Scientific Journal Impact Factor (SJIF): 1.711

e-ISSN: 2349-9745
p-ISSN: 2393-8161

International Journal of Modern Trends in Engineering and Research
www.ijmter.com

# A Review on Clustering Techniques used in Web Usage Mining

**Seema Sheware[1], A.A. Nikose[*2]**

[1] *Department of Computer Sci & Engg , Priyadarshini Bhagwati College of Engg Nagpur,Maharashtra, India*

**Abstract**— The World Wide Web continues to grow repository of web pages and links at an exponential rate which makes exploiting all useful information a standing challenge. It has recently a wide range of applications in E-commerce web site and E-services such as building interactive marketing strategies, Web recommendation and Web personalization. Web usage mining is the process of extracting useful usage patterns from the web data. Web personalization uses web usage mining technique for the process of knowledge acquisition done by analysing the user navigational patterns interest. Nowadays, the Web is an important source of information retrieval, and the users accessing the Web are from different backgrounds. The usage information about users is recorded in web logs. Analysing web log files to extract useful patterns is called Web Usage Mining. Web usage mining approaches include clustering, association rule mining, sequential pattern mining etc. This article provides a survey of the available literature on Web usage mining and reviews the research and application issues in web usage mining.

**Keywords**— Web usage mining, server log file, web logs, clustering, fuzzy logic.

## I. INTRODUCTION

Web has become an unstoppable part of world and web surfing is an important activity for customers who make purchases online. Web mining is the application of data mining techniques used to extract useful patterns from the web. According to analysis objective, web mining can be divided into three different types, which are web usage mining, web content mining and web structure mining [1, 2].

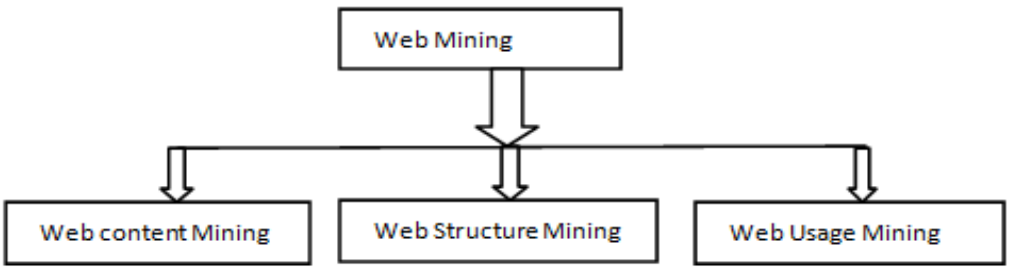


**Fig.1: Web Mining Categories**

### A. *Web Content Mining*

Web content mining is a process of extracting information from texts, images and other contents. The technologies that are mainly used in web content mining are NLP (Natural language processing) and IR (Information retrieval).

 

### B. *Web Structure Mining*

Web structure mining is a process of extracting information from linkages of web pages. Web structure mining is the process of using graph theory to analyse the node and connection structure of a web site. This graph structure can provide information about ranking and enhance search results of a page through filtering.

### C. *Web Usage Mining*

Web usage mining is a process of extracting information from user how to navigate web sites. Web usage mining also known as web log mining, aims to discover interesting and frequent user access patterns from web browsing data that are stored in web server logs, proxy server logs or browser logs.

### D. *Applications of Web Usage Mining*

1) *Personalization:* Reconstruct the website based on users profile and usage behaviour.
2) *System Improvement:* Provide help to understand web traffic behavior. There are some benefits of it like web load balancing, data distribution or policies for web caching.
3) *Adjustment of Website:* Understanding visitor's behavior in a web site provides hints for adequate design and update decision.
4) *Business Intelligence:* It occupies the application of intelligent techniques in order to help certain businesses, mainly in marketing.
5) *Effective:* Valuing the effectiveness of advertising by analyzing large number of access behavior patterns.
6) Improving the design of e-commerce web site according to users browsing behavior on site in order to better serve the needs of users

## II. RELATED WORK AND LITERATURE SURVEY

S. park et al, proposed a framework in which the performances of the algorithms are compared in terms of whether the clusters (groups of Web users who follow the same Markov process) are correctly identified using a replicated clustering approach. A series of experiments is conducted to investigate whether clustering performance is affected by different sequence representations and different distance measures as well as by other factors such as number of actual Web user clusters, number of Web pages, similarity between clusters, minimum session length, number of user sessions, and number of clusters to form. A new, fuzzy ART-enhanced K means algorithm is also developed and its superior performance is demonstrated in this paper [3].

X. Zhang et al, describes a toolset that exploits web usage data mining techniques to identify customer Internet browsing patterns. These patterns are then used to underpin a personalized product recommendation system for online sales. Within the architecture, a Kohonen neural network or self-organizing map (SOM) has been trained for use both offline, to discover user group profiles, and in real-time to examine active user click stream data, make a match to a specific user group, and recommend a unique set of product browsing options appropriate to an individual user [4].

Z. Li et al, present a novel ontology based Web usage mining framework that leverages search engine queries to improve the accuracy of unemployment rate prediction. The proposed framework is underpinned by a domain ontology which captures unemployment related concepts and their semantic relationships to facilitate the extraction of useful prediction features from relevant search engine queries. In addition, state-of-the-art feature selection methods and data mining models such as

 

neural networks and support vector regressions are exploited to enhance the effectiveness of unemployment rate prediction [5].

M. Belk et al, focuses on modelling users' cognitive styles based on a Web usage mining techniques on client navigation patterns and click stream data. Main aim is to inspect whether exact clustering techniques can group user of particular cognitive style by measures obtained from psychometric test and content navigation behaviour [6].

M. Wu. et al, proposes an approach based on web mining to analyse product usability. This approach uses the massive online customer reviews on analogous products and features as data source, which are easy to get from Web and can reflect the most updated customer opinions on product usability. Association rule mining techniques are adopted to extract customer opinions on the usability of product features [7].

S. G. Matthews et al, presented genetic algorithm (GA)-based solution is described that uses the elastic nature of the 2-tuple linguistic illustration to discover rules that occur at the intersection of fuzzy set borders. The GA-based advance is enhanced from previous work by including a graph illustration and a better fitness function [8].

Y. T. Wang et al, introduced the concept of throughout- surfing patterns (TSP) and then present an competent method for mining the patterns. Authors propose a compact graph structure, term a path traversal chart, to record information about the navigation paths of website visitors. The graph contains the frequent surfing paths that are required for mining TSPs [9].

X. Wang et al, propose a concurrent neuro-fuzzy model to discover and analyse useful knowledge from the available Web log data. We made use of the cluster information generate by a self organizing diagram for pattern analysis and a fuzzy inference system to capture the chaotic movement to provide short-term (hourly) and long-term (daily) Web traffic movement predictions [10].

G. Castellano et al, proposed NEWER (NEuro-fuzzy Web Recommendation), a usage-based Web advice system that exploits the possible of Computational cleverness techniques to dynamically advise interesting pages to user according to their preference. NEWER employs a neuro- fuzzy move toward in order to conclude categories of users distribution similar interests and to determine a recommendation model as a set of fuzzy rules express the associations between user category and relevances of pages [11].

C. C. Aggarwal et al, designed an algorithm which combine classical partition algorithms among probabilistic models in order to produce an effective clustering approach. Then show how to enlarge the approach to the categorization problem [12].

## III. PROPOSED WORK

We are using data mining techniques such as clustering in data mining and we are expecting the prediction of web usage mining. Web usage mining is the process of finding most important pages or sections from web which being highly visited by user or predicting the user's preference.

 

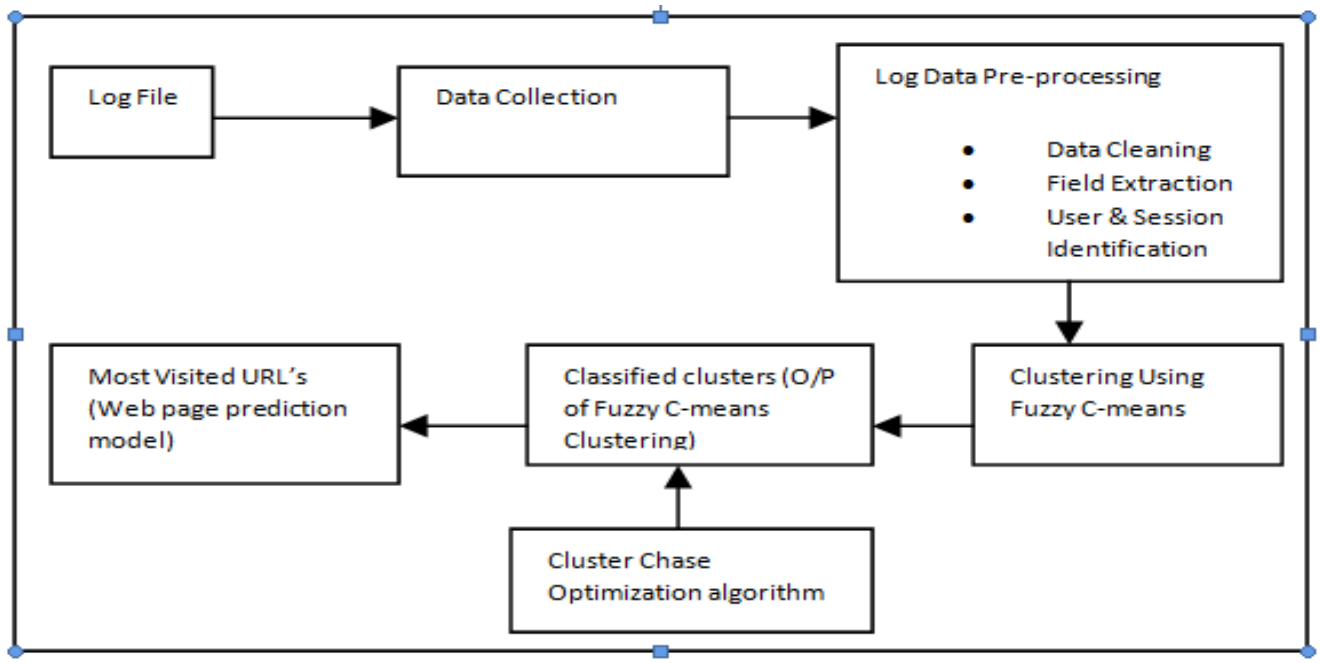


**Fig.2: Proposed Architecture System**

In the above figure architecture of our proposed system is shown. The working of this model is discussed in detail in our next paper where algorithm is explained based on Fuzzy C-means clustering.

## IV. WEB USAGE MINING TECHNIQUES

The important data mining techniques functional in the web domain contain Association Rule, Sequential pattern detection, clustering, path analysis, classification and outlier discovery [13].

### A. *Association Rule Mining*

Predict the association and relationship among set of items "wherever the presence of one set of objects in an operation implies (with a certain degree of assurance) the presence of extra items [14].

### B. *Sequential Pattern Discovery*

It is applied to web access server transaction logs. The purpose is to determine sequential patterns that specify user visit patterns over an assured period. That is, the order in which URLs lean to be accessed [15].

*1)* ***Advantages:*** Useful user trend can be discovered. Predictions concerning stay pattern can be made. To improve website steering. Personalize advertisements Dynamically reshuffle link structure and adopt web site inside to individual client necessities or to provide clients by automatic recommendations that best costume customer profiles.

### C. *Clustering*

It groups together items (like users, pages, etc.,) that have parallel characteristics [16]. Page clusters consists of groups of pages that appear to be conceptually correlated according to users' perception. User Cluster consists of groups or user that seem to be behave equally when navigating through a web site.

 

### D. *Classification*

It maps a data item into one of some predetermined classes. Example: describing every user's category via profiles. Classification algorithms can be decision tree, naive Bayesian classifier, neural networks, Support Vector Machine, K- Nearest Neighbor Classifier [17].

### E. *Path Analysis*

A technique that involve the generation of some type of graph that represents relations defined on web pages. This would be able to be the physical layout of a web site into which the web pages be nodes and links among these pages are directed edges. The majority graphs are involved in formative frequent traversal patterns more frequently visited paths in a web site [18].

## V. CONCLUSION

In this paper we have tried to deliver a survey of the rapidly rising area of Web usage mining, which is the order of current technology. In this paper a common overview of Web usage mining is offered. Web usage mining is used in various fields. We studied various techniques for pattern discovery. We can further work on web usage mining with the combination of these techniques because we need to design algorithm using Fuzzy C-means clustering, which can help to better understand the mined knowledge.

## ACKNOWLEDGMENT

I would be grateful to my guide for her cooperation in studying various clustering techniques used in web usage mining.

## REFERENCES

[1] R. Kosala, H. Blockeel, Web mining research: a survey, ACM SIGKDD Explorations Newsletter 2 (1) (2000)pp, 1–15.
[2] F.M. Facca, P.L. Lanzi, Mining interesting knowledge from weblogs: a survey, Data and Knowledge Engineering 53 (3) (2005)pp, 225– 241.
[3] Park, Sungjune, Nallan C. Suresh, and Bong-KeunJeong. "Sequence- based clustering for Web usage mining: A new experimental framework and ANN-enhanced K-means algorithm." Data & Knowledge Engineering 65.3 (2008)pp, 512-543.
[4] Zhang, Xuejun, John Edwards, and Jenny Harding. "Personalised online sales using web usage data mining." Computers in Industry 58.8 (2007)pp, 772-782.
[5] Li, Ziang, et al. "An ontology-based Web mining method for unemployment rate prediction." Decision Support Systems 66 (2014) pp,114-122.
[6] Belk, Marios, et al. "Modeling users on the World Wide Web based on cognitive factors, navigation behavior and clustering techniques." Journal of Systems and Software 86.12 (2013) pp, 2995-3012.
[7] Wu, Mingxing, et al. "An approach of product usability evaluation based on Web mining in feature fatigue analysis." Computers & Industrial Engineering 75 (2014) pp, 230-238.
[8] Matthews, Stephen G. et al. "Web usage mining with evolutionary extraction of temporal fuzzy association rules." Knowledge- BasedSystems 54 (2013) pp, 66-72.
[9] Wang, Yao-Te, and Anthony JT Lee. "Mining Web navigation patterns with a pathtraversal graph." Expert Systems with Applications 38.6 (2011) pp,7112-7122.
[10] Wang, Xiaozhe, Ajith Abraham, and Kate A. Smith."Intelligent web traffic mining and analysis."Journal of Network and ComputerApplications28.2 (2005) pp, 147-165.
[11] Castellano, Giovanna, Anna Maria Fanelli, and Maria AlessandraTorsello. "NEWER: A system for NEuro-fuzzy WEb Recommendation." Applied Soft Computing 11.1(2011) pp,793-806.
[12] Aggarwal, C., Yuchen Zhao, and P. Yu. "On the use of Side Information for Mining Text Data." (2012) pp, 1-1.
[13] Cooley, R.; Mobasher, B.; Srivastava, J.; "Web mining: information and pattern discovery on the World Wide Web".In Proceedings ofNinth IEEEInternational Conference., 3-8 Nov. (1997)pp, 558 – 567.
[14] Peng, Huiping. "Discovery of interesting association rules based on web usage mining." Multimedia Communications (Mediacom), 2010 International Conferenceon.IEEE, (2010) pp, 272-275.
[15] Masseglia, Florent, DoruTanasa, and Brigitte Trousse. "Web usage mining: Sequential pattern extraction with a very low support." Advanced Web Technologies andApplications.Springer Berlin Heidelberg, (2004) pp,513-522.

 

[16] Varghese, NayanaMariya, and Jomina John. "Cluster optimization for enhanced web usage mining using fuzzy logic." Information andCommunication Technologies (WICT), 2012World Congress on.IEEE, (2012) pp,948-952.
[17] Raghavendra, Prakash S., Shreya Roy Chowdhury, and SrilekhaVedulaKameswari. "Comparative study of neural networks and k-means classification in web usage mining." Internet Technology and Secured Transactions (ICITST), 2010 International Conference for.IEEE, (2010) pp,1-7.

 