Prof. Gajendra S.Chandel, Prof. Ravindra Gupta, Mr. Hemant k. Dhamecha / International Journal of Engineering Research and Applications (IJERA) ISSN: 2248-9622
www.ijera.com Vol. 2, Issue 3, May-Jun 2012, pp.114-117

# Dynamic user profiles using fusion of Web Structure ,Web content and Web Usage Mining

**Prof. Gajendra S.Chandel , Prof. Ravindra Gupta , Mr. Hemant k. Dhamecha**

HOD (I.T Dept)
SSSIST , Sehore(M.P)
RGPV , BHOPAL

A.P, (CSE Dept)
SSSIST , Sehore(M.P)
RGPV , BHOPAL

MTECH(I.T),4 SEM
SSSIST, Sehore(M.P)
RGPV , BHOPAL

**Abstract** --- ***Web mining is the application of data mining techniques to extract knowledge from Web. Web mining has been explored to a vast degree and different techniques have been proposed for a variety of applications that includes Web Search, Classification and Personalization etc. . In this paper, we highlight the significance of studying the evolving nature of the Web personalization. Using fusion of web structure , web content and web usage mining for Dynamic user profiles***

***Keywords : Web logs ,Web usage . Web Structure Web content, pre-processing , Web Personalization.***

## I. INTRODUCTION

Web mining technology provides techniques to extract knowledge from web data. Researchers on web mining have already identified three main important areas, namely web content mining, web usage mining and web structure mining [1].Web structure mining deals with the discovery of structures from the web topology. Due to the properties of the huge, diverse, dynamic and unstructured nature of Web data, Web data research has encountered a lot of challenges, such as scalability, multimedia and temporal issues etc. As a result, Web users are always drowning in an "ocean" of information and facing the problem of information overload when interacting with the web. A user interacts with the Web there is a wide diversity of user's navigational preference, which results in needing different contents and presentations of information. To improve the Internet service quality and increase the user click rate on a specific website, thus, it is necessary for a Web developer or designer to know what the user really wants to do, predict which pages the user is potentially interested in, and present the customized Web pages to the user by learning user navigational pattern knowledge

## WEB MINING TECHNIQUES

### *A. Web Content Mining:*

Web Content Mining is the process of extracting useful information from the contents of Web documents. Content data corresponds to the collection of facts a Web page was designed to convey to the users. It may consist of text, images, audio, video, or structured records such as lists and tables. Research activities in this field also involve using techniques from other disciplines such as Information Retrieval (IR) and natural language processing (NLP).

### *B. Web Structure Mining*

The structure of a typical Web graph consists of Web pages as nodes, and hyperlinks as edges connecting between two related pages. In addition, the content within a Web page can also be organized in a tree-structured format, based on the various HTML and XML tags within the page. Thus, Web Structure Mining can be regarded as the process of discovering structure information from the Web. This type of mining can be performed either at the (intra-page) document level or at the (inter-page) hyperlink level (Figure 1).

### *C. Web Usage Mining*

Web Usage Mining is the application of data mining techniques to discover interesting usage patterns from Web data, in order to understand and better serve the needs of Web-based applications. Usage data captures the identity or origin of Web users along with their browsing behavior at a Web site. Some of the typical usage data collected at a Web site include IP addresses, page references, and access time of the users.

## II. EXISTING WORK

Web personalization is a strategy, a marketing tool, and an art. Personalization requires implicitly or explicitly collecting visitor information and leveraging that knowledge in your content delivery framework to manipulate what information you present to your users and how you present it. Correctly executed, personalization of the visitor's experience makes his time on your site, or in your application, more productive and engaging. Personalization can also be valuable to you and your organization, because it drives desired business results such as increasing visitor response or promoting customer retention Web personalization include modeling of Web objects (pages, etc.) and subjects (users), categorization of objects and subjects, matching between and across objects and/or subjects, and determination of the set of actions to be recommended for personalization. Figure1 depicts a general architecture for Web personalization based on usage and content mining. The overall process is divided

into two components: the offline component which is comprised of the data preparation and specific Web mining tasks, and the online component which is a real-time recommendation engine. The data preparation tasks result in aggregate structures containing the preprocessed usage and content data to be used in the mining stage. These structures include a user transaction file capturing meaningful semantic units of user activity. Furthermore, Only relevant page views are included in the transaction file. Each page view represents a collection of HTTP requests contributing to a single display in the user browser. The inverted file structure contains the relevant features extracted from text or meta-data associated with the content of the site and their associated weights for each page view.

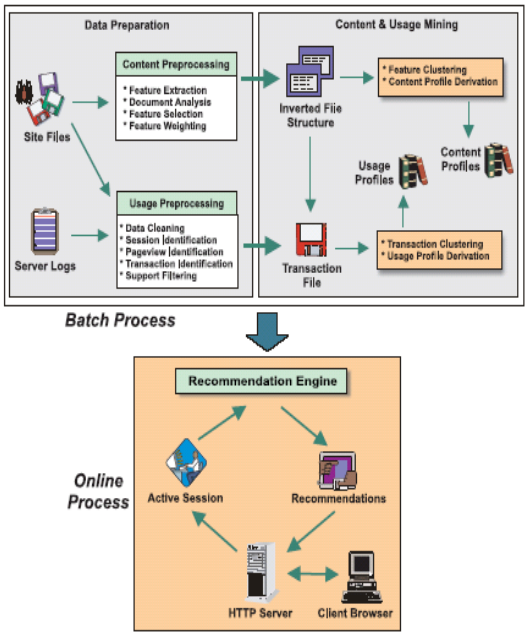


Figure 1. A general framework for Web personalization based on usage and content mining. The data preparation and mining tasks are performed offline resulting in a uniform representation for both usage and content profiles. The real-time recommendation engine uses these profiles and the current active user session to provide recommendations.

Given the pre processed data, a variety of data mining tasks can be performed. For example, the
usage mining tasks can involve the discovery of association rules [1,2], sequential patterns [3], page view clusters, transaction clusters, or any other pattern discovery method from user transactions. The content mining tasks may involve feature clustering (based on occurrence patterns of features in page views),page view clustering based on content or meta-data attributes, or the discovery of (content-based)association rules among features or page views..
In the online component of the system, the Web server keeps track of the active server session as
the user's browser makes HTTP requests. The recommendation engine considers the active server session in conjunction with the discovered patterns and profiles to provide personalized content. The personalized content can take the form of recommended links or products, targeted advertisements, or text and graphics
tailored to the user's perceived preferences as determined by the matching usage and content profiles.

## III. PROPOSED WORK

In this paper we have focused on the effectiveness of that particular web data by evaluating it by means of all three web usage , web structure and web content mining.
Process

Step 1: Get the log file (From server)

Step 2: Extract the web pages and file structure from preprocessing of web pages .

Step 3: Get the content file from file sites

Step 4: Combining the all three usage , content and web pages for effective result

In the above algorithm by combining all three web mining techniques we calculate the time and show more accurate results.

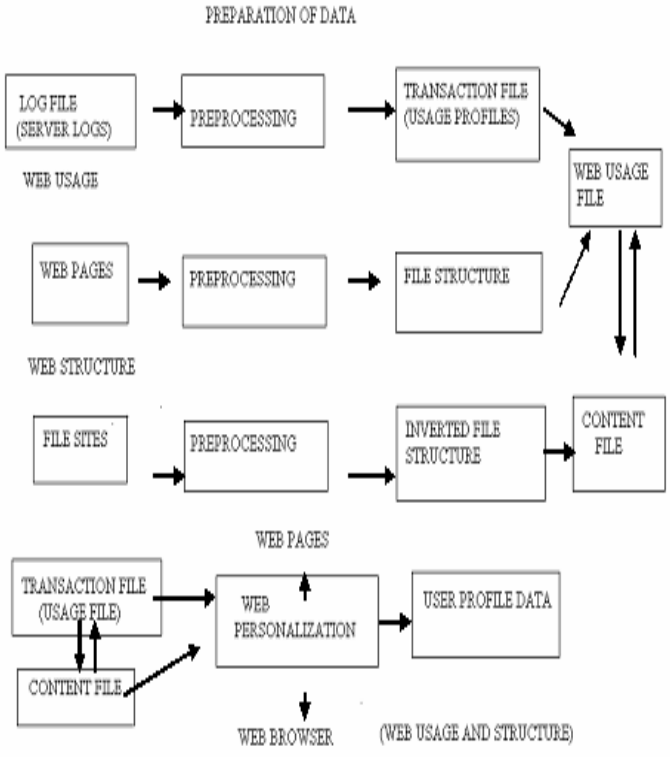


Fig -2

In this algorithm we are matching the results on the basis of all three mining techniques.

## IV. EXPERIMENTAL RESULTS

By conducting the series of experiments with the data from web which contains various information regarding no. of conferences and meetings held . by means of :

1.Usage and Structure preprocessing we generate usage as well as Structured data .

2. We evaluate this on basis of recommendations we get of that particular data.

3. Content table explains the recommendations results on the basis of score held.

4. The minimum score is 0.5 above and equal results are calculated via recommendations.

TABLE 1. Recommendations based on Usage and structure profiles

| Active Session | Recommendations | Score |
|---|---|---|
| • Conference Meeting | NO RECOMMENDATION | - |
| • Conference Update | BCL BOARD OF DIRECTOR MEETING | 0.61 |
| | ANNUAL CONFERENCE | 0.60 |
| | MARKETING | 0.59 |
| | DEVELOPMENT | 0.54 |
| • Conference on Global Marketing | BCL BOARD OF DIRECTOR MEETING | 0.57 |
| | ANNUAL CONFERENCE | 0.55 |
| | GLOBAL MARKETING | 0.53 |
| | DEVELOPMENT | 0.52 |

TABLE 2. Recommendations based on Content profiles

| Active Session | Recommendations | Score |
|---|---|---|
| • Annual Conference meeting | SPECIAL LETTERS AND ASSIGNMENT TOPICS | 0.70 |
| • Conference Update | BCL BOARD OF DIRECTOR MEETING | 0.61 |
| | ANNUAL CONFERENCE | 0.60 |
| | MARKETING | 0.59 |
| | DEVELOPMENT | 0.54 |
| • Conference on Global Marketing | BCL BOARD OF DIRECTOR MEETING | 0.57 |
| | ANNUAL CONFERENCE | 0.55 |
| | GLOBAL MARKETING | 0.53 |
| | DEVELOPMENT | 0.52 |

The recommendations engine was used for this sample user session .It is clear from the table 1 and table 2 the combination of transaction profile and content profile will provide the desired effective data to the user.

## V. CONCLUSION

With the use of this three web mining techniques we can sort out the most effective and efficient results will evaluate on the basis of combination recommendations we get .Discovery of authorities and link discovery. Domains include the world-wide web, bibliographic citations, criminology and bio-informatics, to name just a few. Learning tasks range from predictive tasks, such as classification, to descriptive tasks, such as the discovery of frequently occurring sub- patterns. There are other different data mining challenges in link mining such as identify of the, Link discovery, common relational patterns, where these topics lie in research area.

## ACKNOWLEDGEMENTS

"Dynamic user profiles using fusion of web usage ,web structure and web content mining" research is supported by the management of **SSSIST ,Sehore(m.p) , India** .We are pleased to acknowledge.

## VI . REFERENCES

[1] "Fusion of Web Structure and Web usage Mining"(IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 2 (3) , 2011, 965-967

[2] An implementation of web personalization using web mining techniques Journal of Theoretical and Applied Information Technology© 2005 - 2010 JATIT.

[3] Combining Web Usage and Content Mining for More Effective Personalization(2006)

[4] B. Liu. Web Data Mining: Exploring Hyperlinks, contents, and Usage Data. Springer, 2006.

[5] Miller, C. S. and Remington, R. W., " Modeling Information Navigation : implications for Information Architecture", Human- Computer Interaction, Vol. 19, No. 3, 2004

[6] "Designing the User Interface", Strategies for Effective Human-Computer Interaction 3rd ed, reading MA: Addison-Wesley, 1998.

[7] Larson, K., & Czerwinski, M., "Web page design: Implications of memory, structure and scent for information retrieval", CHI'98: Human Factors in Computing Systems, New York: ACM press 1998, pp.25- 32.

[8] R. Cooley, M. Deshpande, and P.-N. Tan. Web usage

mining: Discovery and applications of usage patterns from web data. SIGKDD.
[9] P. P. J. E. Pitkow Mining longest repeating subsequences to predict World Wide Web surfing. In USENIX, 1999.
[10] S. E. Schechter, M. Krishnan, and M. D. Smith. Using path profiles to predict http requests. Computer Networks, 30(1-7):457–467, 1998.
[11] S.G¨und¨uz and M. T. ¨Ozsu. A web page prediction model based on click-stream tree representation of user behavior. In KDD, pages 535–540, 2003.
[12] E. Frias-Martinez and V. Karamcheti. A customizable behavior model for temporal prediction of web user sequences. In WEBKDD, pages 66–85, 2002.Explorations, 1(2):12–23, 2000.
[13] Y. M. A. Nanopoulos, D. Katsaros. Effective prediction of web-user accesses: A data mining approach. In WEBKDD, 2001.
[14] M. Spiliopoulou. Web usage mining for web site evaluation. Commun. ACM, 43(8):127–134, 2000.
[15] R. Autonomous Agents, Minneapolis/St. Paul, May 9-13, pp132-139
[16] Human-Computer Interaction 3rd ed, reading MA: Addison-Wesley, 1998.
[17] R. Srikant and Y. Yang. Mining web logs to improve website organization. In WWW, pages 430–437, 2001.
[18] B. Mobasher, R. Cooley, and J. Srivastava. Automatic personalization based on web usage mining. Commun. ACM, 43(8):142–151, 2000.
[19] B. Mobasher, H. Dai, T. Luo, and M. Nakagawa.Discoveryandevaluation.

## AUTHORS

**Prof .Gajendra S. Chandel** received Master of Technology in Computer Science & Engineering from SSSIST, Sehore Under Rajiv Gandhi Proudyogiki Vishwavidyalaya, State Technological University of MadhyaPradesh He is working as an Asst.Professor and HOD(I.T) In SSSIST, Sehore,Madhya Pradesh He has publish many papers on data mining..

**Prof .Ravindra Gupta** Master of Technology in Computer
Science & Engineering from SSSIST, Sehore Under Rajiv Gandhi Proudyogiki Vishwavidyalaya, State Technological University of MadhyaPradesh He is working as an Asst.Professor in SSSIST, Sehore,Madhya Pradesh He has publish many papers on data mining..

**Mr .Hemant Kumar Dhamecha** received bachelor degree in Information Technology from Rajiv Gandhi Proudyogiki
Vishwavidyalaya, State Technological University of Madhya
Pradesh and pursuing Master of Technology in Information Technology from SSSIST, Sehore under Rajiv Gandhi Proudyogiki Vishwavidyalaya. He has publish papers on data mining and participated in many workshops.

.