ISSN No:-2456-2165

Performance Analysis of User Behavior Across a Web for User Location Tracking

N . Ulaganathan
Ph.D. (Part-Time) Research Scholar
Department of Computer Science
Nandha Arts and Science College, Erode, Tamil Nadu, India

Dr. S. Prasath
Research Supervisor
Department of Computer Science
Nandha Arts and Science College, Erode Tamil Nadu, India

Abstract:- Dominance Fuzzy Clustering and Distributed Probability Graph (DFC-DPG) framework is introduced with the goal of attaining effective web data usage analysis by achieving higher clustering efficiency with less latency. At the beginning process of proposed DFC-DPG framework, the web user information collection phase is deployed to collect the information of all users from weblog database by using server log files. Following this, the Dominance Rank model is presented for dividing the relevant and irrelevant data with respect to the web user by the consideration of Spearman rank correlation between the data of the web user. In the next part Map Reduce Pearson Correlation Fisher's Linear Discriminant Classifier (MPC-FLDC) technique is developed to provide better results on the web traffic pattern mining by enhancing classification efficiency with the reduction of prediction time. MPC-FLDC technique, the preprocessing is carried out with the help of Map Reduce framework to group the web patterns from weblog database into different sessions depending on the access time. In the final part, proposed Poisson Fragment Frequency based Web Pattern Clustering (PFF-WPC) technique is introduced with the purpose of tracking the web user location through effective performance of web traffic pattern mining with improved clustering efficiency with less time complexity. In the beginning of proposed PFF-WPC technique, the Poisson fragment process is carried out for session identification in order to accumulate the web patterns at diverse sessions depending on the access time. Through the use of Apache log samples dataset in the experimental evaluation, web traffic patterns are effectively mined with the goal of tracking the location of web user. The proposed techniques are compared with existing methods while conducting the experiments. The experimental results, it is clearly identified that proposed techniques are successfully mining the web traffic patterns by performing clustering or classification process in a significant manner. The performance of proposed techniques are verified through the metrics such as true positive rate, prediction time, space complexity, accuracy level and computational complexity. Comparatively, the proposed PFF-WPC technique provides effective web traffic pattern mining for web user tracking and attains enhanced results in the above mentioned metrics than the other proposed and existing methods.

Keywords:- *Web Mining, Web Tracking, FLDCMPC, PFF, WPC.*

I. INTRODUCTION

In general, web mining is the process of detecting and extracting the data information from documents stored on web using data mining techniques. In web mining, web usage mining is one of the classes to extract the exciting usage patterns of user from web. For extracting the patterns, the web analytics is carried out by collecting and examining the web data from database with the aim of amending the web usage. Through web analytics, the behavior of user is examined because it offers the information of web user such as the number of user.

Normally, the web user searches for a specific term over the internet by giving the request to the web. After getting the request from the user, the web finds and extracts the particular web page on website and the user clicks (visited) on extracted site and spends a few seconds on the homepage. Further, the user browses few other web pages and spends varied amounts of time on each web page. After obtaining the search results, the user logs out from the website. The number of web pages and kinds of web pages visited by every user are stored on weblog database in a sequential manner. The storing of this browsed information of each web user helps to effectively perform the user behavioral analysis on web.

The web user behavior analysis is performed through collecting and examining several activities of every user on web. Through the successful web user behavior analysis, the location of web user is identified in a significant manner. The achieving of an effective web user behavior analysis is a challenging task due to vast development of web as well as the web user counts. The activities of users on web are successfully analyzed with the implementation of clustering and classification technique in the web traffic pattern mining. Many research works are developed to mine the web traffic patterns from weblog database. The previous developed existing techniques are not sufficient to track the location of web user while performing the behavior analysis of web user.

Let us consider the web with huge amount of data which is utilized for more number of web users in order to access the required information through the internet. The web users are allowed to access the web pages by sending the request to the web. The web provides the required web page related to the request information from the user. The

ISSN No:-2456-2165

user stays on that obtained web page for few minutes and move to several numbers of web pages and the information about the accessed web page with time duration of all users for each web page is stored on weblog database. With the stored information on weblog database, the behaviors of web users are examined by mining the web traffic patterns for finding the location of particular web user. Fig. 1. shows the overview of the web user behavioral analysis for web user location tracking.

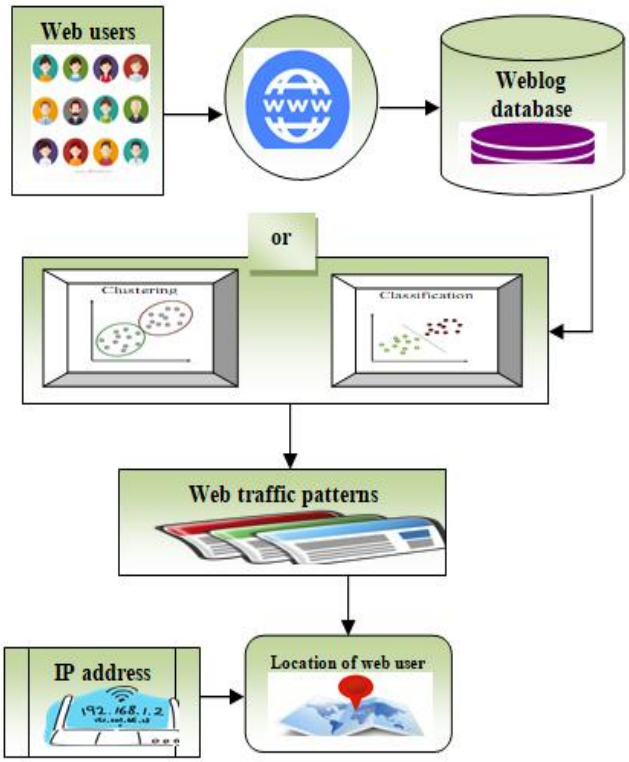

Fig. 1:- Web User Behavior Analyzation Based Web User Location Tracking

II. RELATED WORKS

Anandhi et al. [1] described clustering methods for identifying patterns namely path detection, page clustering, fuzzy clustering, ant-depended clustering and graph portioning. A comparative investigation of these clustering methods was performed by identifying the upcoming visit of probable user in web server log. Fuzzy clustering protocol provided higher accuracy in user navigation pattern identification when compared to other clustering approaches and latency in pattern discovery increased.

Binu Thomas et al. [2] developed web classification protocol. The basic fuzzy association rule mining was used to categorize the web pages into various web types based on their performance in user sessions. The outcomes are characterized in type of classification rules. These rules are compared with outcomes of Boolean Apriori association rule mining method. However, the classification accuracy was not enhanced to desired level.

Cheng Fang et al. [3] discussed a stream algorithm for recognizing user click requests. The user-browser interactions are reconstructed through influence of Spark Streaming framework. A massive real HTTP traffic records is occupied in verifying the stream algorithm obtained from a cellular core network by means of high-performance observing devices. A statistical study is carried out on the reconstructed data set for obtaining the entire features of mobile web traffic. The essential mobile web traffic model enhancement is progressed and main factors disturbing web performance are identified. Though these web traffic models are utilized for mobile network operators to recognize the mobile web traffic and subscriber actions the web traffic pattern prediction time did not get effectively minimized.

Doddegowda et al. [4] developed web personalization algorithm through a combination of data from user profiles and behavioral patterns. A number of dynamic behavioral patterns and user profiles together with discovery period were taken as input. Similarity between user profiles and behavioral patterns had been estimated for identifying the pages. The important behavioral patterns and user profiles were chosen for evaluating the rank for each page. The top n-pages with maximum rank were suggested. The important data for discovery and modified web pages could be achieved for Web user with minimum latency and traffic even though a search scheme is not included for better behavior discovery process.

Gajendra Singh Chandel et al. [5] explained Fuzzy C-Mean (FCM) Clustering protocol. FCM protocol was designed with Clustering Technique in Web Usage Mining to identify the user patterns. The transformations changed the data storage in Web Servers Log files by FCM Algorithm but high dimensionality of user session was not addressed.

Geetharamani et al. [6] proposed Apriori Prefix tree (PT) algorithm for finding the frequent usage of web pages to handle their performance. Apriori algorithm is used with Boolean association rules with frequent itemsets extraction. The rules produced from prefix tree algorithm are based on support, confidence and lift evaluation measures. The huge amount of rules is filtered regarding the minimum support threshold. The support and lift measures and the final findings about the frequent items are explored. Apriori prefix tree algorithm increased the mining efficiency but increased the running time of rule generation.

Giorgos Kollias et al. [7] designed the multidamping for ranking. An algorithmic reformulation occupied the functional rankings like LinearRank, Total Rank and Generalized Hyperbolic Rank. The rankings are estimated with finite series demonstrations. The polynomials of stochastic matrices are denoted as products of Google matrices. Individual matrices are parameterized through many damping factors. Multidamping included many features like locating highest ranked pages and multidamping with inexact keys. An instinctive analysis of functional rankings has been described with the surfing patterns of web users and the latency is increased.

Guosheng Kang et al. [8] developed web service ranking approach. The ranking approach depended on Collaborative Filtering (CF) with the help of user behavior. The query history is used for gathering the probable user behavior. Compared invocations and queries between users are utilized in estimating CF-depended user similarity. The characteristics of web services namely functional significance, score depended on CF and QoS utility have enabled Web service ranking through aggregation method depended on rank positions but the computational complexity increased in Web service ranking approach.

Jianping Zeng et al. [9] established an integration framework for the examination of user activity on an interactive website. A hidden Markov model (HMM) was considered in characterizing user activity model and user interest evaluation approach was provided. User activity investigation termed as user group detection is also performed. The integration framework utilized user activity investigation and user interest on an interactive website for a well-organized measure in examining user actions. Traffic pattern prediction rate is minimized in integration framework.

Khanchana et al. [10] discussed Fuzzy Possibilistic algorithm for clustering Web Usage Mining system. Hybrid Extreme Learning Machine (HELM) was developed to categorize user navigation patterns to identify users requests depending on clustering of users browsing behavior data. Enhanced classification accuracy was achieved when compared to other conventional approaches but HELM failed in achieving increased traffic pattern prediction rate.

Mamoun et al. [11] considered Markov model and all-Kth Markov model in Web prediction. A modified Markov model is originated for overcoming the problems of scalability and a two-tier prediction framework is implemented for producing classifier that depended on training examples. The prediction time is increased and number of paths is decreased with enhanced accuracy. The space complexity is minimized by using two-tier prediction framework and cache utilization did not get minimized.

Manohar et al. [12] established summarization and aggregation method developed by web log, web ranking, web rating and web review approaches for recognizing success rate of different web pages. The value was summarized to find the exact success rate of every web page. The success rate is normalized and aggregated into three forms to personalize the web user. Personalization of web user determined clustering significant web access actions with decreased calculation complexity. However, the identification of user satisfaction could not be enhanced.

Manoj Swami et al. [13] designed web personalization with web usage mining. Web is a large storehouse of information in internet technology. The inability of humans in recognizing the content was resolved by presenting personalized web services for personalized web experience to users. The essential technique in presenting the web personalization is Web Usage Mining. Personalization process employed the Web personalization for tailored recommendations to user. The exact prediction of user behavior was not achieved in web personalization.

Mohammed Asad et al. [14] explained various clustering methods of web usage mining process. Clustering methods namely k-Means, k-Medoids, Leader and DBSCAN has been analyzed for grouping web usage but the accuracy of user navigation pattern identification decreased.

Nazli Mohd Khairudin et al. [15] addressed the temporal attribute problem in relational rule mining. The impact of diverse temporal attributes is examined by integrating time characteristics during rule mining process. Apriori and FP-Growth algorithms are used to compare the rules. Integration of time with temporal attribute decreased the set of rules generated. But, memory consumption is high by using temporal attribute.

Orit Raphaeli et al. [16] analyzed online consumer behavior of an e-retailer website with aid of web usage mining method. Online consumer behavior is represented by engagement evaluations and identification of navigation patterns. This representation was carried out by integration of footstep graph visualization with sequential association rule mining. The session in mobile devices is task-oriented actions and the sessions in PC devices are exploration-oriented browsing actions. Specific sequence rules are linked with an improved likelihood of attainment in both mobile and PC sessions. The tracking of defenses was not included during operational caches.

Padmaja et al. [17] designed to improve K-means clustering algorithm to discover internet user behavior. Web data incorporated the conversion and recognition of web log data to acquire the information, patterns and knowledge. An enhanced log data investigation was used for finding internet user behavior. Clustering efficiency remains did not get addressed in K-means clustering algorithm.

Rajesh et al. [18] presented web personalization techniques. A personalization process and web personalization techniques were designed for executing the personalization systems. A number of techniques were in place to emphasize the prominent features for presenting the web personalization. The identification of traffic patterns was not performed for Web Personalization techniques.

Rahul Mishra et al. [19] analyzed Apriori and frequent pattern tree algorithm. Apriori is a general algorithm selected for extracting frequent patterns from transaction database. Apriori algorithm is a huge item set property and simple in construction with repeated scanning. Frequent pattern tree (FP-tree) is occupied for accumulating compressed and essential frequent patterns to extract the repeated patterns in huge databases. FP algorithm utilized divide and conquer method and consumed minimum time. The identification of frequent patterns is not performed efficiently.

RaviBabu et al. [20] proposed field mining and data cleaning approaches for improving the learning abilities with minimum computation intensity. In this, K means algorithm is occupied in cleaning log file for providing effective clustered outcomes by eliminating irrelevant items and unsuccessful requests for web investigation. Multi-layered network architecture was developed with back propagation learning method for identifying and examining significant information from accessible Web log data. Yet, clustering efficiency is minimized by employing K means algorithm.

Shivaprasad et al. [21] developed Neuro-fuzzy based hybrid model for identifying hidden patterns in Web Log of polytechnic web site. Web Log Pre-processing methods depended on dimensionality reduction approaches has been utilized. The preprocessing stage restricted all unrelated and noisy data to produce Web Log size. Then, neural networks and the fuzzy set theory are integrated for producing patterns and rules repeatedly. The neuro-fuzzy clustering collected the users with identical browsing patterns into clusters. The information was obtained through website manager for attaining a well-organized administration and personalization. Still, space complexity is higher in Neuro-fuzzy based hybrid model.

Tomasz Bujlow et al. [22] analyzed the web service techniques for identifying the users online, suggestions and probable user securities. Essential categories of approaches were depending on sessions, client storage, client cache and fingerprinting for user detection. The mechanisms utilizing web caches, operational caches and fingerprinting were also considered for diverse creative approaches. The process of predicting users on web with their real names, e-mail addresses, phone numbers or even street addresses was also performed. Each tracking approach was provided with defenses but cache utilization did not decrease to desired level.

Vedpriya Dongre et al. [23] described system architecture for considering the issues of prediction techniques. The data were obtained from proxy server and preprocessed to store the data in an access log database. K-means clustering algorithm was implemented to store the data for producing clusters i.e., web pages used by targeted user and other users. Subsequently, operated web pages weights were estimated using regression analysis algorithm with frequencies. Regression analysis became prediction technique utilized in identifying exact data over numeric values. The process of finding relevant user behavior was not performed by using log data investigations.

Vijaya Kumar et al. [24] have proposed the Clustering and Visualization of Web Usage Data by SOM and XML. Web Usage data by SOM and XML are designed with the help of application language. The sessions are framed by concept hierarchy and link information. The framed sessions are changed into Extensible Markup Language (XML) format and the clusters representing sessions with same patterns are attained through SOM. However, the computational complexity has not been minimized to desired level.

Xiaozhe Wang et al. [25] proposed a concurrent neuro-fuzzy model for predicting and examining essential information from accessible Web log data. The cluster data are produced through a self governing map for pattern investigation and a fuzzy inference system was utilized in extracting the chaotic trend for offering short-term and long-term (daily) Web traffic trend recognitions. An efficient mining and Web server traffic identification were attained by using concurrent neuro-fuzzy model. The process of pattern identification in concurrent neuro-fuzzy model is not sufficient for efficiency.

A Mountain Density Function (MDF)-based fuzzy clustering framework was proposed by Zahid Ansari et al. [26] for estimating the clusters in web log data. The key process comprised of web log preprocessing, Mountain Density Function (MDF)-depended discovery of fuzzy user session clusters and confirmation of clusters. For high dimensionality of user session data, fuzzy approach was implemented for allocating the weights to user sessions. Fuzzy C-Means (FCM) and fuzzy C-Medoids (FCMed) algorithms were used to group the user sessions. The choice of cluster centers is a demanding one in which the Mountain Density Function (MDF)-based Fuzzy C-Means (FCM) (MDFCM) and FCMed (MDFCMed) algorithms were designed. Classification accuracy became minimized due to fuzzy clustering framework.

Zhen Liao et al. [27] discussed the effectiveness of task trails in different search applications for finding user satisfaction, user search interests and signifying associated queries. The task trail provided enhanced performance in finding user satisfaction and improved web page necessities of end users. Task trails are sensitive in evaluating various ranking functions. Query terms from similar task became more stable to each other and Query suggestion provided on task trail became better for query proposals. The requirement of mining task trails from web search logs also performed to maximize applications in search and recommendation schemes. The allocation of executions is not considered for achieving accurate results in search applications.

Zheng Xu et al. [28] suggested personalized web search using semantic context. The technique collected user context to present accurate preferences of users in personalized search. The short-term query context was generated to identify related concepts of query. The user context was produced depended on click through the data of users. A forgetting factor was developed for combining the self-governing user context in user session to preserve the evolution of user preferences. Clustering and classification methods of web pages were not included to get accurate outcomes.

III. METHODOLOGY

A. Web Usage Mining

Web servers monitor the important information accessible on user–server interaction. The log data, otherwise termed as web user access or click stream data to search model predict user behavior. Web directories are the Website records that are arranged by human reviewers while search engine catalogs are combined through the automated systems. Web directories are the computerized process that has the personalized information of user. User communities are created by data collected from Web proxies while users browse the Web. Many hybrid representations are designed over time as search engines integrated directory features to address the problems like categorization and site quality. The key objective is to recognize the behavioral patterns in collected usage data and implement community Web directories depending on patterns. The method of collecting the patterns from data to Web directories is called Usage Data Preparation.

Usage Data Preparation includes the collection, cleaning of usage data and recognition of user sessions. Web Directory Initialization presents the characterization of Web pages in usage data with types of Web directory. Two types of techniques are designed for characterization of Web pages. The first technique categorizes the Web pages into Web directory through hierarchical document clustering and the second technique categorizes the Web directory called ODP. Community Web Directory Discovery is the key method for creating the user from data by machine learning methods and developed representations to plan the community Web directories. Web usage mining is the relevance of data mining procedures on web data to identify user access patterns for diverse applications. Web Usage Mining is an influential tool for analyzing, designing and modifying the organization of website and to recognize by investigating the site user behavior.

B. Map Reduce Pearson Correlation Fisher's Linear Discriminant Classifier Framework

The web usage mining approach was implemented to predict the online navigational behavior of web users but it failed to perform the effective prediction of web traffic patterns at the required level. However, a novel method was implemented with the objective of providing better results in the web usage pattern detection by the implementation of client-side logging. It failed to minimize the time consumption for detecting the web usage patterns. Hence, the proposed Map Reduce Pearson Correlation Fisher's Linear Discriminant Classifier (MPC-FLDC) technique is introduced with the objective of effectively predicting the web traffic patterns from weblog database with improved accuracy and less time. In the proposed MPC-FLDC technique, the frequent or the non frequent web patterns on weblog database are effectively classified with higher accuracy by using Fisher's Linear Discriminant (FLD) Classifier. Thus, the performance of Pearson Correlation Analysis effectively predicted the web traffic patterns with minimized time consumption.

Then, the proposed MPC-FLDC technique is carried out to analyze the web traffic pattern analysis within three phases such as preprocessing, Fisher's Linear Discriminant (FLD) Classifier and Pearson Correlation Analysis. During the web usage mining, the Sequence-based clustering was presented for estimating the elder self-care behavior patterns. It was not effective to provide better improvement in the performance of web usage mining. The proposed MPC-FLDC technique is carried out for achieving effective web traffic pattern mining. In the initial phase, the proposed MPC-FLDC technique takes the weblog database (i.e. Apache log samples dataset) which is considered as input. The preprocessing step is carried out for grouping the web patterns from weblog database according to various sessions and the Fisher's Linear Discriminant Classifier is introduced for classifying the web patterns as frequent or non-frequent which helps to achieve improved accuracy in the classification process. Finally, the Pearson Correlation Analysis is carried out for predicting the web traffic patterns with less amount of time in an effective manner. In the design of proposed MPC-FLDC technique, the preprocessing is performed as a beginning process. Depending on the diverse sessions (i.e. Time Interval), the web patterns from the weblog database is grouped by Map Reduce framework through the process of preprocessing. By using the Map Reduce framework, the web user activities are separated into a sequences (sessions) with the consideration of access time. The performance of Map Reduce framework is carried out within two phases such as Map and Reduce which are termed as divide-and-conquer program model. The Map Reduce framework separates the input into a list of key/value pairs. The performance of map and reduce task of data is carried out on all nodes which is stored in a local machine. In the proposed MPC-FLDC technique, the MapReduce framework is presented for segmenting the original web logs into a number of sessions and it depends on the access time.

In general, the access activities of web users are stored by web server on the weblog database. The weblog database has client IP address, time, requested URL, HTTP status code, referrer etc. By partitioning the web patterns based on the different sessions, the web pages are browsed by web users within a particular period of time. From this, the process of web traffic pattern mining is effectively carried out through the proposed MPC-FLDC technique.

C. Poisson Fragment Frequency Based Web Pattern Clustering (PFF-WPC) Technique

The flexible methodology is developed with aim of examining the effectiveness of various variables on diverse department at a particular time session. But, the tracking of web user location remained unaddressed. The Linear-Temporal Logic (LTL) model has been developed with the objective of analyzing the user behavior in e-commerce websites. It could not become a sufficient model to perform effective traffic pattern mining for web user tracking. The proposed Poisson Fragment Frequency based Web Pattern Clustering (PFF-WPC) technique is designed with the main goal of tracking the web user location by the successful performance of web traffic pattern mining.

The proposed PFF-WPC technique performs web user tracking with three processes such as session identification, clustering and location tracking. During the web traffic pattern analysis, the session identification and clustering process are conducted with the deployment of Poisson fragment process and frequency based web pattern clustering technique respectively. The Poisson fragment process helps to perform web user tracking in an effective way. Through the frequency based web pattern clustering technique, the clustering efficiency and the true positive rate had improvement in clustering the frequent and non-frequent web patterns. The location tracking is performed by using temporal similarity measurement with the reduction of computational and space complexity. The proposed PFF-WPC technique is performed for discovering the user location of web traffic patterns. In the initial step, the proposed PFF-WPC technique takes the weblog database (i.e. Apache log samples dataset) which is considered as input and the session's identification is carried out with the implementation of Poisson Fragment Process. In the next step, the clustering process is performed by using Frequency Based Web Patterns Clustering to group the web pages i.e. web patterns as frequent or non-frequent web pages with improved clustering efficiency. After clustering the web patterns, the web traffic patterns are discovered through the Temporal Similarity Measurement. Finally, location of user is effectively identified with the aid of public IP address. Following this, the proposed PFF-WPC technique performs the web user behavior analysis with higher clustering efficiency, true positive rate and less computational complexity.

D. Proposed Methodology

The web traffic pattern mining is the process of extracting the web pages i.e. web patterns which are browsed from the weblog database by the web users for more number of times. By the extraction of frequent web traffic patterns from weblog database, the web user behavior is analyzed in order to identify the location of web users. The proposed Dominance Fuzzy Clustering and Distributed Probability Graph (DFC-DPG) framework is introduced with the aim of performing successful clustering of web patterns with less latency for web data usage analysis. Also, the web data usage analysis is done in three phases such as web user information collection, dominance rank model, fuzzy clustering approach and Distributed Probability Graph Arc (DPG) model. In web user information collection, the user information is collected from the weblog database through the server log files. The information of web user is extracted from common log format and access log file which contains User (Client) IP address, User ID, Access date, Access time, status code and HTTP. Thus, the information of all users in web who visits the web pages on website is collected. With the collected information, the Dominance Rank model is carried out to separate the relevant and the irrelevant data to the web user. The relevant and the irrelevant data regarding the web user are identified by the estimation of spearman rank correlation between the data of the web user.

As long as the output result of spearman rank correlation is +1, the data is identified as relevant. Otherwise, the data are identified as irrelevant, if the output result of spearman rank correlation is -1. Further to this, the relevant data of the web user are extracted and the fuzzy clustering approach is performed on that data to cluster the similar user interest web pages from weblog database. The clustering process is conducted with higher clustering efficiency by measuring the fuzzy membership for each web page. Following this, the DPG model is deployed with heuristic probabilistic framework to minimize the latency and space complexity for analyzing the web user behavior through the web patterns and it extracts the web pages from web sites which are visited more number of times by forming the adjacent matrix from the directed graph. Thus, the proposed DFC-DPG framework successfully performs the web pattern mining through the effective performance of web user behavior analysis.

The proposed Map Reduce Pearson Correlation Fisher's Linear Discriminant Classifier (MPC-FLDC) technique has been introduced with the aim of improving the prediction with less time consumption for performing the web traffic pattern mining. So, the performance of web traffic pattern prediction in proposed MPC-FLDC technique is efficiently made by the implementation of Map Reduce framework based preprocessing, Fisher's Linear Discriminant (FLD) Classifier and Pearson Correlation Analysis. In the initial phase, the proposed MPC-FLDC technique performs preprocessing with the help of Map Reduce framework for grouping the web patterns from weblog database according to different sessions. Based on the access time, the user activities of web patterns are separated into sequences through two phases i.e., Map and Reduce of Map Reduce framework. After the completion of preprocessing, the proposed MPC-FLDC technique uses FLD Classifier to obtain the frequent patterns with improved classification efficiency. This leads the FLD classifier determine hit ratio and optimal projection direction for each web page to group the web pages visited by the users as frequent or non-frequent patterns regarding their sessions.

Using Pearson Correlation Analysis frequent patterns for effectively mining the web traffic pattern prediction (daily/hourly traffic) in weblog database are identified. Through the Pearson Correlation Analysis, Pearson correlation coefficient value for each frequent web page is computed in order to obtain the prediction of traffic web patterns within less amount of time. As a result, the proposed MPC-FLDC technique is able to mine the web traffic patterns for performing the prediction of web traffic patterns in an effective manner.

Finally, the proposed Poisson Fragment Frequency based Web Pattern Clustering (PFF-WPC) technique is implemented for performing effective web user tracking by achieving the successful web pattern traffic mining with higher clustering efficiency and less computational complexity. Therefore, the three processes such as Poisson fragment process, frequency based web patterns clustering

and temporal similarity measurement have been deployed with the proposed PFF-WPC technique for getting better results in web user tracking. Initially, the session identification is presented with the help of Poisson fragment process to collect the web pages for different session from the weblog database according to the access time and the frequency based web patterns clustering technique is performed for attaining the frequent web patterns with the improvement of clustering efficiency.

Through the performance of frequency based web patterns clustering technique, the web patterns are clustered by the measurement of frequency for all web pages from weblog database. Here, the web patterns with higher frequency are clustered as frequent and then the web patterns with lower frequency are clustered as non-frequent. Following this, the temporal similarity is determined from the attained frequent web patterns with the objective of extracting the web traffic patterns with higher accuracy. By the measurement of temporal similarity for each session of frequent web patterns, the web traffic pattern is detected which leads to identify the location of traffic web patterns by using IP address. Finally, the location of web user is extracted in an effective way by the implementation of proposed PFF-WPC technique in the web traffic pattern mining.

From the performance of three proposed techniques, it is clearly observed that the web traffic pattern is efficiently extracted by analyzing the behaviors (activities) of users on web. From the stored weblog database, the web pages which are frequently browsed by the users as frequent web patterns are successfully obtained by performing the clustering or classification on the web patterns from weblog database. To carry out this, there are three techniques such as fuzzy clustering approach, FLD classifier and frequency based web patterns clustering technique. Then, the separation of frequent or non-frequent patterns aids to track the location of web user by using IP address. Through the three proposed techniques, the mining process of obtaining web traffic patterns from weblog database is significantly achieved with tracking web user location by the effective analysis of user behaviors on web.

In order to analyze the performance of proposed DFC-DPG framework, MPC-FLDC technique and PFF-WPC technique are deployed in Java language using Apache log samples dataset. For the purpose of extracting the web traffic patterns, the Apache log samples dataset is used to analyze the behaviors of user such as IP address, Date, Time of Access, Port Number and accessed Webpage on web. During the experiment, three proposed techniques are compared with the existing Web usage mining approach and Web service ranking approach methods. While carrying out the experiment, the number of web patterns is considered from 30 to 300 range which is taken as input. Through the three proposed techniques, the mining of web traffic pattern is effectively performed out with higher accuracy level with less prediction time for tracking the web user location with minimized computational complexity.

IV. PERFORMANCE EVALUATION

An effective Clustering framework is implemented in Java language using Apache log samples dataset. The Apache log samples datasets identifies the access activities of several web users namely IP address, Date, Time of Access, Port Number and accessed Web page. The tables and the graphs generated depend on the performance values obtained from experiments to assure the effectiveness of the proposed technique.

❖ *Performance Analysis for DFC-DPG Framework, MPC- FLDC Technique and PFF-WPC Technique*

The performance of proposed Dominance Fuzzy Clustering and Distributed Probability Graph (DFC-DPG) framework, Map Reduce Pearson Correlation Fisher's Linear Discriminant Classifier (MPC-FLDC) technique and Poisson Fragment Frequency based Web Pattern Clustering (PFF-WPC) technique are implemented in Java language using Apache log samples dataset. During the experiment, the effectiveness of proposed techniques are verified by comparing existing Web usage mining approach and Web service ranking approach. The performance of proposed techniques is evaluated with the aid of parameters such as true positive rate, prediction time, space complexity, accuracy level and computational complexity. The complete result analysis of these metrics is elaborately described in further section.

A. Performance Analysis of True Positive Rate

The true positive rate is measured as the ratio of number of web pages i.e. web patterns regarding the web user are correctly identified as frequent web traffic patterns to the total number of web patterns.

The true positive rate is measured in terms of percentage (%). When the true positive rate is high, then the technique is said to be more efficient.

ISSN No:-2456-2165

Number of web patterns	True Positive Rate (%)				
	Web usage mining approach	Web service ranking approach	DFC-DPG framework	MPC-FLDC technique	Proposed PFF-WPC technique
30	55	50	60	69	71
60	62	58	68	72	75
90	63	59	69	76	79
120	65	60	70	80	83
150	71	66	76	82	85
180	75	70	80	86	88
210	81	76	86	89	91
240	82	77	87	91	92
270	84	79	89	93	95
300	85	80	90	95	97

Table 1:- Tabulation for True Positive Rate

From the above Table 1 illustrates the comparative result analysis of true positive rate with respect to the web patterns. The comparison analysis is carried out by using proposed DFC-DPG framework, MPC-FLDC technique and PFF-WPC technique with the existing methods. From the experiment, the number of web patterns is considered in the range of 30 to 300 which is taken as input.

As shown in the above table, the performance analysis of true positive rate in the proposed and existing methods significantly improved during mining the web traffic patterns from weblog database. Comparatively, the proposed PFF-WPC technique has provided better results in the improvement of true positive rate than the other proposed and existing methods. The graph is plotted in Fig. 2 by utilizing the table values from table 1.

In the Fig. 2 shows the measure of true positive rate for three proposed and existing methods. The experiment is done by comparing DFC-DPG framework, MPC-FLDC technique and PFF-WPC technique with the state-of-the-art methods.

As illustrated in Fig. 2, it clearly described that the proposed PFF-WPC technique improved true positive rate for web traffic pattern mining when compared to other available methods. This is due to the performance of frequency based web patterns clustering technique in the proposed PFF-WPC technique. The frequency based web patterns clustering technique computed the frequency for all web pages in each session.

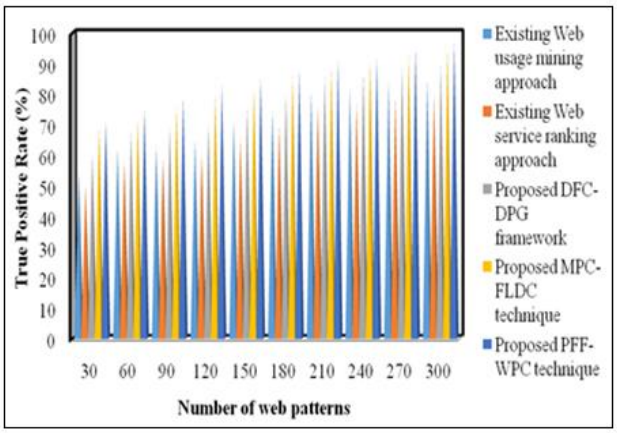

Fig. 2:- Measure of True Positive Rate

According to the values of estimated frequencies, the web pages in different session are clustered in an effective manner. The web pages with higher frequency are clustered as frequent as well as the web pages with lower frequency are clustered as non-frequent. Thus, the proposed PFF-WPC technique accurately identified the frequent web patterns from weblog database with the help of frequency based web patterns clustering technique and enhanced the true positive rate up to 24% when it is compared to existing methods.

Similarly, the proposed DFC-DPG framework and proposed MPC-FLDC technique performs the fuzzy clustering approach and Fisher's Linear Discriminant (FLD) Classifier respectively for obtaining the frequent web patterns from weblog database and the proposed DFC-DPG framework and MPC-FLDC technique improved the true positive rate up to 11% and 20% when compared to existing methods respectively. The experimental results show the proposed PFF-WPC technique provided higher true positive rate than the other two proposed techniques.

B. Performance Analysis of Prediction Time

The prediction time is defined as the amount of time taken to detect the web traffic patterns on weblog database. The prediction time is measured in terms of milliseconds (ms). During web traffic pattern mining, the reduction of prediction time ensures better results for the technique.

ISSN No:-2456-2165

Number of web patterns	Prediction Time (ms)				
	Web usage mining approach	Web service ranking approach	DFC-DPG frame work	MPC-FLDC technique	Proposed PFF-WPC technique
30	15	18	11	9	7
60	31	34	17	15	13
90	32	36	28	26	24
120	44	48	30	28	26
150	50	54	36	34	32
180	51	55	47	45	43
210	52	56	48	46	44
240	56	60	52	50	48
270	57	61	53	51	49
300	59	65	55	53	51

Table 2:- Tabulation for Prediction Time

The above table 2 illustrates the comparative result analysis of prediction time with respect to the web patterns. The comparison analysis has been carried out by using proposed DFC-DPG framework, MPC-FLDC technique and PFF-WPC technique with the state-of-the-art-methods. Web patterns in the range of 30 to 300 are taken as input during the experiment. As shown in above table, the performance analysis of prediction time in the proposed and existing methods are gradually got minimized while mining the web traffic patterns from weblog database. While comparison, the proposed PFF-WPC technique took less time to predict the web traffic patterns than the other proposed and existing methods. The graph (Fig. 3) is obtained by using the table values from the table 2.

From the Fig. 3 gives the measurement of prediction time for three proposed techniques and existing methods. The experiment is conducted by comparing the proposed DFC-DPG framework, MPC-FLDC technique and PFF-WPC technique with the other available methods.

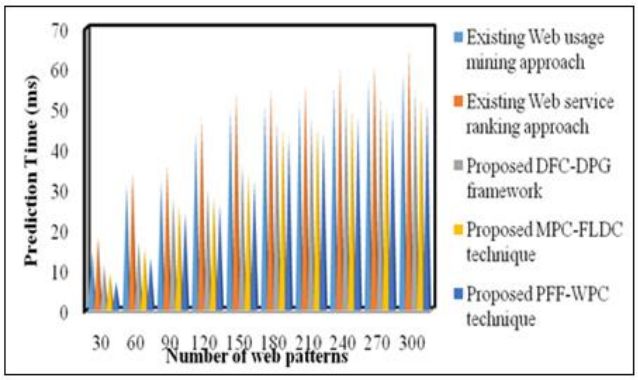

Fig. 3:- Measure of Prediction Time

In the Fig. 3 clearly explained that the proposed PFF-WPC technique has taken less time to predict the web traffic patterns when it is compared with other proposed and existing methods. This is due to the introduction of temporal similarity in the proposed PFF-WPC technique. The Temporal Similarity measurement is determined for all obtained frequent patterns at each session. When the value of temporal similarity measurement of web pattern lies between 0 and +1, the pattern is identified as traffic web pattern. Otherwise, the pattern is identified as non traffic web pattern if the value of temporal similarity measurement of web pattern is between -1.00 and 0. Thus, the prediction of web traffic patterns is achieved with less amount of time according to the temporal similarity value in the proposed PFF-WPC technique. Therefore, the proposed PFF-WPC technique reduced the prediction time by 32% than the existing methods.

Similarly, the proposed DFC-DPG framework and MPC-FLDC technique consumed less time to predict the web patterns with the introduction of Distributed Probability Graph Arc (DPG) model and Pearson Correlation Analysis. Thus, the proposed DFC-DPG framework and MPC-FLDC technique minimized the prediction time by 22% and 27% which are better than the existing methods respectively. As a result, during the web user behavior analysis, the proposed PFF-WPC technique consumed minimized time to predict the web traffic patterns than the other proposed techniques.

C. Performance Analysis of Space Complexity

The space complexity is the amount of space consumed for storing the web traffic patterns on weblog database. The space complexity is measured in terms of Mega Bytes (MB). It is measured as the difference between the total memory space and the unused memory space on weblog database during the web pattern mining. As long as the space complexity is less, the technique is said to be more efficient.

ISSN No:-2456-2165

Number of web patterns	Space Complexity (MB)				
	Web usage mining approach	Web service ranking approach	DFC-DPG framework	MPC-FLDC technique	Proposed PFF-WPC technique
30	20	22	16	14	11
60	25	27	21	19	16
90	26	28	22	20	17
120	27	29	23	21	18
150	32	34	28	26	23
180	33	35	29	27	24
210	34	36	30	28	25
240	35	37	31	29	26
270	36	38	32	30	27
300	37	39	33	31	28

Table 3:- Tabulation for Space Complexity

In the above Table 3 has the comparative result analysis of space complexity with respect to the web patterns. The comparison analysis is completed using proposed DFC-DPG framework, MPC-FLDC technique and PFF-WPC technique with the existing methods. For the experiment, the number of web patterns considered is in the range of 30 to 300(input).

From the above table, the performance analysis of space complexity in the proposed and existing methods got gradually minimized when mining the web traffic patterns from weblog database. From the result, the proposed PFF-WPC technique required less space complexity than the other proposed and existing methods. The graph is plotted in Fig. 4 and is obtained using the table values from table 3.

In the Fig. 4.4 describes the measurement of space complexity for three proposed techniques and existing methods. The experiment has been done by comparing the proposed DFC-DPG framework, MPC-FLDC technique and PFF-WPC technique with other available methods.

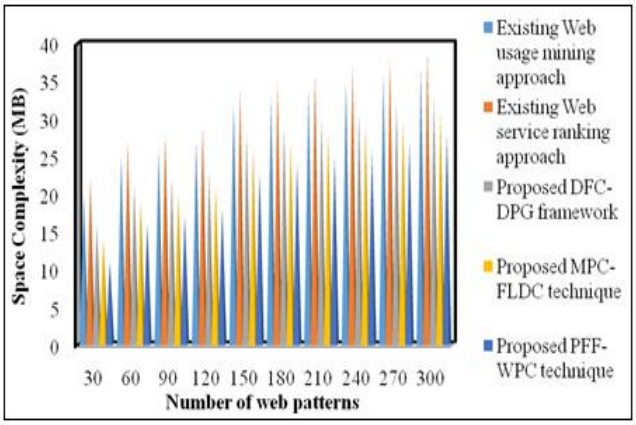

Fig. 4:- Measure of Space Complexity

From the above Fig. 4 clearly describes that the proposed PFF-WPC technique consumed less memory space to store the web traffic patterns when comparison is made with other proposed and existing methods. This is because of the development of frequency based web patterns clustering technique and temporal similarity in the proposed PFF-WPC technique. The frequency based web patterns clustering technique determined the frequency for all web pages in different sessions. According to the frequency, the web pages are clustered as frequent or non-frequent patterns which reduce the space complexity to store web patterns. Further to this, the temporal similarity is measured on frequent patterns to extract the web traffic patterns. Thereby, the PFF-WPC technique required less memory space for storage purposes during the web traffic pattern mining. Hence, the proposed PFF-WPC technique minimized the space complexity by 33% after comparing with other existing methods.

Similarly, the DFC-DPG framework and MPC-FLDC technique required less memory space to store the web traffic patterns by using Distributed Probability Graph Arc (DPG) model and Fisher's Linear Discriminant (FLD) Classifier. Thus, the proposed DFC-DPG framework and MPC-FLDC technique reduced space complexity by 16% and 23% than the other existing methods. As shown in comparison results, the proposed PFF-WPC technique comparatively minimized the space complexity.

D. Performance Analysis of Computational Complexity

The Computational Complexity is measured as the amount of time consumed for extracting the location of the web user of web traffic patterns with respect the number of web patterns. The computational complexity is measured in terms of milliseconds (ms). When the computational complexity is less, then the technique is said to be efficient.

Number of web patterns	Computational Complexity (ms)				
	Existing Web usage mining approach	Existing Web service ranking approach	Proposed DFC-DPG framework	Proposed MPC-FLDC technique	Proposed PFF-WPC technique
30	17	20	15	13	11
60	20	23	17	15	13
90	24	27	20	18	16
120	28	31	24	22	20
150	33	36	29	27	25
180	35	38	32	30	28
210	41	44	39	37	35
240	46	49	43	41	39
270	52	55	50	48	46
300	60	63	56	54	52

Table 4:- Tabulation for Computational Complexity

In the Table 4 brings out the comparative result analysis of computational complexity with respect to the web patterns. The result of the proposed DFC-DPG framework, MPC-FLDC technique and PFF-WPC technique are compared with the existing methods. In the experiment, web patterns in the range of 30 to 300 inputs are considered. It is observed from the result that the performance analysis of computational complexity in the proposed and existing methods is gradually reduced while tracking the location of web users. The result is that the proposed PFF-WPC technique minimized the computational complexity than the other proposed and existing methods. The graph is plotted in Fig. 5 from the table values of 4.

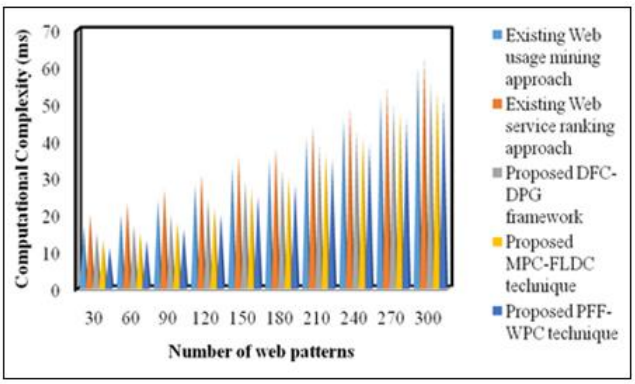

Fig. 5:- Measure of Computational Complexity

In the Fig. 5 shows the measurement of computational complexity for three proposed techniques and existing methods. After the experiment, the results of the proposed techniques such as DFC-DPG framework, MPC-FLDC technique and PFF-WPC technique are compared with the other available methods.

From the Fig. 5, it is clearly noticed that the proposed PFF-WPC technique had less computational complexity from the other proposed and existing methods. This is due to the fact that the proposed PFF-WPC technique is tracking the location of web user by effectively performing the web traffic pattern mining from weblog database. The clustering process performs frequency based web pattern clustering technique through the frequency measurement of all web pages in diverse sessions. So, the web pattern is clustered as frequent web pattern with less amount time. The temporal similarity on the frequent web pattern helps to extract the web traffic patterns. With the identified traffic pattern, the location of web user is extracted with less time duration by using IP address. Hence, the proposed PFF-WPC technique had minimized the computational complexity by 26%.

Similarly, the performance of Distributed Probability Graph Arc (DPG) model and Pearson Correlation Analysis in the proposed DFC-DPG framework and MPC-FLDC technique respectively minimized the computational complexity of web traffic pattern mining. Observation shows that the proposed DFC-DPG framework and MPC-FLDC technique could effectively reduce the computational complexity by 20% and 14%. Hence, the result is that the computational complexity is effectively reduced in the proposed PFF-WPC technique than the other two proposed techniques.

V. CONCLUSION

In the performance evaluation, the proposed Dominance Fuzzy Clustering and Distributed Probability Graph (DFC-DPG) framework, Map Reduce Pearson Correlation Fisher's Linear Discriminant Classifier (MPC-FLDC) technique and Poisson Fragment Frequency based Web Pattern Clustering (PFF-WPC) technique are compared with existing methods. As described in this research, three proposed techniques are utilized with effective clustering and classification process with the goal of attaining web traffic pattern mining in a significant manner.

In the first part, proposed Dominance Fuzzy Clustering and Distributed Probability Graph (DFC-DPG) framework is introduced for web user behavior mining in

an effective manner. In proposed DFC-DPG framework, the user information is collected from the weblog database through the web user information collection phase. By performing Dominance Rank model in proposed DFC-DPG framework, the relevant and the irrelevant data regarding the web user are separated based on spearman rank correlation and then the irrelevant data is removed. The proposed DFC-DPG framework performs the fuzzy clustering approach to cluster the similar user interest web pages. With the development of Distributed Probability Graph Arc (DPG) model, the web patterns are mined with less latency and space complexity.

The second part is concentrated on the web traffic pattern mining with the introduction of proposed Map Reduce Pearson Correlation Fisher's Linear Discriminant Classifier (MPC-FLDC) technique. The preprocessing is completed through the Map Reduce framework to group the web patterns at different sessions according to the access time. Through the Fisher's Linear Discriminant (FLD) Classifier, the frequent patterns or non-frequent patterns are classified based on the hit ratio. Then, the Pearson Correlation Analysis is employed on frequent web patterns to get the prediction of traffic web patterns with less time.

In third part, the proposed Poisson Fragment Frequency based Web Pattern Clustering (PFF-WPC) technique is introduced for achieving effective performance of web user tracking. With the performance of Poisson fragment process, the session identification is deployed for collecting the web pages at different session based on access time. By performing the frequency based web patterns clustering technique in proposed PFF-WPC technique, the frequent web patterns are obtained by the consideration of frequency of web pages and the temporal similarity is computed for all frequent web patterns to obtain web traffic patterns for identifying the web user location with the aid of IP address.

From the comparison and result analysis, it is clearly known that the third work i.e., the proposed PFF-WPC technique could achieve better performance during web user tracking by effective prediction of web traffic pattern mining. This is achieved with the improvement in parameters such as true positive rate, prediction time, space complexity, accuracy level and computational complexity than the other proposed and existing methods.

REFERENCES

[1]. D. Anandhi and M. S. Irfan Ahmed, “Prediction of user’s type and navigation pattern using clustering and classification algorithms", Cluster Computing, Springer, Pp.No.1-10, 2017.
[2]. Binu Thomas and G. Raju, “A Novel Web Classification Algorithm Using Fuzzy Weighted Association Rules”, Hindawi Publishing Corporation, ISRN Artificial Intelligence, Vol. 2013, Pp. No.1-10, 2013.
[3]. Cheng Fang, Jun Liu and Zhenming Lei, “Fine-Grained HTTP Web Traffic Analysis Based on Large-ScaleMobile Datasets”, IEEE Access, Vol. 4, Pp. No. 4364 – 4373, 2016.
[4]. Doddegowda B J, Sunil Kumar S Manvi and G T Raju, “A Novel Algorithm for Web Personalizationthrough Integration of Web User Profiles and Behavioral Patterns”, International Journal of Computer Science and Information Technology & Security, Vol. 7, Iss. No:2, Pp. No.4-13, 2017.
[5]. Gajendra Singh Chandel, Kailash Patidar and Man Singh Mali, “A Result Evolution Approach for Web usage mining using Fuzzy C-Mean Clustering Algorithm”, IJCSNS International Journal of Computer Science and Network Security, Vol. 16, Iss. No:1, Pp. No.135-140, 2016.
[6]. R. Geetharamani, P Revathy and Shomona G Jacob, “Prediction of users webpage access behaviour using association rule mining”, Indian Academy of Sciences, Vol. 40, Iss No:8, Pp. No. 2353–2365, 2015.
[7]. Giorgos Kollias, Efstratios Gallopoulos and Ananth Grama, “Surfing the Network for Ranking by Multidamping”, IEEE Transactions on Knowledge and Data Engineering, Vol. 26, Iss. No:9, Pp. No. 2323-2336, 2014.
[8]. Guosheng Kang , Jianxun Liu, Mingdong Tang , Buqing Cao and Yu Xu, “An Effective Web Service Ranking Method via Exploring User Behavior”, IEEE Transactions on Network and Service Management , Vol. 12, Iss. No:4, Pp. No. 554 – 564, 2015.
[9]. Jianping Zeng, Shiyong Zhang and Chengrong Wu, “A framework for WWW user activity analysis based on user interest”, Knowledge-Based Systems, Elsevier, Vol. 21, Pp. No. 905–910, 2008.
[10]. R. Khanchana and Dr. M. Punithavalli, “A Web Usage Mining Approach Based On New Technique in Web Path Recommendation Systems”, International Journal of Engineering Research & Technology (IJERT), Vol. 2, Iss. No:1, Pp. No.1-6, 2013.
[11]. Mamoun A. Awad and Issa Khalil, “Prediction of User's Web-Browsing Behavior: Application of Markov Model”, IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), Vol. 42, Iss. No:4, Pp. No. 1131 – 1142, 2012.
[12]. Manoj Swami and Prof. Manasi Kulkarni, “Understanding Web personalization with Usage Mining and its Application: Recommender System”, International Journal of Emerging Technology and Advanced Engineering, Vol. 3, Iss. No:5, Pp. No. 726-730, 2013.
[13]. E. Manohar and D. Shalini Punithavathani, “Hybrid Data Aggregation Technique to Categorize the Web Users to Discover Knowledge about the Web Users”, Wireless Personal Communications, Springer, Pp. No.1-15, 2017.
[14]. Mohammed Asad and Girish P. Potdar, “A Survey on Different Clustering Techniques for Web Usage Mining”, International Journal of Computer Science

and Information Technology & Security, Vol. 6, Iss. No:2, Pp. No. 200-204, 2016.
[15]. NazliMohd Khairudin, AidaMustapha and Mohd Hanif Ahmad, "Effect of Temporal Relationships in Associative Rule Mining for Web Log Data", Hindawi Publishing Corporation, The Scientific World Journal, Vol. 2014, Pp. No.1-10, 2014.
[16]. OritRaphaeli, AnatGoldstein and LiorFink, "Analyzing online consumer behavior in mobile and PC devices: A novel web usage mining approach", Electronic Commerce Research and Applications, Elsevier, Vol. 26, Pp. No. 1-12, 2017.
[17]. S.Padmaja and Ananthi Sheshasaayee, "Clustering of User Behavior based on Web Log data using Improved K-Means Clustering Algorithm", International Journal of Engineering and Technology (IJET), Vol. 8, Iss No:1, Pp. No. 305-310, 2016.
[18]. Rahul Mishra and Abha choubey, "Comparative Analysis of Apriori Algorithm and Frequent Pattern Algorithm for Frequent Pattern Mining in Web Log Data", International Journal of Computer Science and Information Technologies, Vol. 3, Iss No:4, Pp. No. 4662 – 4665, 2012.
[19]. Rajesh K Shukla, Sanjay Silakari and P K Chande, "Existing Trends and Techniques for Web Personalization", IJCSI International Journal of Computer Science Issues, Vol. 9, Iss. No:4, pp. No. 430-439, 2012.
[20]. K.RaviBabu, "Performance of web traffic analysis using feed forward neural networks and K-means algorithms", International Journal of Engineering Research & Technology (IJERT), Vol. 1, Iss. No.:6, Pp. No.1-13, 2012.
[21]. G. Shivaprasad N. V. Subba Reddy, U. Dinesh Acharya and Prakash K. Aithal, "Neuro-Fuzzy Based Hybrid Model for Web Usage Mining", Procedia Computer Science, Elsevier, Vol. 54, Pp. No. 327 – 334, 2015.
[22]. Tomasz Bujlow, Valentín Carela-Español, Josep Solé-Pareta and Pere Barlet-Ros, "A Survey on Web Tracking: Mechanisms, Implications, and Defenses", Proceedings of the IEEE, Vol. 105, Iss. No:8, Pp. No. 1476 – 1510, 2017.
[23]. Vedpriya Dongre and Jagdish Raikwal, "An Improved User Browsing Behavior Prediction Using Web Log Analysis", International Journal of Advanced Research in Computer Engineering & Technology (IJARCET), Vol. 4 Iss. No:5, Pp. No. 1838- 1842, 2015.
[24]. S.Vijaya Kumar, A.S.Kumaresan and U.Jayalakshmi, "Frequent Pattern Mining in Web Log Data using Apriori Algorithm", International Journal of Emerging Engineering Research and Technology, Vol. 3, Iss No:10, Pp. No. 50-55. 2015.
[25]. Xiaozhe Wang, Ajith Abraham and Kate A. Smith, "Intelligent web traffic mining and analysis", Journal of Network and Computer Applications, Vol. 28, Iss. No:2, Pp. No. 147-165, 2005.
[26]. Zahid Ansari, Syed Abdul Sattar, A. Vinaya Babu and M. Fazle Azeem, "Mountain density-based fuzzy approach for discovering web usage clusters from web log data", Fuzzy Sets and Systems, Elsevier, Vol. 279, Pp. No. 40–63, 2015.
[27]. Zhen Liao, Yang Song, Yalou Huang, Li-wei He and Qi He, "Task Trail: An Effective Segmentation of User Search Behavior", Transaction on Knowledge and Data Engineering, Vol. 26, Iss. No:12, Pp. No. 3090 – 3102, 2014.
[28]. Zheng Xu, Hai-Yan Chen and Jie Yu, "Generating Personalized Web Search Using Semantic Context", The Scientific World Journal, Hindawi Publishing Corporation, Vol. 2015, Pp. No. 1-10, 2015.