An Implementation of Web Recommendation System using Web Usage Mining Technique

Dimple Trivedi
Department of Computer Science and Engineering
M.Tech in computer Science
Medi-Caps University, Indore, Madhya Pradesh, India

Swati Tahiliani
Department of Computer Science and Engineering
Assistant Professor in Computer Science
Medi-caps University,Indore, Madhya Pradesh, India

ABSTRACT
Recommendation system is a classical approach of e-commerce application for suggesting the products to the users. But the use of recommendation system is performed in various other applications such as caching and pre-fetching, lead generation, CRM systems and others. In this presented work an efficient recommendation system is proposed for design and implementation. The proposed recommendation system inherits the properties of web usage mining and web content mining for recommending the most relevant user next web page for user. The web usage mining evaluates the user behavior and the content mining extract the user interest. In first step the web navigation history is analyzed for obtaining user web browsing behavior.

Therefore a list of web URLs are extracted from the web access log. In next the user current search requirement is considered and correlated with their past navigational pattern. In final step the user query semantics are measured and the rank based recommendation is produced. The implementation of the proposed technique is performed using JAVA technology. In addition of that for demonstrating the superiority a collaborative filter based recommendation system is compared with the proposed approach. According to the experimental evaluation the proposed technique is found efficient and accurate as compared to the classical recommendation model in terms of accuracy and resource usage.

Keywords
Recommendation System, Web Usages Mining, Web Content Mining, KNN, Correlation Coefficient, Data Mining

1. INTRODUCTION
E-commerce web-sites are providing new business portals and large amount of product information, so customers spend more and more time browsing the net in order to find the right information or product. One solution to overcome with this problem is to develop a personalized Recommendation system. This system retrieves the information desired by the customer and helps him in determining which product to buy. In recent years recommendation systems have changed the way of communication between both websites and users. Recommendation system sorts through massive amounts of data to identify interest of users and makes the information search easier. For that purpose many methods have been used [1] [2].

In order to adapt functionality to their individual users, systems need information about these users. The Social Web provides opportunities to gather user data from outside the system itself for recommender systems to be successful; they need to achieve a certain level of accuracy in their recommendations that is acceptable to the users. The continuous growth in the size and use of the World Wide Web imposes new methods of design and development of online information services. Most Web structures are large and complicated and users often miss the goal of their inquiry, or receive ambiguous results when they try to navigate through them. Therefore, the requirement for predicting user needs in order to improve the usability and user retention of a Web site can be addressed by personalizing it. The huge and ever increasing amount, complexity and heterogeneity of available digital information overwhelm the human processing capabilities in a wide array of information seeking and e-commerce tasks. To cope with information overload recommender systems have been introduced to filter those items –Web pages, images, videos, audio– that are of low relevance or utility for the user, and present only a small selection better suiting the user's tastes, interests, and priorities. Often these suggestions are presented while the user is browsing an information service, and without requiring her to launch explicit search queries, as is usually done in information retrieval systems [3] [4] [5].

2. BACKGROUND
The background of a study is an important part of our research work. It provides the context and purpose of the study. Hence there is need for background study that contributes to prepare proposed system.

2.1. What is Web Usage Mining?
The World Wide Web is growing continuously and huge amount of data is generated due to users' numerous interactions with web sites. Web Usage Mining is the application of data mining techniques to discover the useful and interesting patterns from web usage data. It supports to know frequently accessed pages, predict user navigation, improve web site structure etc. The web usage data consists of the data from web server logs, browser logs, proxy server logs and user profiles. In Web Usage Mining, data mining techniques are applied to pre-processed web log data in order to find interesting and useful patterns. Visitors' browsing behavior is recorded into web server log [6].Web usage mining is also called as web log mining which is used to analyses the behaviour of online users. It fed into two types of tracking; one is general access tracking and another one is customize usage tracking. The general access tracking is used to predict the customer behaviour on the web and it identifies the user while the user interacts with the web. It can store the data automatically when the web server log and application log [7].

2.2 What is Recommendation System?

Web recommendation systems are one kind of recommender systems, which can automatically recommend Web-pages that are most interesting to a particular user based on the user's current Web navigation behaviour.

Web Page recommendation has become increasingly popular, and is shown as links to related stories, related books, or most viewed pages at websites. Web Recommendation system is a specific type of information filtering system technique that attempts to predict the user next browsing activity then recommend to the user web pages items that are likely to be of interest to the user. A recommender system is a typical software solution used in e-commerce for personalized services. Based on the customer preferences, it helps to find the products they would like to purchase by providing recommendations and is particularly useful in ecommerce sites that offer millions of products for sale [8].

2.3 Categories of Web Page Recommendation System

The domain of Recommender System consists of a number of techniques, each with their own set of advantages and disadvantages, which make them suitable for applications in different domains. The image below gives an overview of the relation between different techniques [9] [10].
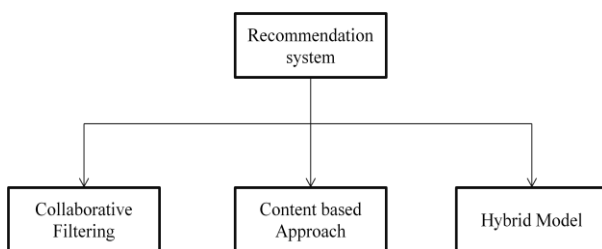

Figure : 1 Categories of Recommendation System

A recommendation system can be developed in a number technique. According to behavior of recommendation system can be classify into following categories.

✓ Content based Recommendation: Similar items to the ones the user preferred in the past are generated as a recommendation

✓ Collaborative Recommendation: Items preferred by the people who have the similar taste to the user are generated as a recommendation.

✓ Hybrid Approach: The above recommendation methods are combined in this approach

3. PROPOSED WORK

This section provides the understanding about the proposed concept of web recommendation system design. Therefore, this section includes description of the methodology and the algorithm steps.

3.1 System Overview

Web mining is a sub-domain of data mining. In this domain the traditional data mining algorithms are applied on web data for recovering the data patterns. The web mining can be categorized in three major classes namely web usages mining, web content mining and web structure mining. The web usage mining is used to analyze the web usage patterns of a web application similarly the web content mining is used to analyze the web pages contents. Finally the likes and the web page organization structures are studied under web structure mining. The proposed work is an application of web usages mining and web content mining for designing an efficient recommendation system.

The proposed recommendation system is an extension of traditional web based recommendation system where the URLs are predicted using the collaborative filtering basis. In other to extend the traditional web URL prediction for improving accuracy the web URLs and their content are also included which speed up the recommendation of URL. In addition of the proposed system incorporate the web user behavioral pattern analysis with the similar behavior analysis which helps to locate the need of the required web object. In order to accomplish this task the web assess log file is used with the KNN (k-nearest neighbor) search algorithm and the correlation coefficient is used for establishing relationship among the web URL and web contents. Both the factors are combined to produce the most suitable web page prediction according to the user past web navigation history. This section provides the overview of the proposed model for design and implementation. In next section the proposed model is explained in detail for demonstrating the functional aspect of the proposed system..

3.2 Methodology

The proposed user next web page prediction system is demonstrated in figure 2. This diagram contents the required components and the intermediate processes involved in this architecture the component of the system are explained in detail as:
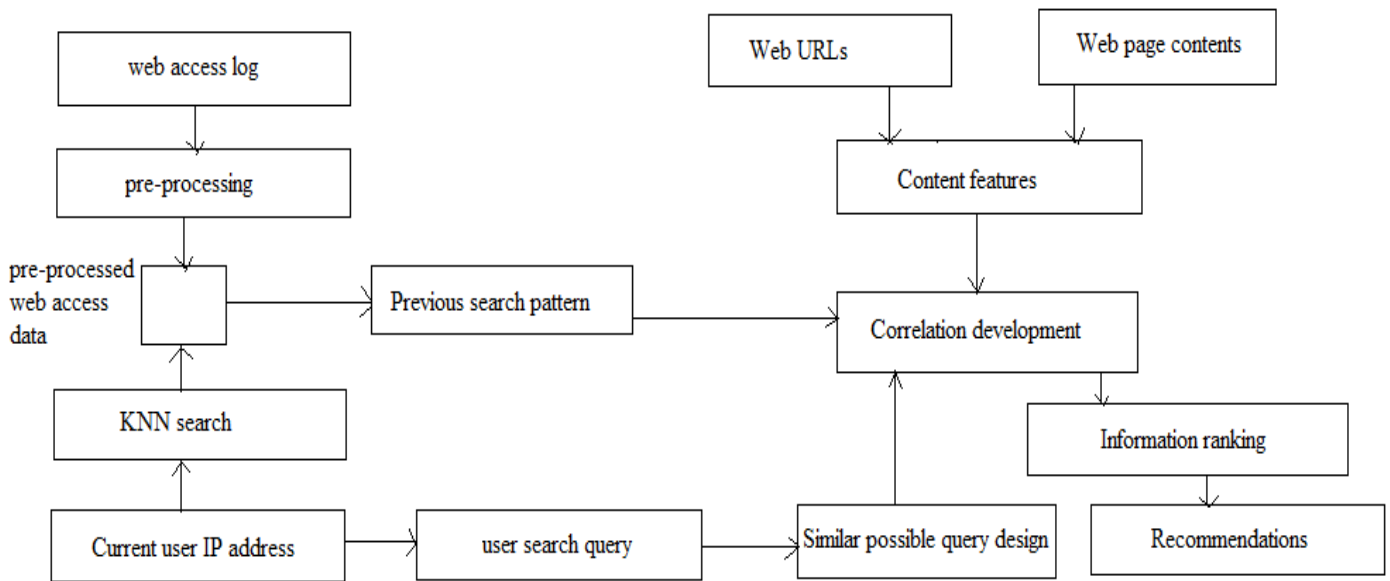
Figure : 2 Proposed System Architecture

Web access log: The web servers are maintaining a log file which contains the information about all the navigation performed by the end user through their own web browsers. This log file contains various information such as request time stamp, user IP address, request URL, protocol and others. This web access log file is initial input for the proposed system for analyzing the user past web browsing behavior.

Pre-processing: Not all the web access log file data is essential for utilizing in the system. Therefore the web access log file is processed for finding the essential information. In this phase of data preprocessing only required content from web access log file is selected and remaining attributes from the web access log is removed.

Pre-processed web access data: the remaining content of web access log after pre-processing of data is preserved in a temporary database. That selected information is used in further processing of user next web page recommendation.

Current user IP address: this user IP address is the current user's IP address for which the recommendation is required to perform.

KNN search: the KNN [6] (k-nearest neighbor) algorithm is a search algorithm which accepts the user query as input and the database for finding the most relevant record from data base according to the query. Here the KNN search is performed on the web access based preprocessed data for finding the user IP address. The search results extract the all the data instance that belongs to the particular user.

Previous search pattern: the list of user data which is belongs to the end user IP address is denoted here as previous user search pattern. Using this list of data two things is recovered:

Web URLs: The list of web URLs, which are available in user past access search patterns.

Web page contents: second component is the web page contents which are available in URLs. Using this web page content further process is taken place.

Content features: for evaluation of the web page contents the term frequency of the web page content keywords are extracted. For finding the web page content term frequency the following formula is used:

$$term\ frequency = \frac{word\ count}{total\ number\ of\ words}$$

On the other hand the occurrence of URLs associated with the contents is also computed.

User search query: that is the user's search query which is searching for currently on the web application.

Similar possible query design: system accepts the end user's query and generates the similar query from the initial query by replacing the keywords by the similar semantic keywords.

Correlation coefficient development: in this phase the web URL occurrence and web page content are used for computing the correlation coefficient. In addition of that this phase also accept the user's possible query keywords. The system matches the user query keywords from the correlation coefficient data for finding the URLs which contains the relevant keywords.

Information ranking: in this phase the URLs and the correlation coefficient values for sorting or indexing of the URLs according to relevancy of contents.

Recommendation: the selected most higher ranked URLs are predicted as recommendation for the particular user.

3.3 Proposed Algorithm
The above given methodology is concluded in this section in which the algorithms are described for summarizing the proposed work.

Table 1 Proposed Algorithm

Input: web access log W, user current IP address IP, user search keywords K
Output: recommended URL

Process:
1. $R = readAccessLog(W)$
2. $P = preProcessData(R)$
3. $[URL_n] = KNN.search(IP, P)$
4. $for(i = 1; i \leq n; i++)$
   a. $C = extractContent(URL_i)$
5. $end\ for$
6. $CC = ComputeCorrelation(C, URL_n)$
7. $K_m = genrateQuery(K)$
8. $for(j = 1; j \leq m; j++)$
   a. $URL = search(K_j, CC, URL_n)$
9. $end\ for$
10. Return URL

4. RESULT ANALYSIS

The given section includes the performance analysis of the implemented algorithms for the web recommendation systems. Therefore the performance of algorithms are evaluated using key parameters.

4.1 Memory Consumption

In order to execute the processes the data is placed in the main memory for computation. This required space is termed as the memory usages or space complexity of the algorithms. In java technology for computing the processes consumed memory or utilized main memory the following formula can be used:

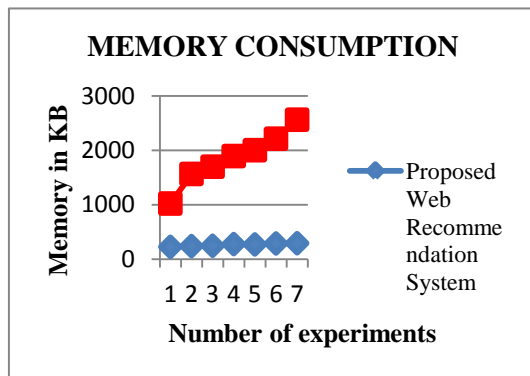Memory Consumption = Total Memory − Free Memory

Figure: 3 Memory Consumption

The comparative memory consumption of the proposed and traditional web based recommendation approach is described using figure 3 and table 2. In order to depict the memory consumption of the proposed approach the blue line is used and the red line is used for demonstrating the performance of traditional approach. The X axis of the diagram shows the number of experiments performed with dataset and the Y axis shows the obtained memory consumed of the system. According to the performance the proposed approach consumes less memory as compared to the traditional technique, thus proposed model is superior to the traditional approach of web based recommendation system.

Table 2 Tabular Form of Memory Consumption

No. of Experiments	Proposed Web Recommendation System	Traditional Web Recommendation System
1	225	1021
2	239	1564
3	245	1698
4	276	1893
5	269	2001
6	287	2212
7	291	2561

4.2 Time Consumption

In computational domain every process of the system need to some execution time to generate the outcomes. This time requirement of process or the algorithm is termed as the time consumption for the system. The total time required for completing the task can be computed using the following formula:
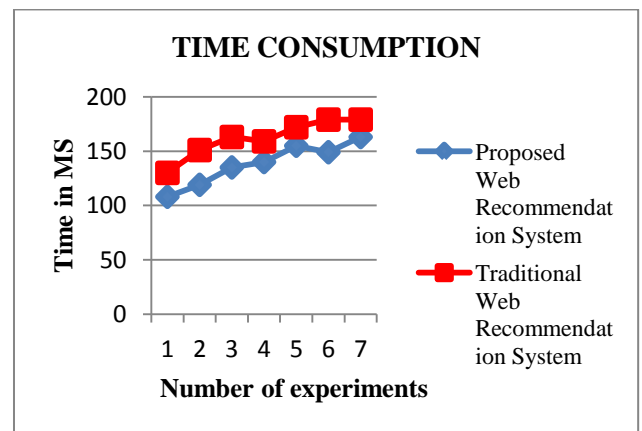
$$\Delta t = T_{end} - T_{start}$$

Figure: 4 Time Consumption

The time consumption of both the algorithms namely traditional and proposed is given using figure 4 and table 3. The measured time of execution is demonstrated in terms of MS (milliseconds). The red line of the diagram shows the performance of Traditional system and proposed techniques performance is visualized using blue line. For preparing the comparative graphs the X axis contains the number of experiments performed and the Y axis shows the time consumption. According to the experimental results the proposed technique improves the performance in terms of time requirements. Thus proposed model is time preserving also. Hence, our web based recommendation system is recommending to e-commerce based web application for end user.

algorithm a fixed amount random URLs are extracted from database and it recommending web URLs.

Table 3 Tabular Values of Time Consumption

No. of Experiments	Proposed Web Recommendation System	Traditional Web Recommendation System
1	108	130
2	119	151
3	135	163
4	140	159
5	155	172
6	149	179
7	163	179

Table 4 Tabular Values of Accuracy

No. of Experiments	Proposed Web Recommendation System	Traditional Web Recommendation System
1	93.21	81.25
2	94.35	80.55
3	94.88	82.54
4	93.01	83.14
5	95.45	82.74
6	95.88	83.65
7	96.41	85.2

4.3 Accuracy

The accuracy of the proposed algorithm is representing in term of accurate extracted patterns. Therefore that can also be defines as the amount of correctly recognized patterns among the total samples produces to test. That can also be evaluated using the following formula:

$$Accuracy = \frac{Total\ Correctly\ Identified\ Patterns}{Total\ Input\ Samples} X\ 100$$

4.4 Error Rate

The error rate provides the information how far the accurate prediction is, thus that can be calculated using the incorrectly identified instances during classification. The error in percentage can be evaluated using the below given formula.

$$Error\ Rate = 100 - Accuracy$$

Or

$$Error\ Rate = \frac{Total\ Incorrectly\ Classifid\ Instances}{Total\ Samples\ to\ Classify} X100$$
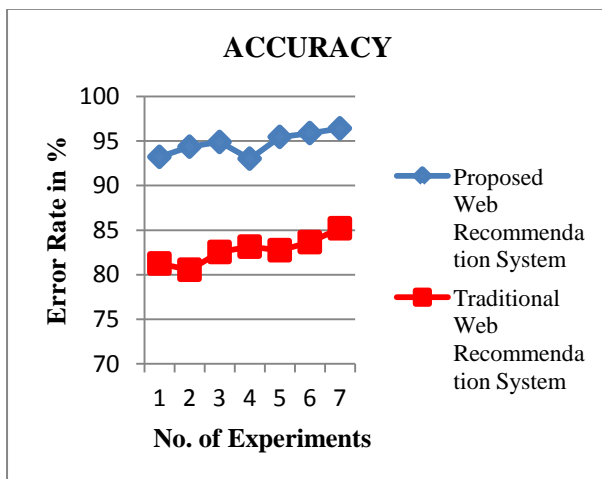

Figure 5 Accuracy

The figure 5 and table 4 contains the evaluated performance of the system in terms of algorithm accuracy of recommendation. The computed accuracy is given here in terms of percentage values. In this diagram the amount of accurate web recommending is given using Y axis in percentage and the X axis shows the different run of experiments. According to the comparative results the performance of the proposed algorithms remains much consistent and increasing as compared to the traditional recommendation system. In order to evaluate the accuracy of
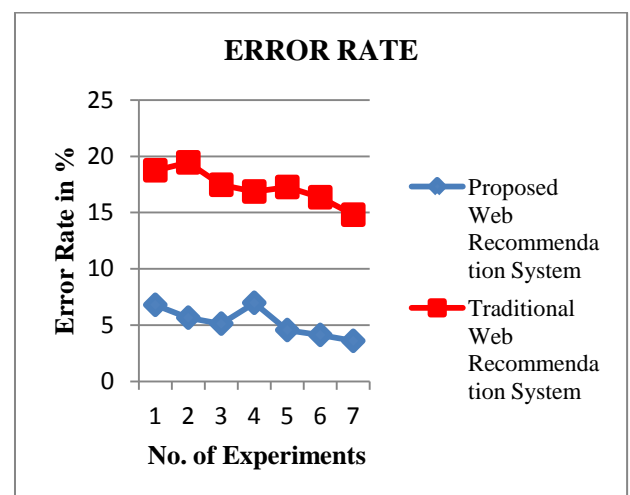

Figure 6 Error Rate

The given figure 6 and table 5 shows the error rate of the proposed web recommendation system, in this given diagram number of experiments are given in X axis and in Y axis the obtained error rate is provided. According to the obtained results the amount of error rate is varied as the number of experiments is increased.

Table 5 Tabular Values of Error Rate

No. of Experiments	Proposed Web Recommendation System	Traditional Web Recommendation System
1	6.79	18.75
2	5.65	19.45
3	5.12	17.46
4	6.97	16.86
5	4.55	17.26
6	4.12	16.35
7	3.59	14.8

5. CONCLUSION AND FUTURE WORK

This section provides the conclusion of the proposed work performed for designing an efficient recommendation system. In addition of that the future extension of the proposed work is also included in this section.

5.1 Conclusion

The data mining and their relevant domains are now in these days applied in various applications for making decisions, recommendations and analyzing reviews. In this presented work an application of web data mining is presented for recommending the user next web page. Basically the web recommendation systems are frequently used in e-commerce application. The recommendation system is basically a web algorithm that understands the pattern of data access from web for a specific user and according to the user data browsing behavior the recommendations are performed. In this context the data mining algorithms are utilized to analyze the web access data and generation of recommendation seeds.

In this proposed recommendation system first the user web history is analyzed first for extracting the user visited web pages. That list of web pages are termed here as user browsing behavior of user. In further the current search and user requirements are considered for finding the similar semantic web contents from the available web data. On the basis of the web content and navigated URLs the correlation coefficient is computed that help to sort the data and URL corresponding the user search requirements. The user query is searched on the available web records and web URLs for finding most relevant URL and data to be recommended. In order to perform this task the KNN (k-nearest neighbor) algorithm and correlation coefficient is used where the KNN is used to finding the data similar to user query and for finding the relevant content and to rank the content and URL the correlation coefficient is used.

The implementation of the proposed recommendation system is performed using JAVA technology. Additionally for obtaining the performance the results analysis is performed on different performance factors. The computed performance is summarized in table 6.

Table 6 Performance Summary

Performance Parameters	Proposed Web Recommendation	Traditional Web Recommendation
Memory Consumption	261.71 KB	1850 KB
Time Consumption	138.42 MS	161.85 MS
Accuracy	94.74 %	82.72 &
Error Rate	5.25 %	17.27 %

The performance of the proposed recommendation system is computed and found optimal for different experimental scenarios. Therefore the proposed recommendation system is acceptable for different application needs.

5.2 Future Work

The main aim of the proposed recommendation system design is accomplished successfully. In near future the following possible extensions are possible for the proposed work.

✓ The proposed work is not considering the cold start problem of recommendation system design in near future for cold start problem the work is extended.
✓ The proposed work only considers the limited factors of web usage and web content mining in near future the web application factors are also included for improving the recommendations.

6. REFERENCES

[1] Paritosh Nagarnaik and A. Thomas, "Survey on recommendation system methods", 2nd International Conference on Electronics and Communication Systems (ICECS), 18 June 2015, Coimbatore, India

[2] Megha Sen and Seema Udgirikar, "Survey Paper on Web Recommendation System", International Journal of Science and Research (IJSR), Volume 4 Issue 11, November 2015

[3] Toshi Sharma and Ritu Tondon, "User Behavior Analysis Based On Predictive Recommendation System for E-Learning Portal", International Journal of Core Engineering & Management (IJCEM) Volume 2, Issue 4, July 2015.

[4] Tuzhilin, Alexander, and Gediminas Adomavicius, "Integrating user behavior and collaborative methods in recommender systems", CHI'99 Workshop, Interacting with Recommender Systems. 1999.

[5] Amarjeet Bathija, Jitesh Budhwani and Surabhi Kanth, "Cross Category Product Recommendation System Using Clickstream Mining", International Journal of

Advanced Research in Computer Science and Software Engineering, Volume 6, Issue 1, January 2016

[6] Srivastava, Jaideep, Robert Cooley, Mukund Deshpande, and Pang-Ning Tan. "Web usage mining: Discovery and applications of usage patterns from web data." ACM SIGKDD Explorations Newsletter 1, no. 2 (2000), pp. 12-23.

[7] P. Nithya, and P. Sumathi, "A Survey on Web Usage Mining: Theory and Applications", International Journal of Computer Technology & Applications, Volume 3, Issue pp. 1625-1629.

[8] AlMurtadha, Yahya, et al. "Ipact: Improved web page recommendation system using profile aggregation based

on clustering of transactions." American Journal of Applied Sciences 8.3 (2011): 277.

[9] Chapter 9- Recommendation Systems, available online at: http://i.stanford.edu/~ullman/mmds/ch9.pdf

[10] Barragáns-Martínez and J. Burguillo, "A hybrid content-based and item-based collaborative filtering approach to recommend TV programs enhanced with singular value decomposition", Inform. Sci., 180: 4290-4311, 2010.

[11] Wu, X., Kumar, V., Quinlan, J.R., Ghosh, J., Yang, Q., Motoda, H., McLachlan, G.J., Ng, A., Liu, B., Philip, S.Y. and Zhou, Z.H., 2008. Top 10 algorithms in data mining. Knowledge and information systems, 14(1), pp.1-37