IJDCST @October Issue- V-1, I-7, SW-3
ISSN-2320-7884 (Online)
ISSN-2321-0257 (Print)

Dynamic Clustering Analysis over Web Logs

[1]Nurunnisa Shaik, [2]U.Tulasi,[3] S.Phani Praveen

[1] Student, DEPT.OF CSE, PVPSIT, KANURU, VIJAYAWADA.

[2,3] Asst.Professor, DEPT.OF CSE, PVPSIT, KANURU, VIJAYAWADA.

Abstract: Data mining is the latest technology for accessing web related data. In this region traditionally more number of techniques was used for accessing web related data aspects present environment. In data mining Web usage mining is the main concept for accessing relevant web data. Web Usage Mining has three aspects for preparing relevant data representation they are data preprocessing, data mining analysis and results analysis. Considering these aspects present in the web data mining, traditionally Longest Common Subsequence analysis and some clustering algorithms were developed in Web Usage Mining process. But these techniques do not refer the particular user information from retrieved web search results. So for increasing the user abilities in mining of usage data. In this paper we propose hybrid clustering algorithms like Principal Components Analysis and multi Classification Analysis are used for analyzing Weblogs. Our experimental shows the web usage results efficiently based on user abilities present in the data sets.

Keywords: Multi classification analysis, Automatic clustering, Web Usage Mining, WWW, Web logs, pre-processing, Data Warehouse, Longest Common subsequence.

I. INTRODUCTION

Data mining is the process of extracting relevant information from different data warehouses present in the mining framework process. Data mining is the computational process identifying patterns present in the web related data processing. In these web discovering different data patterns in large data sets involving methods at the intersection of database systems. The goal of the data mining process is to retrieve information from a data set and transfer it into an understandable feature structure in data processing.

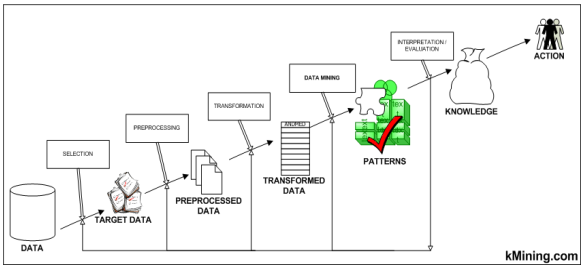

Figure 1: Data mining process on retrieving information.

In data mining Web mining is the research topic present in days we are focused on application development of related web content. Web mining deals with three main areas that are Web Usage Mining, Web structure Mining, Web Content Mining. Web usage Mining is the process of extracting useful information from server logs. Web Usage mining is the process of fiding out what users are looking for internet. 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. Usage data captures the identity or origin of Web users along with their browsing behavior at a Web site. In Web Usage Mining Traditionally developed techniques are processed i.e. In this region classification algorithms are used like Longest Common Subsequence. The WUM analysis, also, allows the webmaster to optimize the response of the Web server (Web caching) and to make recommendations to the user. In this paper propose to extend the longest common Subsequence algorithm, for dynamic clustering; it is the hybrid clustering process for arranging web related data in sequential

order. Our experimental results shows efficient data accessing from web content mining. We perfect the architecture and lets it servers for accessing relevant user information. These results are based on the user ability present in the original data sets.

II. RELATED WORK

Data mining concepts are the basic data extracting process. Data extracting is the main aspect in present days. In data web usage mining many numbers of techniques was developed previously. In this aspect WUM are specifically designed to carry out the analyzing data representation usage data about a particular Web site. In this aspect web usage mining can model user behavior and their failures. To provide online prediction efficiently, we advance architecture. After that online prediction in web usage mining with novel technology i.e. Longest Common Subsequence clustering algorithm. It improves the accuracy of the data accessing in classification in the architecture. But these results are not sufficient for accessing web related data content. So in this paper we propose to extend the data extracting process for relevant web data content based on structure or usage of user information. Our proposed work provides data accessing results efficiently.

III. EXISTING APPROACH

A web server page is quadratic in the number of pages. During the online phase, when a new request arrives at the server, the URL requested and the session to which the user belongs are identified, the underlying knowledge base is updated, and a list of suggestion is appended to the requested page.

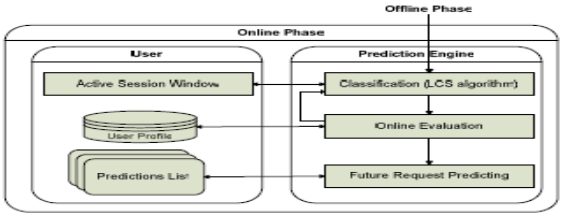

Figure 2: Data phasing process based on longest common subsequences.

Typically, the WUM prediction process is structured according to two components performed online and off-line with respect to the Web server activity. The off-line component is aimed at building the knowledge base by analyzing historical data, such as server access log files, that is then used in the online component. In this project. In this architecture we propose novel approach for classifying user navigation patterns by using Longest Common Subsequence (LCS) algorithm. The LCS algorithm exploits for improving accuracy of classification in the architecture.

Classification Algorithms with Longest Common Subsequence:

We are taking input as data sets like np = {np1,np2,…..npn} with pattern recognization like (p1,p2,…..pn). We consider the data pattern matching present in original data set. Firstly we are taking preprocessing on original data sets and then we are apply clustering algorithms based longest common subsequence *W.* This technique can be applied for minimizing data sequences present in the original data sets. Next user activity also provides using classification process. Examples processing data results are as follows:

Number List	Navigational Pattern
1	(p1,p4,p5,p10)
2	(p12,p56,p26)
3	(p12,p35,)
4	(p25,p35,p65)

Table 1: Navigational results present in the datasets with pattern matching.

A user choose a prediction list, prediction engine rebuild the predictions. This model prediction results access the prediction list that strictly related those determine the classification process.

IV. PROPOSED APPROACH

The web logs files are the input data in web usage data process. For this category we assigning the web data extracting present in the data base.

Preprocessing: Typically, the WUM prediction process is structured according to two components performed online and off-line with respect to the Web server activity. The off-line component is aimed at building the knowledge base by analyzing historical data, such as server access log files, that is then used in the online component. In this project. In this architecture we propose novel approach for classifying user navigation patterns by using Longest Common Subsequence (LCS) algorithm. The LCS algorithm exploits for improving accuracy of classification in the architecture. It exploits the web usage process in the data navigation process.

Data Clustering Process: Dynamic clustering is the web process for extracting efficient results. Regarding to the web usage data representation static analysis and dynamic analysis is the assurance present in the original data sets. In this region it will display the data efficient results with user accessing. Based on URLS and data usage present in the data sets we are providing efficient data process. In this region user register and login with potential credentials.

Building WUM Data Warehouses: The Web server usually registers all user's access activities of the website as Web server logs. Due to different server setting parameters, there are many types of web logs, but typically the log files share the same basic information, such as: client IP address, request time, requested URL, HTTP status code, referrer, etc. Generally, several pretreatment tasks need to be done before performing web mining algorithms on the Web server logs. The main objective of prediction engine in this part of architecture is to classify user navigation patterns and predicts user's future requests. For this purpose we propose a novel approach to classify current user activity. Client side verification implemented by using a remote agent by modifying source code to existing code aspects.

V. PERFORMACE ANALYSIS

The hybrid clustering uses Principal Component Analysis and Dynamic clustering method is to find the group of homogeneous navigation results as follows:

Step 1: Let x be the objects of a set of "navigational "at probabilistic, be the "representation" of the x clusters. Let us call them B1,B2,B3,...Bk.
Step 2: All navigations Xi are assigned to cluster x i® d(Xi; Ak) is minimum.
Step 3: A new representation Bx is computed, it is the average of the elements of the cluster x
Step 4: Stability sequences.

Algorithm 1: Hybrid clustering process.

Hybrid clustering method applied on quantitative variables; Analysis of the navigational results.

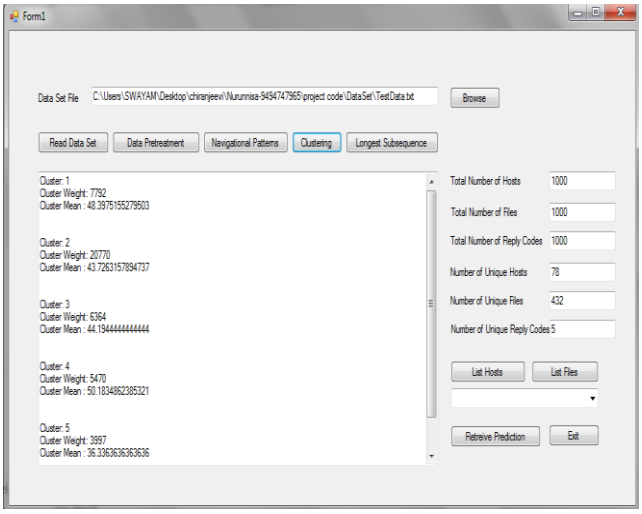

Figure 3: Data clustering results can be applied in original data sets.

Data results present in the original data sets we will apply data reading process and then regarding patterns and real data sets with longest sequence.

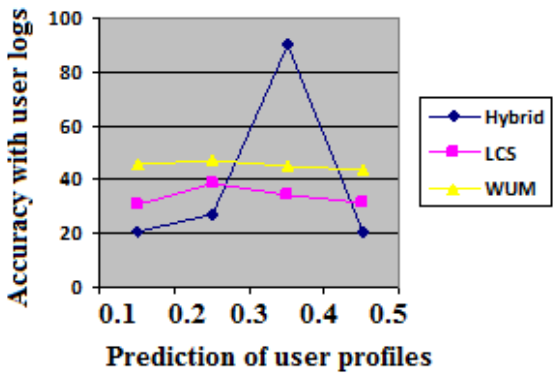

Figure 4: Accuracy results comparison with Web Usage mining and LCS and Hybrid Clustering Processes.

Data accuracy is the main aspect in present days of web usage mining process. Those results can be accessed in above diagram.

VI. CONCLUSION

Web usage mining is the main concept for accessing relevant web data. Web Usage Mining has three aspects for preparing relevant data representation they are data preprocessing, data mining analysis and results analysis. So for increasing the user abilities in mining of usage data. In this paper we propose hybrid clustering algorithms like Principal Components Analysis and multi Classification Analysis are used for analyzing Weblogs. Our experimental shows the web usage results efficiently based on user abilities present in the data sets. Further achievement of the our proposed work we have to develop WUM with most efficient clustering based on heuristic model, those results are accessed by the efficient user information based on user profile histories.

VII.REFERENCES

[1] 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, 1550-6037/08 $25.00 © 2008 Crown Copyright DOI 10.1109/IV.2008.40.

[2] E. Frias-Martinez, V. Karamcheti,"Reduction of user perceived latency for a dynamic and personalized site using web-mining techniques", *WebKDD*, 2003.

[3] R. Liu, V. Keselj," Combined mining of Web server logs and web contents for classifying user navigation patterns and predicting users' future requests", *Data & Knowledge Engineering*, Elsevier, 2007, pp.304-330.

[4] Mireille ARNOUX, Yves LECHEVALLIER," Automatic Clustering for the Web Usage Mining", Analele Universit¸at»ii din Timi»soara Vol. XLI, Fasc. special, 2003 Seria Matematic¸a{Informatic¸a.

[5] M. Jalali, N. Mustapha, A. Mamat, Md N. Sulaiman ,"OPWUMP An architecture for online predicting in WUM-based personalization system", *In 13th International CSI Computer Science*, Springer Verlag,2008.

[6] R. Liu, V. Keselj," Combined mining of Web server logs and web contents for classifying user navigation patterns and predicting users' future requests", *Data & Knowledge Engineering*, Elsevier, 2007, pp.304-330.