IJRSET OCTOBER 2016 Volume 3, Issue 10 Pages: 1-11

ISSN 2394-739X

International Journal for Research in Science Engineering and Technology

WEB PREDICTION USING FCM AND KFCM ALGORITHM

[1] Rubab Sharma, [2] Akanksha sambyal
[1] M.Tech (CSE), [2] Astt. Prof. (CSE),
[1, 2] SRI SAI UNIVERSITY,
[1, 2] Palampur
[1, 2] Himachal Pradesh, India

ABSTRACT- In present days, Internet is playing such a significant role in our day-to-day life. We have witnessed the evermore- interesting and upcoming publishing medium is the World Wide Web (WWW). The rapid growth in the volume of information available over the WWW and number of its' potential users' has leads to difficulties in providing effective search service for users', resulting in decrease in the web performance. Web Usage Mining is an area, where the navigational access behaviour of users' over the web is tracked and analyzed. So that websites owner can easily identify the access patterns of its users'. By collecting and analyzing this behavior of user activities, websites owner can enhance the quality and performance of services to catch the attention of existing as well as new customers. Web prediction is a clustering problem which attempts to predict the most likely web pages that a user may visit depending on the information of the previously visited web pages. In this research work focus is on web page prediction. For web page prediction we use algorithms FCM and KFCM and comparison between them on the basis of conditional variance parameter. But to remove the fuzziness and to reduce the time complexity we have introduced a term Gaussian under KFCM and named it as GKFCM.

Keywords- web mining, clustering, FCM and KFCM

1. INTRODUCTION

Web is an album of billion of travel document. The web is very diverse, enormous, dynamic, and flexible. The World Wide Web continues to increase both in the gigantic quantity of traffic and the bulk and intricacy of Web sites. It is tricky to identify the applicable information present in the web. Mostly the contents are unstructured in nature, but very minute work deals with heterogeneous and unstructured information on the Web. The up-and-coming field of web mining focus at discovering and extracting required information which is out of sight in Web-related data, in particular in text documents which are published. Data Mining introduces the thought of extraction useful and expensive information from large capacity of data. Web mining discover useful information or facts from the Web hyperlink structure, usage data and page content. Even though Web mining uses many data mining techniques, it is not merely an application of data mining due to the heterogeneity and semi structured or unstructured nature of the Web data. Many new mining

responsibilities and algorithms were made-up in the past decade. So web mining is disintegrating into various sub-tasks.

Resource discovery: It is used to extract the data from online text resources available on web.

Information choice and pre-processing: This process changes the original retrieved data into Information.

Generalization: Individual web sites as well as across multiple sites.

Analysis: It involves the interpretation and validation of the mined patterns.

The considerable web mining applications are web search, website design, information retrieval, network management, search engines, business and artificial intelligence, e-commerce, web market places and web communities. Online business breaks the obstacle of space and time in comparison to the physical office business. Huge companies approximately around the world are realizing that e-commerce is not just buying and selling over Internet, somewhat it improves the efficiency to contend with other giants in the market. This application consists of the temporal issues for the users. Web mining is an essential region in data mining where we extract the appealing patterns from the contents. Generally three kinds of information are handled in web site namely Content Structure Log data.

Web mining has three further division explicitly, web content mining, web structure mining and web usage mining. Each categorization is having its own tools and algorithms. Web content mining is nothing but the detection of precious information from web documents and these web documents may contain manuscript, picture, metadata, structured records and hyperlinks. It is used to look at the information by search engine or web spiders i.e. Google, Yahoo. It is the procedure of retrieving the valuable information from the web documents or web content. Web structure mining is also a process of discovering structured information from the websites. The structure of a graph contains web pages and hyperlinks where the web pages are measured as nodes and the hyperlinks as edges and these are associated between related pages. Web usage mining is also said as web log mining. It depicts the user's behaviour which can grasp the meaningful patterns from one or more web localities [9]. Fig.1 represents web mining and its types.

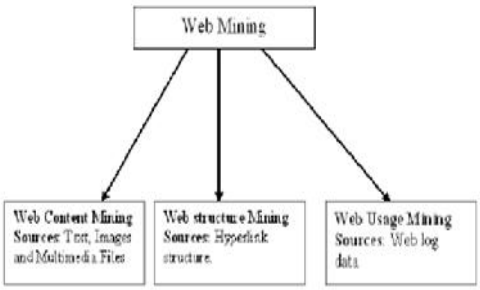

Figure 1- The types and source of web mining

2. WEB CONTENT MINING

Web content mining is also recognized as text mining is generally used to deals with raw data. Content mining handles text, pictures and graphs of a Web page to conclude the significance of the content to the search query. Content mining search the results lists to search engines in order of uppermost importance given to the keywords in the query. Web text mining is very useful when used in next of kin to a content database which deals with definite topics. For example online universities use a library system to remember articles associated to their common areas of study. The ability to carry out Web content mining permits results of search engines to take full advantage of the stream of customer clicks to a Web site, or particular Web pages of the particular site, to be accessed abundant times in importance to search queries. It is the method of getting the information from the web into more structured forms and indexing the information so as to retrieve quickly or finding valuable information from web documents or web contents. Web content mining consists the web documents which may consist of manuscript, html, multimedia documents i.e., pictures, auditory, video and sound

etc. It further consists two approaches; those are Agent based Database Approach.

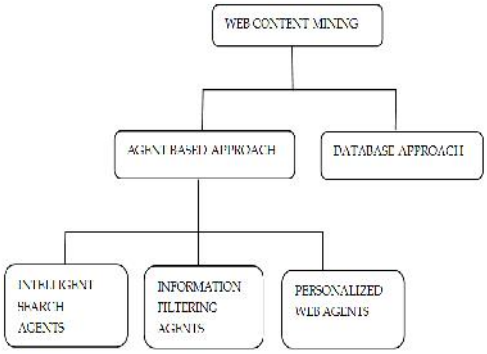

Figure 2- Web Content Mining

3. WEB STRUCTURE MINING

Web structure mining is the learning of data unified to the structure of a exacting website. It contains web graph which has the web documents or web pages the same as nodes and hyperlinks like edges which are concerning between two linked pages [7]. Figure 3 represent the web graph structure.

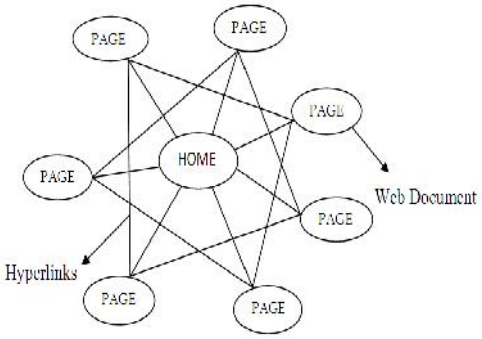

Figure 3- Web Graph Structure

Web structure is valuable supply for extracting information. Web structure, which extracts several attractive web graph patterns like complete bipartite graphs, social choice, co-citation, etc [1]. It further divides the web page on a variety of topics and concluding which web page is to be considered into the compilation of web pages. Web structure mining can be done either at inter-page level or intra-page level. A hyperlink which links to a various part of the similar page is known as intra-page hyperlink.

4. WEB USAGE MINING

When client browses various websites then browsing actions of the client automatically set aside into web log file. Web usage mining handles with such log files and take out information related to client browsing activities on internet. This taken out information is used in improving the website design, Personalization, predicting the user future requests, Business intelligence. Getting better the prediction process can drop off the client's access times during browsing and it can also relieve network traffic by trimming to visit the needless pages. client future request prediction is a web usage mining method for predicting user next request. For this idea, firstly web log files are checked and client's future requests are predicted in accordance to the previous related actions. Web usage mining on the whole aims on the methods which predict the client actions when the client get in touch with the World Wide Web. So at any time new web site is produced the main spotlight is reserved on client interest. It is an commotion which includes the automatic detection of patterns as of one or more servers. The usage data keep the records of every client action on web server when the client makes communication. It is also describes as the finding of user access patterns from web server logs, which organise an account of every client. The exposed patterns are shown as collection of pages, objects etc which are accessed by various clients having common area of interest. In this web usage mining, the subsequent activities are observes that are : browsing actions prediction of the client actions inside the site, adjustment of the Web site in accordance to the interests of its clients, relationship between expected and actual Web site usage. The web usage mining is completed in 2 steps:

The first step sends the usage data of the www server into the relational data prior to the data mining technique is done.

The second step uses the data straight by applying pre-processing method to mine the data.

Web usage mining is also known as web log mining which analyzes the manners of online clients [2]. It is divided into two types of tracking; customize usage tracking and general access tracking [3]. The general access tracking assures the customer deeds on the web and it recognize the client while the client co-operate with the web. It can handles the data without human intervention in the web server log and application log [15]. The web log is positioned in three various locations they are web server log, web proxy server and client browser and it includes only plain text file (.txt). The huge amounts of unrelated data are offered in the web log file as it deals with large amount of incomplete, noisy data, eroded and needless information [6]. Web usage mining is to take out the data which are lay up in server access logs, agent logs, referrer logs and error logs. Web usage mining generally utilize basic data mining algorithms such as association rule mining, clustering, sequential rule mining and classification. It has quite a lot of tools to analyze the performance of the client.

5. ARCHITECTURE OF WEB USAGE MINING

Architecture of the web usage mining is revealed in Fig 1.4. Originally the data is composed from different sources. Now server logs are unsoiled by data cleaning process which eliminate the monotonous data like images, needless links etc which are stored inside the Clean log. Then concerned transactions are recognized from the clean log by making use of Transaction Identification process and supplies into the Transaction Data.

Then it go away throughout the Integration process by integrating user registration data, transaction data, documentations and usage attributes into integrated data. Then alteration process is applied on this data to normalize the data. A variety of blue print can be taken out from this formatted data like clustering, classification, association analysis etc.

Ultimately Pattern analysis is applied on these patterns by applying different tools to dig out the useful information. The tools which are used at the final stage are Knowledge Query Mechanism to mine the necessary knowledge as of from the final output by indicating the query, OLAP/Visualization tools to envisage the output in different formats, Intelligent agents pull out automatically useful knowledge from the patterns created.

6. WEB LOG FILE

Log files are files that list the actions that have been occurred. These log files reside in the web server. Computers that deliver the web pages are called as web servers. The Web server stores all of the files necessary to display the Web pages on the users computer. All the individual web pages combines together to form the completeness of a Web site. Images/graphic files and any scripts that make dynamic elements of the site function. , The browser requests the data from the Web server, and using HTTP, the server delivers the data back to the browser that had requested the web page. The browser in turn converts, or formats, the files into a user viewable page. This gets displayed in the browser. In the same way the server can send the files to many client computers at the same time, allowing multiple clients to view the same page simultaneously. Log files contain information about User Name, IP Address, Time Stamp, Access Request, number of Bytes Transferred, Result Status, URL that Referred and User Agent. The log files are maintained by the web servers. By analysing these log files gives a neat idea about the user.

7. USING LOG FILE DATA in WEB USAGE MINING

The filling of the Log files are worn in this type of mining. Web usage mining also contains three main steps:

Pre processing

The data there in the log file cannot be worn as it is for the mining process. [9] Consequently the filling of the log file should be dirt-free in this pre-processing

step. The unnecessary data are separated and a reduced log file is obtained. The pre-processing of Web logs is usually time demanding and complex. It encompasses four different tasks:

Data cleaning

Identification and the rebuilding of clients sessions Retrieving of information about page content and structure, and
Data formatting

Data cleaning

This step consists of take away all the data followed in Web logs which are useless in mining intention e.g.: requests for any another file which might be incorporated into a web page; desires for graphical page content (e.g., jpg and gif images); or even direction-finding sessions performed by Web crawler and robots. While requirements for graphical contents and files are easily removed, robots and Web crawler direction-finding patterns must be clearly identified. This is usually done for illustration by referring to the user agent, by referring to the remote hostname, or by examination of the access to the robots.txt file. However, some robots in fact send a fake client agent in HTTP request. In these cases, a heuristic based on direction-finding behaviour can be used to break up robot sessions from definite client's sessions.

CONTENT AND STRUCTURE RETRIEVING

The enormous mainstream of Web Usage Mining applications makes use of the most visited URLs as the main supply of information for mining intention. URLs are however a deprived source of information since, for illustration, they do not express any information regarding to the actual page substance. To utilize content based information the Web log data is enriched. Introducing an additional labelling step in which Web pages are confidential according to their substance type; this additional information is then broken down during the mining of Web logs. If an enough classification is not identified in progress, Web Structure Mining techniques can be engaged to expand one. As in search engines, Web pages are confidential in accordance to their semantic areas by way of Web Content Mining methods; this classified information can be used to improve information taken out from logs. In Semantic Web for Web Usage Mining: Web pages are record on top of ontology's to add sense to the often observed paths. Specified page in the Web site, we must be capable to take out domain-level structured objects as semantic entities enclosed in this page. This task may engross the automatic taking out and classification of objects of various types into programme based on the fundamental domain anthologies. The domain anthologies may be scholarly automatically from existing training data or may be pre-specified. Given this aptitude, the transaction data can be altered into a representation which integrates multifaceted semantic entities admittance by client during a call to the site. Concept-based paths is an another option to the usual client direction-finding paths; concept-based path are a lofty level overview of standard path in which frequent concepts are mined by means of roundabout of raw client paths and resemblance measures.

DATA FORMATTING

This is the last step of pre-processing. Once the earlier stage have been finished, data are appropriately formatted before using mining techniques holds data obtained from Web logs into a relational database by using a click fact plan, so to make available better hold to log querying conclude to numerous pattern mining.

8. CLUSTERING

In the field of software data analysis is considered as a very useful and important tool as the task of processing large volume of data is rather tough and it has accelerated the interest of application of such analysis. To be precise data

mining is the analysis of datasets that are observational, aiming at finding out unsuspected relationships among datasets and summarizing the data in such a noble fashion that are both understandable and useful to the data users [9]. It also makes data description possible by means of clustering visualization, association and sequential analysis. Data clustering is primarily a method of data description which is used as a common technique for data analysis in various fields like machine learning, data mining, pattern reorganization, image analysis and bio-informatics.

Cluster analysis is also recognised as an important technique for classifying data, finding clusters of a dataset based on similarities in the same cluster and dissimilarities between different clusters [13]. Putting each point of the dataset to exactly one cluster is the basic of the conventional clustering method where as clustering algorithm actually partitions unlabeled set of data into different groups according to the similarity as compare to data classification, data clustering is considered as an unsupervised learning process which does not require any labelled dataset as training data and the performance of data clustering algorithm is generally considered as much poorer.

Although data classification is better performance oriented but it requires a labelled dataset as training data and practically classification of labelled data is generally very difficult as well as expensive. As such there are many algorithms that are proposed to improve the clustering performance. Clustering is basically considered as classification of similar objects or in other words, it is precisely partitioning of datasets into clusters so that data in each cluster shares some common trait. The hierarchical, partitioning and mixture model methods are the three major types of clustering processes that are applied for organising data. The choice of application of a particular method generally depends on the type of output desired, the known performance of the method with particular type of data, available hardware and software facilities and size of the dataset [13].

Generally clustering is classified into hard cluster and soft cluster [13]. If the data present in only one cluster then it comes under Hard cluster, if the data is allowed to present in more than one set then it comes under Soft cluster. K-means algorithm comes under hard cluster and fuzzy clustering algorithm comes under soft cluster. Sometimes the clustering is classified as Exclusive clustering, overlapping clustering, hierarchical clustering, and probabilistic clustering. Exclusive clustering is also defined as hard clustering and the best example is k-means clustering. Overlapping clustering is soft clustering and the example for it is Fuzzy C-Means clustering.

Similarly the example for probabilistic clustering is mixture of Gaussians. In Clustering, one of the most widely used algorithms is fuzzy clustering algorithms. Fuzzy Clustering is also called as soft clustering i.e. data elements belong to more than one cluster. Sometimes fuzzy cluster is defined as the soft version of k-means so it is also called as Fuzzy C-Means Clustering (FCM). Fuzzy set theory was first proposed by Zadeh in 1965 & it gave an idea of uncertainty of belonging which was described by a membership function.

The use of fuzzy set provides imprecise class membership function. Applications of fuzzy set theory in cluster analysis were early proposed in the work of Bellman, Zadeh, and Ruspini [2]. Integration of fuzzy logic with data mining techniques has become one of the key constituents of soft computing in handling challenges posed by massive collections of natural data. The central idea in fuzzy clustering is the non-unique partitioning of the data into a collection of clusters. The data points are assigned membership values for each of the clusters and fuzzy clustering algorithm allow the clusters to grow into their natural shapes [3].

9. FUZZY C-MEANS CLUSTERING

Bezdek [5] introduced Fuzzy C-Means clustering method in 1981, extend from Hard C-Mean clustering method. FCM is an unsupervised clustering algorithm that is applied to wide range of problems connected with feature analysis, clustering and classifier design. FCM is widely applied in agricultural engineering, astronomy, chemistry, geology, image analysis, medical diagnosis, shape analysis and target recognition [16]. With the development of the fuzzy theory, the FCM clustering algorithm which is actually based on Ruspini Fuzzy clustering theory was proposed in 1980's.

This algorithm is used for analysis based on distance between various input data points. The clusters are formed according to the distance between data points and the cluster centers are formed for each cluster. Infact, FCM is a data clustering technique [11, 7] in which a data set is grouped into n clusters with every data point in the dataset related to every cluster and it will have a high degree of belonging (connection) to that cluster and another data point that lies far away from the center of a cluster which will have a low degree of belonging to that cluster.

Fuzzy clustering is a powerful unsupervised method for the analysis of data and construction of models. In many situations, fuzzy clustering is more natural than hard clustering. Objects on the boundaries between several classes are not forced to fully belong to one of the classes, but rather are assigned membership degrees between 0 and 1 indicating their partial membership. Fuzzy c-means algorithm is most widely used. The FCM employs fuzzy partitioning such that a data point can belong to all groups with different membership grades between 0 and 1.

This algorithm works by assigning membership to each data point corresponding to each cluster center on the basis of distance between the cluster center and the data point. More the data is near to the cluster center more is its membership towards the particular cluster center. Clearly, summation of membership of each data point should be equal to one. After each iteration membership and cluster centers are updated according to the formula. The fuzzy c-mean clustering is classified under Soft Clustering i.e., overlapping clustering. It is also indicated as the soft version of K-Means clustering [11].

10. Kernel zed Fuzzy C-Means (KFCM) Algorithm

KFCM is an algorithm which is generated from FCM by modifying the objective function using Kernel induced distance matrix instead of Euclidean distance in FCM. KFCM overcomes all the limitations of FCM algorithm but to remove fuzziness we have introduced term Gaussian into the system which reduces the fuzziness as well as reduce the time complexity. And thus the corresponding algorithm is derived and called as the Gaussian kernel fuzzy c-means (GKFCM) algorithm, as KFCM is more robust than FCM. The main motives of using the kernel methods consist in:

Inducing a class of robust non- Euclidean distance measures for the original data space to derive new objective functions and thus clustering the non-Euclidean structures in data;

Enhancing robustness of the original clustering algorithms to noise and outliers

Retaining computational simplicity. [10]

KFCM is a new approach to user future request prediction. KFCM is better than FCM because KFCM use kernel induced function instead of Euclidean distance function. The results show that KFCM pick maximum data that has highest probability and it makes center point at that place where the data points are more. But under KFCM we have introduced a term name Gaussian which reduces the fuzziness as well as reduces the time complexity. Thus naming this algorithm as GKFCM. As the clusters of KFCM are better than FCM clusters and prediction is also better and by using GKFCM we can reduce the time complexity too.

11. METHDOLOGY USED

Step1: Read web log file.
Step2: Pre-processing
Select required attribute from log file like IP Address, User Requests, Request Method, Date, Time, and Status Code.
Remove irrelevant entries like all log entries with file name .jpg, .gif, .jpeg, robots files, error code, Request method HEAD, POST.
Cleaned log file obtained. From cleaned log file identify unique users according to IP Address and unique web pages.
Session identification step is performed after user identification. In this step sessions are identified for all users by taking 30 minute time threshold value. Pages visited by user less than or equal to 30 minute time put into one session and another pages which are visited after 30 minute put into another session.
Assigning the unique session id to all sessions
Step3: Clustering
Put the whole data of user session ids and Webpage visited by each user in an array to make clusters.
Divide the data into clusters using Fuzzy C-Means and Kernelized Fuzzy C Means algorithm. Gaussian term is also used under KFCM so as to remove the fuzziness and to reduce the time complexity.
Find the web pages with highest grade of membership in each cluster.
Step4: Prediction
Assign weightage to each webpage according to grade of membership, page with highest Weightage has higher membership and page with low weightage has low membership.
Predicting user future webpage using Fuzzy C Means and Kernelized Fuzzy C-Means algorithms according to each user in particular session. The webpage which has more weight has more probability for opening in future by user.

12. SIMULATED RESULTS

In this section, the proposed system is evaluated via computer simulation using MATLAB simulator. Firstly the result for FCM algorithm are shown in figure 3 and figure 4. In this parameters namely as daily predict, 60 day rolling window and predict metrics are taken. Firstly the result between these parameters of each algorithm are compared and after that these two algorithms are compared with one another to find out which algorithm gives the best result.
The fig.4 and 5 shows the result for FCM algorithm, in this 2 of taken parameters ie predict metrics and 60 day rolling window have been seprated as a result of which the variance of 60 day rolling window and predict metrics is less as compared to daily predict.

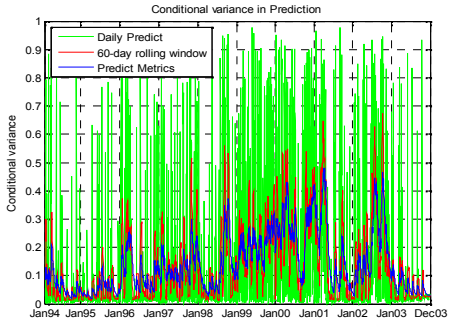

Figure 4- Daily Prediction under FCM Algorithm

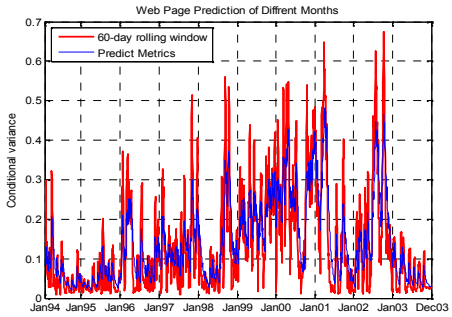

Figure 5 - 60 Day Prediction under FCM Algorithm

Similarly, fig 6 and fig 7 shows the result for KFCM algorithm. In this the same parameters are taken and comparison of their prediction result to one another is done.
After comparing the result of these parameters the variance of prediction of 60 day rolling window and predict metric is less as compared to the daily prediction made by the users. Under KFCM a new

term Gaussian which removes the fuzziness is introduced. The main motive of introducing the Gaussian is to reduce the time complexity during predicting thus named as GKFCM.

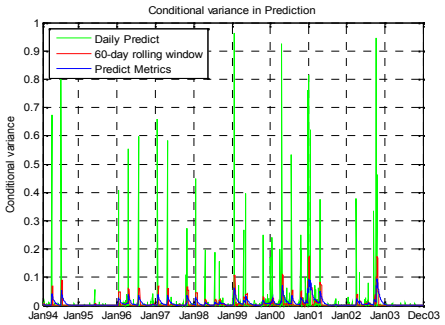

Figure 6- Daily Prediction Under KFCM Algorithm

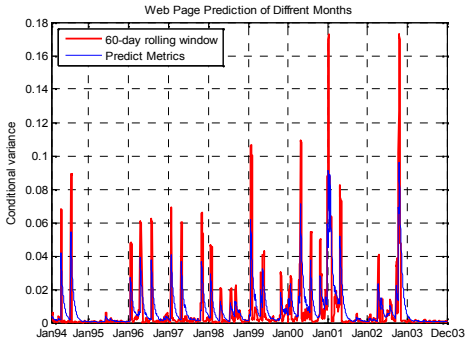

Figure 7- 60 Day Prediction Under KFCM Algorithm

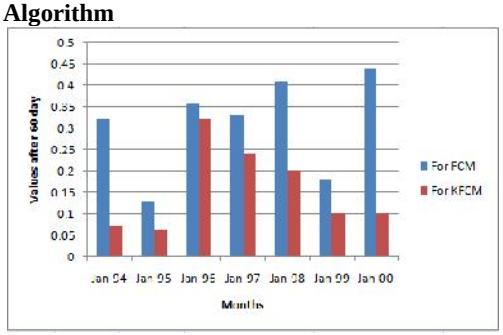

Figure 8 - Comparison Between FCM and KFCM

CONCLUSION

Web usage mining is application of data mining. Web Usage Mining is the automatic discovery of user access pattern from web servers. Web usage mining is extracting the information from web browser log file which access by users. Web usage mining is done using three main steps pre-processing, clustering. Firstly collect the web log file then pre-processing is implemented on web log file. In pre-processing module unformatted web log data is converted into formatted web log data. Fuzzy clustering is also known as soft clustering. In this data elements can belong to more than one cluster and associated with each element is a set of membership levels.

Fuzzy clustering methods are FUZZY C-means and KERNELIZED FUZZY C-means. Under KFCM Gaussian is used to remove fuzziness and to reduce the time complexity so naming it as GKFCM. The performance of web page prediction system concludes on the basis of parameter called conditional variance. Efficient web page prediction requires minimum conditional variance. From the above results calculation of the web page prediction value is recorded on the basis of daily predict value and 60 day rolling window value.

And we calculate these values by applying the FCM and KFCM algorithms. From the comparison we conclude that the KFCM gives better results as compare to FCM. In future, this proposed work is useful in prediction and can be applied on large data sets. This proposed work can be applied on various websites so as to estimate the effectiveness and performance too.

ACKNOWLEDGEMENT

I express my deep sense of gratitude to Sri Sai University, Palampur for giving an opportunity to prepare this paper. we would like to express a deep sense of gratitude and thanks to Department of Computer Science and Engineering, Sri Sai University, Palampur for generous guidance, help and useful suggestions. I express my thanks to my advisor, who guided me and provided with all the useful information for presenting this paper.

REFFERENCES

[1] Virendra R. Rathod and Govind V. Patel “Prediction of User Behavior using Web log in Web Usage Mining” International Journal of Computer Applications Volume 139 No.8, April 2016.

[2] E. A. Vimal and Dr. S. Chandramathi "Topic Trend Prediction for Learning Objects Using Kernel Fuzzy C Means Clustering and Autoregressive Integrated Moving Average Model" International Journal of Applied Engineering Research Volume 11, Number 4,2016.
[3] Dr.S. Vijiyarani and Ms. E. Suganya "RESEARCH ISSUES IN WEB MINING" International Journal of Computer-Aided Technologies Vol.2, No.3, July 2015.
[4] Megha P. Jarkad and Prof. Mansi, Bhonsle "Improved Web Prediction Algorithm Using Web Log Data" International Journal of Innovative Research in Computer and Communication Engineering Vol. 3, Issue 5, May 2015.
[5] P. Sampath and Prabhavathy M. "WEB PAGE ACCESS PREDICTION USING FUZZY CLUSTERING BY LOCAL APPROXIMATION MEMBERSHIPS (FLAME) ALGORITHM" VOL. 10, NO. 7, APRIL 2015.
[6] Parth Suthar and Prof. Bhavesh Oza "A Survey of Web Usage Mining Techniques" (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 6 No. 6, 2015.
[7] Anitha Talakokkula "A Survey on Web Usage Mining, Applications and Tools" Computer Engineering and Intelligent Systems Vol.6, No.2, 2015.
[8] B. Rosiline Jeetha "APPROACHES ON FUTURE REQUEST PREDICTION IN WEB USAGE MINING USING DATAMINING TECHNIQUES" VOL. 9, NO. 10, OCTOBER 2014
[9] Payal Gupta and Dr. Sukhvir Singh "Web Usage Mining: Optimize Web Site Navigation Through Log Files" International Journal of Advanced Research in Computer Science and Software Engineering Volume 4, Issue 8, August 2014.
[10] M. Vengateshwaran and E. V. R. M Kalaimani "Web Mining Research Direction and Open Source Tools" International Journal of Advanced Research in Computer Science and Software Engineering Volume 4, Issue 7, July 2014.
[11] D. Jayalatchumy and Dr. P.Thambidurai " Web Mining Research Issues and Future Directions – A Survey" IOSR Journal of Computer Engineering Volume 14, Issue 3 Sep. - Oct. 2013.
[12] Dilpreet Kaur and A.P. Sukhpreet Kaur "User Future Request Prediction Using KFCM in Web Usage Mining" International Journal of Advanced Research in Computer and Communication Engineering Vol. 2, Issue 8, August 2013.
[13] Dr. Sanjeev Dhawan and Mamta Lathwal " Study of Pre-processing Methods in Web Server Logs" ISSN: 2277 128X International Journal of Advanced Research in Computer Science and Software Engineering Volume 3, Issue 5, May 2013.
[14] Soumi Ghosh and Sanjay Kumar Dubey " Comparative Analysis of K-Means and Fuzzy C-Means Algorithms" (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 4, No.4, 2013.
[15] Yinghua Lu, TinghuaiMa, Changhong Yin, Xiaoyu Xie, Wei Tian andShuiMing Zhong "Implementation of the Fuzzy C-Means Clustering Algorithm in Meteorological Data" International Journal of Database Theory and Application, pp.1-18 Vol.6, No.6 2013.
[16] R.Suganya and R.Shanthi "Fuzzy C-Means Algorithm- A Review" International Journal of Scientific and Research Publications, Volume 2, Issue 11, November 2012.
[17] Pooja Mehtaa, Brinda Parekh, Kirit Modi, and Paresh Solanki "Web Personalization Using Web Mining: Concept and Research Issue" International Journal of Information and Education Technology, Vol. 2, No. 5, October 2012.
[18] Chhavi Rana "A Study of Web Usage Mining Research Tools" Int. J. Advanced Networking and Applications 1422 Volume:03 Issue:06 Pages:1422-1429 (2012).
[19] L.K. Joshila Grace, V.Maheswari, Dhinaharan Nagamalai "ANALYSIS OF

WEB LOGS AND WEB USER IN WEB MINING" International Journal of Network Security & Its Applications (IJNSA), Vol.3, No.1, January 2011.
[20] QINGYU ZHANG and RICHARD S. SEGALL "WEB MINING: A SURVEY OF CURRENT RESEARCH, TECHNIQUES, AND SOFTWARE" International Journal of Information Technology & Decision Making Vol. 7, No. 4 (2008).
[21] Rachit Adhvaryu "A Review Paper on Web Usage Mining and Pattern Discovery" JOURNAL OF INFORMATION, KNOWLEDGE AND RESEARCH IN COMPUTER ENGINEERING.
[22] Raulji Jitendrasinh G "A Review on Fuzzy C-Mean Clustering Algorithm" International Journal of Modern Trends in Engineering and Research