

A Survey on Web Usage Mining and Pattern Discovery(A Review)

Neha Soni

CSE Dept, RIMT, MandiGobindgarh, Punjab , India

ruheesoni@yahoo.co.in

Abstract:-Data Mining is the process used by Companies to turn raw data into useful information. Web usage mining is the application of data mining techniques which is used by Business or Website administrator to extract the usage patterns of user. The Pattern information is extracted in the form of the web server logs, proxy server logs, browser logs, user profiles..This paper deals with web log record or user information and details on the Pattern matching.

Keywords: Web usage mining, proxy server logs, browser logs, user sessions, Pattern.

I. INTRODUCTION

As growing Popularity of World Wide Web, many organizations rely on the internet to extract required information of user pattern to conduct the business. This extracted information can then be used in a variety of ways such as, buying user pattern, visitor browsing behavior, checking of fraudulent elements etc. Web Usage Mining is used as tool for business Strategies by gathering important information from the user. Knowledge obtained from the Web usage patterns could be directly applied to efficiently manage activities related to e-Business, e-CRM, e-Services, e-Education, e-Newspapers, e-Government, Digital Libraries etc. The major problem with Web Usage Mining is that its need proper Preprocessing, data Cleaning or filtering before extracted proper information.

Preprocessing, pattern discovery, and pattern analysis are the three phases of Web Usage Mining .After passing these three phases raw data become information.

II. DATA SOURCES

The Web Usage Mining used various data sources for gathering required information. The Various data sources are:

1. Web Server Logs – Web Server Log is used as a data source in Web Usage Mining. Basically Server Log having collection of data include User IP Address, date or time

information on which particular page request, These data can be saved in Single or Multiple file. Such as access log, error log or referrer log. Only Administrative person is authorized to access these types of files. It can help to provide pattern of browsing user by doing analysis on time and week of day. This type of Analysis is also used by the Website Administrator for better web designing; marketing administrator used this for better market approach by knowing interest of User.

2. Proxy Server Logs – Proxy server Log is another data source for Web Usage Mining. A Proxy server is act as Server. Proxy Server Log contain the information of user HTTP request having same content which its save in Caches so to reduce the load time of server it serve the request of user from Caches which having same content of data. By this way it also serve as source to finding pattern of browsing User.

3. Browser Logs – Web Browser is act as the medium for Web User to access http page request from Web Server. Various browsers like Mozilla, Internet Explorer etc. This kind of log files can be made to collect information of HTTP Cookies used by Client.

III. Web Usage and Pattern Discovery

Preprocessing: Preprocessing convert raw data into useful information which is necessary for pattern discovery. The different types of preprocessing in Web Usage Mining are:

A Usage Pre-Processing: The main goal of Usage Preprocessing is to identify user Session, In this user identification is only done by using IP address, agent and server side click. Data is also not available completely in this phase so it's difficult to manage.

B Content Pre-Processing: In this information is represented by page view. It can also classified by conveyor gather information of user.

C Structure Pre-Processing: Pre-Processing related to structure of the website. It is created by the hypertext link between page views.

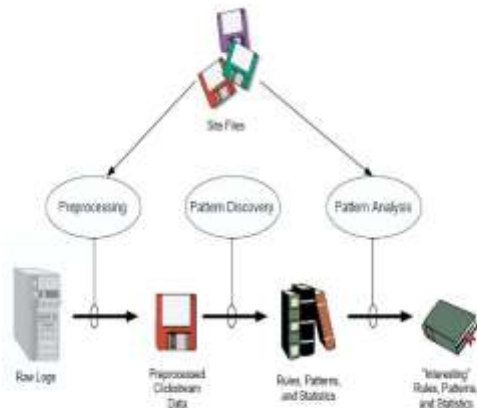


Figure 1: High Level Web Usage Mining Process

Pattern Discovery:

In Web Usage Mining, Pattern discovery uses various method and algorithms to access data from web access log. The access data is represented in the form of rules ,tables ,charts and graphs etc so that one can easily analysis pattern used by web browsing users.

The pattern discovery methods are described below:

- 1) **Statistical Analysis:** It is most common method to extract the information about visitor of website by statistical analysis on page view, navigation path, viewing time.
- 2) **Association Rules:** It can be used to relate pages that are most often referenced by single session but not connected with one hyperlink.
- 3) **Clustering :**It can be used for group having similar pattern characteristics It is of two types Usage Cluster having group of similar browsing user, page cluster: group of pages having similar contents.
- 4) **Classification:** It is used to develop profile of user's belongings to particular class such as class of user having interest in online music products etc.
- 5) **Sequential Patterns:** It is used by Web Marketers for placing advertisements for particular user group by predicting future visit patterns.

Pattern Analysis: The final Phase of web usage mining is pattern analysis where the validation of obtained data is analysed, after passing from preprocessing and pattern discovery phase. This type of analysis is process by

methods like SQL (Structured Query Language) processing and OLAP (Online Analytical Processing) .

IV. Application

The general goal of Web Usage Mining is to gather interesting information about users browsing patterns .Web Usage Mining is used in the following area:

- (i) **Personalization of Web Content:** Web Usage Mining provides Personalization for a user by determining the user pattern which were extracted from keeping track of previously accesses pages.
- (ii) **Prefetching and Caching:** Web Usage Mining improves user navigation through prefetching and caching.
- (iii) **Support to Design:** Web Usage mining improve web design; or in e-commerce sites with finding visiting browsing user pattern or behavior.
- (iv) **Ecommerce:** Web usage Mining improve the customer satisfaction. With the use of gather information of user interest.

V. Conclusion

As growing need of internet for marketing and advertising increased, the quality of web user browsing is need to be evaluated which is done by web usage Mining for which techniques of pattern discovery alike sequential patterns, association rules, classification, Clustering, Sequential Pattern and Pattern analysis is discussed above.

References

- [1] **RACHIT ADHVARYU** Student M.E CSE, B. H. Gardi Vidyapith, Rajkot, Gujarat, India
- [2] Artificial Intelligence and Robotics Laboratory Dipartimento di lettronica e Informazione Politecnico di Milano.
- [3] Dr Ahmed Rafea Web Usage Mining Reference : <http://maya.cs.depaul.edu/~classes/ect584/papers/srivastava.pdf>.
- [4] J.W. Han, M .Kamber, Data Mining--Concepts and Techniques, Elsevier Science & Technology Books, 2006.
- [5] E raju "Aanalysis of social Networks using the techniques of web mining"
- [6] Pooja Mehtaa,Brinda "Web personalization using web mining " Concept and Research Issue"
- [7] Dr. Antony Selvdoss Dava mani "A survey on preprocessing methods for web usage data ."
- [8] Rajni Pamnnani , Pramila Chawan , "Web Usage Mining: A Research Area in Web Mining"