WEB USAGE DATA CLUSTERING USING NEURAL NETWORK LEARNING

Vinita Shrivastava*

Abstract

Due to the increasing amount of data available online, the World Wide Web has becoming one of the most valuable resources for information retrievals and knowledge discoveries. Web mining technologies are the right solutions for knowledge discovery on the Web. The knowledge extracted from the Web can be used to raise the performances for Web information retrievals, question answering, and Web based data warehousing. In the present work, we propose a new technique to enhance the learning capabilities and reduce the computation intensity of a competitive learning multi-layered neural network using the K-means clustering algorithm. The proposed model use multi-layered network architecture with a back propagation learning mechanism to discover and analyse useful knowledge from the available Web log data.

KEYWORDS - *Data mining, Unsupervised Learning algorithm, neural network, k-mean clustering, web usage mining.*

*M.Tech.(IT)Technocrat Institute of Technology, Bhopal, India

1. INTRODUCTION:

Web mining the application of machine learning techniques to web-based data for the purpose of learning or extracting knowledge. Web mining encompasses wide variety techniques, including soft computing. Web mining methodologies can generally be classified into one of three distinct categories: web usage mining, web structure mining, and web content mining examine web page usage patterns in order to learn about a web system's users or the relationships between the documents. In web usage mining the goal is to examine web page usage patterns in order to learn about a web system's users or the relationships between the documents. For example, the tool presented and creates association rules from web access logs, which store the identity of pages accessed by users along with other information such as when the pages were accessed and by whom; these logs are the focus of the data mining effort, rather than the actual web pages themselves. Rules created by their method could include, for example, "70% of the users that visited page A also visited page B examines web access logs. Web usage mining is useful for providing personalized web services, an area of web mining research that has lately become active. It promises t o help tailor web services, such as web search engines, to the preferences of each individual user. In the second category of web mining methodologies, web structure mining, we examine only the relationships between web documents by utilizing the information conveyed by each document's hyperlinks.

In web content mining [7] we examine the actual content of web pages (most often the text contained in the pages) and then perform some knowledge discovery procedure to learn about the pages themselves and their relationships. Most typically this is done to organize a group of documents into related categories. This is especially beneficial for web search engines, since it allows users to more quickly find the information they are looking for in comparison t o the usual "endless" ranked list.

Data mining is a set of techniques and tools used to the no trivial process of extracting and present implicit knowledge, no knowledge before, this information is useful and human reliable; this is processing from a great set of data; with the object of describing in automatic way models, no knowledge before; to detect tendencies and patterns [1,2] The Web Mining are the set of techniques of Data Mining applied to Web [7]. The Web Usage Mining is the process of applying techniques to detect patterns of usage to Web Page [3,5]. The Web Usage Mining use the data storage in the Log files of Web server as first resource; in this file the Web server register the access at each resource in the server by the users [4,6].

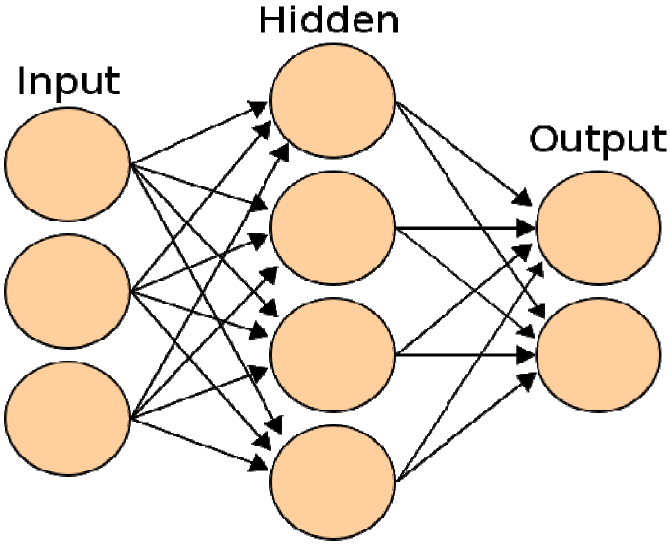
The World Wide Web (WWW) is continuously growing with the information transaction volume from Web servers and the number of requests from Web users. Providing Web administrators with meaningful information about users' access behaviour and usage patterns has become a necessity to improve the quality of Web information service performances. The evolution of the Internet has lead to an enormous proliferation of the available information and the personalization of this information space has become a necessity. The knowledge obtained by learning web users' preferences can be used to improve the effectiveness of their web sites by adapting the web information structure to the users behaviour. Automatic knowledge extraction from web log files can be useful for identifying such reading patterns and infer user profiles [2,4]. However, it is hard to find appropriate tools for analysing raw web log data to retrieve significant and useful information.

Recently, the advent of data mining techniques for Discovering usage patterns from web data (web log mining or web usage mining) made it possible to mine typical user profiles from the vast amount of access logs. Web usage mining can be viewed as the extraction of usage patterns from access log data containing the behaviour characteristics of users [4,6].the statistical data available from the normal Web log files or even the information provided by most conventional Web server analysis tools including commercial Web trackers could only provide explicit information due to the natural limitation of statistic methodology used. Computational Web Intelligence (CWI), a recently coined paradigm, is aimed to improve the quality of intelligence in the Web technology and includes Web mining as one main stream [1]. Generally, the Web analysis relies on three general sets of information given: past usage patterns, degree of shared content and inter- memory associative link structures corresponding to the three subsets in Web mining namely: Web usage mining, Web content mining and Web structure mining [1] Web usage mining, the pattern discovery consists of several steps including statistical analysis, clustering, and classification and so on. Most of the current research is focusing on finding patterns but with little effort on the detailed pattern/trend analysis that varies with the Web environments and the intelligent paradigms considered [7,10].

2. NEURAL NETWORK:

An artificial neural network (ANN), usually called neural network (NN), is a mathematical model or computational model that is inspired by the structure and/or functional aspects of biological neural networks. A neural network consists of an interconnected group of artificial neurons, and it processes information using a connectionist approach to computation. In most cases an ANN is an adaptive system that changes its structure based on external or internal information that flows through the network during the learning phase. Modern neural networks are non-linear statistical data modelling tools. They are usually used to model complex relationships between inputs and outputs or to find patterns in data.

Figure 1:An artificial neural network is an interconnected group of nodes, akin to the vast network of neurons in the human brain.



An Artificial Neural Network (ANN) is an information-processing paradigm that is inspired by the way biological nervous systems, such as the brain, process information. The key element of this paradigm is the novel structure of the information processing system. It is composed of a large number of highly interconnected processing elements (neurons) working in unison to solve specific problems. An ANN is configured for a specific application, such as pattern recognition or data classification, through a learning process [9].

2.1 Learning paradigms

There are two major learning paradigms, each corresponding to a particular abstract learning task. These are supervised learning and unsupervised learning.

2.1.1 Supervised learning

In supervised learning, we are given a set of example pairs $(x,y),x\in X,y\in Y$ and the aim is to find a function $f:X\rightarrow Y$ in the allowed class of functions that matches the examples. In other words, we wish to infer the mapping implied by the data; the cost function is related to the mismatch between our mapping and the data and it implicitly contains prior knowledge about the problem domain.

A commonly used cost is the mean-squared error, which tries to minimize the average squared error between the network's output, f(x), and the target value y over all the example pairs. When one tries to minimize this cost using gradient descent for the class of neural networks called multilayer perceptions, one obtains the common and well-known backpropagation algorithm for training neural networks.

Basically supervised learning are classified in two types. These are error connection gradient descent and stochastic. Error connection gradient descent are also classified into least mean square and backpropagation.

2.1.2 Unsupervised learning

In unsupervised learning, some data x is given and the cost function to be minimized, that can be any function of the data x and the network's output, f.

The cost function is dependent on the task (what we are trying to model) and our a priori assumptions (the implicit properties of our model, its parameters and the observed variables).

As a trivial example, consider the model $f(x)=a$, where a is a constant and the cost $C=E[(x-f(x))^2]$. Minimizing this cost will give us a value of a that is equal to the mean of the data. The cost function can be much more complicated. Its form depends on the application: for example, in compression it could be related to the mutual information between x and y, whereas in statistical modelling, it could be related to the posterior probability of the model given the data. (Note that in both of those examples those quantities would be maximized rather than minimized).

Tasks that fall within the paradigm of unsupervised learning are in general estimation problems; the applications include clustering, the estimation of statistical distributions, compression and filtering.

2.2 Architecture of neural networks

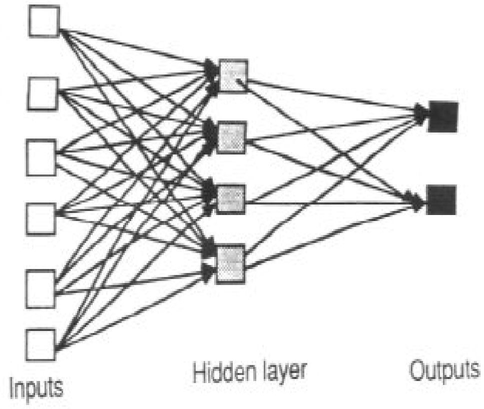
2.2.1 Feed-forward networks

Feed-forward ANNs allow signals to travel one way only; from input to output. There is no feedback (loops) i.e. the output of any layer does not affect that same layer. Feed-forward ANNs tend to be straightforward networks that associate inputs with outputs. They are extensively used in pattern recognition. This type of organization is also referred to as bottom-up or top-down.

2.2.2 Feedback networks

Feedback networks can have signals travelling in both directions by introducing loops in the network. Feedback networks are very powerful and can get extremely complicated. Feedback networks are dynamic; their 'state' is changing continuously until they reach an equilibrium point. They remain at the equilibrium point until the input changes and a new equilibrium needs to be found. Feedback architectures are also referred to as interactive or recurrent, although the latter term is often used to denote feedback connections in single-layer organizations.

Figure 2: General Neural Network



3. MINING WEB USAGE DATA:

Web usage mining is defined as the process of applying data mining techniques to the discovery of usage patterns from web logs data, to identify web users' behaviour. In Web mining, data can be collected at the server-side, client-side and proxy servers. The information provided by the data sources described above can be used to construct several data abstractions, namely users, page-views, click-streams, and server sessions. A user is defined as a single individual that is accessing file web servers through a browser. In practice, it is very difficult to uniquely and repeatedly identify users. A page-view consists of every file that contributes to the display on a user's browser at one time and is usually associated with a single user action such as a mouse-click. A click-stream is a sequential series of page-views requests. A server session (or visit) is the click-stream for a single user for a particular Web site. The end of a server session is defined as the point when the user's browsing session at that site has ended [3, 10]. The process of Web usage mining can be divided into three phases: pre-processing, pattern discovery, and pattern analysis [3,8].

Pre-processing consists of converting usage information contained in the various available data sources into the data abstractions necessary for pattern discovery. Another task is the treatment of outliers, errors, and incomplete data that can easily occur due reasons inherent to web browsing. The data recorded in server logs reflects the (possibly concurrent) access of a Web site by multiple users, and only the IP address, agent, and server side click-stream are available to Identify users and server sessions. The Web server can also store other kinds of usage information such as cookies, which are markers generated by the Web server for individual client browsers to automatically track the site visitors [3, 4]. After each user has been identified (through cookies, logins, or IP/agent analysis), the click-stream for each user must be divided into sessions. As we cannot know when the user has left the Web site, a timeout is often used as the default method of breaking a user's click-stream into sessions [2].

The next phase is the pattern discovery phase. Methods and algorithms used in this phase have been developed from several fields such as statistics, machine learning, and databases. This phase of Web usage mining has three main operations of interest: association (i.e. which pages tend to be accessed together), clustering (i.e. finding groups of users, transactions, pages, etc.), and sequential analysis (the order in which web pages tend to be accessed) [3, 5]. The first two are the focus of our ongoing work. Pattern analysis is the last phase in the overall process of Web usage mining. In this phase the motivation is to filter out uninteresting rules or patterns found in the previous phase. Visualization techniques are useful to help application domains expert analyse the discovered patterns.

4. CONVENTIONAL METHOD USED IN WEB MINING:

4.1 Clustering

Clustering and classification [11] have been useful and active areas of machine learning research that promise to help us cope with the problem of information overload on the Internet. With clustering the goal is to separate a given group of data items (the data set) into groups called clusters such that items in the same cluster are similar to each other and dissimilar to the items in other clusters. In clustering methods no labelled examples are provided in advance for training (this is called unsupervised learning). Under classification we attempt to assign a data item to a predefined category based on a model that is created from preclassified training data (supervised learning). In more general terms, both clustering and classification come under the area of knowledge discovery in databases or data mining. Applying data mining techniques to web page content is referred to as web content mining, which is a new sub-area of web mining, partially built upon the established field of information retrieval.

When representing text and web document content for clustering and classification, a vector-space model is typically used. In this model, each Possible term that can appear in a document becomes a feature dimension. The value assigned to each dimension of a document may indicate the number of times the corresponding term appears on it or it may be a weight that takes into account other frequency information, such as the number of documents upon which the terms appear. This model is simple and allows the use of traditional machine learning methods that deal with numerical feature vectors in a Euclidean feature space. However, it discards information such as the order in which the terms appear, where in the document the terms appear, how close the terms are to each other, and so forth. By keeping this kind of structural information we could possibly improve the performance of various machine-learning algorithms. The problem is that traditional data mining methods are often restricted to working on purely numeric feature vectors due to the need to compute distances between data items or to calculate some representative of a cluster of items, both of which are easily accomplished in a Euclidean space. Thus either the original data needs to be converted to a vector of numeric values by discarding possibly useful structural information (which is what we are doing when using the vector model to represent documents) or we need to develop new, customized methodologies for the specific representation.

Clustering the process of partition a set of data in a set of meaning full subclasses known as clusters. It helps users understand the natural grouping or structure in a data set. Clustering is an unsupervised learning technique which aim is to find structure in a collection of unlabeled data. It is being used in many fields such as data mining, knowledge discovery, pattern recognition and classification [3].

A good clustering method will produce high quality clusters in which similarity is high known as intra-classes and inter-classes where similarity is low. The quality of clustering depends upon both the similarly measure used by the method and it, s implementation and it is also measured by the it's ability to discover hidden patterns.

4.2 K-Mean Algorithm

The K-Means algorithm is one of a group of algorithms called partitioning clustering algorithm [4]. The most commonly use partition clustering strategy is based on square error criterion. The general objective is to obtain the partition that, for a fixed number of clusters, minimizes the total square errors.

Suppose that the given set of N samples in an n-dimensional space has somehow been partitioned into K-clusters {C1, C2, C3... CK}. Each CK has nK samples and each sample is in exactly one cluster, so that ∑ nK = N, where k=1… K. The mean vector Mk of cluster CK is defined as the centroid of the cluster

$$M_K = (1/n_k) \sum_{i=1}^{n_k} x_{ik} \qquad (1)$$

Where xik is the ith sample belonging to cluster CK. The square-error for cluster CK is the sum of the squared Euclidean distances between each sample in CK and its centroid. This error is also called the within-cluster variation [5]:

$$e_k^2 = \sum_{i=1}^{n_k} (x_{ik} - M_k)^2 \quad (2)$$

The square-error for the entire clustering space containing K cluster is the sum of the within-cluster variations

$$E_k^2 = \sum_{k=1}^{K} e_k^2 \quad (3)$$

The basic steps of the K-mean algorithm are:

1. Select an initial partition with K clusters containing randomly chosen sample, and compute the centroids of the clusters,
2. Generate a new partition by assigning each sample to the closest cluster centre,
3. Compute new cluster centre as the centroids of the clusters,
4. Repeat steps 2 and 3 until optimum value of the criterion function is found or until the cluster membership stabilizes.

4.3 Problem identification:

Problems with k-means

•In k-means, the free parameter is k and the results depend on the value of k. unfortunately; there is no general theoretical solution for finding an optimal value of k for any given data set.

•It take more time for calculating the data set.

•It can only handled the Numerical data set.

•The Result depend on the Metric used the measure ǁ x-miǁ.

5. PROPOSED APPROACH:

In the present work, the role of the k-means algorithm is to reduce the computation intensity of the neural network, by reducing the input set of samples to be learned. This can be achieved by clustering the input dataset using the k-means algorithm, and then take only discriminate samples from the resulting clustering schema to perform the learning process.

By doing so, we are trying to select a set of samples that cover at maximum the region of each class in the N-dimensional space (N is the size of the training vectors). The input classes are clustered separately in such a way to produce a new dataset composed with the centroid of each cluster, and a set of boundary samples selected according to their distance from the centroid. Reducing the number of used samples will enhance significantly the learning performances, and reduce the training time and space requirement, without great loss of the information handled by the resulting set, due to its specific distribution. The Figure.3 illustrates an example of the application of this selection schema to a 2-dimentional dataset. One promising research in Web mining concerns the application of the Neural Network techniques. The proposed technique used in web mining is following

Figure 3. An Illustrative example on the application of the proposed method to a 2-dimensional training set;(a) Initial distribution,(b) Selected samples after clustering

Figure3 (a)

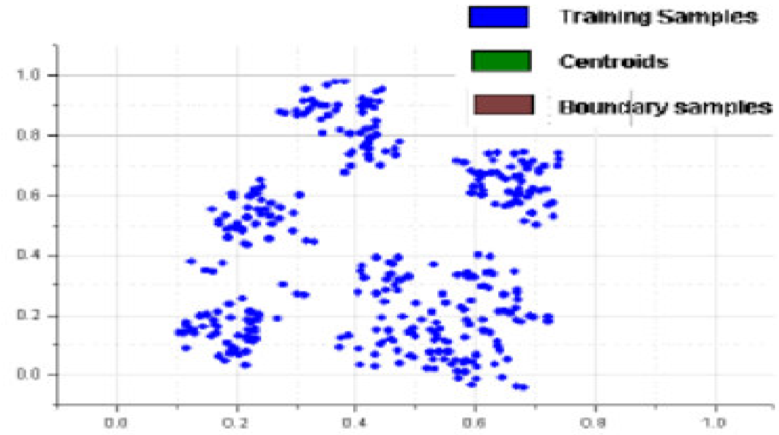
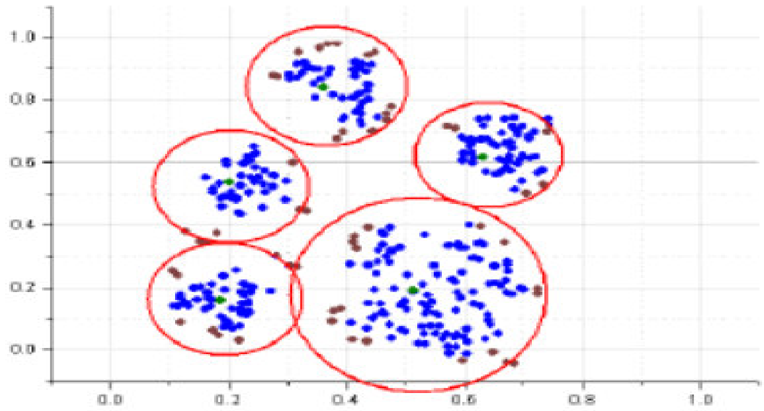


Figure3 (b)



The number of fixed clusters can be varied to specify the coverage repartition of the samples. The number of selected samples for each class is also a parameter of the selection algorithm. Then, for each class, we specify the number of samples to be selected according to the class size. When the clustering is achieved, samples are taken from the different obtained clusters according to their relative intraclass variance and their density. The two measurements are combined to compute a coverage factor for each cluster. The number of samples taken from a given cluster is proportional to the computed coverage factor. Let A be a given class, to witch we want to apply the proposed approach to extract S sample. Let k be the number of cluster fixed to be used during the k-means clustering phase. For each generated cluster cli, (i:1..k), the relative variance is computed using the following expression:

$$\mathbf{Vr}(\mathbf{cl}_i) = \frac{\frac{1}{\mathrm{Card}(\mathrm{cl}_i)} * \sum_{x \in \mathrm{cl}_i} \mathbf{dist}(\mathbf{x}, \mathbf{c}_i)}{\sum_{j=1}^{k} \left(\frac{1}{\mathrm{Card}(\mathrm{cl}_j)} * \sum_{x \in A} \mathbf{dist}(\mathbf{x}, \mathbf{c}_j) \right)} \tag{4}$$

When Card(X) give the cardinality of a given set X, and dist(x,y) give the distance between the two points x and y.

The most commonly used distance measure is the Euclidean metric which defines the distance between two points x=(p1,….pN) and y=(q1,….,qN) from RN as:

$$\mathbf{dist}(\mathbf{x}, \mathbf{y}) = \sqrt{\sum_{i=1}^{N} (\mathbf{p}_i - \mathbf{q}_i)^2} \tag{5}$$

The density value corresponding to the same cluster cli is computed like the following:

$$\mathrm{Den}(\mathrm{cl}_i) = \frac{\mathrm{Card}(\mathrm{cl}_i)}{\mathrm{Card}(A)} \quad (6)$$

The coverage factor is then computed by:

$$\mathrm{Cov}(\mathrm{cl}_i) = \frac{\left(\mathrm{Vr}(\mathrm{cl}_i) + \mathrm{Den}(\mathrm{cl}_i)\right)}{2} \quad (7)$$

We can clearly see that: $0 \leq \mathrm{Vr(cli)} \leq 1$ and $0 \leq \mathrm{Den(cli)} \leq 1$ for any cluster cli. So the coverage factor Cov(cli) belong also to 1-Cluster the class A using the k-means algorithm into k cluster.the [0,1] interval. Furthermore, it is clear that:

$$\sum_{i=1}^{k} \mathrm{Vr}(\mathrm{cl}_i) = 1 \quad \text{and} \quad \sum_{i=1}^{k} \mathrm{Den}(\mathrm{cl}_i) = 1$$

We can so deduce easily that:

$$\sum_{i=1}^{k} \mathrm{Cov}(\mathrm{cl}_i) = 1$$

Hence, the number of samples selected from each cluster is determined using the expression

Num_samples(cli)=Round(S*cov(cli)

Let A be the input class;

k: the number of cluster;

S: the number of samples to be selected ($S \geq k$);

Sam(i): the resulting selected set of samples for the cluster i;

Out_sam: the output set of samples

Selected from the class A;

Candidates: a temporary array that contain the cluster points and their respective distance from the centroid.

i,j,min,x: intermediates variables;

ε: Neiberhood parameter.

The proposed selection model algorithm is

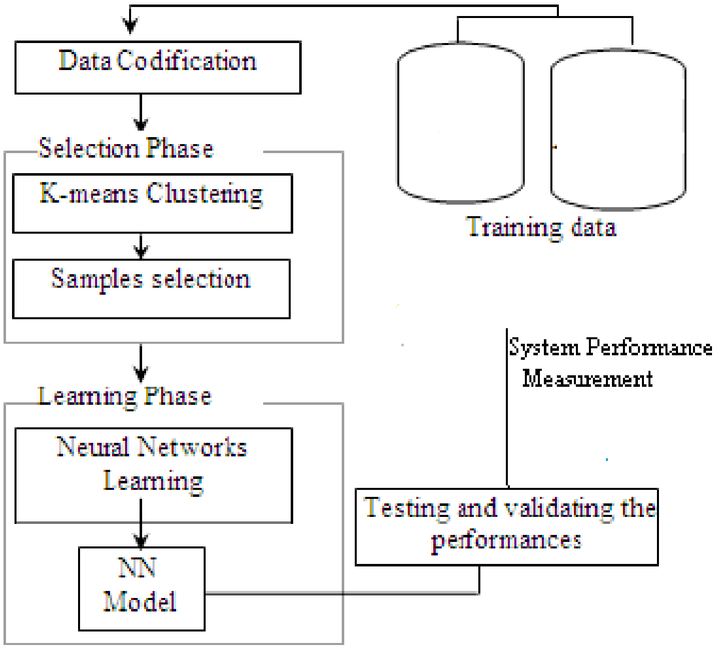
1-Cluster the class A using the k-means algorithm into k cluster.

2-For each cluster cli (i:1..k) do

{Sam(i) :={centroid(cli)};

```
        j:=1;
For each x from cli do
{ Candidates [j].point :=x;
Candidates [j].location :=dist(x, centroid(cli)) ;
j:=j+1 ;};
Sort the array Candidates in descending order with Hence, the number of samples selected from
each cluster is respect to the values of location field;
  j:=1;
While((card(Sam(i)))<Num_samples(cli))
and (j<card(cli)) do{min:=100000;
For each x from Sam(i) do
  {if dist(Candidates[j].point,x)<min
          then min:= dist(Candidates[j].point,x) ;
             }
            if (min > ε) then
        Sam(i):=Sam(i) U{Candidates[j].point};
j:=j+1; }
      if card(Sam(i)) < Num_samples(cli) then
repeat{Sam(i):=Sam(i)UCandidates[random].point
}until (card(Sam(i)) = Num_samples(cli));
3-For i=1 to k do Out_sam:=Out_sam U Sam(i);
```

Figure 3.(c) The general operating mechanism of the proposed method



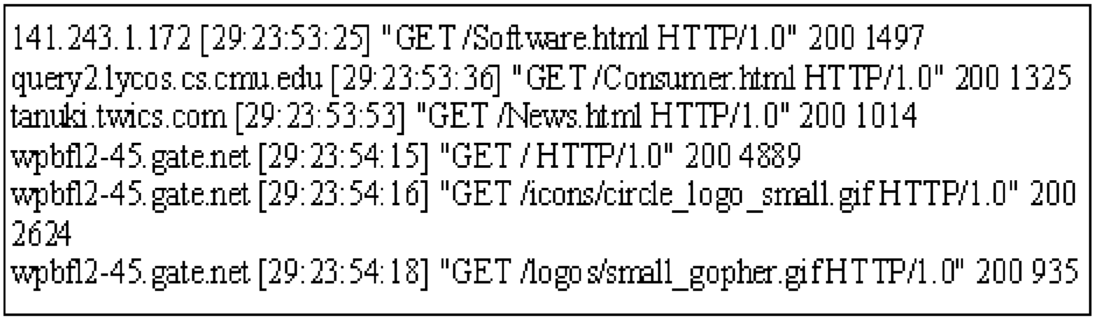
.6. Result Analysis:

6.1 Data Set Description

Web data clustering is the process of grouping Web data into "clusters" so that similar objects are in the same class and dissimilar objects are in different classes [2, 7]. Its goal is to organize data circulated over the Web into groups / collections in order to facilitate data availability and accessing.

We can broadly categorize Web data clustering into (i) users' sessions-based [3,6,17,20,27] and (ii) link-based. [12,14,19] The former uses the Web log data and tries to group together a set of users' navigation sessions having similar characteristics. In this framework, Web-log data provide information about activities performed by a user from the moment the user enters a Web site to the moment the same user leaves it [8]. The records of users' actions within a Web site are stored in a log file. Each record in the log file contains the client's IP address, the date and time the request is received, the requested object and some additional information -such as protocol of request, size of the object etc. Figure 1 presents a sample of a Web access log file from a Web server.

Figure 4: A sample of Web Server Log File



```
141.243.1.172 [29:23:53:25] "GET /Software.html HTTP/1.0" 200 1497
query2.lycos.cs.cmu.edu [29:23:53:36] "GET /Consumer.html HTTP/1.0" 200 1325
tanuki.twics.com [29:23:53:53] "GET /News.html HTTP/1.0" 200 1014
wpbfl2-45.gate.net [29:23:54:15] "GET / HTTP/1.0" 200 4889
wpbfl2-45.gate.net [29:23:54:16] "GET /icons/circle_logo_small.gif HTTP/1.0" 200
2624
wpbfl2-45.gate.net [29:23:54:18] "GET /logos/small_gopher.gif HTTP/1.0" 200 935
```

.

6.2 Data Preprocessing

We need to do some data processing, such as invalid data cleaning and session identification [10]. Data cleaning removes log entries (e.g. images, java scripts etc) that are not needed for the mining process. In order to identify unique users' sessions, heuristic methods are (mainly) used [8], based on IP and session time-outs. In this context, it is considered that a new session is created when a new IP address is encountered or if the visiting page time exceeds a time threshold (e.g. 30 minutes) for the same IP-address. Then, the original Web logs are transferred into user access session datasets for analysis. Clustering users' sessions are useful for discovering both groups of users exhibiting similar browsing patterns and groups of pages having related content based on how often URL references occur together across them. Therefore, clustering users' sessions is more important in some Web applications, such as on-line monitoring user behaviour, on-line performance analysis, and detecting traffic problems.

6.3 Web URL

Resources in the World Wide Web are uniformly identified by means of URLs (Uniform Resource Locators). The syntax of an http URL is: '

http://' host.domain [':'port] [abs path ['?' query]]

Where {host.domain [: port] is the name of the server site. The TCP/IP port is optional (the default port is 80),{ abs path is the absolute path of the requested resource in the server file system. We further consider abs path of the form path '/' filename ['.' extension], i.e. consisting of the file system path, filename and file extension. { query is an optional collection of parameters, to be passed as an input to a resource that is actually an executable program, e.g. a CGI script.

On the one side, there are a number of normalizations that must be performed on URLs, in order to remove irrelevant syntactic differences (e.g., the host can be in IP format or host format {131.114.2.91 is the same host as kdd.di.unipi.it). On the other side, there are some web server programs that adopt non-standard formats for passing parameters. The vivacity. it web server program is one of them. For instance, in the following URL:

Cluster	K-Mean Algorithm (SSE)	Modified K-Mean Algorithm (SSE)
20	324	119
40	445	122
60	516	154
80	633	176
100	769	208

6.4 Empirical Setting

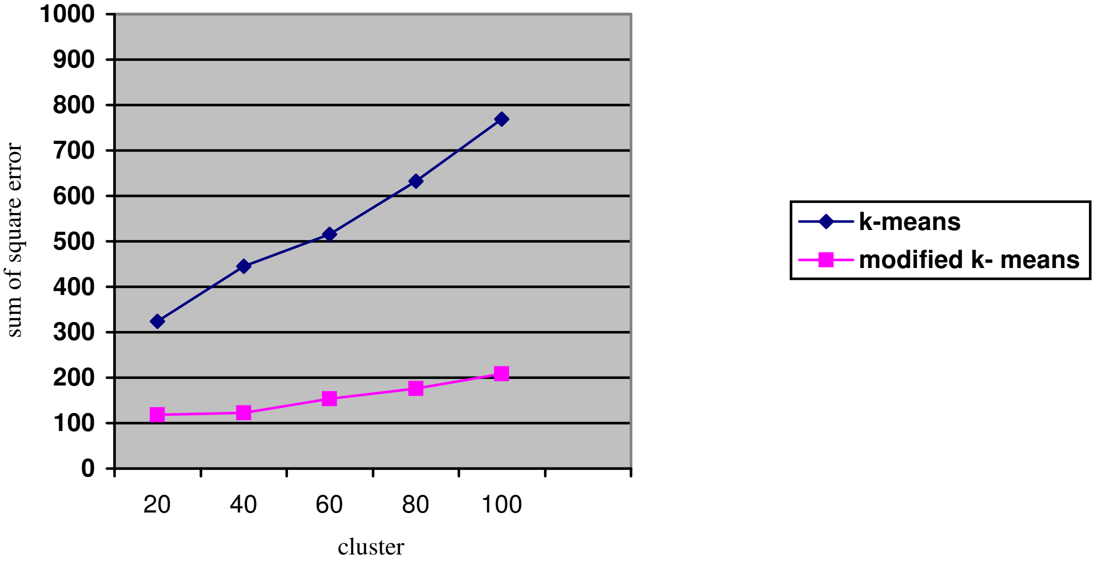
The K-Means and Modified K-Mean algorithms are written in visual basic 6.0 as front-end and MS-Access used as Backend and compiled into mix files. K-Mean algorithms are relatively efficient due to vectored programming and active optimisation. All experiments are run on a PC with a 3.06GHz Pentium-4 CPU with 1GB DRAM and running Windows XP. For the Modified K-Mean Algorithm, the learning rate follows m = 0.5.

In order to study the effect of the total number of clusters on the web mining results, we performed empirical studies with 100 total numbers of clusters.

We compare the aforementioned clustering algorithms on the whole data set with 5000 data set The Computation time results for the clustering algorithms with 100 clusters are shown

Table1: A comparison between K-Mean and Modified K-mean algorithm

Graph between SSE and Cluster to compare the results of k-means and modified k-means



As it is seen from the figure, as the number of cluster increase, then the value of SSE in k-mean algorithm will be increase but in modified algorithm the values of SSE is slow increased compared to k-mean algorithm. As a summary, we can easily say that our new Modified K-Mean algorithm performs much better than the K-Mean algorithm in discovering user sessions for all kinds of parameters.

7.CONCLUSION:

In this work, we study the possible use of the neural networks learning capabilities to classify the web traffic data mining set. The discovery of useful knowledge, user information and server access patterns allows Web based organizations to mining user access patterns and helps in future developments, maintenance planning and also to target more rigorous advertising campaigns aimed at groups of users. Previous studies have indicated that the size of the Website and its traffic often imposes a serious constraint on the scalability of the methods. As popularity of the web continues to increase, there is a growing need to develop tools and techniques that will help improve its overall usefulness.

REFERENCES:

[1] W.J. Frawley, G. Piatetsky-Shapiro, and C.J. Matheus,(1991) “Knowledge Discovery in Databases: An Overview,º Knowledge Discovery in Databases”, G. Piatetsky-Shapiro and W.J Frawley, eds., Cambridge, Mass.: AAAI/MIT Press, pp. 1-27.

[2] Mika Klemettinen, Heikki Mannila, Hannu Toivonen(1997) A Data Mining Methodology and Its Application to Semi-automatic Knowledge Acquisition. DEXA Workshop 670-677.

[3] R. Kosala, H. Blockeel,(2000) Web Mining Research: A Survey, SIGKKD Explorations, vol. 2(1),.

[4] Borges-Levene,(2000) “An average linear time algorithm for web usage mining:”.

[5] J. Srivastava, R. Cooley, M. Deshpande, P.N. Tan,(2000) Web Usage Mining: Discovery and Applications of Usage Patterns from Web Data, SIGKKD Explorations, vol.1.

[6] P. Batista, M. J. Silva, (2002), “Mining web access logs of an on-line newspaper,” http://www.ectrl.itc.it/rpec/RPEC-apers/11-batista.pdf.

[7] Cernuzzi, L., Molas, M.L. (2004). Integrando diferentes técnicas de Data Mining en procesos de Web Usage Mining. Universidad Católica "Nuestra Señora de la Asunción". Asunción. Paraguay.

[8] R. Iváncsy, I. Vajk,(2005) Different Aspects of Web Log Mining. 6th International Symposium of Hungarian Researchers on Computational Intelligence. Budapest.

[9] Chau, M.; Chen, H., (2007)“Incorporating Web Analysis Into Neural Networks: An Example in Hopfield Net Searching”, IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews, Volume 37, Issue 3, May 2007 Page(s):352 – 358

[10] Raju, G.T.; Satyanarayana, (2007) P. S. "Knowledge Discovery from Web Usage Data: Extraction of Sequential Patterns through ART1 Neural Network Based Clustering Algorithm", International Conference on Computational Intelligence and Multimedia Applications, 2007, Volume 2, Issue , 13-15 Dec. 2007 Pages :88 -92

[11] Jalali, Mehrdad Mustapha, Norwati Mamat, Ali Sulaiman, Md. Nasir B. (2008) " A new classification model for online predicting users' future movements", in International Symposium on Information Technology, 2008. ITSim 2008 26-28 Aug. 2008, Volume: 4, On page(s): 1-7, Kuala Lumpur, Malaysia