Web Usage Mining with Personalization on Social Web

Rutvija Pandya#1
Lecturer at Diploma Computer engineering, GTU

***Abstract*— *web mining is the technique to extract the useful knowledge from web. For mining the web, three categories -Web Content Mining, Web Structure Mining and Web Usage Mining is used. Web usage mining is used to discover interesting patterns from the web server log files. These user navigation patterns can be applied to many real-world problems, such as improving Web sites, product recommendations, user or customer behaviour studies, etc. Web Personalization is the area of the Web usage mining. Web personalization is defined as suitable content delivery to a particular user and takes the advantage of that knowledge to manipulate how and what information you present to your users. In this paper, we describe web usage mining techniques. The research also represents web personalization techniques in Social media.*

Keywords: *Usage Mining, Pattern Analysis, Content Mining, Structure Mining, Personalization*

I. INTRODUCTION

The growth of information is uncontrollable over the Internet; web data search has encountered a lot of challenges. Web users are always get lots of information and having the problem of overloaded by information. To improve the Internet service quality, web mining is used. Web mining can be classified into three different categories: Web content mining, Web structure mining and Web usage mining.

Web content mining: The process of Web Content Mining is getting useful information from the Web documents. This technique summarizes, classify and cluster the web content. It can provide useful and interesting patterns about user needs and contribution behaviour. The content on the internet, usually semi structured, un-structured, and structured. The web pages may consist of text, images, audio, video, and tables.

Web structure mining: Web Structure Mining is the process of discovering structure of the hyperlinks within the Web. The Web Structure Mining interprets information about pages and enhances search results through filtering. With the help of web structure mining, categorizing the web pages based on the hyperlinks and generate the information. For particular domain of website, structure mining is useful to discover the hierarchy of hyperlinks. This type of mining can be performed either at the document level or at the hyperlink level.

Web usage mining: Web Usage mining is used to mine the usage characteristics of the Web applications users. This retrieved information can be used for different purpose like checking of fake elements and application improvement. Web mining systems analysing the web usage data and recognize knowledge about users' interest. This knowledge is used in decision support, network traffic flow analysis, creating adaptive web sites, etc.

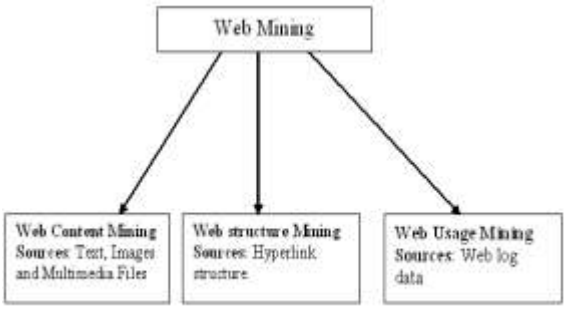

Fig. 1. The types and sources of Web mining

II. FUNCTIONS OF WEB USAGE MINING

A Web usage mining system must be able to perform five major functions:

A. Gathering Data from Different Sources
B. Pre-processing of Data
C. Discovery of Pattern
D. Analysis of Pattern
E. Applications of Pattern

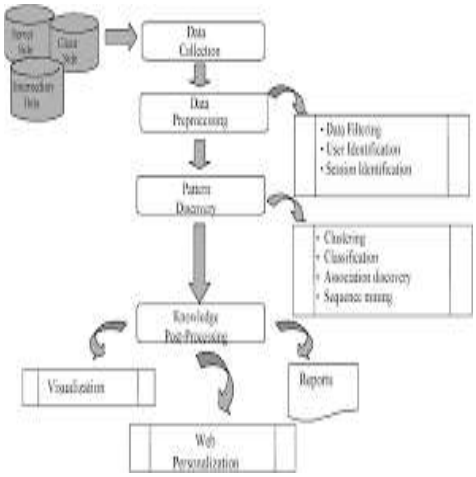

Fig:2 process of web usage mining

A. Gathering Data from Different Sources:

The navigation patterns of usage data which is collected from different sources, represent the overall web traffic, ranging from single user, single site browsing Behavior to multiuser, multi-site access patterns. The primary data sources in web usages mining are Server log files. When a user requests a

Web server, Web log file records information about users' activity.
A log file is located at web server, client browser and web proxy servers. For web usage mining, the usage data collected from different sources like,

1) Server data
2) Client data
3) Proxy data.

1) Server Data:

Web server collect User logs that include IP address, page reference and access time. Web server can also store cookie information, data generated by online visitors while searching relevant information, etc. Web Usage mining retrieved useful patterns from these all data, which is used further for decision making.

2) Client Data:

For Client-side data collection, modify the source code of browser which enhances the data collection capabilities of browser. The web usage mining uses this Client-side data collection to improve the caching and session identification problems.

3) Proxy Data:

An intermediate level of caching between client browsers and Web servers is known as Web proxy. Proxy caching is used to reduce the loading time of a Web page because it reduce network traffic at the server and client sides. Proxy element may acknowledge the actual HTTP requests from multiple clients to multiple Web servers. Web Usage mining gets useful patterns from these proxy data and used it for required purpose.

B. Pre-processing of Data:

The data is converted in the format which is easily understood by the users. User can use this when it is required to them. In web usage mining, required pattern is generated from web log files. The web log data is used for the identity of the user. These log files give details of sites requested by user and time spent by user on each site.
These data processed again and give required information about customer's need which is helpful in marketing.
The preprocessing includes:

1) Data Cleaning
2) Data Integration
3) Data Conversion
4) Data Reduction

1) Data Cleaning

In Data cleaning method, the data which is not used for analysis process is eliminated from log files.

2) Data Integration

After the data cleaning, one or a series of transaction identification modules is used to partition the log entries into logical clusters. It is applied in two ways; either many page references and a single transaction, or single page reference and many transactions.

3) Data Conversion

The log file contain client IP address, client name, date, time, instant name, server name, server IP, status codes, method and page name. [5] In data conversion method, table is created using algorithm which helps to derived useful patterns from the Log files, which is used for data mining.

4) Data Reduction

Here, Digital information is converted into simple form. It diminishes the amount of data to gain required information from log files.

C. Pattern Discovery:

Main goal of Web usage mining is to divulge interesting patterns from log files. These give important information about the users of a system. For pattern discovery and analysis, generic machine learning and data mining techniques, such as association rule mining, classification, and clustering, can often be applied.

D. Pattern Analysis:

Final stage of Web usage mining is Pattern Analysis. This process eliminates inapplicable patterns and extract the required patterns or rules from the output of the pattern discovery process. Main aim of this analysis is to discover "how people are using the site?", "Which of them are accessed most frequently?" etc. To get this kind of information, the contents of the page and structure of hyperlinks are analysed. Some methodologies and analysis tools are used for this analysis. Few most common techniques for pattern analysis are OLAP Techniques, Data and Knowledge Querying, Data and Knowledge Querying, Visualization Techniques and Usability Analysis.

E. Pattern Application:

There are various purposes for which Web usage mining has been used. For example, mining marketing intelligence from Web data, Web traffic patterns also can be produced from Web usage logs to maximize the performance of a Web site.

II. PERSONALIZATION ON SOCIAL WEB

Nowadays drastic growth in involvement of media content such as blogs, social sites, web personalization is used to recommend user only useful content to them. Personalization requires collecting visitor information and gets knowledge that how is the user's behavior on particular site which helps website administrator to decide "what information present to which user and how to present it".

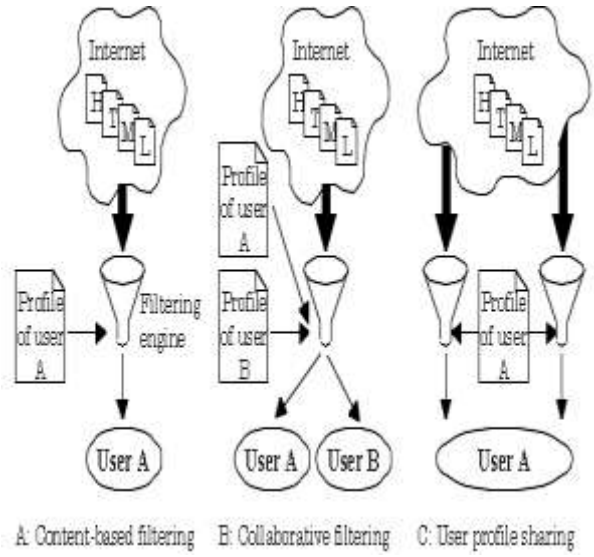

Fig:3 Different methods of filtering data

To make the visitor's time more productive and engaging on the site, personalization of data is very much helpful. It helps administrator of the website to Increase visitor response towards their sites. Personal data of users used by Social Network websites to provide relevant advertisements for their users. To give better services Google and Facebook sites are using account information of users. There are three categories of personalization:

1) Profile / Group based
2) Behavior based
3) Collaboration based

Rules-based filtering is include in Web personalization models, based on "if this, then that" rules processing, and collaborative filtering, which serves relevant material to customers by combining their own personal preferences with the preferences of like-minded others. Collaborative filtering works well for books, music, video, etc. However, it does not work well for a number of categories such as apparel, jewelry, cosmetics, etc. Recently, another method, "Prediction Based on Benefit", has been proposed for products with complex attributes such as apparel. [10]

To differentiate between different users or groups of users, the system should be able personalize a web site. This is known as user profiling process. There are few websites which require user's registration on the website then the information situated on the web log data can be merged with the users' data, as well as with their individual ratings or purchases. In order to classify users on web, web personalization is widely use .it uses Web usage mining to extract log files which contains user's navigation information.

Other approaches also consider as Collaborative filtering. These techniques are based on relationships concluded from users' profile. To categorize classes of profile, Implicit filtering process observes user's behavior and activities. Other approaches consider information semantics which assume that reading the content and interacting with a document takes much time for the user. Reading time, scrolling over the same page and interacting with the system consider as three sources of implicit feedback which gives approximate user's interest for a given web page. For Web information retrieval and mining, Web pages are the component to be analyzed, organized and presented to the user. Personalization process has been improved at the semantic level, based on user modeling and on log files analysis.

III. PERSONALIZATION STRATEGIES

Personalization is categories in four basic categories:

A. ***Memorization*** – In this method, name and browsing history of users stored information which is used to acknowledge that user. Generally, Web server is used to implement this.
B. ***Customization*** – In this method, Input is given by users from registration forms which are useful in customization of the content and structure of a web page. Yahoo and Google are typical examples of web personalization.
C. ***Guidance or Recommender Systems*** –This is a guidance based system recommend hyperlinks as per user's interests which gives facility to user to access needed information on large website. To recognize users interest Web server logs and questionnaire of registration form is used.
D. ***Task Performance Support*** –client side personalization systems involves a personal assistant executes actions on behalf of the user, which helps users to get required information. It requires access, installation, and maintenance of the personal assistant software. Limitation is, it cannot use information about other users with similar interests.

IV. CONCLUSIONS

In this paper, first I describe web mining types then focus on web usage mining and its functions. Web usage mining is helpful to users for reorganization of sites, recommendation of required product or site. I describe here useful area of web usage mining, web personalization. Personalization helps to improve visitor response at the sites. It helps website administrator to present required information to their user and achieve user satisfaction.

Web personalization is useful in social sites. Personalization can helpful to the administrator to know about users choice on the basis of profile or group which developed by user as well as data of media or product shown by on the site. To gather such kind of information few personalization strategies are useful. This paper gives overview that how web personalization is facilitate user with their required data from wide variety of knowledge which is available on internet.

ACKNOWLEDGMENT

This research paper is developed with the help of everyone's support and Guidance. I want to show special gratitude towards God. He gives me enough Wisdom for my research. I would like to thank my family and friends for giving me support and encouragement. This gives me the good experience for research. It teaches me how to participate in a serious project.

REFERENCES

[1] Agrawal R. and Srikant R. (2000).Privacy preserving data mining, In Proc. of the ACM SIGMOD Conference on Management of Data, Dallas, Texas, 439-450.

[2] Berners-Lee J, Hendler J, Lassila O (2001) The Semantic Web. Scientific American, vol. 184, pp34-43.

[3] Berendt B., Bamshad M, Spiliopoulou M., and Wiltshire J. (2001). Measuring the accuracy of sessionizers for web usage analysis,
In *Workshop on Web Mining*, at the First SIAM International Conference on Data Mining, 7-14.

[4]International Journal of Computer Applications (0975 – 8887) Volume 69– No.8, May 2013 39 Web Mining Techniques in E-Commerce Applications Ahmad Tasnim Siddiqui College of Computers and Information Technology Taif University Taif, Kingdom of Saudi Arabia Sultan Aljahdali College of Com puters and Information Technology Taif University Taif, Kingdom of Saudi Arabia

[5]K.S.R. Pavan Kumar et al, / (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 3 (5) , 2012,5276-5279 A Critique on Web Usage Mining ,K.S.R. Pavan Kumar V.V. Sreedhar L. Manoj Chowdary

[6]A Social Network Based Approach to Personalized Recommendation of Participatory Media Content
Aaditeshwar Seth and Jie Zhang School of Computer Science University of Waterloo, ON, Canada

[7] AN IMPLEMENTATION OF WEB PERSONALIZATION USING WEB MINING TECHNIQUES A.JEBARAJ RATNAKUMAR Professor and Head, Department of Computer Science and Engineering, Apollo Engineering College, Chennai, Tamil Nadu, India

[8]New Approaches to Web Personalization, Ph.D. Thesis, Magdalini P. Eirinak, Athens University of Economics and Business Dept. of Informatics May 2006

[9] An Implementation on Web Log Mining Bhaiyalal Birla*, Sachin Patel *IT & PCST, Indore India*

[10] Personalization: Collaborative Filtering vs Prediction Based on Benefit Theory, November 05, 2007,

[11] International Journal of Information and Education Technology, Vol. 2, No. 5, October 2012 Web Personalization Using Web Mining: Concept and Research Issue

[12] International Journal of Engineering Research & Technology (IJERT) ISSN: 2278-0181Vol. 2 Issue 9, September - 2013A Study of Web Usage Mining Applications and its Future Trends
1 D. Uma Maheswari, 2 Dr. A. Marimuthu D.Uma Maheswari, Research Scholar,Karpagam University,Coimbatore, Tamil Nadu.
Dr. A. Marimuthu, Associate Professor, Department of Computer Science, Government Arts College, Coimbatore. Tamil Nadu