

# INCREASE THE PERFORMANCE OF ONLINE WEB USAGE RECOMMENDATION SYSTEM DATA BY USING IMPROVED HYBRID CLUSTERING TECHNIQUE

Hiral M Shastri<sup>1</sup>, Indr Jeet Rajput<sup>2</sup>

Department of Computer Engineering, Hasmukh Goswami College of Engineering, Vahelal Gujarat Technological University

**Abstract:** *The amount of information over internet has been growing last few years. And it has caused risk of information problem of accessing related data to the users. The information demand of the online users can be figured out by evaluating user's web navigation behavior. Web Usage Mining (WUM) is used to extract knowledge from Web user's access logs by using Data Mining Techniques. One of the applications of WUM is Web Sites Recommendation system which is personalized information filtering technique used to either determine whether a certain user will approve a given item or to identify a list of items which can be of significant importance to the user. In this paper the modified architecture that integrates item information with user's access log data and then find pattern and make pattern clustering. There after generates a set of recommendations for the user. So execution time and fetching time is reduced.*

**Keywords:-** *Web Usage Mining, Online Web Recommendation System, Pattern Matching, K-Means.*

## I. INTRODUCTION

Web mining is the utilization of information mining systems to concentrate learning from web information, including web archives, hyperlinks between reports, use logs of sites, and so forth. Web mining can be comprehensively separated into three unmistakable classifications, as indicated by the sorts of information to be mined. Web Mining has three types that we mention below.

### *Web Content Mining*

Web content mining is the way toward separating valuable data from the substance of web archives. Content information is the accumulation of certainties a website page is intended to contain. It might comprise of content, pictures, sound, video, or organized records, for example, records and tables. Use of content mining to web content has been the most generally examined. Issues tended to in content mining incorporate point revelation and following, removing affiliation designs, grouping of web archives and arrangement of website pages.

### *Web Structure Mining*

The structure of a run of the mill Web chart comprises of Web pages as hubs, and hyperlinks as edges interfacing related pages. Web Structure Mining is the way toward finding structure data from the Web. This can be further isolated into two sorts in light of the sort of structure data utilized.

### *Web Usage Mining*

Web Usage Mining is the use of information mining systems to find intriguing utilization designs from Web information, keeping in mind the end goal to comprehend and better serve the necessities of Web-based applications [4]. Utilization information catches the personality or beginning of Web clients alongside their perusing conduct at a Web website.

## II. BACKGROUND AND RELATED WORKS

In first paper they used two tier architecture for integrating semantic information with web usage mining. The architecture for recommending the list of products to the users can be partitioned into two main phases; offline phase and online phase [14]. These two phases work together strongly. In the offline phase there are two main modules, data pre-processing and semantic knowledge base. The main module of the online phase is Recommendation engine. They have used LCS algorithm to generate a list of recommended products to the user. In second paper author proposed a structural engineering which incorporates semantic data with web utilization mining. Longest Common Subsequence is utilized to create a rundown of suggested items. This system improved the performance of the existing Recommender systems by overcoming the new item problem. The system comprised of online and offline phase. RDF model is used for incorporating Semantic information. In third paper author proposed an Online recommendation System utilizing LCS Algorithm. The system involved two phases that work in conjunction with each other i.e. the online and offline phase. Data pretreatment and navigation pattern mining is carried out in offline phase while predictions are generated in the online phase. The accuracy of expectation is 73%. In fourth paper author have utilized an architecture which uses semantic data in pattern generation phase for Web Usage Mining. Spade, is utilized to create regular access successions. Here sequential association rule mining is used. Work with classical Web usage mining with ontological terms and concepts. In fifth paper they have researched and proposed a model which integrates Semantic Web and Agent Personalized Recommendation System (SWARPS) in E-Commerce. It involves implementation of AI-Multi Agent technique. The architecture is divided into various agents (monitor agent, web analysis agent, semantic conversion agent, data mining agent, preference analysis agent, semantic recommendation agent and recommendation engine agent) that carry out specific functionalities. In Sixth paper they have put forth a customer end based system where at

customer end web action of the customer framework is assessed. This assessment helps in developing the transactional patterns that helps in personalization and behavioral analysis. A frequent access pattern algorithm is used to divide the data into time slices. This discovering the pattern in data which aids data is formulated using Hidden Markov Model (HMM) for in generating much accurate and efficient recommendations. In seventh paper system implemented the significant elements of web recommendation and personalization like Modelling of web pages or web products and customers, Mapping between appropriate customers and products and determining the set of recommendations. This system made use of Ontology and OWL (Web Ontology Language) for web product classification. This system avoided false positive recommendations i.e. the product will be recommended even though it would not be relevant to the customer.

### III. PROPOSED SYSTEM

Current recommendation systems exhibits certain limitations such as intelligence limited accuracy. These disadvantages can be overcome by implementing a new architecture. This system handles most of the drawbacks and gives more efficient and more accurate result than previous systems.

#### A. Architecture Overview

The architecture of an online web recommendation system based on web usage mining basically consists of three phases : Data Preprocessing, Pattern detection and generating recommendations. Data preprocessing and Pattern detection phases are performed offline and the recommendations are generated online. Data preprocessing involves transforming the web access logs and user profiles into format appropriate for the system. Pattern detection involves using data mining techniques like clustering, sequential pattern mining or association rule mining. Lastly the detected patterns are used to generate recommendations which provide customized links or data to the user.

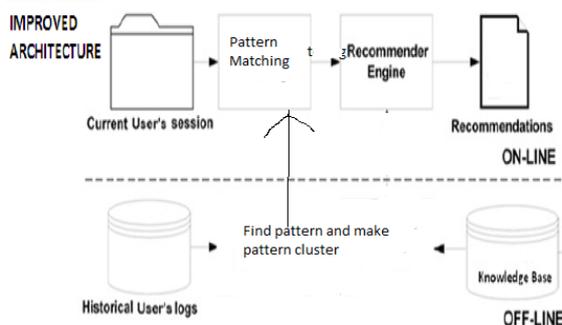


Figure 1.Improved Architecture for Online web-based recommendation system

The abovementioned improved architecture of recommender system involves integrating additional information about the users (such as user profiles). This system uses k-means clustering algorithms like clustering and pattern matching algorithms. Thus the users exhibiting common behavior are first clustered and then patterns for each cluster are discovered. This type of recommendation systems generates personalized recommendations. The new users are first classified in one of the clusters and then, the patterns of the

corresponding clusters are used to customize the recommendations based on user's current navigation and other analogous users in the cluster. The above architecture is divided in two main phases; offline phase and online phase. These two stages cooperate unequivocally in conjunction to one another.

#### B. Offline Phase of the Architecture

This stage comprises of two significant modules: Data preprocessing and Knowledge Base of the Products. In offline stage we begin with the essential Web-Access- Log Preprocessing which includes extraction of client session and entering the important data into the information base.

1) Data Preprocessing: In this phase the original web logs are reformatted to discover the web access sessions. The web server logs that are generated by web server stores all the user's web access activities. There are different types of web logs depending on different server settings. These logs consist of information like client IP address, URL, etc. Different preprocessing functionalities such as data cleaning, session identification is carried out prior to employing web mining algorithms on the Web server logs.

Clustering is also carried out in preprocessing stage. Here k means clustering algorithm is used. In a recommendation system, k means can be used in preprocessing stage for identifying groups of users who appear to have similar preferences. It is used for aggregating the user profiles.

2) Knowledge Base: Following the data preprocessing step, various features of the products are combined with the extracted user session data from the logs. These features include price, brand, etc. The user and transaction details are recorded in form of tables in the database.

#### Online Phase of the Architecture

During this phase when the user logs in at the server, the recommendation engine checks with the knowledge base for the user's earlier transactions. The list of recommended products will be generated based on the user's previous history and the pattern in the cluster to which the user belongs.

1) Generating Recommendation: The essential usefulness of the recommendation system is generating recommendations by utilizing some refining parameters like value, brand, rating and so forth. The refining parameters support to produce exact set of prescribed items from the information base. To produce rundown of proposal we utilize Boyer-Moore Pattern Matching algorithm.

Pattern search algorithm could be used to discover client interest items focused around the current client exercises to anticipate and suggest client's future appeal. Pattern search algorithm, Boyer-Moore is utilized within the recommendation piece of the architecture.

### IV. ALGORITHMS USED

Clustering is an unsupervised classification or dividing of patterns into groups or subsets (i.e. clusters). Here the objects

are grouped into classes of similar objects based on their location and connectivity within an n dimensional space. Mainly the principle of clustering is to maximize the similarity inside a cluster and to minimize the similarity between the clusters. Although there are many clustering algorithms available, one of the most widely used is the k means algorithm. It aims at minimizing the distance of the objects with respect to the centroid of each cluster.

**A. CLUSTER Algorithm for Web Recommended System**

- 1: In data source DS read N no of items
- For i= 1 to i<=N
- Next
- 2: In DS for each item I find frequent access item F
- 3: Read frequency user access items F.
- 4: If I=F frequent items then count C
- 5: Save for clustering frequent user access records in frequency access data source CDS
- 6: Make cluster from frequency user access records
- 7: Else not select records
- End if
- Next record

**B. Boyer Moore Pattern Matching Algorithm**

- 1: Inspect the window from right to left.
- 2: Recognize the possibility of large shifts in the window without missing a match.
- 3: “Bad Character shift rule”: It avoids repeating unsuccessful comparisons against a target character.
- 4: “Good suffix shift rule”: It aligns only matching pattern characters against target characters already successfully matched. [15]

**V. EXPERIMENT**

**Implementation**

The Figures and shows the products available and the recommended products based on user’s interest respectively. Asp.Netis used for developing the front end and MySQL database for the backend.

ID	Products	Price	Product_Request
1	COOLPAD MEGA	6999	www.amazon.in/Coolpad-Mega-2-5D-Royal Gold
2	MOTO G	12000	www.flipkart.com/moto-g
4	LENOVO VIBE NOTE 5	20000	www.amazon.in/lenovo-vibe-note-5
7	MOTO G ADVANCE	15000	www.amazon.in/moto-g-advance
8	MICROSOFT LUMIA 640 XL	48000	www.amazon.in/microsoft-lumia-640-xl
11	XIAOMI REDMI NOTE 3	18000	www.amazon.in/xiaomi-redmi-note-3

Figure 2. List of Available Products



Figure 3. Recommended Products

**C. Results**

The system is tested for accuracy and performance. The new System increase the accuracy and performance up to 36 % where old system provides only 30 %.New system also

decrease the complexity upto 5%. The testing is performed for 100 products in the system. The span of user transaction history is set as 30daysThe following graphs illustrates the results

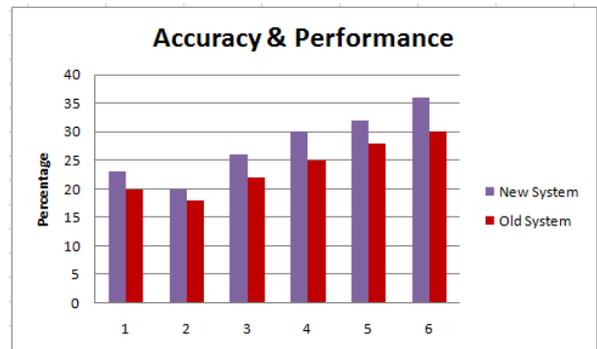


Figure 4. Accuracy Chart

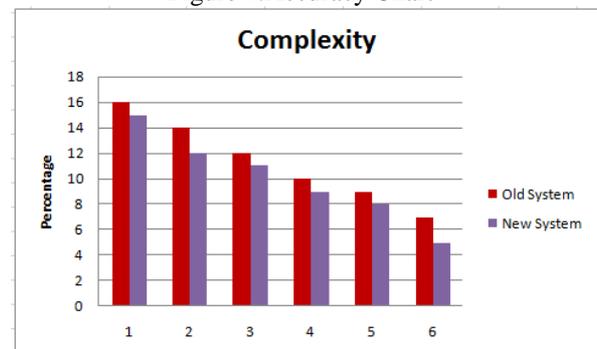


Figure 5. Complexity Chart

**VI. CONCLUSION**

This Advance System displays a list of recommended items. It involves clustering of pattern items which leads to searching of patterns in clusters rather than searching whole items from databases. It reduces execution time and thus increasing performance of the overall system. The system doesn’t suffer from searching items in whole database .it just find from pattern cluster which is to be recommended. Future scope includes using K-Means clustering the performance values obtained as a new parameter which is going to improving the performance of this advance recommendation system.

**REFERENCES**

- [1] Sneha Y.S, G. Mahadevan, Madhura, ”An Online Recommendation System based on web usage mining and semantic web using LCS Algorithm”, IEEE 2011
- [2] Sneha Y.S, G. Mahadevan, Madhura, ”A Personalized Product Based Recommendation System Using Web Usage Mining and Semantic Web”, International Journal of Computer Theory and Engineering (IJCTE) Vol. 4, No. 2, April 2012
- [3] Mehrdad Jalali1, Norwati Mustapha, Md. Nasir B Sulaiman, Ali Mamat, “A Web Usage Mining Approach Based on LCS Algorithm in Online Predicting Recommendation Systems” 12th International Conference Information Visualisation IEEE 2008.

- [4] SuleymanSalim et al “Using Semantic Information for web usage mining based recommendation” International Conference IEEE 2009.
- [5] Xin Sui, Suozhu Wang, Zhaowei Li “Research on the Model of Integration with Semantic Web and Agent Personalized Recommendation System “Proceedings of the 2009 13<sup>th</sup> International Conference on Computer Supported Cooperative Work in Design.
- [6] HimangniRathore, HemantVerma, “Analysis on Recommended System for Web Information Retrieval Using HMM”, International Journal of Engineering Research and Applications ISSN: 2248-9622, Vol. 4, November 2014.
- [7] Hadi Khosravi Farsani, and Mohammadali Nematbakhsh “A Semantic Recommendation Procedure for Electronic Product Catalog”, World Academy of Science, Engineering and Technology 22 2006.
- [8] Hiral Y Modi, MeeraNarvekar .”Enhancement of Online web Recommendation System Using A Hybrid Clustering And Pattern Matching Approach”: 2015 International Conference on Nascent Technologies in the Engineering Field (ICNTE-2015)
- [9] RVVSV Prasad, V ValliKumari “A Categorical Review of Recommender Systems”, International Journal of Distributed and Parallel Systems (IJDPS) Vol.3, No.5, September 2012
- [10] Jiawei Han and MichelineKamber, “Data Mining: Concepts and Techniques”, Morgan kaufmann, 2nd Edition.
- [11] Jaideep Srivastava, PrasannaDesikan, VipinKumar ”Web Mining - Concepts, Applications & Research Directions”
- [12] J. Srivastava, R. Cooley, M. Deshpande and P.-N. Tan “Web usage mining: discovery and applications of usage patterns from web data” ACM SIGKDD Explorations, vol. 1, no. 2, pp. 12-23, 2000.
- [13] [www.w3c.org](http://www.w3c.org)
- [14] BamshadMobasher, Robert Cooley, Jaideep Srivastava “Automatic Personalization Based on Web Usage Mining”, Communications of the ACM Volume 43 Issue 8, Aug. 2000.
- [15] [15] T. H. Cormen, C. E. Leiserson, and R. L. Rivest, “Introduction to Algorithms”, MIT Press, 1990.
- [16] Mr. DushyantsinhRathod, Dr. Samrat Khanna ,Mr. Vijay Gadhavi “Improve quality in user pattern mining approach using two level k-means clustering methodology”, IJSRD - International Journal for Scientific Research & Development| Vol. 4, Issue 01, 2016.
- [17] Mr. Dushyantsinh Rathod, Dr. Samrat Khanna, Mr. Mayank Parekh “Implemented Two Level k-means clustering Algorithm to Improve Quality in User Pattern Mining”, IJSART - Volume 2 Issue 4 – APRIL 2016.