Research article Available online www.ijsrr.org ISSN: 2279–0543

International Journal of Scientific Research and Reviews

Fuzzy Particle Swarm Optimization Based Feature Learning Vector

P. Lavanya

Department of Information Technology Kg College of Arts and Science Coimbatore.

ABSTRACT:

Data mining is the process of analyzing data from the different huge amount of data to increasing the cost, revenue and useful information from the databases. Data mining is applied successfully in the business background, weather forecast, medicine, transportation, healthcare, insurance, and government. The clustering is an unsupervised learning without predefined class identifiers and it is a process of partition the data set into several groups based on the similarity which done by their distance. Most researchers were used the Euclidean distance that obtained by differentiating between centroid and data item. The data sets have many features that's all features are not important to solving the clustering problems. In this circumstance, assign weights to their features that improving the performance of clustering accuracy and reducing its computational time. In this present work, Feature Weighted Fuzzy Particle Swarm Optimization (FW-PSO) algorithm used to solve unsupervised classification and the this algorithm produced superior result than Fuzzy Particle Swarm Optimization.

KEYWORDS: Data Mining, Euclidean Distance, Clustering, Feature Weighted Fuzzy Particle Swarm Optimization

***Corresponding author:**

P.Lavanya

Department of Information Technology

Kg College of Arts and

ScienceCoimbatore.

1. INTRODUCTION

Data mining is one of the most[1] important research fields that are due to the development of both computer hardware and software technologies, which has imposed organizations to depend heavily on these technologies.

Knowledge Discovery in Databases process or KDD is an interdisciplinary subfield of computer science, is the computational process of discovering patterns in large data sets to connect with the algorithms and methods of KDD[2], artificial intelligence, machine learning, statistics, and database systems.

1.1 Clustering in Soft computing Methods

This section further discusses the important usefulness of soft-computing methods in clustering tasks.

1.1.1 Fuzzy Clustering

Conventional clustering approaches are generating partitions. In a partition, each instance belongs to one and only one cluster. Hence, the clusters in a hard clustering[9] are disjointed. Fuzzy clustering extends this notion and suggests a soft clustering schema. In this case, each pattern is associated with every cluster using some sort of membership function, namely, each cluster is a fuzzy set of all the patterns.

Larger membership values indicate higher confidence in the assignment of the pattern to the cluster. A hard clustering can be obtained from a fuzzy partition by using a threshold of the membership value. The most popular fuzzy clustering algorithm is the fuzzy c-means (FCM) algorithm[3].

Even though it is better than the hard K-means algorithm at avoiding local minima, FCM can still converge to local minima of the squared error criterion. The design of membership functions is the most important problem in fuzzy clustering; different choices include those based on similarity decomposition and centroids of clusters. A generalization of the FCM algorithm has been proposed through a family of objective functions.

1.1.2 Evolutionary Techniques

Evolutionary Techniques are called as the optimization technique that's stochastic general purpose methods for solving optimization problems. Since, Clustering problem can be defined as an optimization problem, evolutionary approaches may be appropriate here. The idea is to use evolutionary operators and a population of clustering structures to converge into a globally optimal clustering.

2. RELATED WORK

The clustering analysis is very big problem specially in large data set. The clustering analysis means it will divide the given data into certain classes according to the main attribute of the dataset, thus the item of class to have the same or similar meaning. The Clustering algorithm has some drawback for example K-mean have drawback in searching for optimal solution[8]. It is very difficult to find the cluster center. Consider the drawbacks and limitation of single clustering technique, research has developed optimization algorithm to overcome the drawbacks of clustering technique.

Due to the limitation of the k-mean algorithm the PSO will used for overcome the k-mean drawback. The drawback of K-mean is to initialization of cluster center. This is one of the disadvantages of the k- mean. To overcome this advantages the PSO algorithm was used , using PSO it will find the cluster center will generated initial piratical are optimized to maintain the final cluster center[4].

Mussels Wandering Optimization algorithm will used for finding the optimal path of clustering. It is one of the important algorithm related to the Swarm intelligence. It is one of the new effective global optimization algorithms[7].

In c-mean clustering the cluster center have used to form the cluster and then it will extended to fuzzy c-mean algorithm that is related to the concept of the fuzziness in order to handling the overlapping class in the real data set and then it will extended to the rough fuzzy c-mean algorithm. It will solved the uncertainty problem of the data set.

In cluster analysis, some of the features of a given data set may fall in higher relevance than others. For avoid this issue, Feature-Weighted Fuzzy C-Means (FWFCM) approaches are helpful in recent years. An Improved FWFCM (IFWFCM)[5] is overcome the certain deficiencies in the existing FWFCMs, e.g., the elements in a feature-weight vector cannot be adaptively adjusted during the training phase and also the update formulas of a feature-weight vector cannot be derived analytically.

3. PROPOSED STUDY:

Fuzzy clustering is a fundamental operation used in unsupervised clustering organization, automatic topic extraction, and information retrieval. This investigates presents an optimization to fuzzy clustering problem with assigning weight values for each important features. An optimization approach involving Swarm intelligence based algorithm, Fuzzy Particle Optimization (PSO) combined with its weight values for important features to evaluate for high dimensional clustering.

Fuzzy clustering Problem can be formally defined as below :

i. A membership value calculated by

$$z_j \frac{\sum_{i=1}^{n} \mu_{ij}^{m} o_i}{\sum_{i=1}^{n} \mu_{ij}^{m}}$$

ii. A desired number of clusters k, and

an objective function that evaluates the quality of a clustering, we want to compute an assignment that minimizes the objective function.

$$J_m = \sum_{j=1}^{c} \sum_{i=1}^{n} \mu_{ij} d_{ij}^{w}$$

3.1 Particle Swarm Optimization:

Particle swarm optimization (PSO) is a population based stochastic optimization technique inspired by bird flocking and fish schooling[6] that was originally designed and introduced by Kennedy and Eberhart in 1994. The particle swarm optimization is uses iterative practice to find several better solution fitness functions in a search space.

In PSO, the potential solutions are called particles and they fly through the problem space by following the in progress optimum. The algorithm flow in PSO starts with a population of particles whose positions corresponds to the potential solutions for the studied problem, and velocities are randomly initialized in the search space.

In each iteration, the search for optimal position is personal best position and global best position. The personal best position, pbest, is the best position the particle has visited and gbest is the best position the swarm has visited since the first time step.

A particle's velocity and position are updated as follows.

$$V(t+1) = w.\ V(t).\ c_1 r_1 (pbest(t) - X(t) + c_2 r_2 (gbest(t) - X(t)) \quad k=1,2\ldots P \tag{3.1}$$

$$X(t+1) = X(t) + V(t+1) \tag{3.2}$$

where X and V are position and velocity of particle respectively, w is inertia weight, c1 and c2 are positive constants, called acceleration coefficients which control the influence of pbest and gbest on the search process, P is the number of particles in the swarm, r1 and r2 are random values range between 0 and 1.

Each particle keeps track of its coordinates in the problem space which are associated with the best solution.Which coincides with the best fitness value called pbest. Find the best solution form whole best values are called global best mentions by gbest. The PSO is a very attractive by means least amount parameters only.

3.2. Fuzzy Particle Swarm Optimization For Fuzzy Clustering:

A modified particle swarm optimization for TSP[10] called Fuzzy Particle Swarm Optimization (FPSO) proposed by Peng et al. In which method, the position and velocity of particles redefined to represent the fuzzy relation between variables. In this sub-section we describe this method for fuzzy clustering problem.

In FPSO algorithm,' X' the position of particle, shows the fuzzy relation from set of data objects, $o = \{o_1, o_2, ..., o_n\}$ to set of cluster centers, $z = \{z_1, z_2, ..., z_n\}$, X Can be expressed as follows:

$$X = \begin{pmatrix} \mu_{11} & \cdots & \mu_{1c} \\ \vdots & \ddots & \vdots \\ \mu_{n1} & \cdots & \mu_{nc} \end{pmatrix} \qquad (3.3)$$

In which μ_{ij} is the membership function of the i_{th} object with the j_{th} cluster with constraints stated Eq. (3.1) and Eq. (3.2) therefore we can see that the position matrix of each particle is the same as fuzzy matrix μ in FCM algorithm. Also the velocity of each particle is stated using a matrix with the size *n* rows and *c* columns the elements of which are in range between -1 and 1. We get the Eq. (3.11) and Eq. (3.12) for updating the positions and velocities of the particles based on the matrix.

$$V(t+1) = w \otimes V(t) \oplus c_1 r_1 \otimes (pbest(t) \ominus X(t)) \oplus c_2 r_2 \otimes (gbest(t) - \ominus X(t)) \quad k=1,2\ldots P \qquad (3.4)$$

$$X(t+1) = X(t) \oplus V(t+1) \qquad (3.5)$$

After updating the position matrix, it may violate the constraints given Eq. (3.1) and Eq. (3.2). So it is necessary to normalize the position matrix. First we make all the negative elements in matrix to become zero. If all elements in a row of the matrix are zero, they need to be re-evaluated using series of random numbers within the interval between 0 and 1, and then the matrix undergoes the following transformation without violating the constraints:

$$Xnormal = \begin{pmatrix} \mu_{11} / \sum_{j=1}^{c} \mu_{1j} & \cdots & \mu_{1c} / \sum_{j=1}^{c} \mu_{1j} \\ \vdots & \ddots & \vdots \\ \mu_{n1} / \sum_{j=1}^{c} \mu_{nj} & \cdots & \mu_{nc} / \sum_{j=1}^{c} \mu_{nj} \end{pmatrix} \qquad (3.6)$$

In FPSO algorithm the same as other evolutionary algorithms, we need a function for evaluating the generalized solutions called fitness function. In this paper below equation is used for evaluating the solutions.

$$J_m = \sum_{i=1}^{c} \sum_{j=1}^{n} u_{ij}^m (d_{ij})^2 \qquad (3.7)$$

$$f(X) = \frac{K}{J_m} \qquad (3.8)$$

Therein K is a constant and J_m is the objective function. The smaller is J_m, the better is the clustering effect and the higher is the individual fitness $f(X)$.

A clustering algorithm that automatically assign the weights to the attributes or even selects a proper subset for working out. The feature weights assigning to dataset is done in the form of the feature selection.

In present study, tries to combine clustering and feature selection by pushing the feature selection method into the clustering algorithm can be turned into a feature selection method by simply applying a weight threshold to the computed feature weights.

Dataset has many features, since, in a large amount of the real world problems, each features are not measured to be equally important. For example, iris dataset has four features such as sepal length, sepal width, petal length, petal width. Standard dataset take more computational time, hence, reducing computing time and increase accuracy of clustering algorithms. In present study solve aforementioned drawback used to feature learning vector which incorporated weight value with its distance function.

Assign the weights to each feature for given dataset to increase the accuracy and reducing the computational time by means of proposed feature learning vector combined with particle swarm optimization[10, 26] is called FW-PSO (Feature Weight-Particle Swarm Optimization).

Euclidean distance referred as follows,

$$d_{ij} = \sqrt{(z_j - o_{ij})^2} \qquad (3.9)$$

The feature learning weight vector performed based on weighted Euclidean distance referred as follows,

$$d_{ij} = \sqrt{w_k (z_j - o_{ij})^2} \qquad (3.10)$$

Hence, we construct objective function based on weighted distance,

$$J_m^w = \sum_{i=1}^{c} \sum_{j=1}^{n} u_{ij}^m (d_{ij}^w)^2 \qquad (3.11)$$

Following constrains are define the weight value into normalization form as follows,

$$\sum_{k=1}^{s} w_k = 1 \qquad (3.12)$$

$$\sum_{k=1}^{s} \sum_{j=1}^{n} u_{ij} = n \qquad (3.13)$$

FW-FPSO Algorithm:

Step 1: Initialize the parameters including population size P, c1, c2, w, and the maximum iterative count.

Step 2: Create a swarm with P particles (X, pbest, gbest and V are n× c matrices).

Step 3: Initialize X, V, pbest for each particle and gbest for the swarm.

Step 4: Calculate the cluster centers for each particle using as follows,

$$z_j \frac{\sum_{i=1}^{n} \mu_{ij}^{m} o_i}{\sum_{i=1}^{n} \mu_{ij}^{m}}$$

Step 5: Calculate the distance value using Eq. (3.10).

Step 6: Calculate the fitness value of each particle using by Eq. (3.8).

Step 7: Calculate pbest for each particle.

Step 8: Calculate gbest for the swarm.

Step 9: Update the velocity matrix for each particle using by Eq. (3.4).

Step 10: Update the position matrix

for each particle using by Eq. (3.5).

Step 11: If terminating condition is not met, go to step 4.

The termination condition in proposed method is the maximum number of iterations or no improvement in gbest in a number of iterations.

4. EXPERIMENTAL ANALYSIS:

4.1 Clustering Validation Technique

The aim of clustering validation is to evaluate the clustering results by finding the best partitions that fits the underlying data. Thus, cluster validity is used quantitatively evaluate the results of clustering algorithms.

4.2 Objective Functions

Clustering is a generally applied for data exploration while least labeled data obtainable. An objective function of clustering[11] algorithm is tries to minimize or maximize of a function such that the clusters are obtained when the minimum or maximum is achieved are all the same. The clustering algorithms its performance generated from the given objective functions with a number of examples.

$$J_m^w = \sum_{i=1}^{c}\sum_{j=1}^{n} u_{ij}^m (d_{ij}^w)^2 \qquad (4.1)$$

The performances of the clustering algorithms are work out by objective function for well known datasets. The objective functions of every clustering algorithm are considered into three values such as worst, average and best. The best performance of the every clustering algorithm is considered if best value is high value otherwise its performance is low. The clustering algorithms, its performance are calculated by average value from overall iteration.

4.3 Partition Coefficient Index

The partitions coefficient index obtained maximum value of cluster structure is considered as optimal value of algorithms. This is defined as follows[9],

$$PCI = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{c} u_{ij}^2 \qquad (4.2)$$

4.4 Partition Entropy Index

The partition entropy index obtained minimum value is optimal solutions that is defined as follows,

$$PEI = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{c} u_{ij}\log_b(u_{ij}) \qquad (4.3)$$

4.5 Parameter Settings

In order to improve the performance of the Fuzzy Particle Swarm Optimization and Feature Weight – Fuzzy Particle Swarm Optimization, need to select their parameters appropriately. The performance of the algorithms produced better outcomes for their best parameter values as following: $c_1 = c_2 = 2.0$, P =10, w=0.9, Maximum iterations =100, $r_1 = r_2 = 0.1$ to 0.9, Weight exponents is set m=2 for all algorithms. (m>=2)

4.6 Datasets

- Fisher's iris data set, which consists of three different species of iris flower. For each species, 50 samples with four features were collected.
- Wine data set, which consists of 178 object and 3 classes, each type has 13 features.

4.7 Experimental Result

The Fuzzy Particle Swarm Optimization and the proposed feature weighted learning- Fuzzy Particle Swarm Optimization its individual performance analysis for each dataset is demonstrate in table 1 and its average performance demonstrate in table 2, that are implemented using MATLAB.

Table: 1 Performance Analysis of Fuzzy Clustering Algorithms

Methods	Iris(150,3,3)			Wine(178,3,2)		
	Best	Average	Worst	Best	Average	Worst
FPSO	61.12	62.70	64.28	74.60	116.58	158.56
FW-FPSO	**60.76**	**61.54**	**62.33**	**53.14**	**55.49**	**56.74**

In table 1, the performance of proposed clustering algorithms such feature weight- fuzzy particle swarm optimization algorithms shows better accuracy then existing algorithm. That is highlighted in bold letter in table 1.

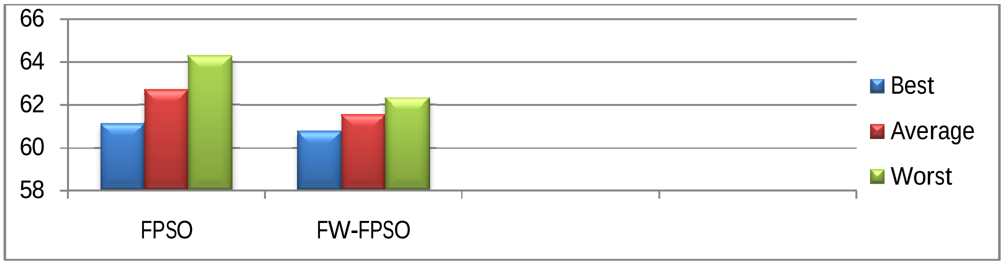

Figure 1: Performance Of Iris Dataset For Clustering

The clustering algorithms have to apply successfully for iris dataset. Fuzzy Particle Swarm Optimization and that the proposed Feature Weighted Fuzzy Particle Swarm Optimization Algorithms its performances are demonstrated in figure 1. That the proposed method is shows the superior result compared with existing algorithms such as fuzzy particle swarm optimization.

Figure 2 is shows performance of clustering algorithms for wine datasets. In which, that the proposed Feature Weight- Fuzzy Particle Swarm Optimization has formed high accuracy than Fuzzy Particle Swarm Optimization.

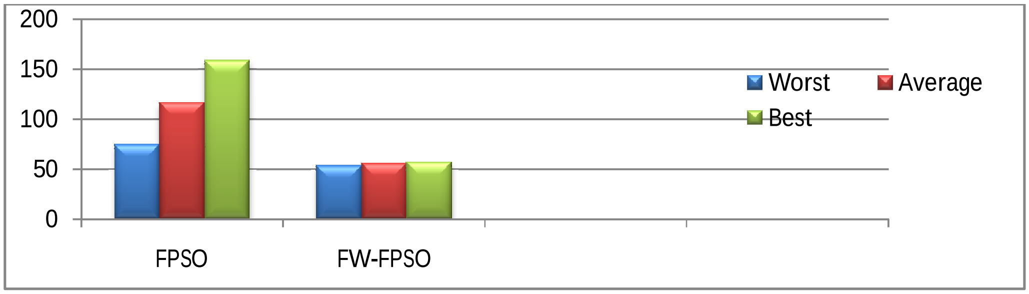

Figure 2: Performance Of Wine Dataset For Clustering

The average performance of the used algorithms are demonstrated in table 2, in case of average performance also produced better results in term of feature weight fuzzy particle swarm optimization. The figure 3 demonstrates average performance of the clustering algorithms in graphical representation form.

Table: 2 Average Performance of Clustering Algorithms

Methods	Worst	Average	Best
FPSO	67.86	89.64	111.42
FW-FPSO	**56.95**	**58.51**	**59.53**

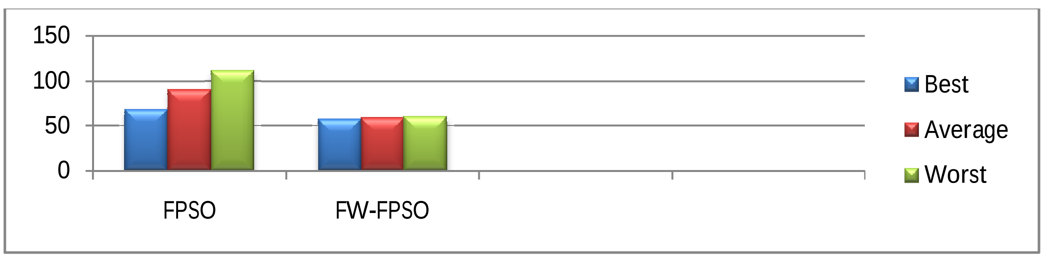

Figure 3: Average Performance Of Fuzzy Clustering Algorithms

Table: 3 Performance of PCI and PEI

Methods/ Datasets	IRIS		Wine	
	PCI	PEI	PCI	PEI
FPSO	0.87	0.31	0.89	0.36
FW-PSO	**0.93**	**0.34**	**0.95**	**0.38**

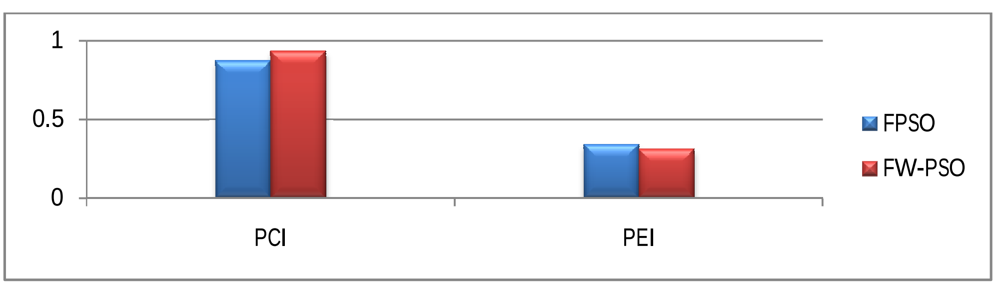

Figure 4: Performance Of The PCI And PEI For Iris Datasets

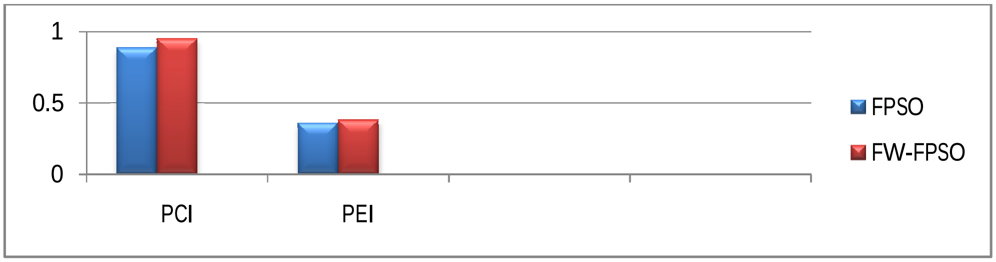

Figure 5: Performance Of The PCI And PEI For Iris Datasets

In table 3, demonstrate performance of the proposed algorithm in order to achieve by Partition Coefficient Index and Partition Entropy Index for all dataset what was used in experimental. Figure 4 and Figure 5 shows graphical representation of the clustering algorithms for all used datasets.

4.8 Discussions

The fuzzy clustering algorithms are applied for well known standard datasets, such as iris and wine data to calculate the performance of the used algorithms such as fuzzy particle swarm optimization and proposed feature weight – fuzzy particle swarm optimization. The clustering algorithms are implemented by MATLAB 2015b.

CONCLUSION

The classification based on unsupervised approach plays vital role in data mining and machine learning community. There are many unsupervised classification algorithms have been introduced by various researcher for different applications. Each approach has its own advantages and disadvantages. In present work, the Feature Weighted - Fuzzy Particle Swarm Optimization algorithm is proposed to overcome the demerits of the Fuzzy Particle Swarm Optimization. Experimental conducted over two well known data sets, Iris and Wine. The computational results show that the proposed Feature Weighted- Fuzzy Particle Swarm Optimization is efficient and can reveal very prominent results in term of their objective values.

REFERENCES

1. C. Borgelt, "Feature weighting and feature selection in fuzzy clustering," in Fuzzy Systems, FUZZ-IEEE (IEEE World Congress on Computational Intelligence). IEEE International Conference on, 2008; 838-844.
2. Y. Yafan, Z. Dayou, and H. Lei, "Improving fuzzy c-means clustering by a novel feature-weight learning," in Computational Intelligence and Industrial Application, 2008. PACIIA'08. Pacific-Asia Workshop on, 2008; 173-177.

3. D. S. Yeung and X. Wang, "Improving performance of similarity-based clustering by feature weight learning," IEEE transactions on pattern analysis and machine intelligence, 2002; 24: 556-561,.
4. J. C. Bezdek, R. Ehrlich, and W. Full, "FCM: The fuzzy c-means clustering algorithm," Computers & Geosciences, 1984; 10:191-203
5. F. Xu, W. Chen, and L. Yang, "Improved particle swarm optimization for realistic portfolio selection," in Software Engineering, Artificial Intelligence, Networking, and Parallel/Distributed Computing, 2007. SNPD 2007. Eighth ACIS International Conference on, 2007; 185-190.
6. E. Mehdizadeh and R. Tavakkoli-Moghaddam, "A Fuzzy Particle Swarm Optimization Algorithm for a Cell Formation Problem," in IFSA/EUSFLAT Conf., 2009;1768-1772.
7. C.-C. Hung, S. Kulkarni, and B.-C. Kuo, "A new weighted fuzzy C-means clustering algorithm for remotely sensed image classification," IEEE Journal of Selected Topics in Signal Processing, 2011; 5: 543-553,.
8. H.-J. Xing and M.-H. Ha, "Further improvements in feature-weighted fuzzy c-means," Information Sciences, 2014; 267: 1-15,.
9. J. Han, J. Pei, and M. Kamber, Data mining: concepts and techniques: Elsevier, 2011.
10. E. Mehdizadeh, "A fuzzy clustering PSO algorithm for supplier base management," International Journal of Management Science and Engineering Management, 2009; 4: 311-320,.
11. KrissnaPriya.R., “A Improved Classification of Network Traffic using Adaptive Nearest cluster Based Classifier”, International Journal of Computer Trends and Technology ISSN:2231-2803, January, 2015;.18(1)