Structure Design for RBF Neural Network Based on Improved K-means Algorithm

Sun Yuqing, Qiao Junfei, Han Honggui

College of Electronic Information & Control Engineering, Beijing University of Technology; Beijing Key Laboratory of Computational Intelligence and Intelligent System, Beijing, 100124.
E-mail:daqing1125@126.com

Abstract: Aiming at the disadvantages of the traditional K-means clustering algorithm, a new algorithm based on density is proposed to remove the noises and outliers in this paper. This algorithm determines whether a point is a noise or not according to the density of the point. Experiments show that this algorithm can effectively eliminate the influence of the noises when the K-means algorithm searches cluster centers in the samples. Then the subtractive clustering algorithm is used to initialize the clustering centers of the K-means algorithm, meanwhile the number of cluster centers is gotten. The improved K-means algorithm is taken to optimize the structure of RBF neural network, and the results of experiments on the typical function approximation show that the proposed algorithm has the better approximation ability.

Key words: K-means algorithm; Density; Subtractive clustering algorithm; RBF neural network

1 INTRODUCTION

RBF (Radial Basis Function) neural network have been widely used in classification, signal processing and function approximation, because it has a simple topology, fast convergence and good robustness [1]. There are several parameters such as the number of hidden neurons, the center and the width of hidden neurons, and the weights needed to be determined in the design of RBF neural network. The core issue to design a RBF neural network is to ensure minimum number of hidden layer nodes, meanwhile the accuracy of the neural network should meet the requirement. Then the RBF neural network will have a better approximation ability.

In recent years, the K-means algorithm is widely used to design the RBF neural network, but the traditional K-means algorithm has some disadvantages[2]: (1) The number of clusters K needs to be given in advance; (2) It depend on the initial value, and often fall into local minima; (3) The Algorithm needs to constantly adjust the sample classification, and calculate new clustering center, so it wastes a lot of time when there is a very large amount of data; (4) The K-means algorithm is sensitive to the noises and outliers; (5) The K-means algorithm is generally used in globular clusters, and so on. Now, aiming at the disadvantages of the traditional K-means clustering algorithm, the intelligent algorithm is used to improve K-means algorithm, in order to determine the parameters of the RBF neural network. Sung et al. proposed a PSO algorithm which was used to initialize the center of the K-means to construct the structure of RBF neural network [2], then used the least squares method to determine the output weights, and the simulation experiments showed that the algorithm had a better approximation ability. Ehsan et al. proposed a method that the algorithm to find the centers of RBF hidden layer is k-means and the algorithm to train the weights of output layer is adaptive variable step-size algorithm [3], and experiments showed that this method was both accurate and fast in comparison with other presented schemes.

This study is based on the shortcomings of the K-means algorithm, firstly eliminating the influence of the noises and outliers according to the density index, then the subtractive clustering algorithm is used to initialize the clustering centers of the K-means algorithm, finally the improved K-means algorithm is taken to optimize the structure of RBF neural network, and experiments is made to verify the

This work was supported by the National Science Foundation of China under Grants 61203099, 61533002 and 61225016, Beijing Nova Program under Grant Z131104000413007, China Postdoctoral Science Foundation under Grant 2014M550017, Ph.D. Program Foundation from Ministry of Chinese Education under Grant 20121103120020 and 20131103110016, Collaborative Innovation Program under Grant ZH14000177, and Beijing Municipal Education Commission Foundation under Grants km201410005001 and KZ201410005002.

validity of this algorithm.

2 RBF neural network

The typical RBF neural network has a three-layer forward structure: the input layer, the hidden layer and the output layer [4]. The structure of RBF neural network is $n-m-o$, the number of input neurons is n, the number of hidden neurons is m and the number of output neurons is o. So the output of the k output neuron is defined as

$$y_k = \sum_{i=1}^{m} w_i \Phi_i(\|x - c_i\|) \qquad (1)$$

Where $x_1=(x_1, x_2 \ldots x_n)^T \in R^T$ is the input, $W \in R^{O\times M}$ is the weights, $y=[y_1, y_2 \ldots y_o]^T$ is the output and $\Phi(j)$ is the activation function of the hidden neuron [5] [6]:

$$\Phi_i(\|x - c_i\|) = e^{-\|x-c_i\|^2/\delta_i^2} \qquad (2)$$

Where c_i represents the center, and δ_i represents the width of the hidden neuron.

3 Improved K-means algorithm optimize the RBF neural network

3.1 Removal of noise points based on density

This paper proposes an algorithm that determines whether a point is a noise or outlier according to the size of its density. For the entire sample data, considering p data points in q dimensional space, expressed as $x_1, x_2 \ldots x_p$, where $x_i = x_{i,1}, x_{i,2} \ldots x_{i,q}$ (i=1, 2…p), the density function [7] for x_i is defined as

$$D_i = \sum_{j=1}^{p} e^{\left[-\|X_i - X_j\|^2 / \left\|\frac{\alpha}{2}\right\|^2\right]} = \sum_{j=1}^{p} e^{\left[-\sum_{k=1}^{q}(x_{i,k}-x_{j,k})^2 \Big/ \left(\sum_{k=1}^{q}\left(\frac{\alpha_k}{2}\right)^2\right)^2\right]} \qquad (3)$$

Where α_i (i=1, 2…q) is the cluster radius in k dimensional, and the greater the value of α_i, the greater the impact range of this point, and that leads to less classification. The smaller the value of α_i, the more the classification. The neighborhood of the point data is defined as $\alpha=[\alpha_1, \alpha_2 \ldots \alpha_q]$. Then finding out the maximum density value D_{max} from the density values of all sample points according to formula (3), and determining whether the point is a noise or outlier according to the following judgment formula

$$D_i < \psi * D_{\max} \qquad (4)$$

Where Ψ is the density coefficient.

If the density value of one point is satisfied with formula (4), the point is the noise or the outlier. In this way all the noise and outlier points can be removed.

3.2 The subtractive clustering algorithm

The subtractive clustering algorithm takes each point as a potential cluster center, and determine the cluster center according to the density of the sample data. This algorithm can effectively reflect the distribution of the data. Its basic ideas are as follows [8]:

Firstly, the density of each sample is calculated accordance to the formula (3). And selecting the data points with the maximum density as the first cluster center, as X_c^1, $D_c^1$is its corresponding density value. While continuing to obtain other cluster centers, in order to avoid the distance between cluster centers is too close, a formula is used to correct the density of each data point.

$$D_i^m = D_i^{m-1} - D_i^{m-1} e^{-\left[\left\|X_i - X_c^{m-1}\right\| \Big/ \left\|\frac{\beta}{2}\right\|^2\right]} \qquad (5)$$

Where D_i^m is the density value of the m point, β is a significantly reduced neighborhood of density value, and if$\beta$is too small, a cluster center may be too close to another cluster center, so$\beta$is generally greater thanα.

After correcting the density of each data point, the new cluster centers are selected. Continuing to repeat this process until the density of new clustering center satisfies the following formula:

$$\frac{D_c^m}{D_c^1} < \delta \qquad (6)$$

Where δ is the empirical value.

Then the cluster number of the sample data can be determined by the subtractive clustering process.

3.3 K-means algorithm

K-means algorithm divides the data set into different categories in accordance with the principle of minimum distance through the iteration process. It is a kind of unsupervised learning method. Because of its reliable theory, simple algorithm and fast convergence speed, it can

effectively deal with large data sets and is used widely [9] [10].The basic steps are as follows:

Step 1: Randomly selected k samples as an initial cluster centers from the total samples.

Step 2: Calculate the Euclidean distance between samples and the clustering centers, then allocate it to the neighboring clustering in accordance with the principle of minimum distance.

Step 3: Use the mean of the each cluster as a new clustering center.

Step 4: Repeat the step 2 and 3, until the clustering center no longer change.

Step 5: End, and obtain k clustering centers.

Because of the initial clustering centers are randomly generated, this algorithm is easy to fall into local minimum, and different values of k will lead to different clustering effect, also the noise and outlier points have a great effect to clustering centers. So K-means algorithm which is based on intelligent algorithm is widely used.

3.4 The algorithm of RBF based on the improved K-means

Firstly, the density value of each point is obtained in this method, and a certain point is considered as a noise or outlier point if its density value is less than the specified threshold. Secondly, the clustering center obtained by the method of subtraction clustering algorithm are as the initial value of K-means algorithm, certainly the numbers of the center are as K values. Finally the clustering centers of the modified K-means algorithm are as the hidden nodes of RBF neural network, and calculate extension constant according to the minimum distance between the clustering centers. Then use the gradient descent algorithm to determine the weights of RBF neural network [11] [12]. Detailed steps are as follows:

Step 1: Calculate the density values of the sample data according to the formula (3), and find out the point with the maximum density value.

Step 2: The noise and outlier points are removed according to the formula (4).

Step 3: The point of the largest density value is as the first clustering centerC_1. Calculate the density value of the rest of the data according to the formula (5), then find out the second highest clustering center and so on. Stop it until meeting the formula (6), and the final clustering centers are as the initial value of K-means algorithm.

Step 4: Calculate the distance between all the input samples and the clustering center, and classify them according to the principle of the minimum distance.

Step 5: Calculate and get new cluster centers.

Step 6: Repeat steps 4, 5, until the clustering center no longer change.

Step 7: The expansion constant of the hidden layer nodes $\sigma_j=\varepsilon * d_j$ Where d_j is the minimum distance between the j data center and other clustering centers, ε is called as overlap coefficient ε= 1.

Step 8: Train the RBF and get the output weight vector.

4 SIMULATIONS

4.1 Removal of the noise points

The entire samples as [0,100], are divided into four categories, and join some noise data of [0,100], such as the following table:

Table1. The sample data

Kind	Range	Number
first	$x\in[0,7]$ $y\in[0,7]$	20
second	$x\in[10,20]$ $y\in[30,40]$	20
third	$x\in[35,45]$ $y\in[25,35]$	20
fourth	$x\in[55,65]$ $y\in[55,65]$	20
noise	$x\in[0,100]$ $y\in[0,100]$	5

According to the experience ψ=0.15, α=6

The simulation effect is as follows:

Note: the blue points are sample; the green point are noise; setting red circle is to remove the noise points.

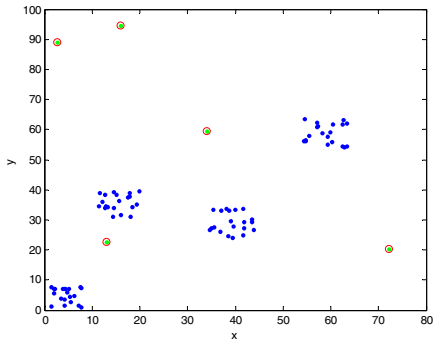

Fig 1. The noise points are all removed

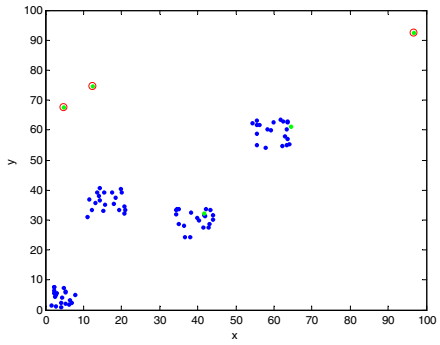

Fig 2. The noise points are not all removed

In Fig. 1, the five points are not belong to the four kinds of cluster, so treat them as noise and remove all the five points. But in Fig. 2, there are two points are belong to the four kinds of cluster, so they will not be removed, meanwhile the other three points should be removed. We can know that it is effect to remove the noise and outlier point based on the density.

4.2 Nonlinear Hermit function

This experiment is designed based on the proposed algorithm to approximate the nonlinear Hermit polynomial function.

$$y = 1.1*(1-x+2x^2)e^{-x^2/2} \qquad (7)$$

Where the training samples $U \in [-4, 4]$, the size of training samples is 100. The test samples are [-4: 0.08: 4], the size of test samples is 101. Also, join the 10 random data as noise points from [0, 20]. Then δ=0.2, Cluster radius α=1, β=1.5*α. The simulation results are as follows:

1. The density coefficient Ψ=0

Note: the red line is expected output; the blue line is actual output.

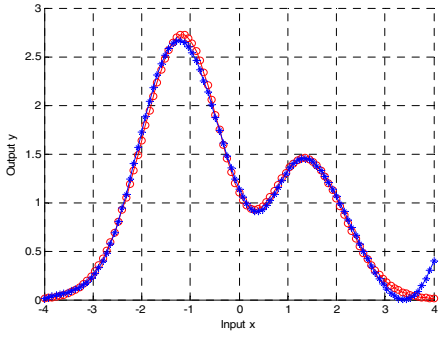

Fig 3. The test output

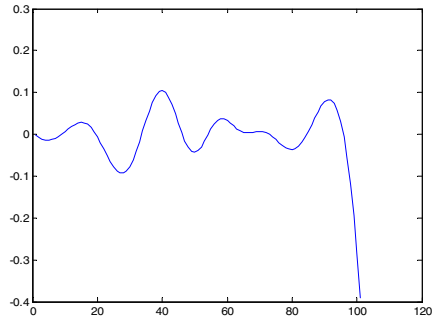

Fig 4. The test error

2. The density coefficient Ψ=0.5

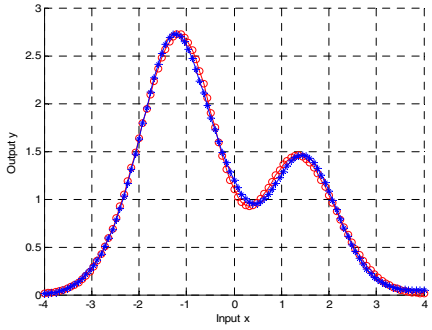

Fig 5. The test output

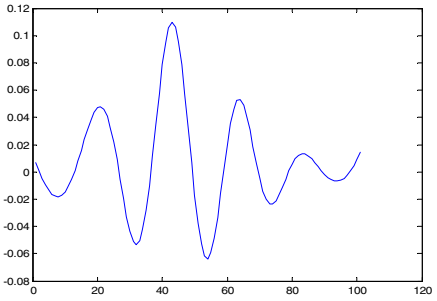

Fig 6. The test error

Table2. The removal of noise points

ψ	Hidden neurons	Noise points	MSE(test)
0	7	0	0.0047
0.5	7	10	0.0014

The test results are shown in Figure.3 and figure.5, and we know that if the noise points are removed the test result will be better. Also in Figure.4 and figure.6, the test error will be better. Based on Table 2, when removing the 10 noise points, the MSE is smaller. That is because the noise points influence the process of K-means clustering. Maybe, the noise points make some cluster centers offset, and destroy

the structure of RBF neural network [13] [14]. In figure.4, at the last of test the test error is about 0.4, but, in figure.6, there is normal, so, it is necessary to remove the noise and outlier point to design a RBF Neural Network.

Table3. The comparison of different algorithms

Algorithms	Hidden neurons	Time(s)	MSE(test)
Proposed algorithm	7	0.2184	0.0014
RBF	10	9.46	0.0028
RAN	14	1.12	0.0073
BIC-OLS	3	*	0.1793
OLS	6	0.80	0.5050

In Table.3, the proposed algorithm has the smallest test MSE and training time. By comparing with BIC-OLS and OLS [5], the proposed algorithm has more hidden neurons, but from the Table.3, we can see the proposed algorithm has smaller test MSE and training time, so design of RBF neural network is not to find the smallest structure but the appropriate[15].

5 CONCLUSION

Aiming at the disadvantages of the traditional K-means clustering algorithm, this paper proposes a new algorithm based on density to remove the noises and outliers, and the validity of this method is proved by the data classification experiment.

Then using the subtractive clustering algorithm to initialize the K-means algorithm, and the final clustering centers can be obtained. In this way, the clustering centers are as the hidden neurons of RBF neural network. And the weights of RBF neural network are trained by the gradient descent algorithm. From the experimental results, we can see that the design of RBF neural network based on the proposed algorithm has a satisfactory generalization ability.

REFERENCES

[1] Y. P. Song, X. Q. Peng. New structure adapting neural network and its training method [J]. Control and Decision, 2010, 25(8): 1265-1268.

[2] Sung-Kwun Oh, Wook-DongKim. Design of K-means clustering-based polynomial radial basis function neural networks realized with the aid of particle swarm optimization and differential evolution [J]. Neurocomputing, 2012, 78(1): 121–132.

[3] Ehsan Mostafapour, Mehdi Panahi, and Morteza Farsadi. A hybrid “k-means, VSS LMS” learning method for RBF network in short-term load forecasting [J]. 2015 9th International Conference on Electrical and Electronics Engineering (ELECO), 2015.

[4] H. G. Han, L. D. Wang, and J. F. Qiao. A Spiking-based Mechanism for Self-organizing RBF Neural Networks [C]. International Joint Conference on Neural Networks (IJCNN), Beijing, China, 2014.

[5] P. Zhou, D. H. Li, H. Wu, F. Cheng. The automatic model selection and variable kernel width for RBF neural networks [J]. Neurocomputing, October, 2011, 74(17): 3628–3637.

[6] T. T. Xie, Y. Hao, H. Joel, “Fast and efficient second-order method for training radial basis function networks,” [J]. IEEE Transactions on Neural Networks and Learning Systems, 2012, 23(4): 609–619.

[7] T. H. Pan, Z. K. Xue, S. Y. Li. An online multi-model identification algorithm based on subtractive clustering [J]. Acta Automatica Sinica, 2009, 35(2): 220-224.

[8] Z. Z, Zhang, J. F. Qiao. Design RBF neural network architecture based on online subtractive clustering [J]. Control and Decision, 2012, 27(7): 997-1002.)

[9] Kanungo, Tapas. An Efficient k-Means Clustering Algorithm: Analysis and Implementation [J]. Pattern Analysis and Machine Intelligence, IEEE Transactions on. 2002, 24(7): 881 – 892.

[10] K. Krishna, M. Narasimha Murt. Genetic K-Means Algorithm [J]. Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on. 1999, 29(3): 433 – 439.

[11] G.E. Tsekouras, J. Tsimikas, On training RBF neural networks using input–output fuzzy Clustering and particle swarm optimization, *Fuzzy Sets System*, 221: 65–89, 2013.

[12] J. J. Yao, J. Yang, L. M. Wang, A HAMPSO-RBF algorithm applied to target localization, AASRI Conference on Computational Intelligence and Bioinformatics, 2012: (183–188).

[13] S. Ferrari, F. Bellocchio, V. Piuri, and N.A. Borghese. A hierarchical RBF online learning algorithm for real-time 3-D scanner. IEEE Trans. on Neural Networks, 21(2): 275–285, 2010.

[14] D. Simon. Training Radial Basis neural networks with the

extended kalman filter, Neurocomputing, 48(4): 455–475, 2002.

[15] M. Chun-tao, L. Xiao-xia and Z. Li-yong, Radial basis function neural network based on ant colony optimization, in: Proceedings of the International Conference on Computational Intelligence and Security Workshops, 2007.