ISSN:0975-9646

Harsheen Kaur et al, / (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 5 (2) , 2014, 1499-1502

Creating Adaptive Interface by Mining Web Log

Harsheen Kaur,Kailash Bahl

Department of Computer Engineering
Patiala Institute of Engineering & Technology for Women ,
Patiala, Punjab-147002, India

***Abstract*- Today, World Wide Web (WWW) has become largest source of information which serves needs of expanding group of users. Some complex web application and bulky data offered by them often lead user to confusion. *Adaptive interface* is means which are area of focus of many researchers handle this problem efficiently. In this paper, we are going to review existing approaches which use web log data for developing adaptive web systems.**

Keywords – Web usage mining, web server log, Preprocessing, Pattern Discovery, Pattern Analysis.

I. INTRODUCTION

Creating complex website is obstacle in user interface design. Websites contains many links, images and facts. Users surfers through website have different goals. Many websites contains complex information which disorientates the user. A good website design may yield information and facts which help user to reach his objective or goal for which he is surfing through the site. In this paper, we are going to review some techniques which are used for developing adaptive web sites. Adaptive websites are sites that automatically improve their organization and presentation according to user navigation pattern.

A. Background

Web mining remains area of focus of many researchers. Basically it combines two research areas i.e. Data mining and WWW.

Data mining is a means of extracting information from large bulky data. When this data mining technique is applied to web data, it is termed as Web mining. Hence Web mining makes use of data mining techniques to discover and extract information automatically from Web [4].As literature concerned, Web mining can be categorized as per as web data which is given as input during data mining process, i.e. Web structure, Web content and Web usage mining.

Web mining includes following subtasks:

1) *Discovering Resources*: it involves fetching desired information from web.
2) *Extracting Information*: to select and process information automatically from fetched web resources.
3) *Generalization*: finds patterns automatically from web.
4) *Analysis*: examine the mined pattern.

Web structure mining considers the organizational structure of web page or web site. Web structure mining takes into account hyperlinks and helps in finding out similarity and relationship between websites by taking benefit of hyperlink topologies. Structural mining proves to be beneficial for navigation purposes and combines or compare various web pages schemes.

Web content mining is applying mining techniques to web content [16]. Web content includes data in the web page which communicates with user. Data can be in the form of unstructured text, graphics, sound, video and hypertext. Meaning of voluminous unstructured web data can be explained by creating conceptual schema [17] in order to manage them. Content mining can be categorized as text mining and multimedia data mining as discussed in [18].When mining algorithm is applied to unstructured text then it is termed as text mining. Whereas applying mining algorithm to various type of unstructured graphics data then it is termed as multimedia data mining.

Web Usage Mining is major area of research as proposed by many researchers. Web usage mining is that category of web mining which uses the information inside the web repository and this information is basically originated from user behavior while he is accessing the web. By using such sort of information one can customize or personalize the site according to his needs and this can help to ease the navigation behavior.

B. Web Usage Mining

When users surfing on the web his activities get recorded in repositories which are located at web servers and proxy servers. By accessing information inside web repositories, user behavior can be learned. Information inside the web repositories include IP address of requestor, date/time on which request is made and source from where request is generated. Using this information one can personalize or customize the sites. By analyzing or processing information inside web logs, developer can develop the interface or sites according to user interests and enhance the performance of their systems.

Further some sites can modify its contents or improve their organization according to user navigation behavior. Such sites are adaptive websites. Process of discovering and examine user's navigation behavior or pattern while he is interacting with websites is called web usage mining.

Web usage mining is the application of data mining techniques which is used to find some useful pattern that can serve the needs of web applications.

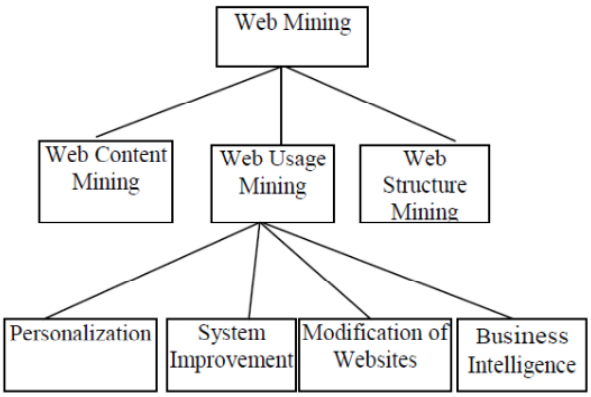

Figure1.Web Usage Mining Categories

Web usage mining consists of three phases, namely preprocessing, pattern discovery, and pattern analysis [2]. These are briefly explained as follows.

1) Preprocessing: Data Preprocessing is preliminary step of data mining. Often log file contains raw, unstructured data consisting of noise; this raw data has to undergo certain stages in order to obtain data which is suitable for mining purpose. This chain of stages is called data preprocessing. It consists of following steps:

(i) Data cleaning: During data cleaning, irrelevant data which is not needed for mining process is deleted from log files i.e. jpg and gif images, navigation performed by spiders or robots etc.

(ii) User and session identification: User is uniquely identified through his IP address. When particular user generates a request his IP address along with source from which request is generated is recorded in web log.

Set of pages visited by a single user within the duration of one particular visit to a website is called session. Session is identified on basis of time or navigation.

(iii) Retrieving of information about page content and structure: Page URLs are considered to be main source of applications used by web usage mining applications but they do not convey any information about contents of page. [19] Content based information is applied to enrich data in the web log. Web structure mining is applied if adequate classification is not known in advance.

(iv) Data formatting: It is last step of data preprocessing. Data is properly formatted before applying data mining process.

2) Pattern Discovery: Used to find patterns using technique like

- Path Analysis
- Association rule
- Classification & Clustering

The pattern discovery stage is applying data mining techniques like path analysis, association rule mining, clustering, classification etc., on preprocessed log data.

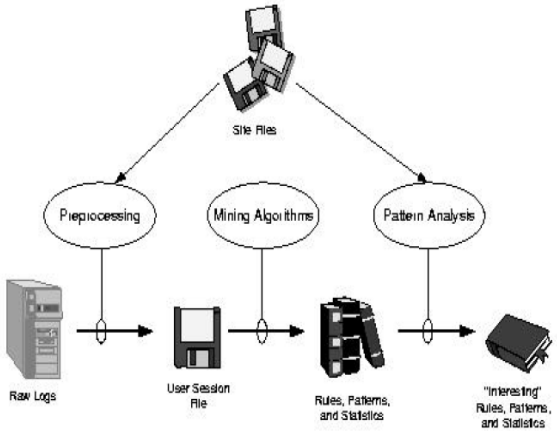

Figure2.web usage mining process

3) Pattern Analysis: Pattern found during pattern discovery phase is analyzed in pattern analysis stage. Pattern Analysis stage make use of-

- OLAP/ Visualization Tool: For multidimensional analysis & Decision making.
- Knowledge Query Management.
- Intelligent Agents.

C. Need of Adaptive Interface

Millions of users interact with websites daily, and every user has different objective and need from other visitors and website administrator. Hence it is needed that website design should be simple, irrelevant material should be prevented to add in the website. So that user can navigate easily through the website.

An adaption can be made by considering following two points:

In first approach of adaptation, website is not going to reorganize its structure or appearance instead some recommendations to the user are suggested.

In second approach of adaptation, website reorganizes itself according to user's interests and needs.

According to Perkowitz & Etzioni [1], adaptive websites are sites that automatically improve their organization and presentation by learning from visitor's access patterns.

Types of adaptations:

- *Customization:* Customization is altering website's contents, so that it can adapt itself to user's needs. Customization creates multiple copies of website i.e. one for each user.
- *Transformation:* Transformation is modifying the site, so that user can navigate easily.
- *Content-based:* In content-based adaptation, content of pages are considered and site organizes and present pages based on their content — what the pages say and what they are about.
- *Access-based:* Access-based uses previous visitor's navigation history in order to guide how information is structured.

Content-based and Access-based adaptations can be used together.

II. RELATED WORK

User can customize website according to his needs or interests. He may add links, reports or quotes etc. He can further enhance customization by observing previous visitor navigation history and suggesting some references and links which can lead new visitor to more interesting pages.

Processing web log and creating adaptive interface based on behaviors of user has remained hot topic in research. Many researchers have worked on the problem of creating adaptive interface according user's interest. For example, *Path Prediction* algorithm determines the next path which user is going to follow in his browsing session. The *WebWatcher* [8] make guess about the links which user follows based on user's interests and predicted link is highlighted on the top of page.

A site can customize itself according to user by making guess about his interests at run time when he enters the browsing session. The AVANTI [7] is online-recommendation system. AVANTI relies on participation of users i.e. when users open particular websites; he enters certain information about them. Based on this information, AVANTI make guess about user's objective and step he will likely to follow next. Then AVANTI will provides the link which it think user will going to follow based on his interest which will take him directly to that page.

Collaborative filtering is also a form of customization. Some recommender system use collaborative filtering. Collaborative filtering automatically determines user's interest by gathering information about user's preference. In this user rate the object based on his preferences. User who gives same rating to same object might have same interest. Hence recommendation system suggests the objects which are highly rated by other users who are having similar interest.Amazon.com uses item based collaborative filtering. Collaborative filtering also finds its application on social web. YouTube, Reddit are example of social media based on collaborative filtering. In case of social media, voting scenario is used .Movies, songs and stories are rated based on user's voting which reflects collaborative filtering approach.

Footprints[10] is an access-based transformation approach. This access-based approach uses previous visitor history i.e. it uses the path the previous visitor has followed while interacted with site. When user navigate through a site, his navigation path get recorded (i.e. links traversed).Then previous user navigation path guides the new user which lead him to some interesting pages. There is no need for visitor to provide information about himself in order to take benefit of the system.

A website able to adapt itself and this adaptation ability can be enhanced if meta-information is given. Meta information includes information about its content, structure, and organization. Meta information can be provided by representing the contents of site in a symmetric (i.e. formal) manner with predefined semantics. Meta-information is used in XML annotations [12], Apple's Meta-Content Format, and other projects [13, 14] in order to customize or optimize the website.

STRUDEL website management system [13] is an example of this approach. STRUDEL's main motive is to separate website's data from graphical presentation. In STRUDEL, firstly developer builds a model of raw data which is present at the site. Secondly, he uses the model of data to define the structure of website and in the final step developer specifies the visual presentation of page.

Previously, we have studied WebWatcher and AVANTI which builds a model based on user's interests by studying navigation history of user when he enters browsing session. Now we are going to discuss *client-side customization*, in which an agent is added by user which learns about his interests and customize website according to his interest. The AiA project [8, 11] adds an agent who can attract user's attention to particular topic of interest and then customize the website accordingly. The agent added by user is 'presentation agent' that maintains a model of individual preferences which helps in deciding what information is going to be highlighted and how to present that information.

Letizia [6] is an intelligent agent. Like WebWatcher and AVANTI, it also develops adaptive website according to user interest.

Some projects are performing customization by using transcoding proxies. Transend [15], for example, is a proxy server at the University of California at Berkeley that performs image compression and allows each of thousands of users to customize the degree of compression, the interface for image refinement, and the web pages to which compression is applied.

III. CONCLUSION

Web sites are of much use for users. Web sites are built, deployed and maintained to serve various functions to user. Complex design and volume of data in a web site confuses its visitors who visits website with different needs and desires. The adaptation of interaction between users and web information space is area of focus. In this paper, we presented and reviewed some existing web systems. We considered a two-dimensional approach for the adaptation of interaction. Here we discussed that web log mining is a means of discovering useful information about the browsing behavior of the users.

The research in the future is expected to focus on developing adaptive interface by exploiting web usage data which makes navigation easier for user.

ACKNOWLEDGEMENT

The author is grateful to Mr. Kailash Bahl, Assistant Professor, Department of Computer Engineering, Patiala Institute of Engineering & Technology, Patiala for their support and guidance.

REFERENCES

[1] Perkowitz M, Etzioni, O. "Adaptive Web Sites: An AI challenge". In: Proceedings of the *IJCAI-97*, Nagoya, Japan, 1997

[2] Mobasher B, Cooley R, Srivastava J. "Creating Adaptive Web Sites Through Usage-Based Clustering of URLs". In: *IEEE Knowledge and Data Engineering Workshop* (KDEX'99).

[3] Paliouras G, Papatheodorou C, Karkaletsis V,Spyropoulos C D. "Clustering the Users of Large Web Sites Into Communities". In: ICML

[4] O. Etzioni, *The world wide web: Quagmire or gold mine?*, Communications of the ACM, 39(11):65-68, 1996

[5] R. Kosala, H. Blockeel, Web mining research: a survey, SIGKDD: SIGKDD explorations: newsletter of the special interest group (SIG) on knowledge discovery & data mining, ACM 2 (1), 1–15, 2000

[6] Lieberman H. "Letizia: An Agent That Assists Web Browsing". In: Proceedings of the *International Joint Conference on Artificial Intelligence*, Montreal, CA, 1995

[7] Fink J, Kobsa A, Nill A. "User-Oriented Adaptivity and Adaptability in the AVANTI project". In: *Designing for the Web : Empirical*, Microsoft Usability Group, Redmond (WA), 1996

[8] E. Andre´, W. Graf, J. Mu¨ller, H.-J. Profitlich, T. Rist and W. Wahlster, AiA: adaptive communication assistant for effective infobahn access, Document, DFKI, Saarbru¨cken, 1996.

[9] Joachims T, Freitag D, Mitchell T. "WebWatcher: ATour Guide for the World Wide Web". In Proceedings of the *International Joint Conference on Artificial Intelligence*, 1997

[10] A. Wexelblat and P. Maes, Footprints: History-rich Web browsing, in: Proc. Conf. on Computer-Assisted Information Retrieval (RIAO), 1997, pp. 75–84.

[11] T. Rist, E. Andre´ and J. Mu¨ller, Adding animated presentation agents to the interface, in: Proc. 1997 Int. Conf. on Intelligent User Interfaces, Orlando, FL, 1997, pp. 79–86.

[12] R. Khare and A. Rifkin, XML: a door to automated Web applications, IEEE Internet Computing 1(4) (1997) 78–87.

[13] M. Fernandez, D. Florescu, J. Kang, A. Levy and D. Suciu, System demonstration — Strudel: a Web-site management system, in: ACM SIGMOD Conf. on Management of Data,1997.

[14] S. Luke, L. Spector, D. Rager and Hendler, Ontology-based Web agents, in: Proc. 1st Int. Conf. on Autonomous Agents, 1997.

[15] A. Fox, S. Gribble, Y. Chawathe and E. Brewer, Adapting to network and client variation using infrastructural proxies: lessons and perspectives, IEEE Personal Communications 5(4) (1998) 10–19.

[16] J. Srivastava, R. Cooley, M. Deshpande, and P. Tan. Web usage mining: Discovery and applications of usage patterns from web data. SIGKDD Explorations, 1(2):12{23, 2000.

[17] K.W. Tan, H. Han, and R. Elmasri. Web data cleansing and preparation for ontology extraction using WordNet. In First International Conference on Web Information Systems Engineering (WISE'00), volume 2.

[18] R. Kosala and H. Blockeel. Web mining research: A survey. ACM SIGKDD, 2(1):1{15, 2000.

[19] Robert Cooley, Bamshad Mobasher, and Jaideep Srivastava. Data preparation for mining world wide web browsing patterns. *Knowledge and Information Systems*,(1):5–32, 1999.