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

Hiral M Shastri[1], Indra Jeet Rajput[2]
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 theonline 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 application of data mining techniques to extract knowledge from web data, including web documents, hyperlinks between documents, usage logs of web sites, etc. Web mining can be broadly divided into three distinct categories, according to the kinds of data to be mined.

Web Content Mining
Web content mining is the process of extracting useful information from the contents of web documents. Content data is the collection of facts a web page is designed to contain. It may consist of text, images, audio, video, or structured records such as lists and tables. Application of text mining to web content has been the most widely researched. Issues addressed in text mining include topic discovery and tracking, extracting association patterns, clustering of web documents and classification of web pages. Research activities on this topic have drawn heavily on techniques developed in other disciplines such as Information Retrieval (IR) and Natural Language Processing (NLP). While there exists a significant body of work in extracting knowledge from images in the fields of image processing and computer vision, the application of these techniques to Web content mining has been limited.

Web Structure Mining
The structure of a typical Web graph consists of Web pages as nodes, and hyperlinks as edges connecting related pages. Web Structure Mining is the process of discovering structure information from the Web. This can be further divided into two kinds based on the kind of structure information used.
Hyperlinks
Document Structure
Web Usage Mining
Web Usage Mining is the application of data mining techniques to discover interesting usage patterns from Web data, in order to understand and better serve the needs of Web-based applications [4]. Usage data captures the identity or origin of Web users along with their browsing behavior at a Web site. Web usage mining itself can be classified further depending on the kind of usage data considered:
Web Server Data
Application Server Data
Application Level Data

II. THEORETICAL REVIEW

Various recommendation systems use different approaches based on the sources of information they utilize. The accessible sources are user information (demographics), the product information (keywords, genres) and the user-item ratings [8].Recommendation system exhibits certain limitations such as intelligence, adaptability, flexibility, Limited accuracy. This paper aims to provide an architecture that integrates product information with users access log data and then generates a set of recommendations for that particular user is presented [8]. These disadvantages can be overcome by implementing a hybrid architecture that integrates product information with user's access log data and then generates a set of recommendations for that particular user using Boyer- Moore Pattern Matching Algorithm and K-Means Clustering Algorithm.

III. ONLINE RECOMMENDATION RELATED ALGORITHMS

3.1 An Online Recommendation System based on web usage mining and semantic web using LCS Algorithm [1]
In this 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 and resolved it using Huffman code and

Wet Paper Codes (WPC). The results show good performance in terms of Precision, Recall and F1 metrics as compared to the existing recommender system. This system doesn't suffer from new item problem. If the user is visiting the server for the first time there will be no list of recommended products available until the user purchases any product.

3.2 A Personalized Product Based Recommendation System Using Web Usage Mining and Semantic Web[2]
In this 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.
The System does not involve clustering of user profiles which leads to searching of patterns by traversing all usage logs thus leading to more time consumption and thus reducing performance of the overallsystem.

3.3 A Web Usage Mining Approach Based on LCS Algorithm in Online Predicting Recommendation Systems [3]
In this 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%. It use a fixed-size sliding window over current active session to capture the current user's activities.

3.4 Using Semantic Information for web usage mining based recommendation [4]
In this 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.Time consuming process.

3.5 Research on the Model of Integration with Semantic Web and Agent Personalized Recommendation System [5].
Author have researched and proposed a model which integrates Semantic Web andAgent 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. The agents work jointly with each other to provide recommendations. The system exhibits limited degree of intelligence, autonomy and flexibility.
3.6 Analysis on Recommended System for Web Information Retrieval Using HMM [6]

This paper 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.

3.7 A Semantic Recommendation Procedure for Electronic Product Catalog [7]
In this 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. The system does not recommend to a customer if the number of earlier recommendations to his/her has not exceeded a dedicated amount of threshold.

3.8Enhancement of Online web Recommendation System Using a Hybrid Clustering and Pattern Matching Approach [8]
In this paper author used K-Means clustering algorithm and boyar more pattern matching algorithm. This Online Web Recommendation System displays a list of recommended products based on the user's recent history. It involves clustering of user profiles which leads to searching of patterns in clusters rather than searching whole user logs, thereby reducing time consumption and thus increasing performance of the overall system [8]. It Provides Recommendation based on previous history. The recommendation will be generated based on feature.

IV. 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.

Paper	Merits	Demerits

2.1	**An Online Recommend ation System based on web usage mining and semantic web using LCS Algorithm.**	The system comprised of online and offline phase. RDF model is used for Incorporating Semantic information. System doesn't suffer from new item problem.	If the user is visiting the server for the first time there will be no list of recommended products available until the user purchases any product. Does not involve clustering of user profiles.	2.6	**Analysis on Recommend ed System for Web Information Retrieval Using HMM**	A frequent access pattern algorithm is used to divide the data into time slices. Data is formulated using Hidden Markov Model (HMM) .Generating much accurate and efficient recommendation s.	Only work with client end system. Use rating to generate the recommendation
2.2	**A Personalized Product Based Recommend ation System Using Web Usage Mining and Semantic Web.**	It solve the new item problem.	System does not involve clustering of user profiles Reduced the performance of system.	2.7	**A Semantic Recommend ation Procedure for Electronic Product Catalogue.**	This system made use of Ontology and OWL (Web Ontology Language) for web product classification. This system avoided *false positive* recommendation s	The system does not recommend to a customer if the number of earlier recommendation s to his/her has not exceeded a dedicated amount of threshold.
2.3	**A Web Usage Mining Approach Based on LCS Algorithm in Online Predicting Recommend ation Systems**	The system involved two phases. Data pre-treatment and navigation pattern mining is carried out in offline. Increase the accuracy up to 73%.	It use a fixed-size sliding window over current active session to capture the current user's activities.	2.8	**Enhancemen t of Online web Recommend ation System Using A Hybrid Clustering And Pattern Matching approach**	It involves clustering of user profiles. Searching of patterns is done in clusters rather than searching whole user logs. It reduce the time. Increase the performance.	Current recommendation system exhibits certain limitations such as intelligence, adaptability, flexibility, limited accuracy
2.4	**Using Semantic Information for web usage mining based recommenda tion**	Sequential association rule mining is used.	Time consuming process.				
2.5	**Research on the Model of Integration with Semantic Web and Agent Personalized Recommend ation System**	It use AI-Multi Agent technique.	The system exhibits limited degree of intelligence, autonomy and flexibility.				

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 13th International Conference on Computer Supported Cooperative Work in Design.

[6] Himangni Rathore, Hemant Verma, "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, Meera Narvekar ."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 Valli Kumari "A Categorical Review of Recommender Systems", International Journal of Distributed and Parallel Systems (IJDPS) Vol.3, No.5, September 2012

[10] Jiawei Han and Micheline Kamber, "Data Mining: Concepts and Techniques", Morgan kaufmann, 2nd Edition.

[11] Jaideep Srivastava, Prasanna Desikan, Vipin Kumar "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

[14] Bamshad Mobasher, Robert Cooley, Jaideep Srivastava "Automatic Personalization Based on Web Usage Mining", Communications of the ACM Volume 43 Issue 8, Aug. 2000.

[15] T. H. Cormen, C. E. Leiserson, and R. L. Rivest, "Introduction to Algorithms", MIT Press, 1990.