search engine (65.2% of all searches). That same report listed Yahoo as second, losing 18% more of its previously reported search share to Google. The percentage of typical daily users has grown to nearly 50%, with users extremely positive about search engines and their search experiences (Fallows, 2008). However, in that same report users are described as generally unsophisticated about how and why they use search [8]. Web Usage Mining is an important type of Web Mining, which deals with extraction of interesting knowledge from the web log files. The lots of research has done in this field but basically this paper emphasize on user future next request prediction using web log record, click streams record and user information. The aim of this paper is to provide past, current evaluation and update in web usage mining- future request prediction. This paper also reports the comparisons and summary of various methods of future request prediction with application, which gives the overview of development in research [4].

Web mining which is a type of data mining is used to extract web data from web pages. As data mining basically deals with the structured form of data, web mining deals with the unstructured and semi structured form of data applications Web mining is an application of data mining which uses data mining techniques to extract useful information from web documents.

IV. TECHNIQUES OF WEB MINING

Web mining consist of three techniques i.e. web content mining, web structure mining and web usage mining for web data extraction [5]

A. Web Content Mining:

Content mining deals with extraction of data from the content of WebPages based upon pattern matching

B. Web Structure Mining:

Describe relational structure of the WebPages and used to extract information from hyperlink structures.

C. Web Usage Mining:

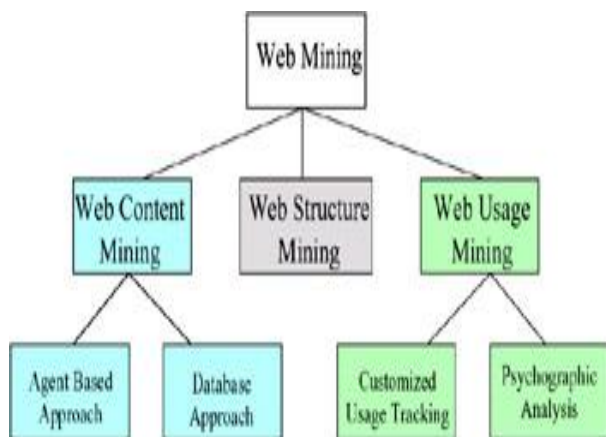
Usage Mining is the application of data mining technique to discover information from the web log data in order to understand and better serve the needs of Web based.

Web usage mining is a process of mining useful information from server logs. This paper describes the methods which already used in past research work for analysing the user behavior and predicting the future request.[7]

In this world of Information Technology, accessing information is the most frequent task. Every day we have to go through several kind of information that we need and what we do? Just browse the web and the desired information is with us on a single click. Today, internet is playing such a vital role in our everyday life that it is very difficult to survive without it. The World Wide Web (WWW) has Influenced a lot to both users (visitors) as well as the web site owners. The web site owners are able to reach to all the targeted audience nationally and internationally. They are open to their customer 24X7. On the other side visitors are also availing those facilities. In the last fifteen years, the growth in number of web sites and visitors to those web sites has increased exponentially. The number of users by June 30 2010 was 1,966,514,816[18] which is 28.7% of the world's population. The number of active web sites is 125,522,259 [19] as on 13-Dec-2010. Due to this growth a huge quantity of web data has been generated. To mine the interesting data from this huge pool, data mining techniques can be applied. But the web data is unstructured or semi structured. So we can not apply the data mining techniques directly. Rather another discipline is evolved called web mining which can be applied to web data. Web mining is used to discover interest patterns which can be applied to many real world problems like improving web sites, better understanding the visitor's behavior, product recommendation etc.[5]

Web mining is the use of data mining techniques to automatically discover and extract information from Web documents/services.

V. CATEGORIES OF WEB USAGE MINING



A. *Content Mining* (Examines the content of web pages as well as results of web Searching)

B. *Structure Mining* (Exploiting Hyperlink Structure)

C. *Usage Mining* (Analysing user web navigation)

Web usage mining is a process of picking up information from user how to use web sites. Web content mining is a process of picking up information from texts, images and other contents. Web structure mining is a process of picking up information from linkages of web pages. Content mining examines the content of web pages, Structure mining exploiting hyperlink structure and Usage mining analysing user web navigation all these types are various categories of WUM[3].

VI. APPLICATIONS OF WEBUSAGE MINING

Web usage mining has various application areas such as web prefetching, site reorganization, web personalization, system improvement, link prediction, Business Intelligence and Usage Characterization.[4]

A) *Web prefetching*

The performance of the system is very important for user satisfaction. Web usage mining is an important research

area for detecting the web traffic. The communications between PC and the server constitute Web traffic. The amount of traffic and the details of each visit are extremely valuable information to a Web-based business. The server computer records every request for a Web page by user, and determines which pages get the most attention.

Web traffic analysis gives businesses concrete, reliable information on the interests of their customers. The more traffic a Web site receives, the more sessions and hits its server processes. Every time a Web server processes a file request, the computer makes an entry in a server log, a dedicated file on the server's hard. To develop the server performance new policies can be used. Downloading files can slow down the user experience.

If a user scrolls through an application screen and has to wait for content to load, the application appears slow to them. To use prefetching effectively, you need to evaluate the content application uses in order to determine meaningful indexes that identify which content is appropriate for prefetching. This prefetching approach is useful for both client and server level web caching, load balancing, transmission of data distributed information are the applications of web mining.

B) *Website reorganization*

Attractiveness of the website is an important one gives good structure of the website. The principle of website reorganization is first need to understand how users interact with web-sites, how they think and what the basic patterns of users' behavior. In website navigation, the structure of the website can rearrange. The relationships between web pages are dynamically updated.

Reorganization can be performed with the extraction of frequent patterns of web usage mining. The web usage information gives the information about the user behaviour's of any website. Both content and structure leads to adaptive web site.

C) *Website personalization*

Website Personalization is gathering the personal user data, some privacy issues is a major anxiety. Personalization Consortium endorses and guide the growth of one-to-one marketing carried out. Behind of personalization technologies include collaborative filtering, in which filtering is applied with different sites

for selecting the relevant information that may apply to the specific group of customers. User profiling information is collected from different web sites can result of personalized web page .By using data analysis can be predict the users future prediction and find the session with multiple websites with interesting links.

C) System improvement

The major system improvement life cycle is planning, analysis, development and implementation. It should support the user demand to build a system. Developing the system with the security can avoid the intrusion and to restrict the user's access to certain online contents. Understand the customer demand and retaining the customized products. Improve some satisfaction with the help of browsing behavior.

D) Link Prediction

Link prediction is used for analysing the nodes in a network, from the large network suggest that information can be extracted from the network topology.

E) Business Intelligence

Web usage mining provides data to improve the customer, sales and marketing field. It is the technology to access the data from various data sources, for business advantage, the data is gathered, stored and analysed in organization can improve the customer needs and demands. Some decision about the business can be made to success. Thus the disciplines of Business intelligence includes decision support, data mining, online analytical processing (OLAP), querying and reporting, statistical analysis and forecasting. Some of the business intelligence tools are Bizz Score Suite, IBM, Cognous, Series10, Web FOCUS, QlikView, Tableau Software, Style Intelligence, Board Management Intelligence Toolkit, AS Enterprise BI Server can retrieve, analyse and generate report.

F) Usage characterization

The usage of web used by the user is for various purposes. By characterizing the data usage of heavy users and normal users, and classify them and clusters according to their usage activities. The user behavior can be observed by usage regularities on the website.

Characterize the users by navigational patterns and agent based approach.

VII. SCOPE OF WEB USAGE MINING

We can characterize the three web mining categories based on the scope of most of the work done in the respective areas: Local scope (spanning an individual website) and global scope (spanning the entire web). Here web content mining and web structure mining are categorized as of global scope, while web content mining (from the DB view) and web usage mining, as of local scope.

In practice, the three web mining categories can be used in isolation or combined in an application; especially in web content and structure mining since links may be considered as part of the content of a web document.

VIII. CONCLUSIONS

In this paper we briefly reviewed the various Web usage mining concepts, its techniques and applications. Web Mining is not a new term, but in the recent years its growth day by day touches great horizons. It has spread in almost all areas nowadays. It is clear that Web Mining tools helps in extracting useful or meaningful knowledgeable attributes or information from the unimaginable massive data. This review would be helpful for the researchers to focus on the various issues of web mining. In future, we will review the popular classification algorithms and significance of their evolutionary computing approach in designing of efficient classification algorithms for data mining.

REFERENCES

- [1] Sandeep Singh Rawat and Lakshmi Rajamani," Discovering Potential User Browsing Behaviors Using Custom-Built Apriori Algorithm", (IJCSIT), Vol.2, No.4, August 2010
- [2] Sujata padmaku mar, Dr.Punithavalli, Dr.Ranjith,"A Web Usage Mining Approach to User Navigation Pattern and Prediction in WebLog Data", IJSR, International Journal of Scientific Research, Volume : 3 | Issue: 4 | April 2014
- [3] S.K. Pani, L. Panigrahy, V.H.Sankar, Bikram Keshari Ratha, A.K.Mandal, S.K.Padhi,"Web Usage Mining: A Survey on Pattern Extraction from Web Logs",

International Journal of Instrumentation, Control & Automation (IJICA), Volume 1, Issue 1, 2011

- [4] M.Selvi Mohana, B.Rosiline Jeetha,"Methodologies on User Behavior Analysis and Future Request Prediction in Web Usage Mining using Data mining Techniques", International Journal of Data Mining Techniques and Applications, Volume: 03, June 2014, Pages: 369-373
- [5] Ujwala Patil, Sachin Pardeshi "A Survey on User Future Request Prediction: Web Usage Mining" ,(IJETA), Volume 2, Issue 3, March 2012
- [6] Lewis, April Ann, "Enterprise Users and Web Search Behavior." Master's Thesis, University of Tennessee, 2010.http://trace.tennessee.edu/utk_gradthes/643
- [7] <https://www.google.co.in/search?q=web+usage+mining>
- [8] <https://www.google.co.in/search?q=user+behavior+analysis>