



**INTERNATIONAL JOURNAL OF ENGINEERING SCIENCES & RESEARCH
TECHNOLOGY**

Web Usage Mining and Internet User Behaviour in Web: A Survey

Dr. Ananthi Sheshasaayee*, Padmaja.S

*Associate Professor and Head, Department of Computer Science, Quaid-e-Millath Government College for women, Chennai, India

#2 Research Scholar, Research and Development Centre, Bharathiar University, Coimbatore, India, India

ananthiseshu@gmail.com

Abstract

In recent years the growth of the World Wide Web exceeded all expectations. Today, there are several billions of HTML documents, Pictures and other multimedia files available via Internet and the number is still rising. In real problems, several studies have been conducted, presented and used on user behaviour in Internet. Various studies on Internet user behaviours have been presented and used in real problems. Data mining techniques are expected to be a more effective tool for analyzing user behaviour. This paper is a survey of recent work in the field of web usage mining for the benefit of research on the Internet user behaviour. A main research area in Web mining focused on learning Web users and their interactions within Websites is Web usage mining. Web usage mining is a main research area in Web mining focused on learning about Web users and their interactions with Web sites. A lot of research has been done already about this area and the obtained results are used in different applications such as recommending the Web usage patterns, personalization, system improvement and business intelligence.

Keywords: Web Usage Mining, Web log file, User behaviour, Application WUM.

Introduction

A lot of research work has been done on Web Usage Mining for web personalization and predicting their web access behaviour. When the user browses the web pages, the user leaves some valuable information in web log files. This web log file information is very helpful to find out the web navigation behaviour of the user. Through his behaviour, we can find out what kind of information user wanted from the web sites. Web usage mining, automatic extract the knowledge from the data collected in the web log file. The collected web log file and pattern analysis click streamed knowledge helpful to web usage mining, which can recommend a set of objects to the active user, possibly consisting of links, ads, text or products, tailored to user perceived preference. This paper did a literature survey on “WUM for the browsing behaviour of a user and subsequently to predict desired pages”.

The growth in online traffic is astronomical and online user behaviour has changed a lot. Web user behaviour is not what it used to be a few years ago. If you have been living in a cave you might have missed the changes that are happening that are causing this. Users today spend a lot of time online (a lot more than they used to a few years ago) either through their desktop or handheld devices.

The total number of internet users is increasing over the time, but perhaps more important is the increased diversity of people who are online. It appears, however, that many website owners and developers are either reluctant, or unable, to make sites that cater to the needs of these different sections of the community.

Related works

The aim of this literature survey is to study and analyze the available technique to predict the user behaviour in interacting with the Web or with a Web site. There are many Web log analysis tools available to mine data from the log record on Web pages. Log record contains plenty of useful information such as URL, IP address and time and so on. In this section we report on existing works on analyses of behaviour patterns

Long Jin, San Diego Yang Chen, Pan Hui, and Athanasios V. Vasilakos [5] presented a survey of the important roles of both providing a systematic exploration of existing research highlights and triggering various potentially significant research. Pablo E. Román and Juan D. Velásquez [3] proposed a new concept in 2009 that was called “Analysis of the Web User Behaviour with a Psychologically-Based Diffusion Model”. They used a new application of a

mathematical theory of psychological behaviour from Usher and McClelland and the random utility model from McFadden, to the web user behaviour.

Luis A. Leiva and Roberto Vivo [6] proposed a new concept in 2013 that was called “Web Browsing Behaviour Analysis and Interactive Hyper video”; they described the design and implementation of SMT2q, a tool for automatically gathering, mining, and visualizing browsing data in an interactive hypermedia presentation. SMT2q collects fine-grained information about user behaviour and allows viewers to control what they watch, when, and how, by selecting diverse types of infographics.

A.Brijesh Bakariya, B.Ghanshyam Singh Thakur [4] proposed a new concept in 2013 that was called “User Behaviour Analysis from Web Log using Log Analyzer Tool”, analyzing of web usage mining has been made with the help if web log data for which web log analyser tool, “Deep Log Analyzer” to find out abstract information from particular server and also tried to find out the user behaviour and also developed an ontology which consist the relation among efficient web apart of web usage mining.

Maryam Jafari, Farzad SoleymaniSabzchi 1, Shahram Jamali [2] presented a survey on some of the existing WUM techniques and it is shown that how WUM can be applied to Web server logs.

Diwakar Shukla and Rahul Singhai[8] recommended for browser owner competitors to reduce their quitting probability Pc as much as possible in order to maintain their customer proportion. However, companies are advised to have a watch on browser failure rate of other competing companies.

David F. Nettleton and Cristina Gonz’alez-Caro [1], proposed a new concept in 2012 that was called” Analysis of User Behaviour for Web Search Success using Eye Tracker data “. They used the number of formulating queries by session, documents clicked, the fixation durations on the documents, and the distribution of the attention in the different areas of the screen, among other aspects.

Data mining

Data mining is the process of analyzing data from different perspectives and summarizing it into useful information. Data are any facts, numbers, or text that can be processed by a computer. The patterns, associations, or relationships among all collected data can provide information. Information can be converted into knowledge about historical patterns and future trends. To maximize user access

and analysis, there needs to be a centralization of data in a data warehouse.

Web mining

The web is a collection of interrelated files on one or more Web servers. Web mining is the application of data mining techniques to extract knowledge from Web data. Web mining may be defined as the use of data-mining techniques to automatically discover and extract information from Web documents and services. It refers to the overall process of discovery, not just to the application of standard data mining tools. Decomposing web mining task into four subtasks:

1. **Resource Finding.** This is the process of retrieving data, which is either online or offline, from the multimedia sources on the web, such as news articles, forums, blogs, and the text content of HTML documentation obtained by removing the HTML tags.
2. **Information Selection and preprocessing.** This is the process by which different kinds of original data retrieved in the previous subtask are transformed. These transformations could be either a kind of pre-processing such as removing stop words and stemming or a pre-processing aimed at obtaining the desired representation, such as finding phrases in the training corpus and representing the text in the first-order logic form.
3. **Generalization.** Generalization is the process of automatically discovering general patterns within individual websites as well as across multiple sites. Different general-purpose machine-learning techniques, data mining techniques, and specific web-oriented methods are used.
4. **Analysis.** This is a task in which validation and/or interpretation of the mined patterns is performed.

Web content, structure, and usage mining

One possible categorization of web mining is based on which part of the web one mines. There are three main areas of web mining: web content mining, web structure mining, and web usage mining. Each area is classified by the type of data used in the mining process.

Web Content Mining

Web content mining uses webpage content as the data source for the mining process. This could include text, images, videos or any other type of content on web pages. Web content mining goal is

gathering data and identifying patterns related to the contents of the web and the searches performed on them. There are two main strategies: Web page mining, extracting patterns directly from the existing contents in web pages. In this case the data in use can

be free text, HTML pages, XML pages, Multimedia elements, and any other type of contents existing in the website. Search results mining, intending to identify patterns in the results of the search engines.

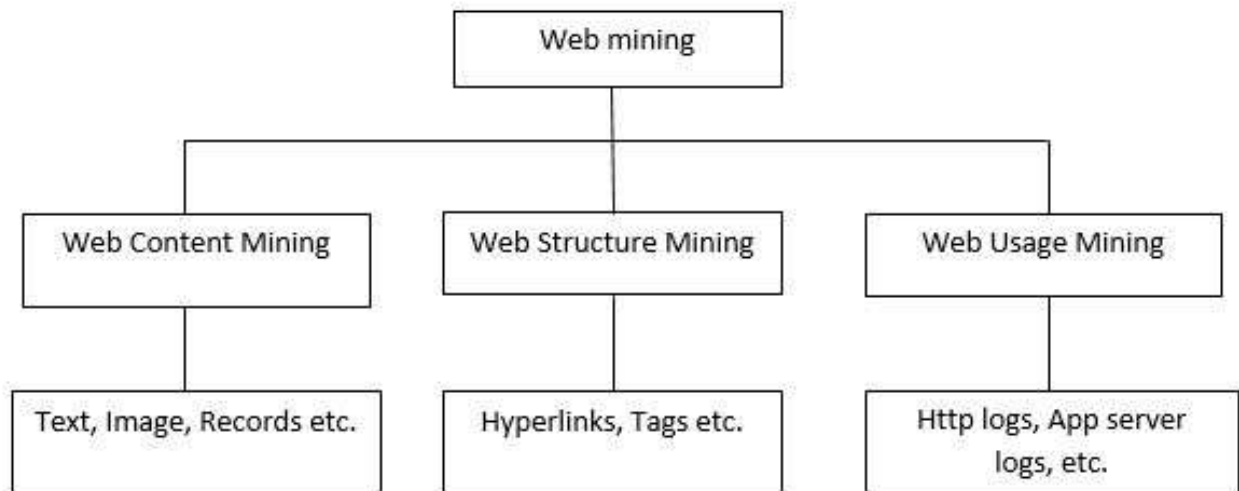


Figure 1: Classification of Web mining

Web Usage Mining

Web usage mining does not use data from the web itself, but takes as input data recorded from the interaction of users using the Internet. Here the goal is to dive into the records of the servers (log files) that store the information transactions that are performed in the web in order to find patterns revealing the usage the customers make of it. For example, the most visited pages, usual visiting paths etc.

General access pattern tracking. Here the interest doesn't rely on the access performs of a particular visitor on the integration of them into trends, allowing us to re-structure the web in order to facilitate our customers access and utilization of our website.

Customised access pattern tracking. Here what we look for is gathering data about the individual visitor's behaviour and their interaction with the website. This way we can establish access/purchase profiles so that we can offer a customised experience to every customer.

3.2.3 Web Structure Mining

Web structure mining mainly focuses on the structure of the link of the web pages. This speciality intends to reveal the real structure of web sites through the gathering of structure related data and mainly about its

connectivity. Typically it takes into account two types of links: static and dynamic.

The behaviour analysis approach

Behaviour modelling

In social learning theory, behaviour modelling is the precise demonstration of a desired behaviour. According to this theory, we learn not only by doing, but by watching what others do. In a therapeutic setting, behaviour modelling is purposeful, positive and teaching clients are the healthier ways of behaving. But behaviour modelling can also be negative, such as a parent passing on a prejudiced way of dealing with others or a friend teaching a child to use drugs. Therapeutic behaviour modelling is often used to help clients change previously learned negative behaviours.

Conclusion

Web usage mining is the process of applying statistical and data mining methods to web log data in order to extract useful patterns concerning the users' navigational behaviour, user and page clusters, as well as possible correlations between Web pages and user groups. The objective of Web usage mining is the extraction of information concerning the navigation behaviour of Web site visitors. On the other hand, this process can also be looked upon as part of the user

profile creation: it is therefore evident that Web page and user group modules overlap and are essentially in the Web personalization process. The extraction of information concerning the navigational behaviour of Web site visitors is the objective of Web usage mining. Nevertheless, this process can also be regarded as part of the creation of user profiles; it is therefore evident that those two modules overlap and are fundamental in the Web personalization process. Web usage mining lately has been used in combination with other technologies, such as user profiling and in some cases content mining, in order to provide a more integrated view of the usage of a Web site, and make personalization more effective.

Pages: A Survey", *IJESIT, Volume 2, Issue 5, September 2013, Pages 101-104.*

10. Sofia Angeletou, Matthew Rowe, Harith Alani," *Modelling and Analysis of User Behaviour in Online Communities*".
11. Mehmed Kantardzic," *Data Mining: Concepts, Models, Methods, and Algorithms*", Wiley publication, Second edition, 2011.

References

1. David F. Nettleton and Cristina Gonz'alez-Caro," *Analysis of User Behaviour for Web Search Success using Eye Tracker data*", *IEEE Conference, 2012, pp.57-63.*
2. Maryam Jafari, Farzad SoleymaniSabzchi1, Shahram Jamali," *Extracting Users' Navigational Behaviour from Web Log Data: a Survey*", *jcsa-1-3-3, May-2013, Vol. 1, No. 3, pp.39-45*
3. Pablo E. Rom'an and Juan D. Vel'asquez," *Analysis of the Web User Behaviour with a Psychologically-Based Diffusion Model* ", *AAAI Fall Symposium*", 200
4. A.Brijesh Bakariya, B.Ghanshyam Singh Thakur," *User Behaviour Analysis from Web Log using Log Analyzer Tool*", *ijcsns, November.2013-Volume 1.No3, pp. 41-52.*
5. Long Jin, San DiegoYang Chen, Pan Hui, and Athanasios V. Vasilakos," *Understanding User Behaviour in Online Social Networks: A Survey*", *IEEE Communications Magazine, September 2013 PP.144-150.*
6. Luis A. Leiva and Roberto Vivo," *Web Browsing Behaviour Analysis and Interactive Hypervideo*", *ACM-Transaction October 3, 2013, pp. 20.1-20.28.*
7. Myriam Abramson and David W. Aha," *User Authentication from Web Browsing Behaviour* ", *AAAI, 2012.*
8. Diwakar Shukla and Rahul Singhai," *Analysis of Users Web Browsing Behaviour Using Markov chain Model*", *IJANA, 2011-Volume: 02, Issue: 05, Pages: 824-830.*
9. Hemant N.Randhir and Ravindra Gupta," *WUM for Browsing Behaviour of a User and Subsequently to Predict Desired*