International Journal of Advance Research in Science and Engineering
Vol. No.10, Issue No. 04, April 2021
www.ijarse.com
IJARSE
ISSN 2319 - 8354

Comparative Analysis of Navigation Patterns by Web Users with Profile

Rajesh Ku.Nigam [1], Dr.Chandikaditya Kumawat [2],

Dr.Manish Shrivastava [3]

1. Research Scholar, CSE Deptt., Mewar University, Rajasthan (India) (rajeshrewa37@gmail.com)

2. Professor, CSE Deptt., Mewar University, Rajasthan (India) (chandikaditya@gmail.com)

3. Professor, IT Deptt LNCT, Bhopal (India) (contct.manishshrivastav@gmail.com)

Abstract: *An Organization need to understand their customers' behavior, preferences and future needs which depend upon past behavior. Web Usage Mining is an active research where customers session clustering is done to understand the customers activities. It investigates the problem of mining frequent pattern and especially focuses on reducing the number of rules using closed sequential pattern technique. It also reduce scans the size of the database using some clustering technique, which is used to compare. It solve the problem through profile based approach. In distributed environment web pages access by the user having some patterns, these patterns are merging and finding closed frequent set of web pages. If user need next request page in advance then it search web data, so it maintain cluster based on some technique. These type different approach or technique are used to find the user's interest with less execution time.Here this paper showing the analysis of previous algorithm for web usage mining.*

Key Words: *Data Mining, Web Mining, Web Usage Mining, Closed Sequential Patterns, Sequence Tree, User Profile, Web Database, Web Services, Self Organizing Map.*

1. INTRODUCTION

Today, organizations are depending more and more on their websites to correlate with customers. Holding recent customers and attracting potential ones push these organizations come across in striking ways to make their websites more useful and efficient. The WWW is a massive source of information that can come either from the Web content, delineated by the billions of web pages openly accessible, or from the Web usage, represented by the registration information collected daily by all the servers around the world. Web Mining is that part of Data Mining which deals with the extraction of interesting information from the WWW. Web usage mining has many applications, e.g., personalization of web substance, support of the design, recommendation systems, pre-fetching and caching, etc. There are miscellaneous benefits of web usage mining, mainly in e-commerce. Customers can be targeted with appropriate advertisement. Also, products which are related to customers can be suggested in real-time while browsing the website. According to, the usage mining process can be divided into three steps. The first step starts with data cleaning and pre-processing. Second step is the pre-processed data are mined for some unseen and productive information. And the last step of the web log mining process ends by analyzing the mining results.

Web usage mining is defined as the process of applying data mining techniques to the discovery of usage patterns from web logs data and to identify web users' behavior. In Web usage mining, data can be collected at the server-side, client-side and proxy servers.

Clustering have been useful and active areas of machine learning research that promise to help us cope with the problem of information overload on the Internet. With clustering the goal is to separate a given group of data items (the data set) into groups called clusters such that items in the same cluster are similar to each other and dissimilar to the items in other clusters.

2. RELATED WORK

In 2016, Fan Muhan [2] proposes a method for mining the frequent closed patterns in a sliding window to capture information timely and accurately when new data stream arrives. Data stream is divided into several basic windows. All possible frequent closed patterns are mined in each basic window and be stored in Closed Pattern-tree in sliding window updates, Closed Pattern-tree can be incrementally updated and the infrequent or unclosed patterns will be deleted from the tree.

In 2016, Minubhai [3] proposed a prefixspan algorithm with GRC constraints which generates sequential patterns by using prefix projected pattern growth approach is implemented. Other than frequency this algorithm also uses gap, compactness and recency constraints during sequential pattern mining process. The gap constraint applies limit on the separation of two consecutive transactions of discovered patterns, recency constraint makes patterns to quickly adapt the latest behaviors and compactness constraint make sure reasonable time spans for the discovered patterns.

In 2016, Doddegowda [4] having approach to personalize the information available on the Web according to user requirements. This is called Web Personalization process that adjusts information/services delivered by a Web to the needs of each user or group of users, taking their behavioral patterns. Frequent Sequential Patterns (FSPs) that are extracted from Web Usage Data (WUD) are very important for analyzing and understanding users' behavior to improve the quality of services offered by the World Wide Web (WWW). User behavioral patterns are required to build profiles of each user, using which Personalization of website is made.

In 2014, Jerry Chun [5] proposed the prelarge concept is adopted to handle the discovered sequential patterns with sequence deletion. An FUSP tree is first built to keep only the frequent 1-sequences from the original database. The prelarge 1-sequences are also kept in a set for later maintenance approach. When some sequences are deleted from the original database, the proposed algorithm is then performed to divide the kept frequent 1-sequences and prelarge 1-sequences from the original database and the mined 1-sequences from the deleted customer sequences into three parts with nine cases. Each case is then processed by the designed algorithm to maintain and update the built FUSP tree. When the number of deleted customer sequences is smaller than the safety bound of the prelarge concept, the original customer sequences are unnecessary to be rescanned, but the sequential patterns can still be actually maintained and updated.

In 2013, Rahul Moriwal [7] - It presented a method for Finding Frequent Sequential Traversal Patterns from Web Logs which is based on Dynamic Weight Constraint, where various frequent sequential pattern mining algorithms have been proposed that mines the set of frequent subsequences pattern which satisfying a min. support constraint in a particular session database. Though, previously sequential pattern mining algorithms gives equal weightage to sequential traversal patterns whereas the pages in sequential patterns have different importance and also have different weightage. Here items are given dissimilar weights and traversal sequences assign a min. and max. weight.

Ketki Muzumdar, Ravi Mante, Prashant Chatur (IJRTE-2013) proposed "Neural Network Approach for Web Usage Mining" in which Web usage mining try to discovers useful knowledge from the secondary data obtained from the connections of the users with the Web. It represent a novel method Self Organizing Map, which is a kind of neural network, in the process of Web Usage Mining to detect user's patterns. It analyze the traditional K-Means algorithm result with comparison to SOM. The process details the transformations necessaries to modify the data storage in the Web Servers Log files to an input of SOM.

In 2013, Omar Zaarour, Mohamad Nagi [8] proposed an improvement to the web log mining procedure and to the prediction of online navigational pattern. Their contribution contains three different components. First they proposed for session identification, a refined time-out based heuristic,. Secondly, suggested the practice for navigational pattern detection by using a specific density based algorithm. Finally, a new method for efficient online prediction is also recommended. The conducted experiment shows the applicability and effectiveness of the proposed method.

3. COMPARATIVE ANALYSIS

Frequent sequence mining is an important part related to web data and now yet a challenging data mining work. The mining frequent sequence has become an important component of many prediction or recommendation systems. The online store every time want customers next item prediction as web pages likely to visit. It also like to buy together which products. The existing algorithms used for frequent sequence mining could be classified either as exact or approximate algorithms. Accurate frequent sequence mining algorithms usually read the whole database several times, and if the database is very large, then frequent sequence mining is not compatible with limited availability of computer resources and real time constraints. So the problem in the current scenario are –

1) Web data partition is not used some conditional parameters just like profile constraint (Income, Age, and Experience etc.) which support as **conditional parameter** for partition of web data.
2) Many previous sequential mining algorithms shows no reflection of **importance of pages** where as every pages have different importance. So the existing methods perform response time is also slow. Website required reasonable approximate methods for analyzing data where the **computation speed** is more important than the precision.
3) Every time the **whole database scan** for searching the frequent pattern not partial database. At the time of program execution **number of cluster** required as a input parameter.

Table 1 Review on Recommender System

Author [Year]	Title of Research Paper	How It Works
IEEE 2014	Web-page recommendation based on web usage and domain knowledge	This paper proposes a method to efficiently provide better Web-page recommendation by integrating the domain and Web usage knowledge. Two models are proposed first model uses ontology and the second model uses automatically generated semantic network. Another model, is conceptual prediction model, it is proposed to generate a semantic network of the semantic Web usage knowledge, which is the integration of domain knowledge and Web usage knowledge. The results have been compared with the results obtained from an advanced Web Usage Mining (WUM) method. The results demonstrate that the proposed method produces significantly higher performance than the WUM.
IJAERD 2015		It has become more difficult for existing web based systems to retrieve any kind of information, due to growth of World Wide Web in terms of information space and amount of users. However, traditional recommender systems result in failure when the data to be used throughout the recommendation process is sparse. This Paper focuses on the development and evaluation of a web based movie recommendation system.
IEEE 2014	Web service recommendation via exploiting location and QoS information	Web services have been employed for service-oriented applications development. Available large amount of web services makes it hard for a user to select a proper Web service. An inappropriate service selection may cause many problems. In this paper, a novel collaborative filtering based Web service recommender system is proposed to help users to select services with optimal Quality-of-Service (QoS). Compared with existing service recommendation methods, the approach achieves considerable improvement on recommendation accuracy.
IEEE 2015	Unified collaborative and content-based web service recommendation	Most existing Web service recommendation methods focus on UDDI registries or keyword based Web service search engines that possess many limitations. Recent research efforts on Web service recommendation center on two prominent approaches: collaborative filtering and content-based recommendation. Unluckily both methods have some drawbacks. In this paper, a novel approach is proposed that unifies collaborative filtering and content-based recommendations. Particularly, this approach considers both rating data and semantic content of Web services. The experimental results show that approach outperforms the state-of-the-art methods on performance.
IRJET 2016	Enhanced Bee Colony Optimization Mechanism In Content Recommendation System	Many internet users are not only information consumers but also providers. There is lots of information in the Web and most people can find what they want by searching the Web. One problem of large quantity of data in Web is that we spend most of our time to find a correct result. A recommendation is based on Bees Algorithm can help for finding appropriate information.

IJAERD 2016	Survey on Hybrid Recommendation System	Recommender systems are used by e-commerce websites to help the customers to find suitable products. It is the technique to deal with large amount of data for efficient data filtering. RS serve right item to user in an automated fashion to satisfy objective. Major task of the recommender system is to present recommendations to users.
IEEE 2016	An advanced recommendation system for Ecommerce users	Over 200 million customers usages online electronic for trade. The internet business model, provides accessibility to suppliers and new stages of efficiency. The development in e-commerce has result in a substantially multiplied demand for understanding and communications. This paper analyzes the influence of e-commerce on markets. In specific, study the retail sector where the development of online procedures and the provision of more than a few resources. Authors have developed a recommender system to analyze effect on competitors. The discounted items are supplied to the users in an effort to expand the client's interest.
Ming Sun et al., (2009) [6]	Mining the association rules on log ontology	Mining the association rules on log ontology by using Inductive Logic Programming (ILP)
Bo Cheng et aal., (2018) [7]	Semantic Web Mining and Indexing	Results to the high precision and recall rate. Thus leads as a solution to the issues of traditional techniques i.e. low searching processing time and high search accuracy for web service discovery field
C. Ramesh et al., (2017) [9]	Sequential Pattern Mining Algorithm, CloSpan	The recommendation of the system was generated on the basis of the semantic knowledge of the data
C S Saravana Kumar et al., (2018) [10]	Semantic web mining for keyword based search	To enhance the relationship establishment of the training sentence along with the high accuracy of information extraction

On the basis of the survey that has been done in this study, it is observed that the ontology based mining is highly preferred in the semantic web mining. Various techniques such as Resource description framework etc are applied to the domain to enhance the accuracy level. The following issues are obtained from the traditional works that are still need to resolve. The vision of the Semantic Web Mining requires information systems that can exchange information and reuse the exchanged information with their intended meanings. Obtaining this interoperability between different information systems is extremely tedious and errorprone in a homogeneous as well as heterogeneous environment like the World Wide Web.

4. CONCLUSION

In this study we had tried to go through the various concepts such as web mining, web content mining, web usage mining and semantic web mining. The major focus behind organizing this study is to analyze and review all of the relative terms of semantic web mining and the researches that had been conducted in this domain to make it more effective and efficient.

On the basis of the review of the related work, it is observed that the previously developed recommendation systems were failed to generate the effective recommendations for the users as it did not considers the recent queries generated by the users. Thus in future an efficient recommendation system can be developed by considering various factors such as the nature of the current queries of the users in the terms of semantics and use these semantics to make recommendation of the web sources to the users.

In this research a novel proposed approach is used to finding closed sequential patterns to generate minimum pattern. After that it scan only partial web data for next item prediction. It filtering large web data by matching user profile similarity based on some attributes and conditions. It fulfil the minimum support criteria. Initially each item suppose as a cluster & merge them in final cluster so that partial web data is scan. Closed frequent pages are very less. It generate useful rules in the form of clustered by clustering technique.

Now multiple scanning of database will be reduced. Scanning only Partial Database not the whole database with increase the response time. It identify user needs by collecting data on the subjects and topics those are most searched by other users. It enables effective tracking for the development and improvement of the user interface and software by analyzing user behaviour.

5. FUTURE WORK

In future work, other data mining algorithms can be implemented in cloud which efficiency handle large web data of e-Commerce website in distributed environment . It is very helpful for finding any critical diseases, so there are many areas just like parallel sequential pattern, grouping of similar type of customers, classification of different diseases in suggestive manner using distributed servers.

REFERENCES

[1] Miss. Vrinda Khairnar and Miss. Sonal Patil, "Efficient clustering of data using improved K-means algorithm", Imperial Journal of Interdisciplinary Research (IJIR), Vol.2, Issue-1, 2016.

[2] Fan Muhan, Shao Sujie, Rui Lanlan, “A Mining Algorithm for Frequent Closed Pattern on Data Stream based on Sub Structure Compressed in Prefix Tree”, IEEE Proceedings of CCIS, pp. 434-439, 2016.

[3] Minubhai Chaudhari, Chirag Mehta, “Extension of Prefix Span Approach with GRC Constraints for Sequential Pattern Mining”, International Conference on Electrical, Electronics, and Optimization Techniques, pp. 2496-2498, 2016.

[4] Doddegowda B. J., G. T. Raju, Sunil Kumar, “Extraction of Behavioral Patterns from Pre processed Web Usage Data for Web Personalization”, IEEE International Conference On Recent Trends In Electronics Information Communication Technology, pp. 494-498, 2016.

[5] Jerry Chun, Wensheng Gan, Tzung Pei Hong, “Efficiently Maintaining the Fast Updated Sequential Pattern Trees With Sequence Deletion”, IEEE Access - The Journal for Rapid open access publishing, Vol. 2, pp. 1374-1383, 2014.

[6] Sheetal Sahu, Praneet Saurabh, Sandeep Rai, “An Enhancement in Clustering for Sequential Pattern Mining Through Neural Algorithm Using Web Logs”. IEEE Proc. of the 2014 Sixth International Conference on Computational Intelligence and Communication Networks, pp 758-764, 2014.

[7] Rahul Moriwal and Vijay Prakash, “An Efficient Algorithm for Finding Frequent Sequential Traversal Patterns from Web Logs based on Dynamic Weight Constraint”, Proceedings of the Third International Conference on trends in Information, Telecommunication and Computing, Vol. 150, 2013.

[8] Zaarour Omar, Nagi Mohamad, “Effective web log mining and online navigational pattern prediction,” International Journal of Knowledge Based Systems, Elsevier, Vol. 49, pp. 50-62, Sept. 2013.

[9] Varghese Nayana Mariya, John Jomina, “Cluster Optimization for Enhanced Web Usage Mining using Fuzzy Logic,” World Congress on Information and Communication Technologies, IEEE, pp. 948-952, 2012.

[10] Raut Ms. Anjali B., and Bamnote Dr. G. R., “Clustering Method based on Fuzzy Equivalence Relation,” International Conference on Computer & Communication Technology (ICCCT), pp. 666-671, 2011.

[11] Azimpour-Kivi Mozhgan, and Azmi Reza, “A Webpage Similarity Measure for Web Sessions Clustering Using Sequence Alignment,” International Symposium on Artificial Intelligence and Signal Processing (AISP), IEEE, pp. 20-24, Jun. 2011.

[12] Giannikopoulos P., Varlamis I., Eirinaki M.., “Mining Frequent Generalized Patterns for Web Personalization in the Presence of Taxonomies”, IJDWM, 2010.

[13] K. R. Suneetha, Dr. K. R. Krishnamoorthy, "Identifying User Behavior by Analyzing Web Server Access Log File", IJCSNS International Journal of Computer Science and Network Security, Vol. 9, No.4, pp. 327, 2009.

[14] Krzysztof D., Wojciech K., Marcin S., "Effective Prediction of Web User Behaviour with User-Level Models", Fundamenta Informaticae , IOS Press , Vol. 89, No. 2-3, pp. 189, 2008.

[15] Khan, J. I. Qingping T., "Webspace surfing patterns and their impact on Web prefetching", Cyberworlds Proceedings, pp 478 – 485, 2003.

[16] Catledge, L. D. and Pitkow, J. E., "Characterizing Browsing Strategies in the World- Wide Web", In Proceedings of the Third International World-Wide Web Conference, Darmstadt, Germany, pp. 1065–1073, 1995.