A Novel Radial Basis Function Neural Network Classifier with Centers Set By Cooperative Clustering

Shaomin Mu, Shengfeng Tian, Chuanhuan Yin

Abstract

The selection of centers and widths has a strong influence on the performance of radial basis function neural network classifier. In this paper, a novel approach of clustering based on Fuzzy C- means clustering is proposed, which is called cooperative clustering, and use it for selection of centers of radial basis function neural network. Experimental results show that the performance of classification using our approach is better than radial basis function neural network.

Keywords: Cooperative Clustering, RBF neural network, Classification.

1. Introduction

The Radial Basis Function (RBF) neural network is a kind of powerful kernel methods, which has been applied to many areas with success [1], such as time series analysis, image processing and bioinformatics [2-4]. In recent years, some improved learning algorithms for RBF neural network have been proposed by several present papers [5-8].

A successful construction of an RBF neural network needs three steps: the selection of the kernel (basis) function, the determination of proper kernel centers, and the learning of the weights connecting the hidden nodes to the output nodes. The key problem is particularly the selection of basis function centers [9], [10]. The most frequently used basis function is the Gaussian kernel function. In the existing literature one often used C-Means or Fuzzy C-Means clustering algorithm [11] to select the centers in the input space.

There are two strategies to determine the basis function centers. The first strategy simply places one center at each sample [7].On the other hand, the second strategy attempts to reduce the number of basis function centers, which can increase the classification efficiency. In the typical procedure, a clustering analysis is employed on the training samples and then allocates one basis function center for each cluster of samples. As pointed by Oyang [6], the distributions of training samples near the boundaries between different classes carry more crucial information than the distributions of samples in the inner parts of the classes. The RBF neural networks constructed with a conventional cluster-based learning algorithm generally are not able to deliver the same level of accuracy as those data classification algorithms that exploit the distributions of samples near boundaries between different classes. Several approaches have been proposed to solve the problem, such as the Orthogonal Least Square (OLS) algorithm, Support Vector Machine (SVM)-based selection [9], class reparability measure based selection, feature subset based selection, data structure preserving selection [12], [13].

After the kernel functions and centers have been determined, the network can be viewed as an equivalent single-layer network with linear output units. Then, the Least Mean Square Error (LMSE) method and regularization theory can be used to determine the weights associated with the links between the hidden layer and the output layer [14-17].

In this paper, we proposed an efficient data clustering method which is called cooperative clustering, and use it for selection of centers of radial basis function neural network. The experimental results show that the approach could not only effectively improve the training speed, but also keep classification accuracies.

The remainder of this paper is organized as follows. In the next section, we briefly review the theory of RBF neural network. In Section 3, the cooperative clustering algorithm is described. In Section 4, we address the problem of how to construct the RBF neural network with cooperative clustering. The experimental results using our algorithm are shown in Section 5. Finally, conclusion and discussion are presented in Section 6.

2. RBF Neural Network

RBF neural network is a type of feed forward network and able to solve non-linear problem by using a special class of functions. Fig.1 illustrates a standard structure of RBF neural network.

Corresponding Author: Shaomin Mu is with the School of Computer and Information Technology, Beijing Jiaotong University, Beijing, 100044, P.R.China.
E-mail: msm@sdau.edu.cn

As revealed in Fig. 1, the structure of a normal RBF neural network consists of three different layers, namely the input layer, the hidden layer, and the output layer. The input layer is made up of source nodes (sensory units) whose number is equal to the dimension of the input vector; the second layer is hidden layer (Gaussian transfer functions typically used) which is composed of nonlinear units that are connected directly to all of the nodes in the input layer. The transformation from the input space to the hidden unit space is nonlinear, usually, the cluster algorithm associates a cluster center to each RBF node in the hidden layer, and the transformation from hidden unit space to output space is linear.

Building a Gaussian RBF neural network for a given learning task mainly involves:

a) Choice of Activation Functions
b) Selection of Centers.
c) Selection of Widths.
d) Solving for the weight coefficients and bias.

In this paper, we mainly studied on how to selection of proper centers for Gaussian RBF neural network.

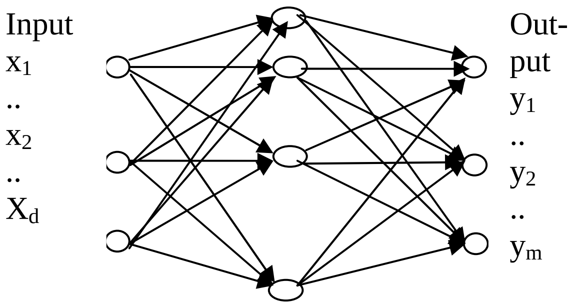


Figure 1. A structure of RBF network with multiple outputs

3. Cooperative Clustering

In this section, the method of cooperative clustering based on Fuzzy C-Means (FCM) algorithm [13] is described, the idea of which is derived from Cooperative Evolution and SVM.

3.1 Fuzzy C-Means

The FCM algorithm is an unsupervised classification and utilized in a wide variety of applications. FCM partitions a given dataset, $X = \{\boldsymbol{x}_1,\ldots,\boldsymbol{x}_n\}\subset R^d$, into c fuzzy subsets by minimizing the following objective function.

$$\sum_{i=1}^{c} u_{ik} = 1,\ k = 1,\cdots,n. \quad (1)$$

Where c is the number of clusters, n the number of data points, u_{ik} the grade of membership of data point $\boldsymbol{x}_k$ in class i, m $(1<m<\infty)$ is the quantity controlling clustering fuzziness, V the set of cluster centers ($\boldsymbol{v}_i \in R^d$). The matrix U with the ik^{th} entry u_{ik} is constrained to the condition.

$$J_m(U,V) = \sum_{i=1}^{c}\sum_{k=1}^{n} u_{ik}^m \|\boldsymbol{x}_k - \boldsymbol{v}_i\|^2 \quad (2)$$

The FCM algorithm consists of the following steps:

(1)Set c, m, and initialize the cluster centers $\boldsymbol{v}_i$, i=1,…, c.

(2)Compute the degree of membership of all feature vectors in the entire cluster, as follow

$$u_{ik} = \left[\sum_{j=1}^{c}\left(\frac{D_{ik}}{D_{jk}}\right)^{\frac{2}{(m-1)}}\right]^{-1} \quad (3)$$

Where $D_{ik}^2 = \|\boldsymbol{x}_k - \boldsymbol{v}_i\|^2$.

If the set $I_k = \{i|1\le i \le c;\ D_{ik} = 0\}$ is not empty, we have

$$u_{ik}=0,\ \forall i\in\{1,\ldots,c\}-I_k, \text{ and } \textstyle\sum_{i\in Ik} u_{ik} =1. \quad (4)$$

(3) Update the cluster centers $\boldsymbol{v}_i$, i = 1,…, c, as follows

$$\boldsymbol{v}_i = \sum_{k=1}^{n}(u_{ik})^m \boldsymbol{x}_k \Big/ \sum_{k=1}^{n}(u_{ik})^m \quad \forall i. \quad (5)$$

(4) Return the matrix U if no cluster center is changed in Step (3), otherwise go to Step (2).

3.2 Cooperative Clustering

In the approach of cooperative clustering, we iteratively compute the cluster centers of two classes simultaneously and draw them towards the boundary of the two classes. Then these cluster centers near the boundary can be used to replace basis function center approximately.

Suppose that there are two-class datasets X^+ and X^-, and each class has c subsets. Let $\boldsymbol{v}^+ = \{\boldsymbol{v}_1^+,\cdots,\boldsymbol{v}_c^+\}$ be the set of cluster centers in class +, $\boldsymbol{v}^- = \{\boldsymbol{v}_1^-,\cdots,\boldsymbol{v}_c^-\}$ be the set of cluster centers in class -, and A be the distance matrix between two sets v+ and v-. The ij-th entry a_{ij} of matrix A can be computed as

$$a_{ij} = \|\boldsymbol{v}_i^+ - \boldsymbol{v}_j^-\|^2 \quad i, j = 1,\cdots,c \quad (6)$$

We take out each pair $(\boldsymbol{v}_p^+,\boldsymbol{v}_q^-)$ of cluster centers with smallest distance from sets $\boldsymbol{v}^+$ and $\boldsymbol{v}^-$ iteratively according to matrix A. The two cluster centers in $(\boldsymbol{v}_p^+,\boldsymbol{v}_q^-)$ should move towards another. Let r_p^+ be the average radius of cluster p in class + and r_q^- be the average radius of cluster q in class -. They are defined by equation (7) and (8)

$$r_p^+ = \sum_{x_k\in X^+} u_{pk} \|\boldsymbol{x}_k - \boldsymbol{v}_p^+\| \Big/ \sum_{x_k\in X^+} u_{pk} \quad (7)$$

$$r_q^- = \sum_{x_k\in X^-} u_{qk} \|\boldsymbol{x}_k - \boldsymbol{v}_q^-\| \Big/ \sum_{x_k\in X^=} u_{qk} \quad (8)$$

Then each pair $(\boldsymbol{v}_p^+, \boldsymbol{v}_q^-)$ is updated as follows

$$\boldsymbol{v}_p^+ = \boldsymbol{v}_p^+ + \lambda \frac{r_p^+}{r_p^+ + r_q^-}(\boldsymbol{v}_q^- - \boldsymbol{v}_p^+) \quad (9)$$

$$\boldsymbol{v}_q^- = \boldsymbol{v}_q^- + \lambda \frac{r_q^-}{r_p^+ + r_q^-}(\boldsymbol{v}_p^+ - \boldsymbol{v}_q^-) \quad (10)$$

Where $\lambda\in(0, 1)$ is the quantity controlling the distance between $\boldsymbol{v}_p^+$ and $\boldsymbol{v}_q^-$.

The whole procedure of the cooperative clustering is as follows.

(1) Initialize the cluster centers $\boldsymbol{v}_i^+$ and $\boldsymbol{v}_i^-$, $i = 1,\ldots, c$.
(2) Compute the matrix U^+ and U^- for classes + and – with equations (3) and (4) respectively.
(3) Partition the datasets X^+ and X^- into clusters X_i^+ and X_i^-, $i = 1,...,c$, respectively.
(4) Compute the cluster centers $\boldsymbol{v}_i^+$ and $\boldsymbol{v}_i^-$, $i = 1,\ldots,c$, for classes + and – with equation (5) respectively.
(5) Computer the distance matrix A with equation (6) and find out the set V^s of cluster center pairs by searching through matrix A.
(6) Update the cluster center pairs in set V^s with equations (9) and (10).
(7) Return the set V^s if it is not changed compared with last iteration, otherwise go to Step (2).

The step (5) can be realized as follows.

1) Compute the matrix $A=\{a_{ij}\}_{1\le i,j\le c}$, with equation (6).
2) Find a smallest element a_{pq} in A.
3) Add $\boldsymbol{v}_p^+$ and $\boldsymbol{v}_q^-$ into the set V^s.
4) Delete the row p and the column q from matrix A.
5) Return the set V^s if no element in A otherwise go to step (2).

With the above procedure, we can find c pairs of cluster centers. Each pair crosses the boundary of two classes. Let the number of total cluster centers be p=2c, the time complexity of the procedure is (*#iterations*) × $O(np)$.

4. Training RBF Neural Network for Classification

We consider the binary classification problems. The general mathematical form of the outputs in an RBF neural network is described as follows:

$$y(\boldsymbol{x}) = \boldsymbol{w} \cdot \phi(\boldsymbol{x}) = \sum_{l=1}^{p} w_l \cdot \phi_l(\boldsymbol{x}) \quad (11)$$

Where p=2c, $\phi(\boldsymbol{x})$ is basis function of RBF neural network, the Gaussian kernel function is selected as basis function of RBF neural network.

$$\phi_l(\mathbf{x}) = \exp\left(-\frac{\|\mathbf{x} - \boldsymbol{v}_l\|^2}{2\sigma_l^2}\right) \quad (12)$$

where $\boldsymbol{v}_l$ is the center of cluster l, σ_l is the bandwidth of its kernel function, Moody and Darken[18] proposed a method which determines the bandwidth of each kernel function, it is as follows:

$\sigma_l = \beta \cdot r_l$, $l = 1,\ldots,p$.

where β is a constant, r_l is the radius of cluster l.

Given a training data set of n points (x_i,y_i), $i = 1,\ldots,n$, where $\boldsymbol{x}_i\in R^d$, and $y_i\in\{-1,1\}$, the solution is to solve a optimization problem. With the least mean square method and regularization theory, the objective function is as follows:

$$\begin{aligned}\min\ L &= \sum_{i=1}^{n}[w\cdot\phi(x_i) - y_i]^2 + \lambda w \cdot w \\ &= \sum_{i=1}^{n}[\sum_{l=1}^{p} w_l\phi_l(x_i) - y_i]^2 + \lambda\sum_{l=1}^{p} w_l^2\end{aligned} \quad (13)$$

where λ is the regularization parameter.

To find the optimal $\boldsymbol{w}$ that minimizes L, we set the gradient of L to be zero:

$$\frac{\partial L}{\partial w_j} = 2\sum_{i=1}^{n}[\sum_{l=1}^{p} w_l\phi_l(x_i) - y_i]\cdot\phi_j(x_i) + 2\lambda w_j = 0 \quad (14)$$

We have

$$\sum_{i=1}^{n}[\sum_{l=1}^{p}\phi_l(x_i)\cdot\phi_j(x_i)\cdot w_l] + \lambda w_j = \sum_{i=1}^{n} y_i\phi_j(x_i) \quad (15)$$

The weight vector $\boldsymbol{w}$ can be found with the following equation:

$$\begin{bmatrix}\sum_{i=1}^{n}\phi_1(x_i)\phi_1(x_i)+\lambda & \cdots & \sum_{i=1}^{n}\phi_1(x_i)\phi_p(x_i)\\ \sum_{i=1}^{n}\phi_2(x_i)\phi_1(x_i) & \cdots & \sum_{i=1}^{n}\phi_2(x_i)\phi_p(x_i)\\ \vdots & \vdots & \vdots\\ \sum_{i=1}^{n}\phi_p(x_i)\phi_1(x_i) & \cdots & \sum_{i=1}^{n}\phi_p(x_i)\phi_p(x_i)+\lambda\end{bmatrix}\begin{bmatrix}w_1\\ w_2\\ \cdots\\ w_p\end{bmatrix} = \begin{bmatrix}\sum_{i=1}^{n} y_i\phi_1(x_i)\\ \sum_{i=1}^{n} y_i\phi_2(x_i)\\ \vdots\\ \sum_{i=1}^{n} y_i\phi_p(x_i)\end{bmatrix} \quad (16)$$

The problem becomes to solve an inverse matrix, the size of which is $p \times p$. The algorithm is efficient.

Next we consider the multiclass classification problems. Given a training dataset of n points $(\boldsymbol{x}_i,y_i)$, $i = 1,\ldots,n$, where $\boldsymbol{x}_i\in R^d$, and $y_i\in\{1, 2,\ldots, m\}$, the m-class classification problem is to find a decision function F: $R^d \to \{1, 2,\ldots, m\}$ mapping an input x to a class label y.

The procedure for determining hidden layer centers is divided into two steps. In the first step, the FCM algorithm is used to partition every class into c clusters. In the second step, cluster centers are moved towards the classification boundaries with cooperative clustering iteratively. In each iteration, two closest cluster centers in two classes are selected and moved towards each other. Let the set of cluster centers be $V = \{\boldsymbol{v}_{11}, \ldots, \boldsymbol{v}_{1c}, \ldots, \boldsymbol{v}_{m1}, \ldots, \boldsymbol{v}_{mc}\}$, where $\boldsymbol{v}_{ij}$ is the jth cluster center of class i.

The selected pair of cluster centers ($\boldsymbol{v}_{ip}$, $\boldsymbol{v}_{jq}$) in class i and j should satisfy the condition:

$$\|\boldsymbol{v}_{ip} - \boldsymbol{v}_{jq}\| = \min_{\boldsymbol{v}_{is}, \boldsymbol{v}_{jt} \in V} \|\boldsymbol{v}_{is} - \boldsymbol{v}_{jt}\| \quad (17)$$

Then the pair ($\boldsymbol{v}_{ip}$, $\boldsymbol{v}_{jq}$) is updated as follows

$$\boldsymbol{v}_{ip} = \boldsymbol{v}_{ip} + \lambda \frac{r_{ip}}{r_{ip} + r_{jq}} (\boldsymbol{v}_{jq} - \boldsymbol{v}_{ip}) \quad (18)$$

$$\boldsymbol{v}_{jq} = \boldsymbol{v}_{jq} + \lambda \frac{r_{jq}}{r_{ip} + r_{jq}} (\boldsymbol{v}_{ip} - \boldsymbol{v}_{jq}) \quad (19)$$

where $\lambda \in (0,1)$ is the quantity controlling the distance between $\boldsymbol{v}_{ip}$, and $\boldsymbol{v}_{jq}$, and r_{ip} is the average radius of the pth cluster of class i.

It is noticed that only the cluster centers in adjacent classes can be moved. To judge if two cluster centers $\boldsymbol{v}_{ip}$, and $\boldsymbol{v}_{jq}$ are in adjacent classes, we search a cluster center $\boldsymbol{v}_{hs}$ which is the closest to the middle of $\boldsymbol{v}_{ip}$, and $\boldsymbol{v}_{jq}$. If $h=i$ or $h=j$, the two cluster centers $\boldsymbol{v}_{ip}$ and $\boldsymbol{v}_{jq}$ are in adjacent classes. The cluster center $\boldsymbol{v}_{hs}$ can be determined by

$$d(h,s,i,p,j,q) = \min_{g,t} d(g,t,i,p,j,q) = \min_{g,t} \left\| \frac{\boldsymbol{v}_{ip} + \boldsymbol{v}_{jq}}{2} - \boldsymbol{v}_{gt} \right\| \quad (20)$$

Where $d(g, t, i, p, j, q)$ is the distance of the cluster center $\boldsymbol{v}_{gt}$ and the middle of $\boldsymbol{v}_{ip}$, and $\boldsymbol{v}_{jq}$.

To make the cluster centers be the members of the dataset, $\boldsymbol{v}_{ip}$ and $\boldsymbol{v}_{jq}$ should be replaced by the nearest data point of the same cluster respectively. Let X_{ij} be the subset of the jth cluster in class i, $i = 1,...,m$, $j = 1,...,c$, the new pair ($\boldsymbol{x}_r$, $\boldsymbol{x}_t$) should be chosen according to

$$\|\boldsymbol{v}_{ip} - \boldsymbol{x}_r\| = \min_{\boldsymbol{x}_k \in X_{ip}} \|\boldsymbol{v}_{ip} - \boldsymbol{x}_k\| \quad (21)$$

and

$$\|\boldsymbol{v}_{jq} - \boldsymbol{x}_t\| = \min_{\boldsymbol{x}_k \in X_{jq}} \|\boldsymbol{v}_{jq} - \boldsymbol{x}_k\| \quad (22)$$

The procedure of the cooperative clustering is as follows.

(1) Initialize sets $V_i = \{\}$, i=1, …, m.
(2) Initialize the cluster center set $V = \{\boldsymbol{v}_{11}, \ldots, \boldsymbol{v}_{1c}, \ldots, \boldsymbol{v}_{m1}, \ldots, \boldsymbol{v}_{mc}\}$.
(3) Do classical clustering procedures for all classes with FCM algorithm.
(4) Partition the dataset into clusters X_{ij}, $i = 1,...,m$, $j = 1,...,c$.
(5) Initialize sets $N_i = \{\}$, i=1, …, m.
(6) For i = 1 to m, do
(7) For $j = i$+1 to m, do
(8) Find a pair of cluster centers ($\boldsymbol{v}_{ip}$, $\boldsymbol{v}_{jq}$), $(i, p) \notin N_j$ and $(j, q) \notin N_i$, with equation (17).
(9) Find $\boldsymbol{v}_{hs}$ for the pair of cluster centers ($\boldsymbol{v}_{ip}$, $\boldsymbol{v}_{jq}$) with equation (20).
(10) If $h=i$ or $h=j$, then
(11) Delete $\boldsymbol{v}_{ip}$ and $\boldsymbol{v}_{jq}$ from V and update them with equations (18)-(19) and (21)-(22).
(12) Add pair (j, q) to N_i and add pair (i, p) to N_j.
(13) End If
(14) End For
(15) End For
(16) If any pair of cluster centers is updated from step (6) to step (15), go to step (6).
(17) If the greatest change of the cluster center pairs is less than a given threshold, go to step (19).
(18) Rebuild cluster center set V and matrix U for all classes with equations (3)-(5). Go to step (4).
(19) For i = 1 to m, do
(20) Add $\boldsymbol{v}_{ij}$, $1 \le j \le c$, to V_i. Add $\boldsymbol{v}_{jq}$, $(j, q) \in N_i$, to V_i.
(21) End For
(22) End

With the above procedure, we obtain the cluster center set V_i for class i, i=1, …, m. The data set V_i includes all cluster centers in class i and cluster centers of other classes, each of which constructs a pair with a cluster center of class i. The time complexity of the procedure is (#*iterations*) × ($O(nmc)$ + $O(m^3c^3)$).

We adopt one-against-rest strategy to solve multiclass classification problems. For one-against-rest approach, m binary classifiers are needed and the training data for each binary classifier include all data in m classes. When an example is classified, m times binary tests are required to make final decision with majority voting. To obtain the ith RBF classifier f_i, the set V_i is used to construct the hidden layer and all examples are used to train the RBF neural network. If the m binary classifiers are combined to a whole RBF neural network, the size of the hidden layer is mc.

5. Experiments

In this section, to verify the performance of our algorithm proposed, the experiments were performed on several synthetic and real-world data sets.

In the experiments, we compare two types of RBF neural networks:

(1) The hidden layer is constructed by cluster centers produced with traditional fuzzy c-means. The network is denoted by traditional RBF neural network.

(2) The hidden layer is constructed by cluster centers produced with cooperative clustering.

A. *Synthetic Data Results*

The synthetic data included two data sets. The first data set has 200 samples in two classes and 2 dimension vector data. We choose RBF kernel $K(\boldsymbol{x},\boldsymbol{x}')=\exp[-\|\boldsymbol{x}-\boldsymbol{x}'\|^2/\sigma^2]$ in RBF network and λ=0.6 in cooperative clustering procedure.

The cooperative clustering results are shown in Fig.2. Tri-angle denotes the sample of synthetic data. 6 pairs of clustering centers are chosen by cooperative clustering, which denoted by large circles. As shown in the Fig.2, the cluster centers selected by cooperative clustering are closely near the boundary of the two classes.

The second data set contains four classes, each of which has 50 points, and 2 dimension vector data. In Fig. 3, the performance of the proposed RBF neural network classifier is compared with the conditional RBF neural network on the data set. It is clear that centers of our method is more near the boundaries between different classes and carry more crucial information as pointed by Oyang.

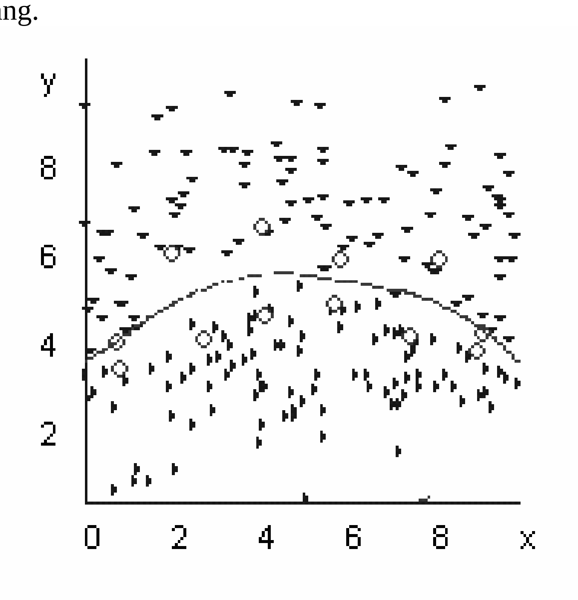
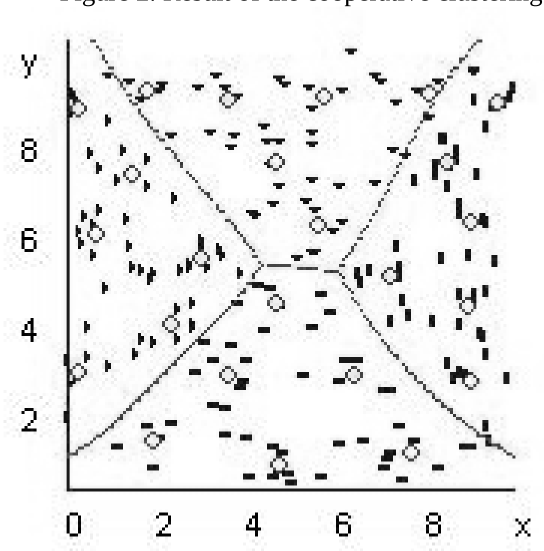
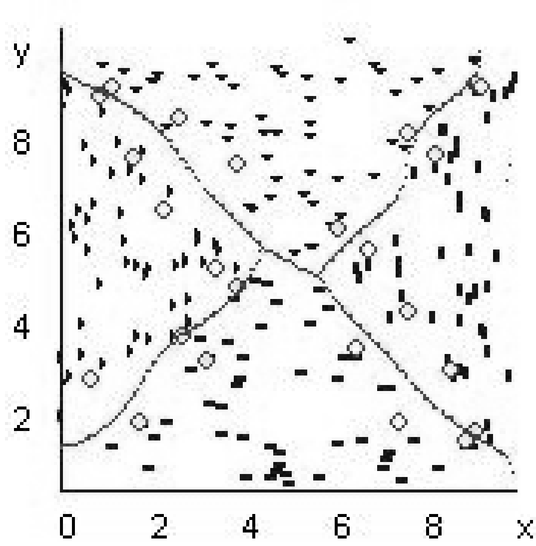


Figure 2. Result of the cooperative clustering



(a) Result of traditional RBF neutral network



(b) Result of our method

Figure 3 Result of the traditional RBF neutral network and our method

B. *Real Data Results*

All these datasets are from the UCI repository [19].Six benchmark data sets are used in the experiments, these data sets include 2 two–class sets and 4 multi-class sets. The features of these datasets are shown in Table 1. For all problems, we linearly scale all training data set to be in [-1, 1] for features with plus and minus values and in [0, 1] for features with only plus values. Then testing data are scaled accordingly. The parameters λ, σ and β are optimized for best accuracy.

Table 1. Benchmark data sets used for testing

Data Set	Classification Accuracy (%)	
	traditional RBF network	Our method
iris	97.5	98
glass	69.16	73.4
wine	96.0	96.7
Pima	74.7	79.6
Vehicle	78.25	79.5
Segment	94.98	97

It has been found that our method outperformed the traditional RBF neural network in the Table 2.

Table 2 Comparison of classification accuracy on six data sets

Problem	#training data	#attribute	#class
iris	150	34	2
glass	214	8	2
wine	178	18	4
Pima	768	19	7
Vehicle	846	4	18
Segment	2310	19	7

6. Conclusions

In this paper, we have proposed an efficient algorithm of clustering which called cooperative clustering; it is used to efficiently select the centers in RBF neural network. The cluster centers near the boundary of two classes can be found with the cooperative clustering. Due to the data points near the class boundary, they can carry more classification information; this method takes advantage to the center selection strategy with the traditional clustering method. The experimental results have been done to validate this idea.

In the future, we will apply the approach proposed in this paper to solve function approximation problems and more practical problems.

Acknowledgment

This work was supported by the National Natural Science Foundation of China by Grants 60503017 and the Science and Technology Foundation of Beijing Jiaotong University (No. 2006XM007).

References

[1] Z.R. Yang, "A novel radial basis function neural network for discriminate analysis," *IEEE Trans.on Neural Networks*, vol. 7, pp. 604-612, 2006.

[2] D.Zhang , L.F.Deng and K.Y. Cai., "Fuzzy nonlinear regression with fuzzified radial basis function network," *IEEE Trans. on Fuzzy Systems,* vol. 13, no. 6, pp. 742-760, 2005.

[3] M. Visani, C.Garcia, J.M.Jolion, "Normalized radial basis function networks and bilinear discriminate analysis for face recognition," *Proceedings. IEEE Conference on Advanced Video and Signal Based Surveillance*, pp.342- 347, 2005.

[4] R. Zhang, Sundararajan, N.Huang,G.B., "An Efficient Sequential RBF Network for Gene Expression-Based Multi-category classification," *Computational Intelligence in Bioinformatics and Computational Biology*, pp. 1-6, 2006.

[5] G. Z. Zhang and D. Huang, "Radial basis function neural network optimized by a genetic algorithm for soybean protein sequence residue spatial distance prediction," *Evolutionary Computation*, vol. 1, pp. 1015-1019, 2004.

[6] G. B. Huang, P. Saratchandran, N. Sundararajan, "A generalized growing and pruning RBF (GGAP-RBF) neural network for function approximation," *IEEE Trans.on Neural Networks,* vol. 16, no. 1, pp. 57-67, 2005.

[7] Y. J. Oyang, S. C. Hwang. Y. Y. Ou C. Y. Chen and Z.-W. Chen, "Data classification with radial basis function networks based on a novel kernel density estimation algorithm," *IEEE Transaction on Neural Networks*, vol. 16, no. 1, pp. 225-236, 2005.

[8] Y. Y. Ou, Y. J. Oyang and C. Y. Chen, "A novel radial basis function network classifier with centers set by hierarchical clustering," *in Proceedings of Int. Joint. Conf. On Neural networks, Montreal*, Canada, pp. 1383-1388, 2005.

[9] B. Schlkopf, K.-K. Sung, C. J. C. Burges, F. Girosi, P. Niyogi, T. Poggio and V. Vapnik, "Comparing support vector machines with Gaussian kernels to radial basis function classifiers," *IEEE Transaction on Signal Processing,* vol. 45, no. 11, pp. 2758-2765, 1997.

[10] K. Warwick, J. D. Mason, E. L. Sutanto, "Neural Network Basis function Center Selection Using Cluster analysis," *American Control Conference*, vol. 5, pp. 3780 - 3781, 1995.

[11] J. C. Bezdek, *Pattern Recognition with Fuzzy Objective Function Algorithms*, New York: Plenum Press, 1981.

[12] K. Z. Mao and G. B. Huang, " Neuron selection for RBF neural network classifier based on data structure preserving criterion," *IEEE Transaction on Neural Networks*, vol. 16, no. 6, pp. 1531-1540, 2005.

[13] K. Z. Mao, "Fast orthogonal forward selection algorithm for features subset selection," *IEEE Transaction on Neural Networks*, vol. 13, no. 5, pp. 1218-1224, 2002.

[14] N. B. Karayiannis, " Reformulated radial basis neural networks trained by gradient descent," *IEEE Transaction on Neural Networks*, vol. 10, no. 3, pp. 657-671, 1999.

[15] H. A. Schwenker, Kestler and G. Palm，"Three learning phases for radial-basis-function networks, *IEEE Transaction on Neural Networks*," vol. 14, pp. 439-458, 2001.

[16] N. B. Karayiannis, "Reformulated radial basis neural networks trained by gradient descent," *IEEE Transaction on Neural Networks*, vol. 10, no. 3, pp. 657-671, 1999.

[17] M. M. Brizzotti, A. C. P.L.F.Carvalho, "Comparing Different Clustering Techniques RBF Networks Training," *VI Brazilian Symposium on Neural Networks*, pp. 225-230, 2000.

[18] J. Moody and C. J. Darken, "Fast learning in networks of locally-tuned processing units," *IEEE Transaction on Neural Computation*, vol. 1, no. 2, pp. 281-294, 1989.

[19] C. L. Blake and C. J. Merz. "UCI repository of machine learning databases," *Dept. Inform. Computalifornia*, Irvine, CA, 1998

Shaomin Mu was born in China ,in 1964. he received the master's degree in computer science in 2000, from School of Information Science and Engineering, Shandong Agriculture University, doctor candidate Beijing Jiaotong University in China. His current research interests in machine learning include supervised learning, kernel methods, and pattern recognition.

Shengfeng Tian was born in 1944. He is a professor with the School of Computer and Information Technology, Beijing Jiaotong University in China. Currently, his research interests include support vector machines and network intrusion detection.

Chuanhuan Yin was born in 1976. He reseived Ph. D from Beijing Jiaotong University in China. He majors in machine learning and network security.