© 2018 IJSRST | Volume 4 | Issue 2 | Print ISSN: 2395-6011 | Online ISSN: 2395-602X
Themed Section: Science and Technology

Well-organized Data Mining Techniques for Clustering of Users on Web Log Data

Avula Chitty
Department of CSE, Assistant Professor, Sri Indu College of Engineering and Technology, Hyderabad, Telangana, India

ABSTRACT

Web usage mining is one among the essential frameworks to find domain data from the interaction of users with the net. This domain data is used for effective management of prognosticative websites, the creation of adaptative websites, enhancing business and net services, personalization, and so on. In nonprofit able organization's web site, it's tough to spot who area unit users, what info they have, and their interest's modification with time. Web usage mining supported log knowledge provides an answer to the present problem. The planned work focuses on weblog knowledge preprocessing, thin matrix construction supported net navigation of every user and clump the users of comparable interests. The performance of net usage mining is additionally compared supported k-means, X-means, and farthest 1st clump algorithms.
Keywords : Web usage mining, sparse matrix, Clustering, Influence degree, K-means, X-means and farthest first algorithm

I. INTRODUCTION

Digitalization of {data|of knowledge} and zoom of {data of knowledge} technology cause monumental data altogether domains within the sort of formats and full data might not be helpful to any or all users because it is. Data processing helps to extract solely relevant data from these giant repositories. The online may be an immense repository of text documents and transmission information. Mining helpful information from net the online the net} is understood as web mining and it's classifieds: online page analysis, net usage mining, and net structure analysis.

Web content analysis is analogous to data processing technique for relative databases. Online page associatealysis is an extraction of helpful domain data or data from net documents. This data could also be audio, video, text, or image. Net document could contain structured or unstructured information. Online page mining is applied to completely different analysis fields supported text and pictures like Text or Image content Retrieval and AI. Net structure analysis is employed to seek out webpage structure or boundary extraction in WebPages accessed by the online users that is employed to project the similarities of online page and to enhance the online structure. Net usage analysis is useful to seek out the frequent net accessed patterns for user analysis.

From net the online the net} analytics and web application purpose of read, extracted domain data obtained from the online usage mining may be directly applied to cross-market analysis for e-market Business, e-Services, e-Education, e-Newspapers-Governance, Digital repositories, etc. as a result of the provision of enormous size of the online information, it's essential to research the online content for the e-business and online applications to avail user's counseled systems [1]. With an oversized range of corporations exploitation the web to distribute and collect data, data discovery on the online has become a crucial analysis space. In net usage pattern analysis the information is extracted from the online log file [2], it contains {the data the knowledge the data} like user browsing information, World Health

Organization area unit visiting the web site, that files the user is accessing, what number bytes of information area unit accessed, style of OS, and different data. There is a unit several techniques of net usage mining are developed by several researchers. The Sect. two covers a review of the online usage mining, Sect. three is regarding existing work and Sect. four provides the main points of planned work and Sect. five describes the results and Sect. 6 mentioned conclusion alongside future work.

II. RELATED WORK

With rapid usage of web by a large number of customers, clustering of these people becomes an important task based on their usage or accessing of web pages. This section gives a brief review of web mining and its literature. It comprises data preprocessing, user and session identification, path completion and clustering is performed using farthest first technique (FFC) used for clustering. But in this method occurrence of outliers is more while clustering and not suitable for smaller datasets.

But this method is applicable only for small repositories, therefore it lacks the scalability. Antony Selvadoss, Thanamani et al. [6] proposed a technique based on fuzzy C-means to cluster the web user activities. The mechanism involves constructing the session matrix, calculating the R index and F-score. But, this technique is highly expensive as it involves large number of computations for clustering. J. HuaXu and H.

Liu [7] developed a web user clustering technique based on vector matrix: Users and URL as rows and columns and values as number of hits. Similarity measure is found by using cosine similarity function to find similarity between the vectors (rows) and then unsupervised k-means model is used to cluster the web user's behavior. But, this model does not work when the variation between hits value is high. K. Santhisree et al.

[8] Implemented a new framework known as rough-set based DBSCAN clustering to find the user access patterns along with inter-cluster similarity identification among the clusters. The implementation is based on the set similarity, sequence similarity values. This is not valid for soft clustering. Xidong Wang et al. [9] proposed a frequent access pattern method to extract the frequent access patterns from users' accessed path for a specific website. Generation of frequent access pattern tree using this technique is difficult when more number of access records exist for each user. Wei et al. [10] implemented a multi-level model for web users clustering based on similarity interests from the user's accessed web pages. This technique is suitable only if the number of sessions is higher than the given threshold measure. No one technique is optimum in clustering the web users. Therefore, novel techniques to achieve optimum clustering and clustering with minimum time are very much necessary.

III. EXISTING SYSTEM

Web user clustering is one of the essential tasks in web usage analysis. The goal of clustering is to group data points that are close (or similar) to each other and identify such groupings (or clusters) in an unsupervised manner [11]. Information of web user clusters has been widely used in many applications, such as solution of website structure design and optimization [12]. There are many clustering techniques employed for finding the interestingness patterns among users. One among such technique is clustering of web users through Matrix Influence Degree (MID), which contains influence degree of each web page corresponding to the user. Simple k-means is applied over the obtained MID to obtain the clusters [13]. Clustering process is pictographically represented in Fig. 1 which is a sequential process. K-means is greedy approach for partitioning n objects into k clusters. Clusters are formed by minimizing the sum of squared distances to the cluster centers. The main task of K-means clustering is described as

- Given a set X of n points in a d-dimensional space and an integer k
- Task: choose a set of k points {c1, c2, ck} in the d-dimensional space to form clusters {C1, C2, Ck} such that the following function (Eq. 1) is minimized.

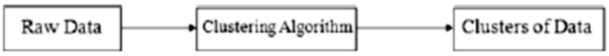

Fig. 1 Clustering of data

$$Cost(C) = \sum_{i=1}^{k} \sum_{x \in C_i} L_2^2(x - c_i) \qquad \text{------------------ (1)}$$

The k-means algorithm:

- Randomly pick k cluster centers {c1, ck}
- For each I, set the cluster Ci to be the set of points in X that are closer to ci than they are to cj for all i ≠ j
- For each i let ci be the center of cluster Ci (mean of the vectors in Ci)
- Repeat until convergence.

In the similar manner simple K-means operates over the generated MID obtained from web log data. J. Xiao et al. [14] developed a cluster analysis method to mine the web, this method clusters the web users based on their profiles. User profile clustering algorithm consists of three modules. First it finds the similarity measure based on Belief function and applies greedy clustering using this function then creates the common user profiles. In greedy clustering, it first randomly selects the users into common profile set and for each user calculate similarity measure, based on this choose best representative and update similarity of each point to the closest representative finally it gives the unique clusters. To generate the clusters using session representatives V. Sujatha and Punitha Valli [15] adopted a method called Hierarchical Agglomerative clustering. In this method first it considers the navigation patterns for each session as a single cluster. Clusters members are added based on similar session pages. And similar navigation patterns, it is calculated by distance similarity measure for several sessions. Density-based algorithm works [16] first searching core objects based on these objects clusters are increased and by looking for objects which are in a neighborhood within a radius of object. The main advantage of this density-based algorithm is it can filter out noise.

Disadvantages of existing systems:

- Time consuming to build the model over large records and increases over the size of MID.
- Web user clusters are dependent on the initial point selected, which have large influence over the cluster sizes.
- More number of outliers can cause problem of ambiguity.
- Does not operate well in large variation between values of the influences degree.

IV. PROPOSED SYSTEM

In this paper, an efficient technique is implemented for clustering web users. It includes the following steps:

1. Data preprocessing of the weblog and extracting the relevant attributes for clustering.
2. Build the Sparse Matrix of every user with sessions and pages visited.
3. Generate Three Tuple matrixes from sparse matrix.
4. Calculate the influence Degree of all web pages for a user.
5. Repeat the Steps 2, 3, and 4 for all web users until all sessions are completed in the web log.
6. Generate the Matrix of Influence Degree (MID) for all web pages.

7. Cluster the MID with X-means and farthest first techniques.

4.1 Data Preprocessing

The raw data of web log is as shown in Table 1, which contains set of attributes and all of these are not useful for clustering the web users. The input raw data is applied to noise removal phase, where missing values and records with error status numbers such as 400,404 are removed and only relevant attributes are selected. The required log data after cleaning is shown in Table 2, it contains relevant features such as network IP address, Time of Request, and URL are extracted to cluster the web users.

The generated three tuple matrices from log data after application of preprocessing are shown in Table 2. From Table 2 each URL is identified with unique number. This data indicates 14 web pages are accessed by web users.

4.2 Building a Sparse Matrix

Sparse matrix is a matrix in which most of the elements are zeros. Web log data is transformed into a sparse matrix. User session is considered as the criterion to group web log data. Web log data is grouped by user session for each web user and the user session is defined as 3600 s.

Let there are 'm' web users Ui, where i = 1, 2,…, m, each web user has 'n' user sessions Sj, where j = 1, 2,…, n, and there are 'p' web pages Pk, where k = 1, 2,…, p then the sparse matrix SMi of Ui is shown in Eq. (2):

$$\mathbf{SM}_i = \begin{bmatrix} s1p1 & s1p2 & s1p3 \\ s2p1 & s2p2 & s2p3 \\ s3p1 & s3p2 & s3p3 \end{bmatrix}, \qquad \text{-------------------(2)}$$

where, Sjpk is the value equal to the times of web page Pk accessed in user session Sj by web user Ui. 50 web log data records are considered for experimentation, the sparse matrix for the 14 users and their session usage of web pages is generated as

Table 1 Sample web log data

IP address	User ID	Time	Method/URL/protocol	Status	Size	Refer	Agent
82.117.202.158	-	[16/Nov/2009:01:12:18 +0100]	GET/galerija.php HTTP/1.0	200	4545	http://www.vtsns.edu.rs/	Moz/5.0
82.117.202.158	-	[16/Nov/2009:01:12:22 +0100]	GET/galerija.php HTTP/1.0	200	3176	http://www.vtsns.edu.rs/	Moz/5.0
82.208.207.41	-	[16/Nov/2009:01:12:43 +0100]	GET/ispit_odbijeni.php HTTP/1.0	200	826	http://www.vtsns.edu.rs/	Moz/5.0
83.136.179.11	-	[16/Nov/2009:01:22:43 +0100]	GET/index.php HTTP/1.0	200	826	http://www.vtsns.edu.rs/	Moz/5.0
83.136.179.11	-	[16/Nov/2009:01:23:43 +0100]	GET/index.php HTTP/1.0	200	826	http://www.vtsns.edu.rs/	Moz/5.0
82.208.207.41	-	[16/Nov/2009:02:12:23 +0100]	GET/ispit_odbijeni.php HTTP/1.0	200	23689	http://www.vtsns.edu.rs/	Moz/5.0
82.208.207.41	-	[16/Nov/2009:02:12:25 +0100]	GET/ispit_odbijeni.php HTTP/1.0	200	5740	http://www.vtsns.edu.rs/	Moz/5.0
82.208.207.41	-	[16/Nov/2009:02:13:43 +0100]	GET/biblioteka.php HTTP/1.1	200	3669	http://www.vtsns.edu.rs/	Moz/5.0
82.117.202.158	-	[16/Nov/2009:02:17:26 +0100]	GET/ispit_raspored_god.php HTTP/1.1	200	3669	http://www.vtsns.edu.rs/	Moz/5.0
82.117.202.158	-	[16/Nov/2009:02:17:39 +0100]	GET/ispit_raspored_god.php HTTP/1.1	200	6554	http://www.vtsns.edu.rs/	Moz/5.0
83.136.179.11	-	[16/Nov/2009:02:17:41 +0100]	GET/ispit_raspored_god.php HTTP/1.1	200	4053	http://www.vtsns.edu.rs/	Moz/5.0
82.208.255.125	-	[16/Nov/2009:02:17:44 +0100]	GET/ispit_raspored_god.php HTTP/1.1	200	1662	http://www.vtsns.edu.rs/	Moz/5.0

Moz/5.0 → Mozilla/5.0 (Windows; U)

Table 2 Web log data after

Table 3 : The Web Log Data After Leaning

IP address	Time	URL
82.117.202.158	1:12:18	0
82.117.202.158	1:12:22	0
82.117.202.158	1:12:43	1
82.208.207.41	1:22:43	2
83.136.179.11	1:23:43	2
83.136.179.11	2:12:23	1
82.208.207.41	2:12:25	1
82.208.207.41	2:13:43	3
82.208.207.41	2:17:26	4
82.117.202.158	2:17:39	4
82.117.202.158	2:17:41	4
83.136.179.11	2:17:44	4
82.208.255.125	2:17:53	3
82.208.255.125	3:12:42	5
82.117.202.158	3:27:23	1
83.136.179.11	3:37:32	1
83.136.179.11	3:37:44	5
82.208.255.125	4:13:43	5
83.136.179.11	4:17:26	1
82.117.202.158	5:17:39	4
82.208.255.125	5:17:41	4
82.208.255.125	5:18:40	5
82.208.207.41	5:37:53	4
82.117.202.158	5:39:42	5
82.117.202.158	6:27:23	4
83.136.179.11	6:37:32	4
83.136.179.11	18:44:25	0
157.55.39.13	18:44:25	3
157.55.39.136	18:08:01	2
199.79.62.54	18:44:15	1
157.55.39.38	18:44:21	6
157.55.39.136	18:47:58	4
40.77.167.52	18:48:04	5
40.77.167.52	18:43:52	7
157.55.39.38	18:49:16	8
123.125.71.70	19:11:33	9
66.249.65.71	19:06:27	10
105.224.92.45	19:06:27	11
105.224.92.45	19:06:27	12
105.224.92.45	19:08:01	13
199.79.62.54	19:06:27	14

(continued)

IP address	Time	URL
105.224.92.45	19:04:32	14
157.55.39.108	19:06:26	9
105.224.92.45	19:06:27	12
105.224.92.45	19:06:29	13
105.224.92.45	19:06:30	0
105.224.92.45	19:06:27	6
105.224.92.45	19:38:05	9
66.249.65.55	19:38:27	13
66.249.65.49	19.37.27	9

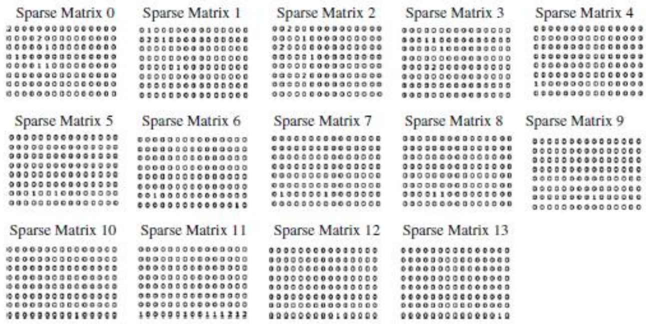

Fig 2 : Spare Matrix For 14 Users

displayed in Fig. 2. It is built for each user with total session count and total number of web pages in the selected website. The values in the cells of matrix indicate the visits of the web page in a particular session (s1p1). Value 0 in the cell indicates that page is not visited in that session with respect to the user. In each sparse matrix we can extract the useful knowledge, whether the web pages are accessed by users or not. This will give the importance or preference of accessed pages by the web users.

From above diagram, consider the sparse matrix-0, the value in the cell s0p0 is 2, it indicates that user-0 visited the page 0 in session-0 twice. The value in cell s4p4 is 1, which indicates that user-0 visited the page-4 in session-4 once. The same procedure is followed for all elements of matrices. Thus, all sparse matrices indicate the Activity of the user in all sessions.

4.3 Generating Three Tuple Matrix

After building the sparse matrix of each web user, we get its 3-tuple representation. 3-tuple representation of sparse matrix is a list of (i, j, v), where i is the row index of nonzero value, j is the column index of nonzero value, and v is the value of nonzero value in the sparse matrix. The three tuple matrices for the 10 users corresponding to their sparse matrices are generated as displayed in Fig. 2. Every tuple User 0 User 1 User 2 User 3

User 0

I	J	V
0	0	2
1	4	2
2	5	1
3	1	1
4	4	1
4	5	1

User 1

I	J	V
0	1	1
1	1	1
1	3	1
4	5	1

User 2

I	J	V
0	2	2
1	4	1
2	1	2
3	5	1
5	4	2

User 3

I	J	V
1	3	1
1	4	1
2	5	1
4	4	2

User 4

I	J	V
6	0	1

User 5

I	J	V
6	3	1
6	6	1

User 6

I	J	V
6	2	1
7	13	1

User 7

I	J	V
6	1	1
6	7	1

User 8

I	J	V
6	4	1
6	5	1

User 9

I	J	V
6	8	1

User 10

I	J	V
7	9	1

User 11

I	J	V
7	0	1
7	6	1
7	9	1
7	10	1
7	11	1
7	12	2
7	13	2
7	4	2

User 12

I	J	V
7	9	1

User 13

I	J	V
7	13	1

Fig 3. Generating three tuple matrix

Table 3. List of Influence degree for 3 tuple array

User 0	2 1 0 0 1 1 0 0 0 0 0 0 0 0 0
1	0 1 0 1 0 1 0 0 0 0 0 0 0 0 0
2	0 2 2 0 1 1 0 0 0 0 0 0 0 0 0
3	0 0 0 1 1 1 0 0 0 0 0 0 0 0 0
4	1 0 0 0 0 0 0 0 0 0 0 0 0 0 0
5	0 0 0 1 0 0 1 0 0 0 0 0 0 0 0
6	0 0 1 0 0 0 0 0 0 0 0 0 0 1 0
7	0 1 0 0 0 0 0 1 0 0 0 0 0 0 0
8	0 0 0 0 1 1 0 0 0 0 0 0 0 0 0
9	0 0 0 0 0 0 0 0 1 0 0 0 0 0 0
10	0 0 0 0 0 0 0 0 0 1 0 0 0 0 0
11	1 0 0 0 0 0 1 0 0 1 1 1 2 1 2
12	0 0 0 0 0 0 0 0 0 1 0 0 0 0 0
13	0 0 0 0 0 0 0 0 0 0 0 0 0 1 0

Efficient Techniques for Clustering of Users … 389

Table 4 Farthest first and X-means clustering results

K	C_1	C_2	C_3	C_4	C_5	C_6	C_7	C_8	C_9	C_{10}	C_{11}	C_{12}	C_{13}	C_{14}	C_{15}	C_{16}	C_{17}	C_{18}	C_{19}	C_{20}	C_{21}	C_{22}	C_{23}	C_{24}	C_{25}
Farthest first																									
10	165	1	5	1	1	1	2	2	4	3															
15	159	1	2	1	1	1	1	1	2	3	1	3	3	2	4										
25	149	1	1	1	1	1	1	1	1	2	1	2	1	2	3	4	2	3	2	1	1	1	1	1	1
X-means																									
10	52	8	10	7	12	6	6	7	12	65															
15	51	8	4	5	7	5	7	2	4	6	7	7	6	15	51										
25	42	2	14	8	10	9	1	1	4	5	30	12	6	5	1	6	6	4	7	5	1	6			

(row) contains three fields: i-row index of sparse matrix; j-column index of sparsematrix; v-value of the cell with index (i, j); This eliminates the zero in sparse matrix and contains only nonzero visits of the webpage with respect to the session and web page index.

From Fig. 3, three tuple for user 0 has 6 rows and three columns which have nonzero values of v (value in the cell). Consider the first row 3-tuple of user-0, it indicates 0th page visited in 0th session two times. The second row in the 3-tuple indicates in session-1 page-4 is visited twice. This is followed to all the three tuple

Obtained from the sparse matrix of all users.

4.4 Influence Degree

Influence degree of 13 web users with 15 accessed web pages is given in Table 3.Influence degree (ID) is the mean value of accessed times of one web page in one web user. We calculated influence degree of each web page from 3-tuple representation using Eq. (3).

$$ID_j = sum(j, v) / N_j, \qquad \text{------------} \qquad (3)$$

Where j is the column index of nonzero value in 3-tuple representation, Sum (j,v) is the sum of all values v with same j value, Nj is the number of value j. After calculation of influence degree of all web users, we get the matrix of influence degree with the row value of each influence degree of each web user. The MID obtained for the data is represented in Table 3.

Table 5 Comparison of clusters for k = 10

Algorithm	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10
K-means	16	8	117	4	7	6	1	10	4	12
Farthest first	165	1	5	1	1	1	2	2	4	3
X-means	52	8	10	7	12	6	6	7	12	65

Table 6 Comparison of clusters for k = 15

Algorithm	C_1	C_2	C_3	C_4	C_5	C_6	C_7	C_8	C_9	C_{10}	C_{11}	C_{12}	C_{13}	C_{14}	C_{15}
K-means	13	8	97	2	7	3	1	8	4	13	7	5	7	4	6
Farthest first	159	1	2	1	1	1	1	1	2	3	1	3	3	2	4
X-means	51	8	4	5	7	5	7	2	4	6	7	7	6	15	51

Table 7 Comparison of clusters for k = 25

Algorithm	C_1	C_2	C_3	C_4	C_5	C_6	C_7	C_8	C_9	C_{10}	C_{11}	C_{12}	C_{13}	C_{14}	C_{15}	C_{16}	C_{17}	C_{18}	C_{19}	C_{20}	C_{21}	C_{22}	C_{23}	C_{24}	C_{25}
Farthest first	149	1	1	1	1	1	1	1	1	2	1	2	1	2	3	4	2	3	2	1	1	1	1	1	1
K-means	13	3	69	2	7	3	1	9	3	11	7	5	6	3	5	6	5	6	2	2	7	2	4	3	1
X-means	42	2	14	8	10	9	1	1	4	5	30	12	6	5	1	6	6	4	7	5	1	6			

4.5 Clustering of MID

Clustering is defined as grouping of similar type of web objects such that the web objects within a same group are similar (or related) to one another group. We performed our proposed approach MID with different clustering algorithms to analyze the effect of grouping the similar web users.

Initially farthest first, X-means clustering algorithms are applied over the obtained MID for required number of clusters. Obtained clustered instances are summarized in Table 4 as follows:

Thus from the results farthest first technique takes less time to build the model and X-means yields optimal number of cluster instances.

V. PERFORMANCE EVALUATION

The performance evaluation of the system gives the clustering of the web users, it is considered over the three clustering techniques: Simple K-means, farthest first and X-Means. Experiments are performed over 1000 records of web log which consist of 185 users, 40 sessions, and 25 web pages. This entire record will be converted into MID by going through many transformations from sparse matrix form.

Final MID obtained is of order 185 by 25. Clustering techniques are applied over the matrix MID for desired k value. Initially, simple k-means is applied for k values 10, 15, 25, and for the same values of k, farthest first, X-means clustering techniques are applied on the same MID and the results are compared which are shown in Tables 5, 6, and 7 as follows:

From the results, compared with existing K-means approach, farthest first technique yields the clusters in less time and optimal clustered instances are obtained through X-means which is shown in Fig. 4.

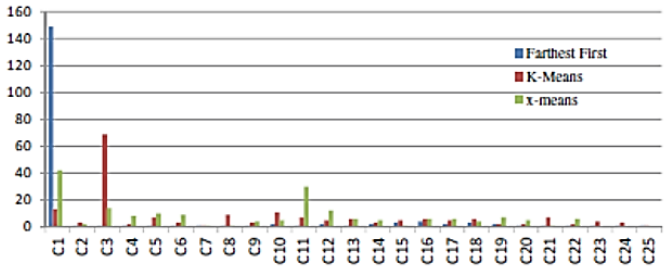

Fig. 4 Graphical representation of clustering techniques (k = 25)

VI. CONCLUSION

We planned an approach to cluster the net users by transforming the weblog information into a thin matrix, and then the three-tuple matrix is built by removing unused webpages. Influence degree of every web content is calculated for one user. Finally, mid is created for all internet users. The planned approach is applied to generate mid to create the best variety of clusters using X-Means and farthest initial technique for clump the net users in less time. In future, our planned approach can be extendable for finding best navigation methods using Association Rule Mining.

VII. REFERENCES

[1]. S. Jagan, Dr. S.P. Rajagopalan. A Survey on Web Personalization of Web Usage Mining, International Research Journal of Engineering and Technology (IRJET), Volume: 02 Issue: 01 | March-2015. pp. 6-12.

[2]. Rachit Adhvaryu, A Review Paper on Web Usage Mining and Pattern Discovery, Journal Of Information, Knowledge And Research In Computer Engineering, Volume - 02, Issue - 02nov 12 To Oct 13, pp. 279-284.

[3]. V. Vidyapriya, S. Kalaivani, An Efficient Clustering Technique for Weblogs, IJISET - International Journal of Innovative Science, Engineering & Technology, Vol. 2 Issue 7, July 2015. pp. 516-525.

[4]. B. Uma Maheswari, Dr. P. Sumathi, A New Clustering and Pre-processing for Web Log Mining, 2014 World Congress on Computing and Communication Technologies, IEEE,pp. 25-29.

[5]. Anupama D. S. & Sahana D. Gowda, Clustering Of Web User Sessions To Maintain Occurrence Of Sequence In Navigation Pattern, Second International Symposium on Computer Vision and the Internet (VisionNet'15), Elsevier 2015, pp. 558-564.

[6]. V. Chitraa, Antony Selvadoss Thanamani. Web Log Data Analysis by Enhanced Fuzzy C Means Clustering, International Journal on Computational Sciences & Applications (IJCSA), Vol. 4 Issue No. 2, April 2014, pp. 81-95.

[7]. J. HuaXu, H. Liu, Web User Clustering Analysis based on KMeans Algorithm, 2010 International Conference on Information Networking and Automation (ICINA) IEEE, pp. 6-9.

[8]. K. Santhisree, Dr A. Damodaram, S. Appaji, D. Nagarjuna Devi, Web Usage Data Clustering using Dbscan algorithm and Set similarities, 2010 International Conference on Data Storage and Data Engineering, IEEE, pp. 220-224.

[9]. Xidong Wang, Yiming Ouyang, Xuegang Hu, Yan Zhang, Discovery of User Frequent Access Patterns on Web Usage Mining, The 8th International Conference on Computer Supported Cooperative Work in Design Proceedings, 2003 IEEE, pp. 765-769.

[10]. LI Wei, ZHU Yu-quan, CHEN Geng, YANG Zhong, Clustering of Web Users Based on Competitive Agglomeration, 2008 International Symposium on Computational Intelligence and Design, IEEE, pp. 515-519.

[11]. Xinran Yu & Turgay Korkmaz, Finding the Most Evident Co-Clusters on Web Log Dataset Using Frequent Super Sequence Mining., 2014 August, 13-15, San Francisco, California, USA. pp. 529-536.

[12]. T. Nadana Ravishankar & Dr. R. Shriram, Mining Web Log Files Using Self-Organizing Map and K-Means Clustering Methods, ICAREM.

[13]. Mohammed Hamed Ahmed Elhiber & Ajith Abraham., Access Patterns in Web Log Data: A Review. Journal of Network and Innovative Computing, ISSN 2160-2174, Volume 1 (2013),pp. 348-355.

[14]. J. Xiao, Y. Zhang, X. Jia & T. Li, Measuring Similarity of Interests for Clustering Web-Users, Proc. of the 12th Australian Database Conference 2001 (ADC'20OI) IEEE, Australia, 29 January - 2 February, 2001, pp. 107-114.
[15]. V. Sujatha & Punitha Valli, Improved User Navigation Pattern Prediction Technique From Web Log Data, International Conference on Communication Technology and System Design, 2011 Published by Elsevier Ltd, pp. 92-99.
[16]. J. Xiao, Y. Zhang, Clustering of Web Users Using Session-based Similarity Measures, Proc.of the 12th Australian Database Conference 2001 (ADC'20OI) IEEE, Gold Coast, Australia, 29 January - 2 February, 2001, pp. 223-228.