Improvement & Analysis of User Classification and Pattern Discovery in WUM

P.S.Shahu¹, D V Jamthe², A. A. Nikose³

Department of CSE

¹PG Student, Privadarshini Bhagwati College of Engineering, Nagpur. ^{2, 3}Assistant Professor, Priyadarshini Bhagwati College of Engineering, Nagpur.

Abstract- On web data, data mining has web mining as an application and web usage mining is an important component clustering algorithm which can of web mining. The foremost objective of web usage mining is to understand the web site usage behaviour that can be achieved through the process of data mining of Web Data and k-means clustering algorithm. The result of experiments web site user's behaviour through the process of data mining of depicts that with the improved k-means algorithm we get the Web Access Data. Facts obtain from the web using mining can be further used to improve web designs, introduce personalisation services and facilitate browsing effectively. An important application of web mining. The study has been done on this paper is to discuss various type of data preparation techniques to identify the unique user and user sessions. The paper shows the result of various clustering techniques for web log data using algorithm such as k-means and bisecting kmean. With respect to the same IP address and packet combination clusters are formed. From the analysis it can be concluded that Bisecting K-means gives the better result than that of the K-means algorithm.

I. INTRODUCTION

Web usage mining is application of data mining technique to discover automatic discover from particular web site. Web mining can be categories into three categories are shown in Fig.1, web content mining, web structure mining and web usage mining. Web content mining is referring to extract information of content of web pages. Web structure mining is used to analyze the links between web pages through the web structure to infer the knowledge. Web usage Mining is extracting the information from web log file which is accessed by users. The main aim of web usage mining is extract information from web pages which access by users. The Kmeans is very standard algorithm for the clustering process. But it has some disadvantages like cluster quality result depends on the selection of initial centroids, clusters produced are of different sizes, by using this method we also get empty 3.1 Bisecting K-means clusters. Bisecting k-means is a enhancement over basic kmeans algorithm. As it is based on k-means, it has the merits of k-means and also has some benefits over k-means.

II. RELATED WORK

In this paper [2] they explained the improved k-means define number of clusters automatically and assign required cluster to un-clustered points. The main focus of this paper to overcome the drawbacks of the better performance.

In this paper [3] they applied the improved k-means algorithm with respect to the various parameters like date, time, user agent and time taken on for the real data sets collected by the group of institution web servers. Also discussed that clustering is done group web users into same cluster based on users browsing behaviour during a specific time interval. The process cannot be completed until Pre-process is done properly.

In this paper [4] they presented a comparative analysis techniques such as support vector machine and artificial neural network with supervise learning technique. The Objective this is the Classification of user"s web access pattern accurately after data cleaning concept for the future improvement. They propose data classification result based on accuracy.

III. PROPOSED WORK

For Web log analysis web usage mining techniques can be applied. By analysing the web access log can understand the user behaviour and the web structure. In this proposed work we cluster the web log data by using the techniques k-mean and bisecting k-mean algorithm and evaluate their performance with respect to the parameters like delay (execution time) and accuracy (infected).

Algorithms and Design Modules

It is a combination of k-Means and hierarchicalclustering. Bisecting k-Means splits into two sub cluster at every bisecting stages instead of partitioning the data into 'k' clusters in each iteration. It is recurring until k clusters are found. Bisecting k-means is more effective when 'k' is large. At first, for the k-means algorithm, the computation includes every Bisecting step of Bisecting k-means, only one cluster expression we extract the IP address. data point and two centroids are arises. Second, Cluster of similar sizes are being produced by the Bisecting k-means, while k-means produce different sizes clusters.

3.2 Algorithm for finding clusters by using k-Means

- 1. Choose a Cluster for splitting.
- 2. Discover two sub clusters by using k-means algorithm.
- 3. Repetition of second step until to get the best matching similar cluster.
- 4. To reach the desire clusters repeat steps first, second and third.

There are many ways to choose the clusters to make it split.

Web log of the dataset are collected in the pdf format of the college which consist of various details .The details are not in the proper format. In the first step extraction of useful information is done .So here first pdf file read and then parsed. Extraction of IP addresses is done by the regular expansion method. To get the same IP address and packet combination at same time bisection of k-means is done, which is the application of clustering technique by k-means.

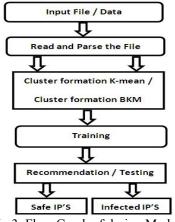


Fig 3: Flow Graph of design Module

IV. EXPERIMENTAL RESULT

Experiment is done by the web log file retrieved. The web log files are in the form of PDF format which is collected from the college.

Step1: Reading and Parsing

The code is executing on the Microsoft Visual Studio. In this c# languages is used. In the first step we have to collect dataset for the pre-processing. For this we maintain a pdf file to

all data point of the data set and k centroids. Besides that in collect all the records of the dataset. By using the regular

State West Leg für verste verste verste der State	0.1 . W.1.7	No. of Clusters 5	K-means	Train System (K-Means)	Apply SVM (K-Means)	
Original Wolk Light (red) In the U In the Open and	Select Web Log file		Pinesting			í
Durby Learners Rameriades Urage Report Park for 2012-01-221 Parent Park Park Park Park Park Park Park Park Park	Read & Parse	Clear DB	Disecting	Train System (BKM)	Apply SVM (BKM)	
Report Date for 2006-722 Senarof Dy Senarof Dy Senarof Dy Backvick by senarof for 2006-723 Backvick by senarof for 2006-723 Backvick by every senarof for 2006-723 Backvick by every senarof at 024 attack point of vioring events have shown a backvick backwick Backvick by every senarof at 024 attack point of vioring events have shown a backvick backwick Backvick by every senarof at 024 attack point of vioring events have senarof backvick backwick backwick Backvick by every senarof at 024 attack backwick backwick backwick backwick Mar Villianian 703.024 MB to sen Output data Dyna comparison of 1	Original Web Log file (pdf)					
177.15.1.90.hda Deer communicated: 1 Deer communicated: 1	Summary Bandwidth Summary for 20 Bandwidth reports display t Bandwidth reports are an id by the hour, day or over a pe	e amount of data transferred (eal starting point for viewing o riod of days. Additionally, you IBytes	verall bandwidth usage.	You can view bandwidth usage view		
Bytes communicated: 1 172, 16:130/inda Pres communicated: 1		lytes				
Bytes communicated: 3651650 172 16 25 50 USA	Max Utilization: 783.282 MI Output data	lytes				

Fig 4.1: Reading and parsing the pdf file

Step 2: Clustering using K-means

		K.means	Train System (K-Means)	Apply SVM (K-Means)	
Select Web Log file	No. of Clusters 5	a-means	man oyaca (n-arcans)	Apply of an (Restream)	
Read & Parse	Clear DB	Bisecting	Train System (BKM)	Apply SVM (BKM)	
Original Web Log file (pdf)					
Powered By Summary					
Bandwidth reports display (Bandwidth reports are an is by the hour, day or over a p Total Utilization: 2194.159	the amount of data transferred t deal starting point for viewing o eriod of days. Additionally, you MBytes	verall bandwidth usage.	You can view bandwidth usage view		
Bandwidth reports are an is	the amount of data transferred t deal starting point for viewing o eriod of days. Additionally, you MBytes	verall bandwidth usage.	You can view bandwidth usage view		,
Bandwidth reports display (Bandwidth reports are an is by the hour, day or over a p Total Utilization: 2194.189 Max Utilization: 783.282 M	the amount of data transferred t deal starting point for viewing o eriod of days. Additionally, you MBytes Bytes	verall bandwidth usage.	You can view bandwidth usage view		
Bandwidth reports display Bandwidth reports are an i by the hour, day or over a p Total Utilization: 2194.189 Max Utilization: 783.282 M Output data CLUSTE	the amount of data transforred t deal starting point for viewing o errod of days. Additionally, you MBytes Bytes R 1]	verall bandwidth usage.	You can view bandwidth usage view		
Bandwidth reports are an i Bandwidth reports are an i by the hour, day or over a p Total Utilization 2194.189 Max Utilization: 783.282 M Output data 	the amount of data transferred t deal tarting politic for viewing o erried of days. Additionally, you MBytes Bytes R 1] 2)	verall bandwidth usage.	You can view bandwidth usage view		

Fig 4.2(a): Formation of clusters using k-means

Fig 4.2(a) shows the result of the k-means clustering algorithm that forms the clusters.

	Pattern discovery :	and user class	ifcation using Web Usa	ge Mining	
Select Web Log file	No. of Clusters	K-means	Train System (K-Means)	Apply SVM (K-Means)	
Read & Parse	Clear DB	Bisecting	Train System (BKM)	Apply SVM (BKM)	
Original Web Log file (pdf)					
Powered By					
Bandwidth reports are an i by the hour, day or over a p Total Utilization: 38.068 M	the amount of data transferred ti leal starting point for viewing or eriod of days. Additionally, you o Bytes	erall bandwidth usage.	You can view bandwidth usage view		
Bandwidth Summary for 20 Bandwidth reports display Bandwidth reports are an i by the hour, day or over a p	the amount of data transferred ti leal starting point for viewing or eriod of days. Additionally, you o Bytes	erall bandwidth usage.	You can view bandwidth usage view		
Bandwidth Summary for 20 Bandwidth reports display Bandwidth reports are an i by the hour, day or over a p Total Utilization: 38.068 ME Max Utilization: 38.068 ME Output data	the amount of data transferred ti leal starting point for viewing or eriod of days. Additionally, you o Bytes	erall bandwidth usage.	You can view bandwidth usage view		
Bandwidth Summary for 20 Bandwidth reports display Bandwidth reports are an i by the hour, day or over a p Total Utilization: 38.068 ME Max Utilization: 38.068 ME Output data [CLUSTE]	the amount of data transferred ti leal starting point for viewing on eriod of days. Additionally, you o Bytes Sytes	erall bandwidth usage.	You can view bandwidth usage view		
Bandwidth Summary for 21 Bandwidth reports display Bandwidth reports are an i by the hour, day or over a p total Utilization: 38.068 ME Output data 	he amount of data transferred ti leal statisting point for viewing or eriod of days. Additionally, you o Bytes ytes 322	erall bandwidth usage.	You can view bandwidth usage view		
Bandwidth Summary for 21 Bandwidth reports display Bandwidth reports display Bandwidth reports display by the hour, day or over a p display display display display Max Utilization: 38.068 ME Output data [CLUSTE] 	he amount of data transferred il leal starting point for viewing ov entition of days. Additionally, you of ytes 22]	erall bandwidth usage.	You can view bandwidth usage view	-	
Badwidth Semmary for 21 Badwidth reports are an i Badwidth reports are an i Total Utilization: 38.069 M Max Utilization: 38.069 M Output data (CLUSTEI 	he amount of data transformed data transformed for for sering cop arried of days. Additionally, you of Systes 12] 13] 13]	erall bandwidth usage.	You can view bandwidth usage view	-	

Fig 4.2(b): Formation of various blank clusters using k-means

Fig 4.2(b) shows the output of the k-means clustering. Here we can see various blank clusters are formed by the algorithm.

Step 3: Clustering using Bisecting K-means

Select Web Log file	No. of Clusters B	K-means	Train System (K-Means)	Apply SVM (K-Means)	
Read & Parse	Clear DB	Bisecting	Train System (BKM)	Apply SVM (BKM)	
Original Web Log file (pdf)					
Daily Internet Bandwidth U: Report Date for: 2016-7-22 2 Powered By Summary	0.30				Ċ
Bandwidth reports display th Bandwidth reports are an id by the hour, day or over a pe Total Utilization: 2194.159 M	he amount of data transferred th eal starting point for viewing ov riod of days. Additionally, you o (Bytes	erall bandwidth usage.	You can view bandwidth usage view		
Bandwidth reports display the Bandwidth reports are an id by the hour, day or over a pe Total Utilization: 2194.159 M Max Utilization: 783.282 ME	he amount of data transferred th eal starting point for viewing ov riod of days. Additionally, you o (Bytes	erall bandwidth usage.	You can view bandwidth usage view		
Bandwidth reports display tl Bandwidth reports are an id by the hour, day or over a pe Total Utilization: 2194.159 M Max Utilization: 783.282 ME Output data	he amount of data transferred th eal starting point for viewing ov riod of days. Additionally, you o (Bytes	erall bandwidth usage.	You can view bandwidth usage view		
Bandwidth reports display (1) Bandwidth reports are an id by the hour, day or over a port Total Utilization: 2194.159 M Max Utilization: 783.282 ME Output data (CLUSTER 173.16.1.90,india(1)	ie amount of data transferred ti eal starting point for viewing ov riod of days. Additionally, you o (Bytes Bytes	erall bandwidth usage.	You can view bandwidth usage view		
Bandwidth reports are an id by the hour, day or over a pa Total Utilization: 2194.159 Max Utilization: 783.282 ME Output data	ie amount of data transferred ti eal starting point for viewing ov riod of days. Additionally, you o (Bytes Bytes	erall bandwidth usage.	You can view bandwidth usage view		

Fig 4 .3(a) : Formation of clusters using BKM

Select Web Log file	No. of Clusters B	K-means	Train System (K-Means)	Apply SVM (K-Means)
Read & Parse	Clear DB	Bisecting	Train System (BKM)	Apply SVM (BKM)
Original Web Log file (pdf)				
Daily Internet Bandwidth Us Report Date for: 2016-7-22 2 Powered By Summary	iage			
andwidth Summary for 201	6-7-22			
andwidth reports display th andwidth reports are an id y the hour, day or over a pe otal Utilization: 2194.159 M	e amount of data transferred eal starting point for viewing o riod of days. Additionally, you (Bytes	verall bandwidth usage.	You can view bandwidth usage view	
Bandwidth reports display th Bandwidth reports are an id-	e amount of data transferred eal starting point for viewing o riod of days. Additionally, you (Bytes	verall bandwidth usage.	You can view bandwidth usage view	
3andwidth reports display th 3andwidth reports are an id- y the hour, day or over a pe otal Utilization: 2194.159 M Max Utilization: 783.282 ME	ie amount of data transferred i eal starting point for viewing o riod of days. Additionally, you Bytes	verall bandwidth usage.	You can view bandwidth usage view	
Sandwidth reports display th Sandwidth reports are an id; y the hour, day or over a pe 'otal Utilization: 2194,159 M fax Utilization: 783.282 ME Dutput data [CLUSTER]	ie amount of data transferred i eal starting point for viewing o riod of days. Additionally, you Bytes	verall bandwidth usage.	You can view bandwidth usage view	-

Fig 4.3 (b) : Formation of clusters using BKM

Select Web Log file Ak at Uniter's a Apply SVM (BLM) Red & Parse Clar DB Bisseling Train System (BKM) Apply SVM (BLM) Priginal Web Log file (afd) Interved Base-bidd Varge Approx flapped for 2016-7-22 Worded By Warnary For 2016-7-22 Worde By Warnary Train System of faits transformed through one or more selected Societ/WALL appliances. Sachidd Argonics 2014-129 MBytes Netper data T216-20138/USA(2019) T216-20138/USA(2019) T216-20138/USA(2019) (CLUSTER 8)		attern uscovery		fcation using Web Usa	
Original Web Leg file (pdf)	Select Web Log file	No. of Clusters	K-means	Train System (K-Means)	Apply SVM (K-Means)
Jinterwet Basdwidd Utage goerd Das for 2016-7-22 words Dy umax7 word Dy coverd By umax7 sachidd rynom are are alsolarating point for forming overall basdwidth ange. You can view basdwidth ange steer audotidd rynom are are alsolarating point for forming overall basdwidth ange. You can view basdwidth ange steer audotidd rynom are are alsolarating point for forming overall basdwidth ange. You can view basdwidth ange steer audotidd rynom are are alsolarating point for forming avabuild rynom are are also avabuild rynom avabuild rynom are are also avabuild a	Read & Parse	Clear DB	Bisecting	Train System (BKM)	Apply SVM (BKM)
spore Ubas for: 2016-7-22 overend By overend By spore Toda	original Web Log file (pdf)				
72 16 20 38:US4(08 019) 72 16 20 38:US4(08 019) 72 16 20 38:US4(08 019) (CLUSTER 8)					
72 16 20 38 USA08 019)	landwidth Summary for 201 landwidth reports display th landwidth reports are an id- y the hour, day or over a pe otal Utilization: 2194.159 M fax Utilization: 783.282 ME	ee amount of data transferred eal starting point for viewing o riod of days. Additionally, you (Bytes	overall bandwidth usage.	You can view bandwidth usage view	
7216 20 38 USAG8 019)	Bandwidth Summary for 201 Bandwidth reports display th Bandwidth reports are an id by the hour, day or over a ge Fotal Utilization: 2194.159 M Max Utilization: 783.282 MB Dutput data	ee amount of data transferred eal starting point for viewing o riod of days. Additionally, you (Bytes	overall bandwidth usage.	You can view bandwidth usage view	
	Bandwidth Summary for 201 Bandwidth reports display th Bandwidth reports are an id by the hour, day or over a ge Fotal Utilization: 2194.159 M Max Utilization: 783.282 MB Dutput data	ee amount of data transferred eal starting point for viewing o riod of days. Additionally, you (Bytes	overall bandwidth usage.	You can view bandwidth usage view	
	Bandridih Summary for 201 Sandridih reports are an ib Sandridih reports are an ib Iotal Utilization: 2194.159 h Isas Utilization: 783.282 MB Dutpat data 172.16.20.38.USA(38.019) 172.16.20.38.USA(38.019)	es amount of data transformed al tatraing point dor viewing ag Bigstess ystes	overall bandwidth usage.	You can view bandwidth usage view	

Fig 4 .3(c): Formation of clusters using BKM

Fig 4.3(a), Fig 4.3(b) and Fig 4.3(c) shows the output executed by the bisecting k-means algorithm. In this we can clearly see that the formations of clusters are uniform. And here we not get any empty clusters.

Step4: Training the system

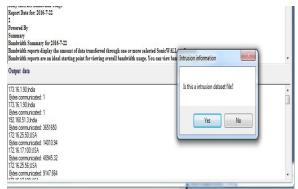


Fig 4.4(a): Training the system

For intrusion data set system asks the end user, as shown in Fig 4.4(a). If user selection is "yes", then the type value is labelled as "1" otherwise it is labelled as "0". Fig. 4.4(b) and Fig 4.4(c) shows respectively, the packets are infected and the normal packets stored in the database while training.

10	PEdt 3	é Copy	Delete	7732	- 1	173.16.1.90.india		+
	PEdt B	Copy	O Delete	7733	1	173.16.1.90:India	-1	1
10	PEdt 3	é Copy	Delete	7734	1	192.168.51.3.india	3651660	t
	P E dt 3	Copy	O Delete	7735	÷,	172.16.25.80,USA	14010.9	1
	Dat 3	Copy	Daleja	7736	1	172.16.25.56 USA	9147 66	1
E.I	PEdi S	Copy	Delete	1737	1	172.16.17.109.USA	18920.6	1
	PEdt 3	Copy	O Delete	7738	1	192.168.51.4.india	137938	1
	P Edit S	Copy	Delets	7739	1	172.16.17.113.USA	7115.65	1
11	PEdt 3	Copy	Delete	7740	1	172.16.25.42;USA	1739.31	1
	Edt.g	& Copy	Delete	7741	1	205.196.122.215.UK	257.052	1
101	PEdt 3	d Copy	Delete	7742	1	199.91.152.43, India	248.052	1
	PEdt 8	& Copy	Dvivte	7743	1	192.168.51.3.indua	3651650	1
	/ Edit 3	& Copy	Delete	7744	1	172.16.25.50,USA	14010.9	+
	Edt a	E Copy	Delete	7745	1	172 16 25 56 UBA	9147 66	1

4.4(b) Infected Packets

🥜 Edit	30	Copy	0	Delete	8135	3	172.16.8.203;USA	14	9
🥜 Edit	-		-				140.98.193.112;India	1.04.4	ġ
🥜 Edit	3.	Сору	0	Delete	8133	3	140.98.193.112;India	14	
🥜 Edit	34	Сору	0	Delete	8132	3	119.255.133.33;India	14	
🥜 Edit	3-6	Сору	0	Delete	8131	2	172.16.17.116;USA	416436	
🥜 Edit	3-6	Copy	0	Delete	8130	2	172.16.17.116;USA	416436	
🥜 Edi	3-6	Сору	0	Delete	8129	2	172.16.17.116;USA	416436	
🥜 Edit	34	Copy	0	Delete	8128	2	172.16.1.7;USA	4	
🥜 Edit	3.	Сору	0	Delete	8127	2	172.16.17.118;USA	4	
🥜 Edit	34	Сору	0	Delete	8126	2	172.16.6.19;USA	4	
🥜 Edit	3-6	Сору	0	Delete	8125	2	172.16.17.118;USA	4	
🥜 Edit	3-6	Сору	0	Delete	8124	2	172.16.17.116;USA	416436	
🥜 Edit	34	Сору	0	Delete	8123	2	172.16.17.118;USA	4	
🥜 Edit	3-6	Copy	0	Delete	8122	2	172.16.17.118;USA	4	

Fig4.4(c): Normal Packets



Fig4.4 (d): List of Safe IP'S

Output data	
NE FINAN KANNAN 172 16.20 46/USA	
172.16.20.46USA	
172.16.20.40.USA 192.168.51.3.India	
72.16.20.46/mda	
Infected Ps	
172.16.8.203;USA	
172.1620.31,USA	
Time needed:147418 ms	

Fig 4.4(e): List of Infected IP's

Fig 4.4(d) and Fig 4.4(e) computes the results of the infected and the safe IP's detected by the system.

V. COMPARISON

The objective of this is to analyze the web usage data by applying techniques such as K-means and Bisecting K-Means clustering algorithms. Here we consider the data file with the related cluster to perform the calculations. We compare the algorithm with respect to the parameter: time and accuracy.

Table 5.1: Comparison of algorithms with respect to time

K-mean(time in ms)	BKM(time in ms)
265045	254634
154288	151488
41622	13160
257444	257784
154818	150618
50722	12490
269765	279405
295326	222492
70654	44472

Fig 5.1: Comparison of K-means and Bisecting K-means (Delay)

Fig shows the comparison of both k-means and bisecting k-means algorithm with respect to the execution time required by them for the given input file. From the graph it can be concluded that the bisecting k-means algorithm give more efficient result then that of the k-means algorithm.

Table 5.2: Comparison of algorithms with respect to infected values (Accuracy)

curacy)
BKM (Accuracy in
%)
81.6
98.9
100
70.53

97.3	93.9
94.7	93.9
61.6	70.53
91.7	99.9
94.7	96.46

Fig 5.2: Comparison of K-means and Bisecting K-means (Accuracy)

From the above graph we can see the result of k-mean and bisecting k-means algorithm in terms of accuracy. We know the accuracy by how much safe IP's we get during the execution in the system. Hence the bisecting gives better accuracy as comparison with the k-mean algorithm.

VI. CONCLUSION

For the data mining, huge volumes of raw data explore with pattern. It refers to the techniques that help us to identify the web content and retrieve user's interest and their needs. This paper focus on the mining algorithms comparison. For the pattern discovery, web usage mining and various criteria are considered. Quality information required to satisfy upcoming needs of user which is allowed by adding huge data in repository. Two clustering algorithm k-means and bisecting k-means used for the formation of clusters. By the result of these two algorithms we conclude that the Bisecting k-means give better performance than that of k-means algorithm. And the execution time required less hence the accuracy get increased.

REFERENCES

- [1] Manoj Kumar, Mrs. Meenu "A Survey on Pattern Discovery of Web Usage Mining" 2017IJARIIT
- [2] Arpit Bansal, Mayur Sharma, Shalini goel "Improved K-mean Clustering Algorithm for Prediction Analysis using Classification Technique in Data Mining" 2017 IJCA
- [3] S.Padmaja ,Dr.Ananthi Sheshasaayee "Clustering of User Behaviour based on Web Log data using Improved K-Means Clustering Algorithm"2016 IJET
- [4] Ramandeep Kaur, Asst. Prof.Gauravdeep "To Study Web Pattern Discovery with Web Usage Mining " 2015 IJIACS
- [5] K.Dharmarajan, M. A. Dorairangaswamy Discovering User Pattern Analysis from Web Log Data using Weblog Expert" November 2016 IJST
- [6] Virendra R. Rathod, Govind V. Patel "Prediction of User Behavior using Web log in Web Usage Mining" 2016 IJCA
- [7] M. Aldekhail "Application and Significance of Web Usage Mining in the 21st Century: A LiteratureReview" 2016 IJCTE

- [8] M.SANTHANAKUMAR,C.CHRISTOPHER COLUMBUS "Web Usage Based Analysis of Web Pages Using RapidMiner"2015
- [9] R. Suganya "Analysis of Implementating Web Usage Mining and I-Miner in E-Business" 2015 IJARCSSE
- [10] Yukai Yao, Yang Liu, Yongqing Yu, Hong Xu, Weiming Lv, Zhao Li, Xiaoyun Chen"K-SVM: An Effective SVM Algorithm Based on K-means Clustering"2013 JCP
- [11] Dhanamma Jagli, Sangeeta Oswal "Web Usage Mining: Pattern Discovery and Forecasting"2012IFRSA
- [12] NeetuAnand, Prof(Dr.) SabaHilal "Identifying the User Access Pattern in Web Log Data"2012 IJCSIT
- [13] K.Poongothai ,M.Parimala, Dr. S.Sathiyabama "Efficient Web Usage Mining with Clustering" 2011 IJCSI
- [14] Jiaqi Wang, Xindong Wu, Chengqi Zhang "Support vector machines based on K-means clustering for realtime business intelligence systems" 2005.