

## Article

# Recommendation System to Predict User Future Movement: A Survey

Swati Bhavsar<sup>1</sup>, Varsha H Patil<sup>2</sup>, Abole H Patil<sup>3</sup>, Radhika Deore<sup>4</sup>

<sup>1</sup>Assistant Professor, <sup>2</sup>Professor, Information Technology Department, Matoshree College of Engineering and Research Centre, Nashik, Maharashtra, India.

<sup>3</sup>Assistant Professor, MIT COE, Pune.

<sup>4</sup>Web Developer, USA.

## I N F O

### Corresponding Author:

Swati Bhavsar, MCERC, Nashik.

### E-mail Id:

bhavsar.swati4@gmail.com

### How to cite this article:

Bhavsar S, Patil VH, Patil AH et al. Recommendation System to Predict User Future Movement: A Survey. *J Adv Res Power Electro Power Sys* 2021; 8(1&2): 17-19.

Date of Submission: 2021-05-03

Date of Acceptance: 2021-05-23

## A B S T R A C T

Web usage mining is one type of web mining used for prediction of users' future movements. Users intended search log is maintained. From obtained web logs, web file is created by server. With the use of web server file future search prediction of user is performed. Significant development has been seen in the last decade for prediction of user's future movement. This paper provides an overview of various predicting systems.

### General Terms

Web usage mining, prediction, future movements, online component, offline component.

**Keywords:** Prediction, System Improvements, Categories

## Introduction

Now days, an important source of information retrieval is a Web. People from different categories are retrieving the information from the web. Their usage information is maintained in the web logs.<sup>1</sup> Analysis of web logs is useful for designing the patterns of information is known as web usage mining. Therefore, web usage mining is known as one of the data mining techniques to discover usage patterns from the web data. The applications of web usage mining are in the field of personalization, system improvements, site modification, business intelligence and usage characterization.

All over the world, large number of users is accessing the web. Large amount of data is generated relate to URLs and collected by web servers. Many times user accessed similar kind of data in terms of web pages and record is maintained in log files. These accessed web pages can be considered as web access pattern and can be used for finding users behaviors information and to predict user's future movement.

The paper is organized as follows. Section 2 discusses literature survey. Section 3 presents system architecture. Section 4 discusses conclusion.

## Literature Survey

Significant development has been seen in web usage mining in the last decade. It mainly focuses on pattern discovery (including Association Rules, sequence pattern, etc) and pattern analysis. Given that high-quality data helps a lot in improving Pattern mining. Precision, studies from this aspects, and proposes the high-effective data pre processing method. Web mining<sup>14</sup> mainly focuses on to provide personalization on the web.

Huan and Kamber presented a new unifying web mining approach for mining of web data.<sup>2</sup> With respect to the research, presented, web mining is classified into three categories, web structure mining which identifies authoritative web pages, web content mining which performs automatic classification of web pages and web usage mining identifies user access pattern while retrieving information.

Web usage mining mainly consists of three phases as preprocessing, pattern discovery, and pattern analysis. Data pre-processing is the process to convert the raw data into the data abstraction necessary for the further applying the data mining algorithm. For pattern mining, user navigation patterns are described as common browsing characteristics among groups of users. The representative user navigation patterns are obtained by clustering algorithms. Clustering of user navigation pattern aims to groups sessions into clusters based on their common properties. Access sessions obtained by clustering are actual patterns of web user activities. These patterns will be used to further classify current user activities. The common approach clustering approach is used for finding collection of related pages are particular web site, relying on visit coherence assumption. The pages that a user visits during one interaction with the site tend to be conceptually related.

The first web usage mining system is proposed by Analog (Yan, Jacobsen, Garcia-Molina, & Dayal,) it mainly based on the concept of offline and online components. The off line component builds users profile based on his past activities and log file is maintained. While online component builds user profile based on his active sessions. It uses geometrical approach for clustering.<sup>11</sup>

Mobasher, Cooley, and Srivastava (2000) and Nakagawa and Mobasher (2003) present web personalizer which uses data mining techniques such as clustering, association rules and sequential pattern discovery in preprocessing phase and to build users profile.<sup>3</sup> The profile building is based on visited pages by the users and their usage characteristics. Web server logs are maintained, and clusters are formed. The active user sessions are used for matching the profiles and to suggest recommendations.<sup>3</sup>

Personalization systems and usage logs are maintained with the help of graph. Perkowitz & Etzioni, 2000b Suggested model for clustering based on connected components and cliques of graph. A new index page is created from each cluster with hyperlinks to all the pages in the cluster.<sup>4</sup> The algorithm developed is named as PageGather. User sessions are termed sets of pages represents vertices and edge between nodes as if the corresponding pages co-occur in more than

a certain number of sessions. The collection of set of vertices and edges forms a graph. It is treated as particular cluster represents a unique behavioral pattern, associating pages in a particular Web site.<sup>4</sup>

Jagan and Rajagopalan focus on web usage mining and algorithms used for providing personalization on the web. Web usage mining (WUM) is a type of Web mining, which exploits data mining techniques to extract valuable information from navigation behavior of World Wide Web users.<sup>5</sup> Aye mainly focus on data pre processing stage of

the first phase of Web usage mining with activities like field extraction and data cleaning algorithms. Data pre processing is considered as an important phase of Web usage mining due to unstructured, heterogeneous and noisy nature of log data.<sup>16</sup>

Srivastava et al worked on data fusion, data extraction and data cleaning steps of preprocessing and proposed an algorithm for data extraction which extracts log data according to analysis of time duration.<sup>17</sup>

Web Mining plays a vital role in research area in the field of data mining. Hence, Par mar study some algorithms are presented which can be used according to one's requirement.

**Table I**

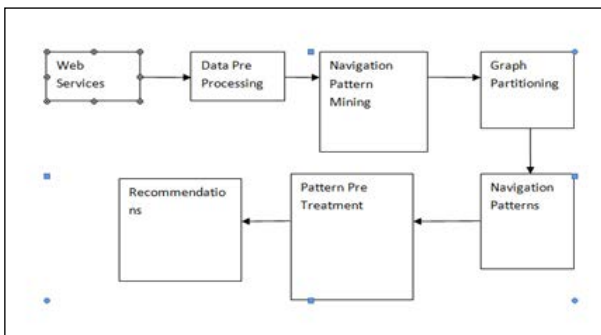
Author	Features	Algorithms Used
P T. Ramya et al. <sup>7</sup>	Time attributes: Time duration, Burst of visit, are considered for finding user behavior Navigation of user pattern and inter visiting time.	K-means
Vijayalakshmi K, et al. <sup>8</sup>	website search for the web personalization	Multilevel association Rule mining classification algorithm.
Nandita Agarwal et al <sup>9</sup>	Web log fetching is used to collect users data	No specific Algorithm
M.Aldekhail <sup>10</sup>	Personalization is based on frequent pages accessed	Multilevel Association Rule Mining and Hierarchical Clustering
P. Dhana lakshmi, et al. <sup>11</sup>	Users behaviors is predicted based on sequence of navigation	(Association Rule)
Goran Matosevic et, al <sup>12</sup>	user usage behavior based on time and content attributes	Clustering Decision tree
Analog et,al. <sup>13</sup>	offline and online component	Geometrical Clustering approach
Mobashe et, al. <sup>3</sup>	Visited log of user	sequential pattern discovery

Perkowitz & Etzioni <sup>4</sup>	connected components and cliques of graph	Linear Common Subsequence
----------------------------------	---	---------------------------

## System Architecture

General framework for web based recommendation system is shown in figure 1. It consists of web services as input. Web server log files are collected from web servers. Data preprocessing is performed to remove irrelevant information.

Navigation and pattern mining will be performed by identifying the number of users and sessions. After that graph partitioning is applied to provide more accuracy with the steps as computation of degree of connectivity between web pages, data set D consisting of set of connected vertices. Which sets are further more forming of adjacency matrix, creating an undirected graph, and finding stronger connected component of it based on graph search algorithm. Finally based on pattern searching more accurate recommendations will be provided.



**Figure 1**

The process of clustering carried out with following steps

- Computer degree of connectivity between web pages and create adjacency matrix. Create an undirected graph corresponding to the adjacency matrix
- Find connected component in the graph based on graph search algorithm
- Find patterns in large graph, it is desirable to analyze, visualize, summarize and mine it. As it is a complex data structure such graphs require excessive processing, more memory for storage and knowledge of a pattern of the graph. Some of the graphs changes with time are known as dynamic graphs. And hence it is very difficult to comment on exact size and pattern of such large graphs.

## Conclusion

For users future movements prediction, commonly used techniques include pattern discovery using graph partitioning, LCS and Classification techniques such as Naive Bayesian. There is still scope for improving the prediction

accuracy by applying graph partitioning technique based on most connected and classification based on association rule mining.

## References

1. Kaur D, Kaur S. A Study on User Future Request Prediction Methods Using Web Usage Mining. *International Journal of Computational Engineering Research* 2013; 3(4): 155-159.
2. Huan J, Kamber M. Data mining: Concept and techniques. San Mateo, CA., 2000.
3. Mobasher B, Cooley R, Srivastava J. Creating adaptive Web sites through usage-based clustering of URLs. Paper presented at the knowledge and data engineering exchange, Chicago, IL, USA, 1999: 19-25.
4. Perkowitz M, Etzioni O. Adaptive Web sites. *Communications of the ACM*, 43(8): 152-158.
5. Jagan S, Rajagopalan SP. A Survey on Web Personalization of Web Usage Mining. *International Research Journal of Engineering and Technology*, Vol.2, No.1 2015; 2(1).
6. Dalvi BB, Kshirsagar M, Sudarshan S. Keyword Search on External Memory Data Graphs 2008; 24-30.
7. Ramya PT, Sanjeev GP. Building web Personalization System with Time Driven Web Usage Mining. *ACM, Proceedings of the Third International Symposium on Women in Computing and Informatics*, August 2015.
8. Vijayalakshmi K, Jena S. Web Usage classification and Clustering approach for Web search Personalization. *ICCT*, ACM 2015; 25-27.
9. Vithya M, Suguna S. Web usage mining techniques for e-commerce applications: a survey, *International Journal of Computer Engineering and Applications*, 2018.
10. Aldekhail M. Application and significance of Web usage mining in the 21 st century: A Literature Review. *International Journal of Computer and Theory and Engineering* 2016; 8(1).
11. Dhanalakshmi P, Ramani K, Reddy BE. The Research of preprocessing and Pattern Discovery Techniques on web log files. 2016 IEEE 6th International Conference on Advanced Computing.
12. Matosevic G, Bevanda V. Mining Customer Behavior in trial period of a web application usage. *Springer International Publishing Switzerland* 2016.
13. Yan, Jacobsen, Garcia-Molina et al. "E-Learning Platform Usage Analysis", Research Gate, 1996.
14. Aye. "A Tool for Web Usage Mining", 8th International Conference proceeding, Birmingham, UK, December 16-19, 2007: 695-704.