# Ensemble Classifier and Clustering for Web Page Prediction

Dr. V. Sujatha[1], Dr. M. Punithavalli[2] and Dr. Renjit Jeba Thangaiah[3]

**Abstract**— Web usage mining is the art of discovering navigation patterns of users from web log data. Next web page prediction, a task of web usage mining, is used to envisage future requirements of the user during surfing. This paper presents an ensemble prediction system, that uses ensemble clustering and ensemble classification. The proposed system uses heterogeneous clustering ensemble model to group similar browsing sequences together, which is then used by a heterogeneous classification ensemble model to predict future requests of the user. The goal of this combination process is to improve the quality of individual data clustering and classification. Experimental results demonstrate that the combination of ensemble is efficient in terms of prediction accuracy and can be used by web masters to attract users.

**Index Terms** - Clustering-based Classification; Ensemble Classification; Ensemble Clustering; Next Page Prediction; User Navigation Pattern Discovery,WebLog file, Navigation Pattern.

## 1. INTRODUCTION

The World Wide Web (WWW) is more dynamic environment, whose evaluations are on par with the advanced technologies of 21st century's. The WWW is envisaging an exponential gr owth both in terms of number of websites and number of users using these websites [1]. The amount of data stored in these websites is enormous, which is increasing in millions of bytes every second. According to http://www.internetlivestats.com, the number of websites has recently exceeded 14.3 Trillion during September 2014 and is being used by more than 2,756,198,420 users. On the other hand, http://www.factshunt.com has reported that, on average, 103 million websites are added per annum in WWW. These statistics show that the number of online businesses (e-market) is growing, along with the number of users using them for their day-to-day purchases and transactions. This evaluation is necessitating the online businesses to adopt more customer-driven initiatives that seek to understand, retain, attract and develop an intimate and friendly long term relationship with their clients [2]. This paradigm shift has undauntedly led to the growing interest in Customer Relationship Management (CRM) initiatives that aim at ensuring customer identification and interactions so as to customize and personalize customer experience so as to achieve total customer satisfaction and retention, thus improving profitability along with additional business benefits [3].

Dr. V. SUJATHA,
Associate Professor, Department Computer Applications, CMS College of Science and Commerce,Coimbatore.
Dr. M. PUNITHAVALLI,
Associate Professor, Department Computer Applications, Bharathiyar University, Coimbatore.
Dr. RENJIT JEBA THANGAIAH
Head, Department of Computer Application,, *Karunya University, Coimbatore.*

These portals, in order to identify user preferences and likes, study the manner in which users' interact and browse. The knowedge gained can be used to improve their experience during a transaction and also to better serve them. Details regarding users browsing details are stored in a special file called "weblogs", whose entries are auto-generated by the server. These files consist of "hidden asset" called "knowledge" that has become the most valuable resource to both web designers and online traders.The knowledge discovery techniques used for these purpose, termed as Web Usage Mining (WUM), is a research field that focuses on analyzing and predicting user's experience / preferences using data mining techniques. The result of such analysis can be used in several applications like personalizing user's web browsing experience [4], predicting intuitive web pages that a user is likely to browse [5], improving user's browsing experience [6] , obtaining traffic trends to perform target marketing [7] and identifying popular pages (globally and regionally) [8].Analysis of user's navigation data from weblog data are mainly used to discover knowledge about the users' access patterns and usage trends. The advantages obtained from such analysis are multifolded. Some of them include (i) better structure and grouping of resources (ii) save browsing and search time (iii) reduce page retrieval time and (iv) improve bandwidth loading time on network. Currently, many such web analysis tools exist. Examples include real time analytic tools like Google Analytics, Open Web Analytics and Mint.

In spite of many such available tools, the field of WUM is considered immature, as many requirements, relating to accuracy of prediction, have yet to be met to reach the state of perfection. This paper proposes technique to improve the accuracy of predicting the users' future web page requests. Given a web log file, and the main objective is to develop a User Navigation Pattern Discovery for Next Page Prediction (NPDNPP) system by using data mining techniques to discover knowledge from weblog data in an accurate and time efficient manner for im-

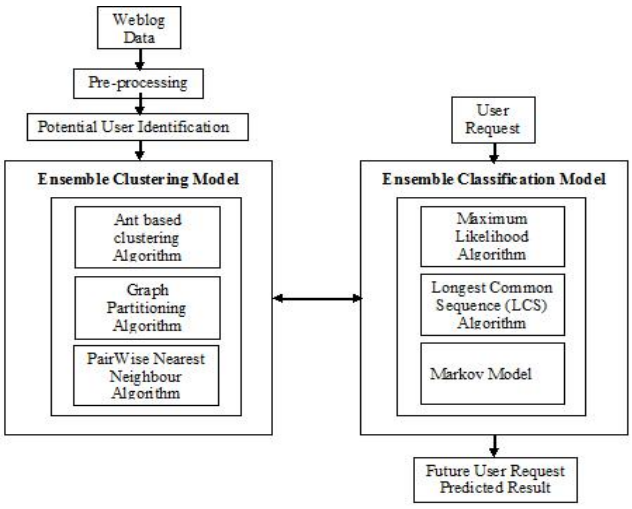
proving the browsing experience of users. The NPDNPP system consists of three main steps, namely, preprocessing, identification of potential users and prediction. The preprocessing step consists of cleaning the raw weblog file by removing irrelevant and unwanted data, user and session identification. The second step separates valuable users from non-potential users. This step, apart from reducing the size of web log file, also helps to focus on only important users. This paper focuses on the third step of NPDNPP system, the development of prediction engine that discovers navigation patterns for predicting next page request of the user. For this purpose, this paper proposes an ensemble cluster-based ensemble prediction algorithm. The rest of the paper includes Section 2 presents the methodology of the proposed prediction system, while Section 3 presents the performance evaluation. Section 4 concludes the work with future research directions.

## 2. Methodology

### 2.1 NPDNPP System

The overall architecture of NPDNPP system is presented in Figure 1. The proposed NPDNPP methodology includes three major stage : i) Preprocessing ii) Ensemble clustering of three heterogeneous clustering algorithm which is used to find the similarity between different users accessing the same web log pattern . iii) Ensemble classification algorithm is used to predict the future navigation pattern by the user in current active session window.

**Fig. 1** Architecture of NPDNPP System




## 2.2 Preprocessing

The unformatted web log data is converted into a form that can be directly applied to the mining process. Cleaning is the process which removes all entries which will have no use during analysis or mining. Details regarding the methods used for preprocessing and potential user identification are presented in our previous publication [9] where the first stage is the cleaning stage, in which the unwanted log entries were removed. In the second stage, cookies were identified and removed. The result was then segmented to identify potential users. From the potential user, a graph partitioned clustering algorithm was used to discover the navigation pattern. An LCS classification algorithm was then used to predict future requests. The present experiment works on the amalgamation of ensemble clustering techniques and ensemble classification in terms of clustering and classification.

### 2.3 Ensemble Selection

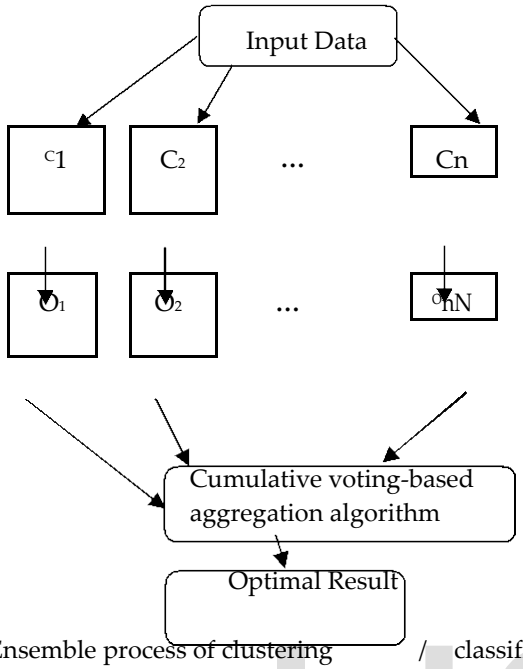
Ensembles can be either homogeneous or heterogeneous. Homogeneous models are considered as those models having the same methodology but different feature vectors, while heterogeneous classifiers are models with each model using a different classification methodology. Cluster Ensemble [10] method is designed based on quality and diversity. In few studies it is sought to use the concept of diversity to improve the performance of ensemble by selecting an ensemble from multiple ensembles [11].

Ensemble clustering combines the results of multiple clustering algorithms into a single consolidated clustering (referred to as consensus solution) with the aim of creating robust and stable clustering results. An ensemble prediction model, on the other hand, consists of a set of independently trained classifiers whose predictions are combined by various statistical or algebraic methods. These clustering and prediction algorithms were chosen because they exhibited the two criterion of ensemble [12] namely, high accuracy and diversity between classifiers (make different error rate). This model performs prediction as a 2-step process.

The first step performs clustering, while the second step performs prediction. When a new user request arrives at the server, the URL requested and the session to which the user belongs are first identified. This information is used to update the underlying knowledge and a list of suggestions is appended to the requested page. This necessitates a search through the whole weblog data file. To reduce this search time, the clustering step is introduced. Thus, while using clustering, the prediction step attempts to find a cluster with high degree of similarity with the user request and predicts possible next page visits using that cluster alone. Classification scheme [13] proved that the success rate or accuracy of a classification or prediction problem can be improved by using multiple classifiers instead of single classifier. Motivated by this, in this paper, both clustering and prediction are performed using ensemble concepts. The main aim of creating an ensemble is to combine the strengths of individual models to

improve the performance of next page prediction.

**Fig. 2** Ensemble process




The Ensemble process of clustering / classification is shown in Figure 2, where *Cs* are the clustering or classification algorithms, *O*s are the outputs created by the clustering and classification algorithms and *n* is number of clustering or classification algorithms.

### 2.4 Ensemble Clustering

The heterogeneous ensemble clustering system is constructed using three algorithms. The first one is the ant based clustering (AC) algorithm, it gathers items to form heaps and when sorting, it discriminates between different kinds of items and spatially arrange them according to their properties, like the behavioral of an ant [14]. The second selection is graph partitioning algorithm to find the correlated pages between each pair of web pages by assigning weights to the graph [15] and the third is the Pair wise Nearest Neighbor algorithm which generates the clustering hierarchically by a sequence of merge operations and at each step two nearby clusters are merged [16].

### 2.5 Ensemble Classifier

The prediction ensemble system is constructed using the following algorithms and the first algorithm is the Maximum likelihood (MLC) algorithm [17] which is used as a statistical decision rule that examines the probability function of a data for each of the classes and assigns the record to the class with the highest probability, the second is the Longest Common Sequence (LCS) [18] is to find the longest subsequence common to all sequences in a set of sequences and to classify current user activities to predict the users' next movement. The final algorithm is the Markov Model (MM) algorithm [19] which is commonly used in the identification of the next page to be accessed by the Web site user based on the sequence of previously accessed pages.

The current active session S , hold the unique web pages like $P_1$, $P_2$, ..., $P_m$. Unique numerical codes were assigned to each page and this number was used to construct S. When the user visits a webpage, the prediction system creates a unique code (if it is a new page) or replaces the URL with its predefined code (already existing page). Ensemble clustering results with a set of navigation patterns, NP = $np_1$, $np_2$...$np_m$ where $np_i$ is a set of k web pages in a navigation pattern, that is, $np_i = P_1, P_2 ... P_k$, where k $1 \le i \le k$ and $P_1...P_k$ Î S. Sequence W' = $P_1$, $P_2$ …Pm is a current active session and m is size of active session window. The navigation patterns $np_i$ and active session S are used as input to the classifier. The main goal of the classification algorithm is to find a cluster that has highest degree of similarity with the user request. The co-occurrence matrix M is built from the user active session. The element $M_{ij}$ of M is defined as conditional probability that the page $P_i$ is visited in the same session where the page $P_j$ is visited. The prediction list is constructed the pages in active session windows are ordered based on values stored in the co-occurrence matrix M. After this step, for building the prediction list, the system finds the cluster based on the classification algorithm.

In LCS Algorithm the prediction engine selects the user request match with the prediction engine selects a cluster in such a way that, if the difference between positions of last elements of longest common subsequence discovered in the cluster and the position of first element of this sequence is minimized, the system chooses this cluster. In this module, if the first page in the next user activity is different with prediction list, it needs again to classify with new user activities. With MLC, the user request is matched with the cluster whose probability is the highest. A similar approach is also used by MM classifier, but here, the probability of visiting a page $P_i$ does not depends on all the pages in a session, but only on a small set of k preceding pages. This model is termed as Kth order Markov model. The aggregation method used for ensemble is the Cumulative voting-based aggregation algorithm.

### 2.6 Consensus Function

The cumulative voting-based aggregation algorithm consists of two steps; the first one is to obtain the optimal relabeling for all partitions, which is known as the voting problem. Then, the voting-based aggregation algorithm is used to obtain the aggregated (consensus) partition. The voting-based aggregation algorithm [20] is modified to be used in this paper.

In figure 3, let $\chi$ denote a set of $n$ data objects, and let a partition of $\chi$ into $k$ clusters be represented by an $n \times k$ matrix **U**

such that $\sum_{q=1} ujq=1$, for ∀ j. Let $u=\{U_i\}_{i=1}$ denote an ensemble of partitions. The voting-based aggregation problem is concerned with searching for an optimal relabeling for each partition $\mathbf{V}^i$ with respect to representative partition $\mathbf{U}^0$ (with $k^0$ clusters) and for a central aggregated partition denoted as $\bar{U}$ that summarises the ensemble partitions. The matrix of coefficients $\mathbf{W}^i$, which is a $k^i \times k^0$ matrix of $w_{lq}{}^i$ coefficients, is used to obtain the optimal relabeling for ensemble partitions.

In this paper, the fixed-reference approach is used, whereby an initial reference partition is used as a common representative partition for all the ensemble partitions and remains unchanged throughout the aggregation process. Instead of selecting random partition, the partition that is generated by the method, which showed high ability to separate active from inactive window in our experiments, is suggested to be the reference partition $U^0$; and this method is the Ward's clustering the cumulative voting-based aggregation algorithm is described as follows:

**Fig. 3.** Cumulative Voting-based Aggregation Algorithm

1. Select a partition $U^i \in u$ which is generated by the Ward's method and assign to $U^0$
2. for $i$=1 to $b$ do
3. $W^i=(U^{iT}U^i)^{-1}U^{iT}U^0$
4. $V^i=U^iW^i$
5. $U0=i-1\ iU0+1iVi$
6. end for
7. $\bar{U} = U^0$.

The results were evaluated based on the effectiveness of the methods to separate active from inactive navigation pattern by using four measures: Precision, Recall, F-measure and accuracy.

## 3. EXPERIMENTAL RESULTS AND DISCUSSION

The proposed system was tested using the weblog data collected from www.microsoft.com, where access information is stored according to the Common Log Format with two log files identified as WL1 (223.7.MB) and WL2 (211.5 MB) and the weblog data collected from NASA produced by the web servers of Kennedspace centre http: // ita.ee.lbl.gov / html / contrib/NASA-HTTP.html), in the Common Log Format and identified as WL3 (205.2 MB) and WL4 (167.8 MB). These four log files are identified WL1, WL2, WL3 and WL4 in this paper.

### 3.1 Pre-Processing Result

In table 1, it is evident that pre-processing log file reduces both the memory used to store the log file and the number of transactions in a tremendous fashion. It records 37,711 randomly selected anonymous users of the site of which 32,711 are given as training set and 5000 as test set. For each user, the data lists all the areas of the web site that user visited in a one-week timeframe. The effect of cleaning log data in terms of number of transactions and amount of memory required to store it is shown in table 1**.**

**Table.1. Effect of Pre-processing result**

| | 1 Day | | 3 Days | | 1 Week | |
|---|---|---|---|---|---|---|
| | Before | After | Before | After | Before | After |
| No. of Transactions | 7000 | 1420 | 11987 | 4320 | 34000 | 11381 |
| Memory Used (MB) | 1.76 | 0.48 | 1.95 | 0.61 | 2.47 | 1.12 |

### 3.2 Performance Measure

The performance of the prediction engine was evaluated using four performance parameters, namely, precision Pr, recall Re, F Measure and accuracy. Each classifier c segments the test set T into four partitions based on both the true label $y_i$ and the predicted label $c(x_i)$ for each example$(x_i, y_i) \varepsilon$ T , it refer to the absolute number of true positive as TP, false positives as FP , true negative as TN and false negatives as FN.

$$Pr = TP / (TP+FP) \quad (1)$$

$$Re = TP / (TP+FN) \quad (2)$$

F-measure combines these two into a single number, which is useful for ranking or comparing methods. F-measure is the harmonic mean between precision and recall.

$$F= \frac{2.\ Pr.\ Re}{Pr + Re} \quad (3)$$

A commonly used measure is the accuracy, which is the

fraction of correct recommendations to total possible recommendations.

$$\text{Accuracy} = \frac{TP + TN}{TP+TN+FP+FN} \quad (4)$$

The experiments were conducted in three stages, where the first stage evaluated the effect of the ensemble clustering algorithm on prediction, second stage evaluated the effect of clustering-based ensemble prediction and final stage of experiments evaluated the proposed ensemble clustering based ensemble classification. For ease of discussion the coding scheme presented in Table 2 was used.

**Table 2 . Coding scheme**

| Description | Code | Description | Code |
|---|---|---|---|
| Ensemble Clustering Model | ECLU | Ensemble Classification Model | ECLA |
| ECLA-Based Models | | ECLU-Based Models | |
| ECLU With ML Classifier | ECLU-MLC | AC Clustering -based ECLA | AC-ECLA |
| ECLU with LCS classifier | ECLU-LCS | GP clustering-based ECLA | GP-ECLA |
| ECLU with MM classifier | ECLU-MM | IP clustering-based ECLA | IP-ECLA |
| Proposed Model | | | |
| ECLU-based ECLA | | ECLU-ECLA | |

Table 3 to 6 show the precision, recall, F Measure and accuracy obtained by the prediction models with respect to the four selected metrics respectively, while using the two selected web log files.

**Table 3 . Precision (%)**

| Model | WL1 | WL2 | WL3 | WL4 | Average |
|---|---|---|---|---|---|
| ECLU-MLC | 85 | 83.53 | 84.89 | 83.64 | 84.27 |
| ECLU-LCS | 86.78 | 85.14 | 85.94 | 85.98 | 85.96 |
| ECLU-MM | 85.71 | 84.39 | 84.1 | 86.01 | 85.05 |
| AC- ECLA | 83.21 | 82.31 | 82.25 | 83.28 | 82.76 |
| GP- ECLA | 84.04 | 82.85 | 82.83 | 84.05 | 83.44 |
| IP- ECLA | 82.59 | 81.45 | 82.91 | 81.12 | 82.02 |
| ECLU-ECLS | 87.7 | 86.01 | 85.99 | 87.75 | 86.86 |

**Table 4 . Recall (%)**

| Model | WL1 | WL2 | WL3 | WL4 | Average |
|---|---|---|---|---|---|
| ECLU-MLC | 90.38 | 88.53 | 90.98 | 87.96 | 89.46 |
| ECLU-LCS | 92.54 | 90.94 | 91.96 | 91.51 | 91.74 |
| ECLU-MM | 91.88 | 89.63 | 91.53 | 89.99 | 90.76 |
| AC-ECLA | 88.9 | 86.24 | 87.09 | 88.03 | 87.57 |
| GP-ECLA | 89.65 | 87.09 | 89.76 | 86.98 | 88.37 |
| IP-ECLA | 86.94 | 85.71 | 88.36 | 84.32 | 86.33 |
| ECLU-ECLS | 93.22 | 91.73 | 92.61 | 92.36 | 92.48 |

**Table 5. F-Measure (%)**

| Model | WL1 | WL2 | WL3 | WL4 | Average |
|---|---|---|---|---|---|
| ECLU-MLC | 87.61 | 85.94 | 84.14 | 89.41 | 86.78 |
| ECLU-LCS | 89.56 | 87.94 | 88.58 | 88.92 | 88.75 |
| ECLU-MM | 88.69 | 86.92 | 87.98 | 87.64 | 87.81 |
| AC-ECLA | 85.96 | 84.24 | 83.61 | 86.6 | 85.1 |
| GP-ECLA | 86.75 | 84.92 | 89.76 | 85.91 | 85.84 |
| IP-ECLA | 84.71 | 83.53 | 83.61 | 84.63 | 84.12 |
| ECLU-ECLS | 90.38 | 88.77 | 89.71 | 89.46 | 89.58 |

**Table 6 . Accuracy (%)**

| Model | WL1 | WL2 | WL3 | WL4 | Average |
|---|---|---|---|---|---|
| ECLU-MLC | 93.17 | 92.38 | 92.49 | 93.07 | 92.78 |
| ECLU-LCS | 95.52 | 94.01 | 94.87 | 94.68 | 94.77 |
| ECLU-MM | 94.89 | 93.77 | 95.73 | 92.94 | 94.33 |
| AC-ECLA | 91.55 | 91.05 | 89.91 | 92.7 | 91.30 |
| GP-ECLA | 92.71 | 91.78 | 92.64 | 91.86 | 92.25 |
| IP-ECLA | 91.06 | 90.09 | 91.54 | 89.61 | 90.58 |
| ECLU-ECLS | 96.47 | 95.33 | 96.8 | 94.98 | 95.90 |

Comparison of the three clustering algorithms revealed that the GP clustering algorithm is more suitable for predicting next page of the user. Similarly, comparison of the three selected classifiers revealed that the LCS classifier produce improved results. However, the ensemble approach, that combined multiple clustering and multiple classifier produced best results.

While comparing with GP-ECLA and ECLU-LCS models, the ECLU-ECLA model produced an average F-Measure efficiency of 4.18% and 0.92% respectively. Similarly, analysis of accuracy revealed that the proposed model gained an average efficiency gain of 3.81% and 1.18% with respect to accuracy over GP-ECLA and ECLU-ELCA. Thus, from the various results, it is evident that the objective of this work has been achieved and can be used to effectively predict the next page requests of the user during surfing.

## 4. CONCLUSION

A next web page system presented in this paper consists of three steps, namely, preprocessing, potential user identification and classification. The preprocessing step transforms the raw web log data into a form that can be directly used by the prediction algorithm. The second step identified important users to improve the performance of prediction and to reduce the size of weblog file. The focus here was to design a prediction step that performs the actual prediction of next page for the user during surfing. For this purpose, an ensemble clustering-based ensemble prediction system was proposed. The clustering step improved the performance of prediction in terms of accuracy. The ensemble clustering model was constructed using three algorithms, namely, ant-based clustering, pair-wise nearest neighbor and graph partitioning. The ensemble clustering algorithm grouped the potential users web log data, thus grouping similar web browsing sequences. Similarly, three classification algorithms (Maximum Likelihood Classification Algorithm, Longest Common Sequence Classification Algorithm and Markov Model based Classification Algorithm) were analyzed for predicting next web page. Using the clustered results and ensemble classifier future user requests were predicted. Experimental results proved that the proposed model that combined multiple clustering and classification algorithm was efficient in next web page prediction in terms of precision, recall, F measure and accuracy. Future plans include combining homogeneous ensemble algorithms with heterogeneous algorithms along with ensemble pruning.

## Acknowledgments

We thank the data collected from www.msn.com and NASA produced by the web servers of Kennedy Center Space which helped us to proceed on with our experiment and it produced desired results.

## 5. References

[1] Yan, Z., Zhang, P. and Vasilakos, A.V. (2014) A survey on trust management for Internet of Things, *Journal of Network and Computer Applications*, Vol. 42, pp. 120-134.

[2] Asiedu, M. and Safo, J.O. A multi-dimensional service delivery among mobile network providers in Ghana : A case of customer satisfaction, *European Scientific Journal*, Vol. 9, No. 23, 2013, pp. 86-101.

[3] Ryals, L. and Knox, S. (2001) Cross-functional issues in the implementations of relationship marketing through CRM, *European Management Journal*, Vol. 19, Issue 5, pp. 534-542.

[4] Pagar, Y.S., Mote, V.R. and Bramhane, R.S. (2012) Web Personalization using Web Mining Techniques, *IJCA Proceedings on Emerging Trends in Computer Science and Information Technology*, Vol. 1, pp. 1-4.

[5] Anitha. A. A New Web Usage Mining Approach for Next Page Access Prediction, *International Journal of Computer Applications,* Vol. 8, No. 11, 2010, pp. 7–10.

[6] Mohanty, B.K. and Passi, K. (2010) Agent based e-commerce systems that react to buyers' feedbacks – A fuzzy approach, *International Journal of Approximate Reasoning*, Vol. 51, Issue 8, pp. 948–963.

[7] Ya, L. (2012) The Comparison of Personalization Recommendation for E-Commerce Physics Procedia, Vol. 25, pp. 475–478.

[8] Langhnoja, S.G., Barot, M.P. and Mehta, D.B. (2013) Web Usage Mining to Discover Visitor Group with Common Behavior Using DBSCAN Clustering Algorithm, *International Journal of Engineering and Innovative Technology*, Vol. 2, Issue 7, pp. 169-173.

[9] Sujatha, V. and Punithavalli, M. (2012) Improved user navigation pattern prediction technique from web log data, *International Conference on Communication Technology and System Design, Procedia Engineering*, Vol. 30, pp. 92-99.

[10]Hu, X. and Yoo, I. (2004). *Cluster Ensemble and Its Applications in Gene Expression Analysis. In Proc. Second Asia-Pacific Bioinformatics Conference (APBC2004), Dunedin, New Zealand. CRPIT,* ***29****. Chen, Y.-P. P., Ed. ACS. 297-302.*

[11] Hadjitodorov, S., Kuncheva, L. I and Todorova, L.P. (2006) Moderate Diversity for Better Cluster Ensembles. *Information Fusion Journal*, pp.264-275.

[12]Kuncheva, L.I. and Whitaker, C.J. (2003) Measures of diversity in classifier ensembles and their relationship with the ensemble accuracy, *Machine Learning*, Vol. 51, No.2, pp. 181-207.

[13] Neeba N.V and Jawahar C.V. (2009) Empirical evaluation of character classification schemes, *Seventh International Conference on Advances in Pattern Recognition (ICAPR),* pp. 310–313.

[14] Neeba N.V and Jawahar C.V. (2009) Empirical Handl, J. and Meyer, B. (2002) Improved ant-based clustering and sorting in a document retrieval interface, *Proceedings of the Seventh International Conference on Parallel Problem Solving from Nature,* Vol. 2439 of LNCS, Springer-Verlag, Berlin, Germany, pp. 913–923.

[15] Jalali, M., Mustapha, M., Mamat, A. and Sulaiman, M.N.B. (2008) A new clustering approach based on graph partitioning for navigation patterns mining, *9th International Conference on Pattern Recognition,* pp. 1-4.

[16] Virmajoki, O and Franti, P (2004) Divide and conquer algorithm for creating neighbourhood graph for clustering *IEEE*, vol.1, pp no.264-267.

[17] Ahmad, A. (2012) Analysis of Maximum Likelihood Classification on Multispectral Data, *Applied Mathematical Sciences,* Vol. 6, No.129, pp. 64256436.

[18] Jalali, M., Mustapha, N., Mamat, A.,Nasir,M.N and Sulaiman, B (2009) ARecommender System for Online Personalization in the WUM Applications *Proceedings of the World Congress on Engineering and Computer Science,* Vol.2, pp.741-746.

[19] Khalil, F., Li, J. and Wang, H. (2007) Integrating markov model with clustering for predicting web page accesses, *Proceedings of the 13th Australasian World Wide Web Conference* (Aus Web 2007), Australia, pp. 1-26.

[20]Ayad HG, Kamel MS: Cumulative voting consensus method for partitions with a variable number of clusters.*IEEETrans Pattern Anal Mach Intell* 2008, 30(1)**:**160-173.