A.Avani ,INDIA / International Journal of Research and Computational Technology
Vol.6 Issue.2 Pages: 25-31
ISSN: 0975-5662, September, 2014 www.ijrct.org

RESEARCH ON ENHANCEMENT OF K-MEANS CLUSTERING ALGORITHM FOR RBF NETWORK

Alla Avani[1], Research Scholar,
Rayala seema university, Kurnool,**A.P** , India
Dr.N.Satya Narayana2 ,Professor &Principal, nsn1208@gmail.com
Nagole Institute Of Technology, Hyderabad,INDIA

Abstract: This is research is for two updating methods to improve the clustering performance of adaptive k-means clustering. The proposed updating methods are suitable for off-line and on-line clustering. The capability of the updating methods are then compared to the existing updating methods using simulated and real data sets. Simulation results showed that the proposed updating methods have significantly improved the overall performance of RBF network. This paper also investigates some properties of adaptation method for on-line adaptive k-means clustering algorithm.

1.0 Introduction

The centre locations will influence the performance of radial basis function (RBF) networks. Poggio and Girosi [14] used all the training data as centres in their regularisation network that is based on RBF network. However, this may lead to network overfitting as the number of data becomes too large. To overcome this problem Poggio and Girosi [14] proposed a network with a finite number of centres. They also showed that a gradient descent approach used to update the RBF centres actually moved the centres towards the majority of the data, suggesting that a clustering algorithm may be used to position the centres.

K-means clustering is the most widely used clustering algorithm to position the RBF centres. Its simplicity and ability to perform on-line clustering may inspire this choice. However, k-means clustering algorithm can be sensitive to the initial centres and the search for the optimum centre locations may result in poor local minima. Many attempts have been made to minimise these problems [5], [6], [9], [11] and [16]. In this paper two updating rules were suggested as alternatives or improvements to the standard adaptive k- means clustering algorithm. The updating methods are proposed to give better overall RBF network performance rather than good clustering performance. However, there is a strong correlation between good clustering and the performance of the RBF network. The sensitivity of the RBF network to the centre locations will also be studied.

1.1 K-means Clustering Problems

K-means clustering algorithm works on the assumption that the initial centres are provided. The search for the final clusters or centres starts from these initial centres. Without a proper initialisation the algorithm may generate a set of poor final centres and this problem can become serious if the data are clustered using an on- line k-means clustering algorithm. In general, there are three basic problems that normally arise during clustering namely dead centres, local minima and centre redundancy.

Dead centres are centres that have no members or associated data. These centres are normally located between two active centres or outside the data range. The problem may arise due to bad initial centres, possibly because the centres have been initialised too far away from the data. Therefore, it is a good idea to select the initial centres randomly from the training data or to set them to some random values within the data range. However, this does not guarantee that all the centres are equally active. Some centres may have too many members and be frequently updated during the clustering process whereas some other centres may have only a few members and are hardly ever updated.

The centres in a RBF network should be selected to minimise the total distance between the data and the centres so that the centres can properly represent the data. A simple and widely used square error cost function can be employed to measure the distance, which is defined as:

$$E = \sum_{j=1}^{n_c} \sum_{i=1}^{N} \left\| v_i - c_j \right\|^2 \quad (1)$$

where N, and n_c are the number of data and the number of centres respectively; v_i is the data sample belonging to

centre c_j. Here, is taken to be an Euclidean norm although other distance measures can also be used. During the clustering process, the centres are adjusted according to a certain set of rules such that the total distance in equation
(1) is minimized. However, in the process of searching for the global minima the centres frequently become trapped at local minima. Poor local minima may be avoided by using algorithms such as simulated annealing, stochastic gradient descent, genetic algorithms, etc. However, these techniques normally involve heavy computation and not suitable for on-line clustering. In the present study, the improvements are made based on the adaptive k-means clustering, which do not require heavy computation.

In order to give a good modeling performance, the RBF network should have sufficient centres to represent the identified data. However, as the number of centre increases the tendency for the centers to be located at the same position or very close to each other is also increased. There is no point in adding extra centers if the additional centres are located very close to the centres that already exist. However, this is the normal phenomenon in k-means clustering and the unconstrained steepest descent algorithm, as the number of centres becomes sufficiently large [4].

1.2 K-means Clustering Algorithm

There are two existing basic versions of k-means clustering, a non-adaptive version introduced by Lloyd [12] and an adaptive version introduced by MacQueen [13]. The most commonly used k-means clustering is the adaptive k-means clustering based on the Euclidean distance [5]. Adaptive k-means clustering can be considered as a special case of the gradient descent algorithm where only the winning cluster is adjusted at each learning step. This paper concentrates only on adaptive k-means clustering as the algorithm can be used for on-line training of RBF network.

Adaptive k-means clustering tries to minimise the cost function in equation (1) by searching for the centre c_j on-line as the data are presented. As the data sample is presented, the Euclidean distances between the data sample and all the centres are calculated and the nearest centre is updated according to:

$$\Delta c_z(t) = \eta(t)\left[v(t) - c_z(t-1)\right] \quad (2)$$

where *z* indicates the nearest centre to the data v(t). Notice that, the centres and the data are written in terms of time *t* where $c_z(t-1)$ represents the centre location at the previous

clustering step. The adaptation rate, $\eta(t)$, can be selected in a number of ways. MacQueen [13] set $\eta(t) = 1/n_z(t)$, where $n_z(t)$ is the number of data samples that have been assigned to the centre up to the time t. Darken and Moody [5] used a constant adaptation rate and a square root method

$\left(\eta(t) = 1/\sqrt{n_z(t)}\right)$. Another method called search-then-converge has been introduced by Darken and Moody [6]. According to this method $\eta(t)$ is updated using:

$$\eta(t) = \eta(0)\frac{1 + \frac{\alpha}{\eta(0)}\frac{t}{\tau}}{1 + \frac{\alpha}{\eta(0)}\frac{t}{\tau} + \tau\frac{t^2}{\tau^2}} \quad (3)$$

The basic idea is to keep $\eta(t)$ approximately constant at times small compared to τ and decrease $\eta(t)$ at the rate of $\alpha/$ as time *t* becomes large compared to τ . This method yields optimally fast asymptotic convergence if $\alpha > 1/2\beta$,
where β is the smallest eigenvalue of the Hessian matrix of the cost function defined in equation (1) [6]. Chen et al. [3] used an adaptation rate that is updated at each step according to:

$$\eta(t) = \eta(t-1) \Big/ \sqrt{1 + int(t / n_c)} \quad (4)$$

where int(.) denotes the integer part of the argument and n_c is the number of centres.

The problem of assigning the adaptation rate to adaptive k-means clustering is very similar to the problem of assigning the learning rate to the back propagation algorithm. Both algorithms are based on the gradient descent method except that in back propagation all the parameters are updated at the same time. Therefore, all the methods that are used to choose the learning rate for the back propagation algorithm may also be applied for the adaptation rate in k-means clustering. These methods include the ones that have been suggested in references [2], [7], [10] and [15]. The usual
approach is to update $\eta(t)$ according to the variation of the cost function during the clustering process, such as [8]:

$$\Delta\eta(t) = \begin{cases} +a & \text{if } \Delta E < 0 \text{ consistently} \\ -b\eta(t-1) & \text{if } \Delta E > 0 \\ 0 & \text{otherwise} \end{cases}$$
(5)

where ΔE is the change in the cost function and, a and b are parameter constants. The term consistently in equation (5)

means a constantly decrease of E for the last few clustering steps. Cater [2] suggested that this kind of adaptive scheme can be made more effective if each parameter (the centre in this case) has a different adaptation rate.

Another method to improve the back propagation algorithm that may be adapted to k- means clustering is the method of momentum that has been introduced by Plaut et al [8]. For k-means clustering, a momentum term can be included as follows:

$$\Delta c_k(t) = \eta(t)\left[v(t) - c_k(t-1)\right] + \alpha \Delta c_k(t-1) \quad (6)$$

The momentum constant α is between 0 and 1, and is often
chosen to be close to 1. In this case, $\eta(t)$ can be a constant or adaptive. Other updating methods such as Newton method, stochastic method and conjugate gradient method may also be adapted to improve the k-means clustering algorithm at the expense of computational time.

In the current study, two updating methods are proposed as alternatives to update $\eta(t)$. The first method update
$\eta(t)$ according to:

$$\eta(t) = \eta(t-1) \Big/ e^{(1/\tau)} \quad (7)$$

where $r = n_c + t$ for off-line clustering and $r = \sqrt{n_c + t}$ for on-line clustering. The updating method uses different r for on-line and off-line clustering because in on-line clustering problems, $\eta(t)$ should be decreased rapidly so that the weights of the network can converge properly. This will not be a problem with the off-line clustering since the weights are estimated after the centres are located.

The second proposed updating method updates $\eta(t)$ according to:

$$\eta(t) = \eta(0)\left[e^{-p\left[t^2/n_c^3\right]} + be^{-\left[n_c\eta(t-1)\right]}\right] \quad (8)$$

where p is a constant, 0 < p 1 and $b = 1/(n_c + n_z(t))$. n_c and $n_z(t)$ are the number of centres and the number of data assigned to centre c_z up to time t respectively. This method involves two terms in the bracket on the right hand side. At the beginning, $\eta(t)$ will be dominated by the first term but as time t becomes large, $\eta(t)$ will converge to the value of b in the second term. The constant term p will determine how long $\eta(t)$ will be dominated by the first term.

In the present study, methods of updating $\eta(t)$ are selected such that the computational time will be minimised, which is beneficial for on-line clustering problems. For this reason the two proposed updating methods (described by equations (7) and (8)) together with the three methods that have been used by Chen et al. [3] and Darken and Moody [5] are studied:

1. $\eta(t) = 1/n_z(t)$, the MacQueen method.
2. $\eta(t) = 1/\sqrt{n_z(t)}$, the square root method
3. $\eta(t) = a/\sqrt{1 + int(t/n_c)}$, Chen?s method, where $a = \eta(0)$ for off-line clustering and $a = \eta(t-1)$ for on-line clustering.
4. $\eta(t) = \eta(t-1)/e^{(1/r)}$, where r = n_c + t for off-line clustering and $r = \sqrt{n_c + t}$ for on-line clustering, the first proposed method.
5. $\eta(t) = \eta(0)\left[e^{-p\left[t^2/n_c^3\right]} + be^{-\left[n_c\eta(t-1)\right]}\right]$, p is a constant, 0 < p 1 and $b = 1/(n_c + n_z(t))$, the second proposed method.

where n_c, $n_z(t)$ are the number of centres and the number of data assigned to centre c_z up to time t respectively. Notice that all these updating methods update the centres based on equation (2).

Simulation Results

The performance of k-means clustering algorithms using the proposed updating methods in previous section were tested using simulated and real data sets. System S1 was a simulated system defined by the following difference equation:

$$y(t) = 0.3\, y(t-1) + 0.6\, y(t-2) + u^3(t-1)$$

$$+ 0.3u^2(t-1) - 0.4\, u(t-1) + e(t) \quad (9)$$

where $e(t)$ was a Gaussian white noise sequence with zero mean and variance 0.05 and the input, $u(t)$ was a uniformly random sequence [-1,+1]. System S1 was used to generate 1000 pairs of data input and output. The second data set, S2 was taken from a heat exchanger system and consists of 1000 samples. A detailed description of the process can be found in Billings and Fadhil [1]. The third data set was taken from system S3 that is a tension leg platform and also consist

A.Avani ,INDIA / International Journal of Research and Computational Technology
Vol.6 Issue.2 *Pages: 25-31*
ISSN: 0975-5662, **September, 2014** **www.ijrct.org**

of 1000 input-output data samples.

Clustering performance was judged based on mean square distance (MSD) defined as:

$$E_{MSD}=\frac{1}{N}\sum_{j=1}^{n_c}\sum_{i=1}^{N}\left(\left\|v_i - c_j\right\|\right)^2 \quad (10)$$

The overall network performance was measured using mean squared error (MSE). The adaptive k-means clustering with updating methods are implemented and tested as part of the RBF network. The weights of the RBF network are updated using Least Squares algorithm as in reference [3]. During the testing, the same structures were assigned to all of the RBF networks. In this way, the performance of the clustering algorithm is measured under the same conditions for each updating method.

In these simulations, all the centres were initialised to the first few data samples. The MSD and MSE plots over the training and testing data sets for systems S1, S2 and S3 are shown in Figures (1a,b,c,d), (2a,b,c,d) and (3a,b,c,d) respectively. The initial updating parameters for the updating methods (3), (4) and (5) for systems S1, S2 and S3 are summarised in Tables (1), (2) and (3) respectively. The parameters are selected to give the smallest MSD for each updating method. Note that all the network models are trained using the off-line method, i.e. the RBF centres are located before the weights are estimated and all MSD and MSE are expressed in dB.

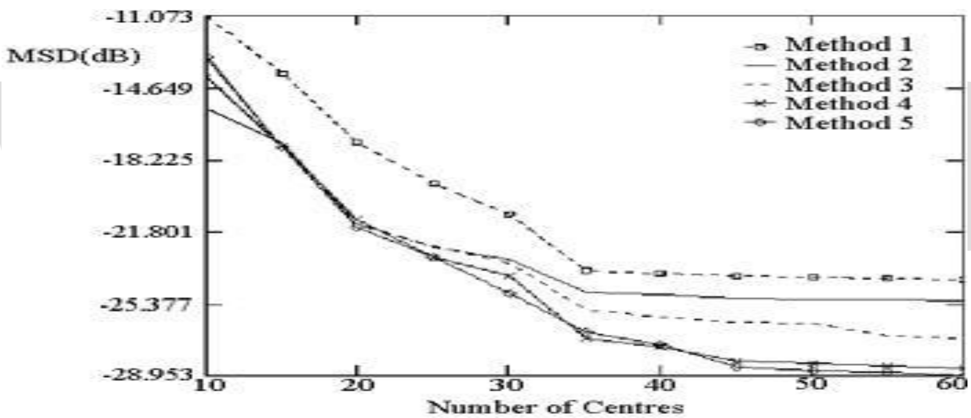

Figure (1a): MSD plots over training data set for data S1

Number of Centres	1	5	10	15	20	25	30	35	40	45	50	55	60
Method (3), (0)	0.1	0.9	0.7	0.9	0.9	0.9	0.95	0.95	0.95	0.95	0.95	0.95	0.95
Method (4), (0)	0.1	0.1	0.2	0.2	0.3	0.5	0.9	0.8	0.8	0.8	0.8	0.8	0.8
Method (5), (0)	0.1 0.1	0.9 0.1	0.9 0.1	0.9 0.1	0.9 0.1	0.9 0.1	0.9 0.1	0.9 0.1	0.9 0.1	0.9 0.1	0.9 0.1	0.85 0.1	0.8 0.1

and p														

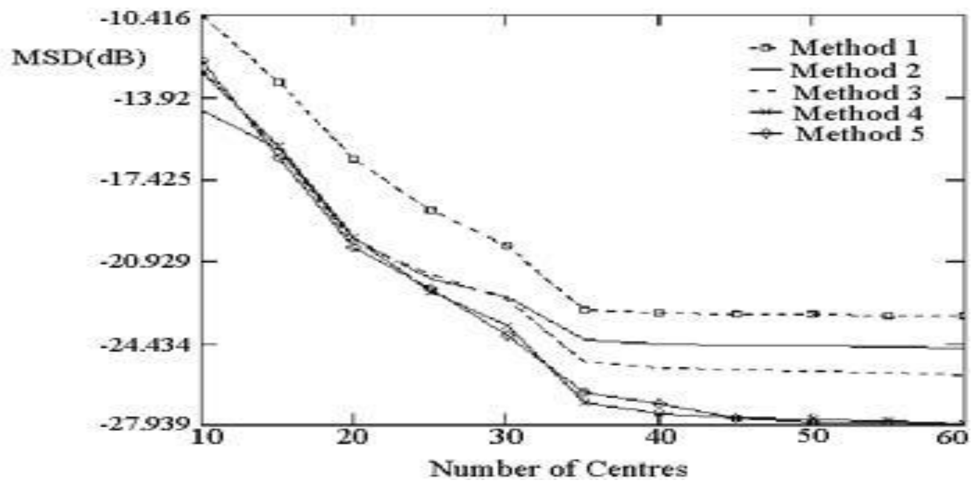

Figure (1b): MSD plots over testing data set for data S1

In all the examples, method 1 produced the worst MSD over both the training and testing data sets and the best MSD was obtained from method 5. Method 4 performs slightly worse than method 5 while method 2 and method 3 are between methods 1 and 4 (refer to Figures 1a,b, 2a,b

Table (1): Initial updating parameters for method (3), (4) and (5) for system S1

Table (2): Initial updating parameters for method (3), (4) and (5) for system S2

3.0 Conclusion

Two updating methods have been proposed and were tested using one simulated and two real data sets. The simulation results showed that the proposed updating methods have significantly improved the performance of k-means clustering algorithm. K-means clustering algorithm that uses both the proposed updating methods offer smaller MSD for the three data sets. Due to the strong correlation between the good clustering and the overall RBF performance, both the proposed updating methods provide significantly better overall performance than the other three updating methods that are considered.

Simulation results also suggested that for on -line clustering the clustering rate and steady state value of adaptation,
$\eta(t)$ should be given extra care. If clustering rate is too small the centres will not be position properly. However, large clustering rate often result in large steady state value.

variation of centre positions. Hence the good updating method is the one that has large clustering rate at the beginning but small steady state value of $\eta(t)$ at the end of training time. The proposed updating method (referred as method 5 in this paper) was designed to satisfy this condition. Thus, this method can offer good overall RBF network performance for both off-line and on-line training.

4.0 References

1. Cater, J.P., 1987,Successfully using peak learning rates of 10 (and greater) in back propagation networks with the heuristic learning algorithm, in IEEE First Int. Conf. on Neural Networks (San Diego 1987), Caudill, M., and Butler, C. (eds.), II, 645-651, IEEE, New York.
2. Chen, S., Billings, S.A. and Grant, P.M., 1992, Recursive hybrid algorithm for non-linear system identification using radial basis function networks, Int. J. of Control, 55, 1051-1070.
3. Cichocki, A., and Unbehauen, R., 1993, Neural Networks for Optimisation and Signal Processing, Wiley, Chichester.
4. Darken, C., and Moody, J., 1990, Fast adaptive k-means clustering: Some empirical results, Int. Joint Conf. on Neural Networks, 2, 233-238.
5. Darken, C., and Moody, J., 1992, Towards fast stochastic gradient search, In: Advance in neural information processing systems 4, Moody, J.E., Hanson, S. J., and Lippmann, R.P. (eds.), Morgan Kaufmann, San Mateo.
6. Franzini, M.A., 1987, Speech recognition with back propagation, In Proc. of the Ninth Annual Conf. of the IEEE Eng. in Medicine and Biology Society, 1702-1703, IEEE, New York.
7. Hertz, J., Krogh, A. and Palmer R.G., 1991, Introduction to the theory of neural computation, Addison Wesley, New York.
8. Ismail, M.A., and Selim, S.Z., 1986, Fuzzy c-means: optimality of solutions and effective termination of the algorithm, Pattern Recognition, 19, 481-485.
9. Jacobs, R.A., 1988,Increased rates of convergence through learning rate adaptation, Neural Networks, 1, 295-307.
10. Kamel, M.S., and Selim, S.Z., 1994, ?New algorithms for solving the fuzzy clustering problem, Pattern Recognition, 27 (3), 421-428.
11. Lloyd, S.P., 1957, Least squares quantization in PCM, Bell Laboratories Internal Technical Report, IEEE Trans. on Information Theory. MacQueen, J., 1967, Some methods for classification and analysis of multi-variate observations,in procs.of Fifth Barkly symposium ,California press 281.

