International Research Journal of Emerging Trends in Multidisciplinary
ISSN 2395-4434
Volume 1, Issue 5 July 2015
www.irjetm.com

Web Log Mining Based on Soft Clustering and Multi-Objective Genetic Algorithm

Mr. Harish Bhabad[1], Prof. Pankaj Kawadkar[2]

[1]Computer Science Engg.Department,
Patel Institute of Engg. Science,
Ratibad,Bhopal,Madhya Pradesh

[2]Assistant Professor,
Head of the Department Computer Science Engg.,
Patel Institute of Engg. Science,
Ratibad,Bhopal, Madhya Pradesh

Abstract:*Web mining can be broadly defined as discovery andanalysis of useful information from the World Wide Web. WebUsage Mining can be described as the discovery and analysis ofuser accessibility pattern, during the mining of log files andassociated data from a particular Web site, in order to realizeand better serve the needs of Web-based applications.*
Web log mining is important area of research for the improvement of web efficiency and log cache enhancement. web log mining various method of data mining is applied one such method is called clustering. Clustering is unsupervised learning technique of data mining. The form of this clustering is k-means and k-median algorithm, but these algorithm are suffered some point of problem now used soft clustering technique such as fuzzy clustering algorithm. Web mining are classify into three domains: content, structure and usage mining.
The purpose of this paper is to provide optimum initial solution for FCM with the help of genetic algorithm For the improvement of FCM clustering technique used multi-objective genetic algorithm for better generation of clustering technique. In this paper discuss FCM algorithm, multi-objective genetic algorithm.

Keywords:*WebMining, Genetic Algorithm,clustering technique, Fuzzy C-means.*

1. INTRODUCTION

The World Wide Web has huge amount information andlarge datasets are available in databases. Informationretrieving on websites is one of possible ways how to extractinformation from these datasets.Web mining is the extraction of interesting and useful knowledge and implicit information fromartifacts or activity related to the WWW. Based on several research studies we can broadly classify Webmining into three domains: content, structure and usage mining.Web mining does not only mean applying data miningtechniques to the data stored in the Web. The algorithmshave to be modified to better suit the demandsof the Web. New approaches should be usedbetter fitting to the properties of Web data. Furthermore,not only data mining algorithms, but alsoartificial intelligence, information retrieval and naturallanguage processing techniques can be used efficiently. The given figure 1 shows the description technique of web mining.

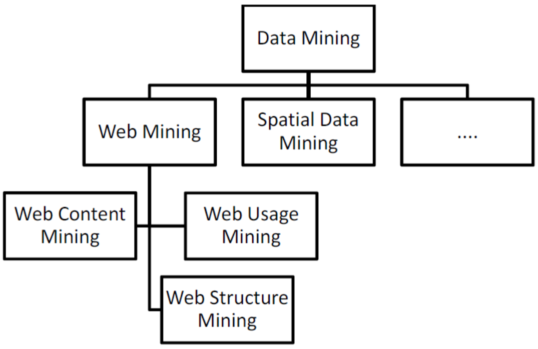


Figure 1. Description of web miming category

1.1 Web Mining: Web mining consists of a set operations defined on data residing on WWWdata servers. Such data can be thecontent presented to users of the web sites such as hyper text markup language(HTML) files, images, text, audio or video.Web mining is mainly categorized into two subsets namely web contentmining and web usage mining.

a)Web Content Mining :
Web content mining describes the automatic search of informationresources available on-line. The focus is on the content of web pagesthemselves.

b)Web Structure Mining :
Web structure mining is theprocess of discovering the structure of hyperlinkswithin the Web. Practically, while Web content miningfocuses on the inner-document information, Webstructure mining discovers the link structures at theinter-document level.

c)Web Usage Mining:
Usage mining as thename implies focus on how the users of websites interact with web site, the webpages

visited, the order of visit, timestamps of visits and durations of them.The main source of data for the web usage mining is the server logs whichlog each visit to each web page with possibly IP, referrer, time, browser andaccessed page link. Although many areas and applications can be cited whereusage mining is useful, it can be said the main idea behind web usage mining is tolet users of a web site to use it with ease efficiently, predict and recommend partsof the web site to user based on their and previous user's actions on the web site. Figure 2.The General architecture of web uses mining.

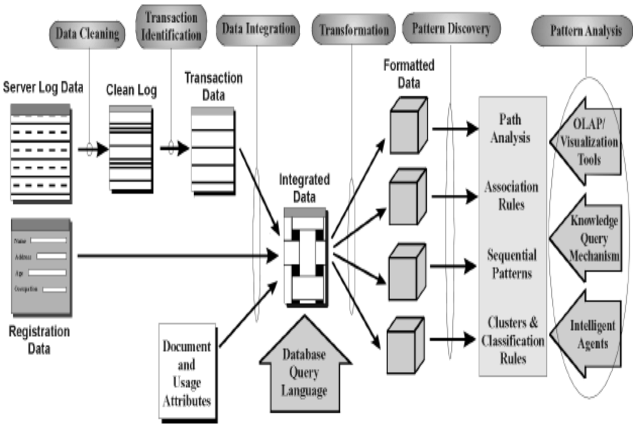


Figure 2: General Architecture for Web Usage Mining

1.2 Web Usage Mining and Pattern Discovery: Web usage mining is the application of data mining techniques to discover usage pattern from Web data, in order to understand and better serve the needs of Web-based applications. Web usage mining consists of three phases, namely preprocessing, pattern discovery, and pattern analysis. A high level Web usage mining Process is presented in Figure 3.

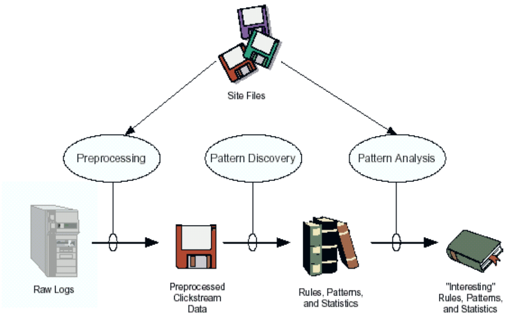

Site Files
Preprocessing
Pattern Discovery
Pattern Analysis
Raw Logs
Preprocessed Clickstream Data
Rules, Patterns, and Statistics
"Interesting" Rules, Patterns, and Statistics

Figure 3. High Level web usage Mining Process

2. RELATED WORK.

There has been an extensive studied of various research paper on fuzzy c-means algorithm. And the various Initialization methods of FCM clustering algorithm have been more emphasized in this literature.

2.1) Initialization method for k-mean algorithm:

Fuzzy clustering plays an important role in solving problems in the areas of pattern recognition and web log mining. A variety of fuzzy clustering methods have been proposed and most of them are based upon distance criteria [6]. One widely used algorithm is the fuzzy c-means (FCM) algorithm. It uses reciprocal distance to compute fuzzy weights. A more efficient algorithm is the new FCFM. It computes the cluster center using Gaussian weights, uses large initial prototypes, and adds processes of eliminating, clustering and merging. In the following sections we discuss and compare the FCM algorithm and FCFM algorithm. The fuzzy c-means (FCM) algorithm was introduced by J. C. Bezdek [2]. The idea of FCM is using the weights that minimize the total weighted mean-square error:

$$J(w_{qk}, \mathbf{z}^{(k)}) = \Sigma_{(k=1,K)} \Sigma_{(k=1,K)} (w_{qk}) || \mathbf{x}^{(q)} - \mathbf{z}^{(k)} ||^2 \quad (1)$$

$$\Sigma_{(k=1,K)} (w_{qk}) = 1 \text{ for each } q$$

$$w_{qk} = (1/(D_{qk})^2)^{1/(p-1)} / \Sigma_{(k=1,K)} (1/(D_{qk})^2)^{1/(p-1)} , p > 1 \quad (2)$$

2.2) Random Sampling Methods:

Probably being most widely adopted in the literature, random sampling methods follow a naive way to initialize the seed clusters, either using randomly selected input samples, or random parameters non-heuristically generated from the inputs. Being one of the earliest references in the literature, Forgy in 1965 [10] adopted uniformly random input samples as the seed clusters. The method, named R-SEL in our study, is formalized below.

R-SEL:
For i= 1.....K, set c_i= $x_r$such that
r =uniRand(1*;N*),
Where Nis the total number of input
Samples and uniRand(min; max) is a uniform random generator producing r*ε*[*min; max*].

2.3) Distance Optimization Methods:

Recognizing the characteristics of many clustering methods is to locally minimize the intra-cluster variances without optimizing the inter-cluster separation; it is a natural consideration to optimize the distances among the seed clusters beforehand towards a satisfactory inter-cluster separation in the output. Among some early practices, the Simple Cluster Seeking (SCS) initialization method [12] is adopted in the FASTCLUS
Procedure, which is a K-Means variance implemented in SAS. The SCS method is summarizes below.

SCS:

(1) Initialize the first cluster centroid with the first input, i.e. $c_1 = x_1$.

(2) For j= 2…….*N*, if $||x_j - c_k|| > p$ for all existing seed clusters c_k,

Where *p*is a threshold, and then add$x_j$as a new seed Cluster.

Stop when *K* seed clusters are initialized.

(3) If after scanning all input samples, there are less than *K* seed

Clusters generated, and then decrease *p*and repeat 1 - 2.

2.4) Density Estimation Methods:

This category of initialization methods is based on the assumption that the input data follow a Gaussian mixture distribution. Hence by identifying the dense areas of the input domain, the initialized seed clusters help the clustering method in creating compact clusters.

3. Problem Statement

The attractiveness of the FCM lies in its simplicity and flexibility. In spite of other algorithms being available, k-means continues to be an attractive method because of its convergence properties. However, it suffers from major shortcomings that have been a cause for it not being implemented on large datasets. The most important among these are

(I)The Number of Cluster is unknown in K-mean clustering algorithm it specifies by user at run time.

(II)The initial center problem because clusters obtained depend heavily on initial centers [9].

(III) K-means is slow and scales poorly with respect to the time it takes for large number of points.

Because of these shortcomings, proposed efficient FCM clustering algorithm is required to overcome above shortcomings

4. PROPOSED METHOD

The process of FCM clustering and multi-objective function takes several process such as population, fitness function, mutation and crossover for propagation of clustering algorithm. Some steps are divided into six phase.

Initializing population

The initial population is done by determining the length of chromosome with size K x d. K is the number of chromosome a lot of d while d is the dimension of the cluster variablesFitness function

K-Means clustering optimization with multi-objective genetic algorithm uses 2 objectives, i.e. minimizing error functions within each cluster (equation 1) and maximizing the centeid value between cluster (equation 2). The calculations used as follows.

$$\sigma_t^2 = \frac{1}{N}\sum_{j=1}^{N} \sigma_{ij}^2 \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots..1)$$

Where i = 1, 2... K; K is the number of cluster

$$\sigma^2 \sum_{i=1}^{K} \sigma_i^2 \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots..2)$$

Whereas:

σ_i^2 : Error in $i -^{th}$ cluster

n_i : number of data in banyak data padai $i -^{th}$ cluster

x_{ij} : Data in i-cluster, $j -^{th}$variabe

z_{ij} : i-cluster average in j-variable

V : Number of variable

The first function is to minimize error average in cluster which formulated as follow:

$$V(w) = \frac{1}{k}\sum_{i=1}^{k} \sigma^2 i \quad ,i = 1,2 \ldots\ldots k \qquad (3)$$

Whereas:

V (w) : Error in cluster

k : Number of cluster

The second function is to maximize inter-cluster error which formulated as follow:

$$V(b) = \frac{1}{k}\sum_{i=1}^{k}\sum_{j=1}^{v}(z_{ij} - \bar{z}_j) \qquad 4)$$

dengan:

V (b) : Error in cluster

z_{ij} : i-cluster average in variable

z̄ : Grand mean of $j -^{th}$ variable

4.1)Proposed Architecturefor processing of clustering

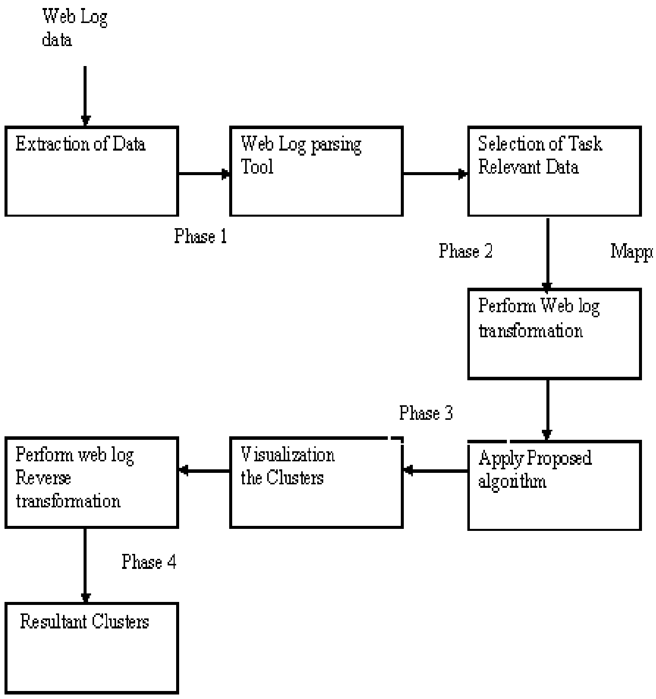

Web Log data
Extraction of Data
Web Log parsing Tool
Selection of Task Relevant Data
Phase 1
Phase 2
Mapp
Perform Web log transformation
Phase 3
Perform web log Reverse transformation
Visualization the Clusters
Apply Proposed algorithm
Phase 4
Resultant Clusters

Fig. 4 Proposed Architecture for Processing of Clustering

Phase 1:
Resource extraction is the process of retrieving the desired web log data files from the web server. These access log files contain information in CERN (Common Log Format).

In this phase of our work:

1. Extract web log data file from web server which contains some common fields:

- User's IP address
- Access date and time
- Request method (GET or POST),
- URL of the page accessed,
- Transfer protocol (HTTP 1.0, HTTP 1.1,),
- Success of return code.
- Number of bytes transmitted.

2. Give this web log data file as an input to web log parsing tool: Web Log Explorer

Web Log Explorer takes your log file, parse it and build reports by grouping or filtering the extracted data. Then we explore page view request in resulting table and export this report in CSV File (Excel).

3. Select task relevant data from excel file. This task relevant data contain 2 fields:

- Web pages
- Frequency of web page

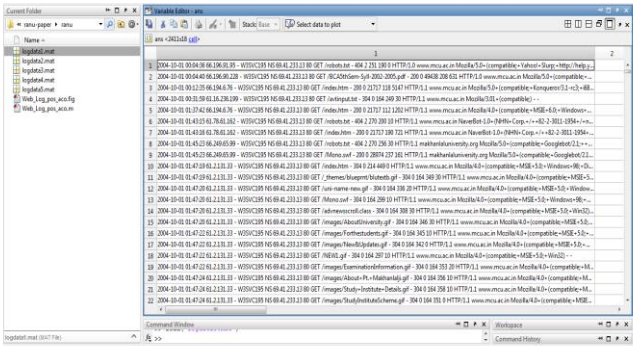
Phase 2:
Proposed algorithm work on only numerical data so mapping is performed on web pages. And assign every web page to unique numerical value.

Phase3:
After phase 2 proposed clustering algorithm is performed .this algorithm is completed in two stage in first stage according threshold value number of cluster and their centroid is generated. And in second stage according the similarity between the clusters they merge into a one cluster.

Phase 4:
After forming the cluster by proposed k-mean clustering algorithm. Remapping the each unique numerical value to web page. So they will form number of cluster according frequency of web pages. And by this algorithm we mining which web pages is highly accessed by client.

5. EXPERIMENT AND RESULT

Web log data used FCM and FCM_MOGA clustering algorithm implement in MATLAB 7.8.0 and found the similar pattern of cluster for IP, session, time and data. The cluster found in color group. The formation of cluster gives the information of valid and invalid cluster according to cluster valid index.For the performance evaluation of FCM technique and our FCM-MOGA used MATLAB software package. MATLAB is a software package for high- performance numerical computation and visualization.For the processing of clustering used university web log files in five sections in different size of data. Result table and graphs are as follows.



```
                                                        1                                                                        2
1  2004-10-01 00:04:36 66.196.91.95 - W3SVC195 NS 69.41.233.13 80 GET /robots.txt - 404 2 251 190 0 HTTP/1.0 www.mcu.ac.in Mozilla/5.0+(compatible;+Yahoo!+Slurp;+http://help.y...
2  2004-10-01 00:04:40 66.196.90.228 - W3SVC195 NS 69.41.233.13 80 GET /BCA5thSem-Syll-2002-2005.pdf - 200 0 49438 208 631 HTTP/1.0 www.mcu.ac.in Mozilla/5.0+(compatible;+...
3  2004-10-01 00:12:35 66.194.6.76 - W3SVC195 NS 69.41.233.13 80 GET /index.htm - 200 0 21717 118 5147 HTTP/1.1 www.mcu.ac.in Mozilla/5.0+(compatible;+Konqueror/3.1-rc3;+i68...
4  2004-10-01 00:31:59 61.16.236.199 - W3SVC195 NS 69.41.233.13 80 GET /avtinput.txt - 304 0 164 249 30 HTTP/1.1 www.mcu.ac.in Mozilla/3.01+(compatible) - -
5  2004-10-01 01:37:42 66.194.6.76 - W3SVC195 NS 69.41.233.13 80 GET /index.htm - 200 0 21717 112 1202 HTTP/1.1 www.mcu.ac.in Mozilla/4.0+(compatible;+MSIE+6.0;+Windows+...
6  2004-10-01 01:43:15 61.78.61.162 - W3SVC195 NS 69.41.233.13 80 GET /robots.txt - 404 2 270 200 10 HTTP/1.1 www.mcu.ac.in NaverBot-1.0+(NHN+Corp.+/++82-2-3011-1954+/+n...
7  2004-10-01 01:43:16 61.78.61.162 - W3SVC195 NS 69.41.233.13 80 GET /index.htm - 200 0 21717 190 721 HTTP/1.1 www.mcu.ac.in NaverBot-1.0+(NHN+Corp.+/++82-2-3011-1954+...
8  2004-10-01 01:45:23 66.249.65.99 - W3SVC195 NS 69.41.233.13 80 GET /robots.txt - 404 2 270 256 30 HTTP/1.1 makhanlaluniversity.org Mozilla/5.0+(compatible;+Googlebot/2.1;++...
9  2004-10-01 01:45:23 66.249.65.99 - W3SVC195 NS 69.41.233.13 80 GET /Mono.swf - 200 0 28874 237 161 HTTP/1.1 makhanlaluniversity.org Mozilla/5.0+(compatible;+Googlebot/2.1...
10 2004-10-01 01:47:19 61.2.131.33 - W3SVC195 NS 69.41.233.13 80 GET /index.htm - 304 0 214 449 0 HTTP/1.1 www.mcu.ac.in Mozilla/4.0+(compatible;+MSIE+5.0;+Windows+98;+Di...
11 2004-10-01 01:47:19 61.2.131.33 - W3SVC195 NS 69.41.233.13 80 GET /_themes/blueprnt/blutextb.gif - 304 0 164 349 30 HTTP/1.1 www.mcu.ac.in Mozilla/4.0+(compatible;+MSIE+5...
12 2004-10-01 01:47:20 61.2.131.33 - W3SVC195 NS 69.41.233.13 80 GET /uni-name-new.gif - 304 0 164 336 20 HTTP/1.1 www.mcu.ac.in Mozilla/4.0+(compatible;+MSIE+5.0;+Window...
13 2004-10-01 01:47:20 61.2.131.33 - W3SVC195 NS 69.41.233.13 80 GET /Mono.swf - 304 0 164 299 10 HTTP/1.1 www.mcu.ac.in Mozilla/4.0+(compatible;+MSIE+5.0;+Windows+98;+...
14 2004-10-01 01:47:20 61.2.131.33 - W3SVC195 NS 69.41.233.13 80 GET /advnewsscroll.class - 304 0 164 308 30 HTTP/1.1 www.mcu.ac.in Mozilla/4.0+(compatible;+MSIE+5.0;+Win32)...
15 2004-10-01 01:47:20 61.2.131.33 - W3SVC195 NS 69.41.233.13 80 GET /images/AboutUniversity.gif - 304 0 164 346 30 HTTP/1.1 www.mcu.ac.in Mozilla/4.0+(compatible;+MSIE+5.0;...
16 2004-10-01 01:47:22 61.2.131.33 - W3SVC195 NS 69.41.233.13 80 GET /images/Forthestudents.gif - 304 0 164 345 10 HTTP/1.1 www.mcu.ac.in Mozilla/4.0+(compatible;+MSIE+5.0;+...
17 2004-10-01 01:47:22 61.2.131.33 - W3SVC195 NS 69.41.233.13 80 GET /images/New&Updates.gif - 304 0 164 342 0 HTTP/1.1 www.mcu.ac.in Mozilla/4.0+(compatible;+MSIE+5.0;+...
18 2004-10-01 01:47:22 61.2.131.33 - W3SVC195 NS 69.41.233.13 80 GET /NEW1.gif - 304 0 164 297 10 HTTP/1.1 www.mcu.ac.in Mozilla/4.0+(compatible;+MSIE+5.0;+Win32) - -
19 2004-10-01 01:47:22 61.2.131.33 - W3SVC195 NS 69.41.233.13 80 GET /images/ExaminationInformation.gif - 304 0 164 353 20 HTTP/1.1 www.mcu.ac.in Mozilla/4.0+(compatible;+M...
20 2004-10-01 01:47:24 61.2.131.33 - W3SVC195 NS 69.41.233.13 80 GET /images/About+Pt.+Makhanlalji.gif - 304 0 164 356 10 HTTP/1.1 www.mcu.ac.in Mozilla/4.0+(compatible;+M...
21 2004-10-01 01:47:24 61.2.131.33 - W3SVC195 NS 69.41.233.13 80 GET /images/Study+Institute+Details.gif - 304 0 164 358 10 HTTP/1.1 www.mcu.ac.in Mozilla/4.0+(compatible;+M...
22 2004-10-01 01:47:24 61.2.131.33 - W3SVC195 NS 69.41.233.13 80 GET /images/StudyInstituteScheme.gif - 304 0 164 351 0 HTTP/1.1 www.mcu.ac.in Mozilla/4.0+(compatible;+MSIE...
```

```
>>
```

Figure 5. Dataset Web log entry

Seed value	Iteration			Error	
	FCM	FCM-MOGA	Number of cluster	Std. deviation	Computed error
3	8	9	6	4.56	6.00
4	7	9	6	3.92	5.00
5	8	6	6	3.53	5.00

Seed value	Iteration			Error	
	FCM	FCM-MOGA	Number of cluster	Std. deviation	Computed error
30	8	6	6	2.85	4.00
40	8	8	6	2.61	4.67
50	8	7	6	3.12	3.05

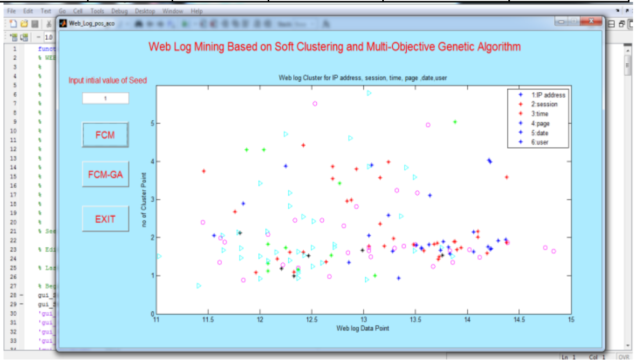

Web Log Mining Based on Soft Clustering and Multi-Objective Genetic Algorithm
Input initial value of Seed
FCM
FCM-GA
EXIT
Web log Cluster for IP address, session, time, page ,date,user
no of Cluster Point
Web log Data Point
1:IP address
2:session
3:time
4:page
5:date
6:user

Figure 6. GUI Cluster putting the value of seed by FCM method

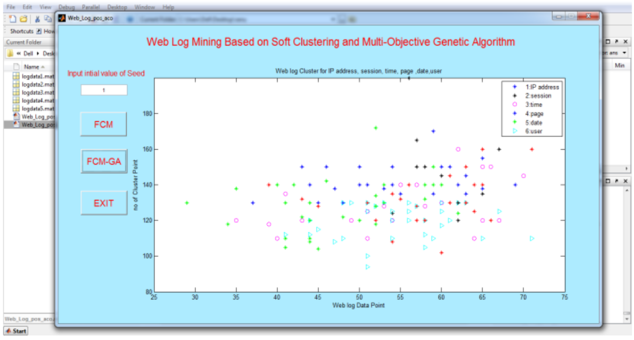
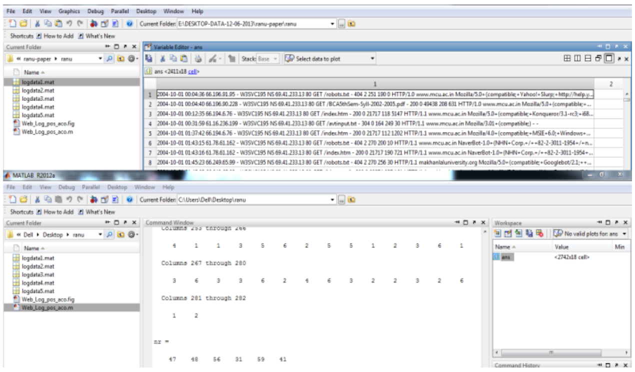

Web Log Mining Based on Soft Clustering and Multi-Objective Genetic Algorithm
Input initial value of Seed
FCM
FCM-GA
EXIT
Web log Cluster for IP address, session, time, page ,date,user
no of Cluster Point
Web log Data Point
1:IP address
2:session
3:time
4:page
5:date
6:user

Figure 7.GUI Cluster putting the value of seed by FCM –MOGA method



```
Current Folder: E:\DESKTOP-DATA-12-06-2013\ranu-paper\ranu
ans <2411x18 cell>
1 2004-10-01 00:04:36 66.196.91.95 - W3SVC195 NS 69.41.233.13 80 GET /robots.txt - 404 2 251 190 0 HTTP/1.0 www.mcu.ac.in Mozilla/5.0+(compatible;+Yahoo!+Slurp;+http://help.y...
2 2004-10-01 00:04:40 66.196.90.228 - W3SVC195 NS 69.41.233.13 80 GET /BCA5thSem-Syll-2002-2005.pdf - 200 0 49438 208 631 HTTP/1.0 www.mcu.ac.in Mozilla/5.0+(compatible;+...
3 2004-10-01 00:12:35 66.194.6.76 - W3SVC195 NS 69.41.233.13 80 GET /index.htm - 200 0 21717 118 5147 HTTP/1.1 www.mcu.ac.in Mozilla/5.0+(compatible;+Konqueror/3.1-rc3;+i68...
4 2004-10-01 00:31:59 61.16.236.199 - W3SVC195 NS 69.41.233.13 80 GET /avtinput.txt - 304 0 164 249 30 HTTP/1.1 www.mcu.ac.in Mozilla/3.01+(compatible;) - -
5 2004-10-01 01:37:42 66.194.6.76 - W3SVC195 NS 69.41.233.13 80 GET /index.htm - 200 0 21717 112 1202 HTTP/1.1 www.mcu.ac.in Mozilla/4.0+(compatible;+MSIE+6.0;+Windows+...
6 2004-10-01 01:43:15 61.78.61.162 - W3SVC195 NS 69.41.233.13 80 GET /robots.txt - 404 2 270 200 10 HTTP/1.1 www.mcu.ac.in NaverBot-1.0+(NHN+Corp.+/++82-2-3011-1954+/+n...
7 2004-10-01 01:43:16 61.78.61.162 - W3SVC195 NS 69.41.233.13 80 GET /index.htm - 200 0 21717 190 721 HTTP/1.1 www.mcu.ac.in NaverBot-1.0+(NHN+Corp.+/++82-2-3011-1954+...
8 2004-10-01 01:45:23 66.249.65.99 - W3SVC195 NS 69.41.233.13 80 GET /robots.txt - 404 2 270 256 30 HTTP/1.1 makhanlaluniversity.org Mozilla/5.0+(compatible;+Googlebot/2.1;++...
```

```
Current Folder: C:\Users\Dell\Desktop\ranu

    4     1     1     3     5     6     2     5     5     1     2     3     6     1

  Columns 267 through 280

    3     6     3     3     6     2     4     6     3     2     2     3     2     6

  Columns 281 through 282

    1     2


nr =

   47    48    56    31    59    41
```

Workspace: ans <2742x18 cell>

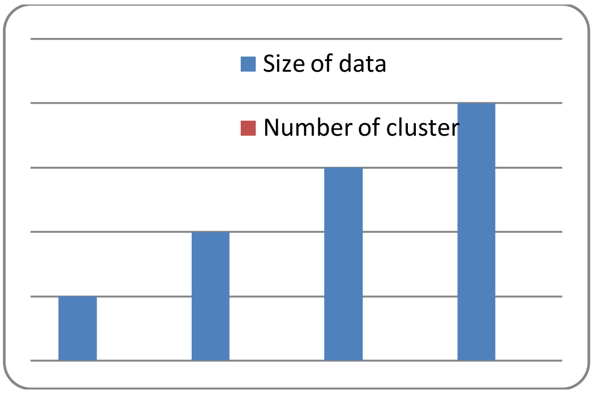
Figure 8. result window of matlab and computed value of number of iteration, Error and standard deviation

5.1 Comparative Result analysis of all data set with two methods.

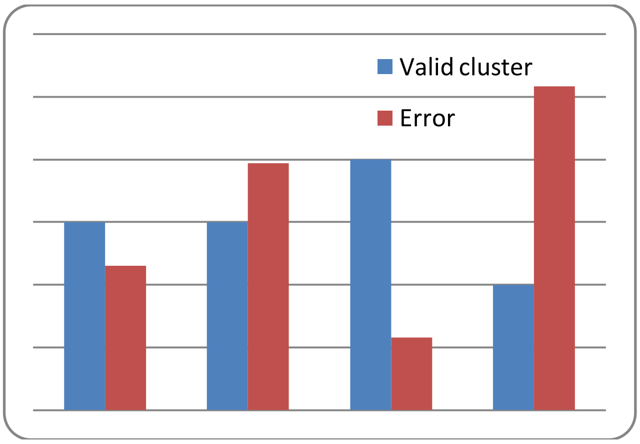
Table1. Performance evaluation of web dataset1

Table2. Performance evaluation of web dataset2

5.2Result Analysis of all web log dataset



Graph 1.shows that comparative cluster generation according to data size



Graph 2.shows that comparative valid cluster generation and generation of error according to cluster size

6. CONCLUSION

The new algorithm for FCM-MOGA clustering is proposed which efficiently overcome the major drawbacks viz. right number of cluster and initial seed (center point) problem. Proposed clustering algorithm is based on two specific factor, threshold factor which initial decide the number of cluster and specific factor which merge the clusters according the similarity. The careful selection of threshold value and specific factor which control merging of clusters yields efficient algorithmic results.

References

[1]Qingtian Han, XiaoyanGao ,Wenguo Wu, “Study On Web Mining Algorithm Based On Usage Mining”,

Computer- Aided Industrial Design And Conceptual Design, 2008.CAID/CD 2008. 9th International Conference On 22-25 Nov. 2008.

[2]Xuan SU ,Xiaoye WANG ,Zhuo WANG ,Yingyuan XIAZO, “An New Fuzzy Clustering Algorithm Based On Entropy Weighting”, Journal Of Computational Information Systems,3319-3326,2010.

[3]Mohanad Alata, Mohammad Molhim, And Abdullah Ramini, “Optimizing Of Fuzzy C-Means Clustering Algorithm Using GA” , World Academy Of Science, Engineering And Technology 39, 2008.

[4]K.Suresh, R.MadanaMohana, A.Ramamohan Reddy, “Improved FCM Algorithm For Clustering On Web Usage Mining”,IJCSI International Journal Of Computer Science Issues, Vol. 8, Issue 1, January 2011.

[5] TasawarHussain, Dr. SohailAsghar “A Hierarchical Cluster Based Preprocessing
Methodology For Web Usage Mining” In Ieee 2010.

[6]TasawarHussain, SohailAsgharAnd NayyerMasood” Hierarchical Sessionization At Preprocessing Level Of WUM Based On Swarm Intelligence” In 2010 6th International Conference On Emerging Technologies.

[7] Alam, S., G. Dobbie, Et Al. (2008). Particle Swarm Optimization Based Clustering Of Web Usage Data. 2008 IEEE/WIC/ACM International Conference On Web Intelligence And Intelligent
Agent Technology 978-0-7695-3496-1/08 2008 IEEE DOI 10.1109/WIIAT.2008.292 2008 IEEE/WIC/ACM International Conference on Web.

[8]Kumar, R. “Mining Web Logs: Applications And Challenges” KDD’09, June 28–July 1, 2009, Paris, France. ACM 978-1-60558-495 2009.

[9] Lu, H. And T. T. S. Nguyen “Experimental Investigation Of PSO Based Web User Session Clustering. 2009 International Conference of Soft Computing And Pattern Recognition “978-0- 7695-3879-2/09 IEEE 2009

[10] Makanju, A., A. N. Zincir-Heywood” Clustering Event Logs Using Iterative Partitioning”. KDD’09, June 28–July 1, 2009, Paris, France. ACM 978-1-60558-495,2009.

[11] Mr.V.K.Panchal, M. H. Kundra, "Comparative Study Of Particle Swarm Optimization Based Unsupervised Clustering Techniques." IJCSNS International Journal Of Computer Science And Network Security, VOL.9 No.10, October 2009.

[12] B. Pavel, “A Survey Of Clustering Data Mining Techniques,” In Grouping Multidimensional Data. Springer Berlin Heidelberg, 2006, Pp. 25–71.

[13] K. Suresh, R. MadanaMohana, A. Rama Mohan Reddy And A. Subramanyam, “Improved FCM Algorithm For Clustering On Web Usage Mining,” International Conference On Computer And Management (CAMAN), Pp. 1 – 4, 2011.

[14] Yanfeng Zhang, XiaofeiXuAndYunming Ye, “An Agglomerative Fuzzy K-Means Clustering Method With Automatic Selection Of Cluster Number,” 2nd International Conference On Advanced Computer Control (ICACC), Vol. 2, Pp. 32-38, 2010.