

# **A Review on Web Recommendation System**

**Animesh Shrivastava**

PG Scholar

Department of CSE

Shri Vaishnav Institute of Technology & Science  
Indore

**Anand Singh Rajawat**

Assistant Professor

Department of CSE

Shri Vaishnav Institute of Technology & Science  
Indore

## **ABSTRACT**

In Web world, there is immense of information available on the internet but user is not capable to find relevant information in short period of time. Therefore, a system called recommendation system developed to assist user in their browsing activities. It analyzes users need and provides relevant information in shorter span. In this work, various recommendation systems reviewed to analyze their problems and solutions. In order to improve the recommendation quality, a new web recommendation system is introduced. This system uses knn and genetic algorithm during web usage mining process to analyze static web access log.

**Keywords:** Recommendation System, Web Usage Mining.

## **1. INTRODUCTION**

Web usage mining is one of the web mining. It uses data mining techniques to extract information about how user interacts with the web. This information provides path to accessed web pages. The web server automatically collects this information into access logs. The web usage mining analyzes behavioral pattern and profile of users. The data generated due to interaction of user with web is analyzed to discover useful patterns and represented as collections of pages, objects, or resources that are frequently accessed by groups of users with common needs or interests[11]. Web usage mining involves three steps. In first step, the usage data is collected in static web log and then preprocessed to remove all the impurities contained in the data. In the second step, the data mining techniques are applied to extract the patterns from the web data. In the third step, the patterns are analyzed to study the behavior of the web user. In this model, knn and genetic algorithm are used to extract patterns from static web log.

The usage mining is used to personalize, modify sites, system improvement, characterize user and recommendations [12]. In this work web usage mining is used for recommendation purpose. It is important to the overall use of data mining for companies and their internet/ intranet based applications and information access. It helps industries to gain information related to business growth. Some of this information can be derived from the collective information of lifetime user value, product cross marketing strategies and promotional campaign effectiveness. Usage mining is valuable not only to businesses using online marketing, but also to e-businesses whose business is based solely on the traffic provided through search engines. This web mining also enables Web based businesses to provide the best access routes to services or other advertisements [13].

Recommendation system suggests items to the web users from the various items available on the web. These suggestions help the users to make proper decision regarding which item to choose and pages to be viewed in

future[4]. Recommendation systems consist of a series of mechanisms and techniques applied to information retrieval with the purpose to solve the problem of data overload on the Internet. This guide user to select item of its interest, items can be a film, music, website etc. Such systems depends upon personalized information filtering that predict whether a particular item liked by user and find set of items of users interest. When it is difficult to find any item from huge number of available items then recommendation is required. There are Collaboration based, Demographic, Content based, Utility based and Knowledge based Systems. The rest of the paper is organized as follows. In section 2, related work on recommendation systems is analysed. Section 3 gives review of previous recommendation model. Section 4 explains the proposed work. Section 5 concludes the paper with future work.

## **2. RELATED WORK**

Every company advertises its product on the internet through Web advertising service but to advertise according to consumer preferences was difficult. So Tung-Yen Lai et al [2010] analyzed this problem and proposed a personalized Web advertisement selection and recommendation system through a membership based advertisement marketing website whose advertisement content is determined by consumer preference. As the consumer preference declines with time, the half-life theory and fuzzy theory is applied to analyze browsing behavior of users. Now the target advertisement is filtered to match with the user preferences. Now only those advertisements shown that satisfy users need. Therefore, this increases the advantages of target marketing and market of advertisement is strengthened.[1]

When a person searches for any tourist spot on the internet then only textual information gets available for that spot. However, person is unable to get geographical view of that place. Then Liangliang Cao et al [2010] observed this problem and maintained a database consisting of geotagged images. They used a clustering algorithm to partition earth area into regions. Every region represented by geotagged image. Whenever a user queries a destination then the system searches for geotagged image corresponding to that destination. The corresponding geotagged regions obtained as the recommended destinations [2]

The user want to buy a mobile phone using online shopping but it was time consuming to find the model of its interest. Deng-Neng Chen *et al* [2010] implemented a recommendation system that help users to choose a mobile phone based on their preferences. Analytic hierarchy processing based recommendation system designed, implemented and evaluated. This system recommends better than rank based and equal weight systems but not better in terms of user satisfaction with the system due to its lengthy process of evoking user preferences [3]

Some of the basic concepts of recommendation system are discussed by Francesco Ricci *et al* [2011] [4].

There is a lot of content available on the web. The web user can rate this content using some explicit feedback methods. In order to find which method is most preferred by the user. Edward Rolando *et al* [2011] reviewed these methods to find the method mostly preferred by user. They implemented a recommendation system using explicit feedback method. [5]

There are several Web services available on the internet. The service quality of Web services is collected from the service users. Zibin Zheng *et al* [2009] employed an effective and novel hybrid collaborative filtering method for Web service recommendation. This system is used to select and recommend Web services [6].

It is easy to provide recommendation to the users with normal interests. But some people have unique interest and to recommend for such people is not easy. Therefore, Aliaksandr Birukou *et al* [2010] presented a multi-agent recommendation system for web search. The agent analyse the behavior of such people to identify their interest. The system adopts the implicit culture approach for recommending web links and exploits social ties existing within the community [7]

Sometimes, a web page pointed by link is disappeared or moved to another location. In this case, the web user tries to find the pages related to the broken link through several resources. However, this work is not easy and time consuming. Therefore, Juan Martinez Romo *et al* [2011] proposed a system that provides suggestion of the pages in order to recover the broken link. A novel methodology is presented to evaluate the system without resorting to user judgments, thus increasing the objectivity of the results and helping to adjust the parameters of the algorithm. A web page collection is compiled with true broken links that has been used to test the full system by humans [8].

The recommendation system suffers from problems such as sparse data structure, browsing history not available, no diversity in set of recommended items and complexity accuracy tradeoffs. In order to overcome these problems Nizar Mabroukeh *et al* [2011] included semantic information in the web log and integrated domain ontology into web usage mining. The recommendation system uses semantic aware sequential pattern mining and integrates with transition probability matrix of Markov models. This increased the efficiency and effectiveness of the recommendation system. [9]

### 3. REVIEW OF PREVIOUS RECOMMENDATION MODEL

It was observed that pages visited by the user in past time were only suggested by the recommended system. However, the pages which the user never visited in past were not recommended by the system. Previously, the user's browsing behavior does not match with the aggregate usage profiles obtained through data mining process. Therefore, a recommendation system was developed to provide recommendation of unvisited pages. Firstly, data is preprocessed to obtain session file consisting of various page requests. Then model based clustering is used to obtain clusters of various sessions. Each cluster shows users interest in particular page. The web usage mining is used to obtain aggregate usage profiles. In the recommendation phase, the current user profile is matched with the aggregate usage profiles. The matched cluster is used for recommending unvisited pages [13].

**Table 1. Some techniques used in Recommendation Systems**

S.No.	Methods and Algorithms Used	Results
1	Collaborative Filtering Approach using knn algorithm with explicit feedback	Limitation of scalability and performance
2	Collaborative Filtering Approach with implicit feedback	Improves Scalability of Collaborative filtering
3	k-means clustering	Improves prediction accuracy
4	Longest Common Subsequence Algorithm	Improves quality of system for recommendation
5	Formal Concept Analysis Approach	Provides personalization and recommendation
6	Model based Clustering Approach	Discovers user's interest in session
7	Integration of clustering, association rules and markov models	Web page prediction accuracy improved

This work helped organizations in various analyses such as web site improvement. Therefore, this model increased the quality of recommendation in large website. Sometimes there is a large data set. Therefore, an algorithm is required to reduce the data set. However, the use of optimization techniques can further improve recommendation quality.

Therefore, a recommendation model is proposed which uses knn and genetic algorithm during web log mining, which will show its effectiveness in terms of memory required, time consumed, and error rate.

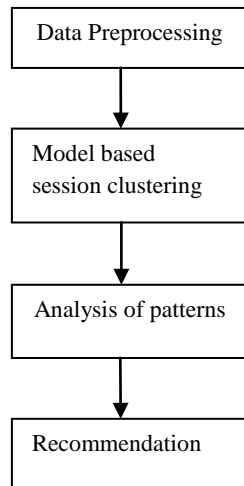


Fig.1 Previous Model

identify a repeat visitor. In data pre-processing, the system accepts the web access log file and cleans them for extracting the desired content from the web access log that is participating in the model.

In this work, the proposed model first consume input data and then pre-process this data using KNN algorithm, the most nearest values are clustered first and the complete dataset is divided in more than one small datasets.

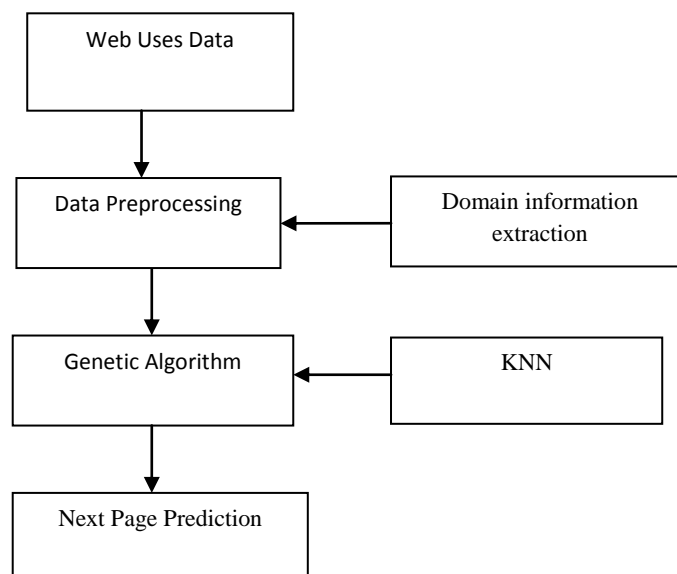


Fig.2 The proposed architecture

#### 4. PROPOSED WORK

In this work, a recommendation model using web access log mining is proposed. A recommendation system is a system that helps web users in their browsing activity by providing suggestions of web items that can be of its interest. This system involves offline and online phase. However, this work is confined to offline phase. The offline phase includes data processing and web usage mining process. The output of this phase passes to the recommendation phase. This phase provides suitable recommendations to the web user. Web server in static web log automatically collects the web usage data. Each client request to the server generates a single entry in the server access log. Each log entry contains the time and date of the client request, the IP address of the client, the resource requested, possible parameters used in invoking a Web application, HTTP method used, the user agent (browser and operating system type and version), the referring Web resource, and, if available, client-side cookies which uniquely

The genetic algorithm processes each cluster data and calculates a fitness value for each input sequence, these fitness values are labels for each input sequence.

Now all the dataset is transposed once, then all data is converted into a new data, and again sequences are produced into genetic algorithm and second time again a fitness value is calculated by the system. Here using row evaluation we found a fitness value says  $f_x$  and using column evaluation we found a fitness value  $f_y$ . Now for each sequence in the clustered data we found a two values and can be represented in the form of  $(f_x, f_y)$  which is plot over 2D hyper plane. The proposed algorithm is summarized using the below given distance and fitness functions. The fitness function measures the performance of the individual string. A fitness function is a problem specific user defined heuristic. After each iteration, the members are given a performance measure derived from

the fitness function and the fittest member of the population will propagate the next iteration.

Fitness  $f_i$ , hit  $\Delta$ , survival  $\phi_i$  and measured distance  $D_i$

$$f_i = 2^{\Delta} - D_i + \sum \phi_i$$

It is expected that web usage mining using genetic algorithm with knn will improve the performance parameters like error rate, time consumed, memory required as compared to web usage mining using only genetic algorithm. So the most accessed page will be obtained more quickly and accurately. This will be used to provide recommendations. These recommendations are usually specialized in predicting the next page of user. Therefore, the proposed model can predict the most probable next page for the user.

## 5. CONCLUSION

Today all information is available on the Internet but it is not easy for every user to find relevant information in short span of time. In order to overcome this problem recommendation system introduced in Web world. In this paper, the problem and solution of some previous recommendation models are analyzed and found that use of optimization techniques can further improve the quality of recommendation. So a new recommendation model using web access log mining proposed. In this work, knn algorithm is used to reduce the data set and then genetic algorithm is applied for classification. These two algorithms analyze the static web log. Therefore, the use of two algorithms one after another can improve the effectiveness of recommendation system.

In future, the implementation of this proposed model will be provided. The log analysis using genetic algorithm and knn with genetic algorithm will be done and their performance will be compared in terms of error rate, memory required, time consumed.

## 6. ACKNOWLEDGEMENTS

It is a pleasure to express obligation and very thankful to all those people who have helped directly or indirectly in successful completion of this review paper.

## 7. REFERENCES

- [1] Tung-Yen Lai, Gwo-Haur Hwang, 'Application of Half-Life Theory and Fuzzy Theory to a Selection and Recommendation System for Web Advertisement Delivery In Consideration of the Time Effect', WSEAS TRANSACTIONS on INFORMATION SCIENCE and APPLICATIONS ISSN: 1790-0832 Issue 1, Volume 7, January 2010.
- [2] Liangliang Cao, Jiebo Luo, Andrew Gallagher, Xin Jin†, Jiawei Han† and Thomas S. Huang, 'A Worldwide Tourism Recommendation System Based on Retagged web Photos', 978-1-4244-4296-6/10/\$25.00 ©2010 IEEE.
- [3] Deng-Neng Chen a, Paul Jen-Hwa Hu b, Ya-Ru Kuo c, Ting-Peng Liang, 'A Web-based personalized recommendation system for mobile phone selection: Design, implementation, and evaluation', Expert Systems with Applications (2010), doi:10.1016/j.eswa.2010.05.066.
- [4] F. Ricci et al. (eds.), 'Recommender Systems Handbook', DOI 10.1007/978-0-387-85820-3\_1, © Springer Science Business Media, LLC 2011.
- [5] Edward Rolando Nuñez-Valdez, Juan Manuel Cueva-Lovelle, Oscar Sanjuan, Carlos E. Montenegro-Marin2,Guillermo Infante Hernandez, 'Social Voting Techniques: A Comparison of the Methods Used for Explicit Feedback in Recommendation Systems', International Journal of Artificial Intelligence and Interactive Multimedia, Vol. 1, No. 4.
- [6] Zibin Zheng, Hao Ma, Michael R. Lyu, and Irwin King, 'WSRec: A Collaborative Filtering Based Web Service Recommender System' 2009 IEEE International Conference on Web Services.
- [7] Aliaksandr Birukou, Enrico Blanzieri PaoloGiorgini, 'Implicit: a multi-agent recommendation system for web search' Auto Agent Multi-Agent Syst (2012) 24:141–174 DOI 10.1007/s10458-010-9148-z © Springer.
- [8] Juan Martinez-Romo, Lourdes Araujo, 'Updating broken web links: An automatic recommendation system' Information Processing and Management 48 (2012) 183–203 doi:10.1016/j.ipm.2011.03.006.
- [9] Nizar R. Mabroukeh "SemAware: An Ontology-Based Web Recommendation System" (2011). Electronic Theses and Dissertations. Paper 415.
- [10] Burke, R.: Hybrid web recommender systems. In: The Adaptive Web, pp. 377–408. Springer Berlin/ Heidelberg (2007)
- [11] Bamshad Mobasher , Web Usage Mining
- [12] Jaideep Shrivastava, Robert Cooley, Mukund Deshpande, PangNing Tan "Web Usage Mining: Discovery and Applications of Usage Patterns from Web Data" Volume 1 ACM, SIGKDD Jan 2000.
- [13] C.P.Sumathi,R.Padmaja Valli,T. Santhanam "Automatic Recommendation of Web Pages in Web Usage Mining" ISSN:0975-3397 Vol, 02,No. 09, IJCS E 2010