A Novel Radial Basis Function Neural Network For Approximation

Shenmin Song[1], Zhigang Yu[2], Xinglin Chen[1]
School of Aerospace, Harbin Institute of Technology

[1]｛songshenmin，chenxl｝@hit.edu.cn
[2]yzglgy@163.com

Abstract

Two difficulties are involved with traditional RBF networks: the initial configuration of an RBF network needs to be determined by a trial-and-error method, and the performance suffers degradation when the desired locations of the center of the RBF are not suitable. A novel RBF network is proposed to overcome these difficulties. A new radial basis function is used for hidden nodes, and the number of nodes is determined automatically by Shannon sampling theorem. The corresponding learning algorithm generally takes far less time for approximation with an optimized parameter setting. The locations of the centers of RBF are fixed. Experimental results have shown that the RBF networks constructed by our method have a smaller number of nodes, a faster learning speed, and a smaller approximation error than the networks produced by other methods.

Keywords: neural networks, radial basis function approximation

1 Introduction

Radial basis function (RBF) networks are widely used for approximating functions from given input–output patterns. Generally, the traditional radial basis function neural network may sometimes be considered prohibitively expensive to implement in computational terms for large amounts of training data. The performance of a trained RBF network depends on the number and locations of the radial basis functions, their shape and the method used for learning the input-output mapping. The most popular existing learning strategies [1-3] for RBF neural networks can be classified as follows: (i) strategies selecting the RBF centers randomly from the training data, (ii) strategies employing unsupervised procedures for selecting the RBF centers, and (iii) strategies employing supervised procedures for selecting the RBF centers. Some approaches are presented, which include orthogonal least squares [4]-[7], resource allocating in[8,9], genetic algorithm in[10,11], gradient decent [12]. These approaches involve searching for a sub-optimal network in a lower dimensional space. However, one common feature of the above methods is they often lead to over-fitting, with a negative effect on the performance of trained RBF neural networks [13]. The biggest obstacle is iterative strategy that often cannot reach an optimal result. To overcome these computational difficulties, the complexity of the network would have to be reduced, which requires an approximation to the regularized network [14]. The design of a supervised neural network may be pursued in a variety of ways.

In this paper, we take a completely different approach by viewing the design of a neural network as an approximation problem in $L^2(R)$ space. The Shannon radial basis functions (SRBF) neural networks are introduced in the solution of the interpolation problem. The simple but effective technique for approximating a continuous function with an SRBF neural network has been presented. We have derived a bound on the generalization error produced by SRBF network, expressed in terms of the size of the hidden layer and the size of the training sample. One example is presented that illustrates the network is concise, effective and accurate. The method speeds up learning, significantly reducing network training and evaluation time. This feature is of significance for many contemporary applications, and has excellent performance on learning convergence. The novel method for efficient construction of SRBF networks can reach the same level of accuracy as conventional radial basis function networks.

2 Interpolation problems

The interpolation problem, in its strict sense, may be stated as follows:
Given a set of N different points $\{x_i \in R^{m_0} \mid i = 1, \cdots, N\}$ and a corresponding set of N real numbers $\{d_i \in R \mid i = 1, \cdots, N\}$, find a function: $R^{m_0} \to R$ that satisfies the interpolation condition:

$$f(x_i) = d_i \qquad i = 1, \cdots, N . \tag{1}$$

The interpolation function is constrained to pass through all the training data point. The RBF technique consists of choosing a function $f(x)$ that has the following form [15]:

$$f(x) = \sum_{i=1}^{N} w_i \phi_{ji}\left(\left\| x_j - x_i \right\|\right) . \tag{2}$$

where $\left\{\varphi_{ji}\left(\left\| x_j - x_i \right\|\right) \mid i = 1, \cdots, N\right\}$ is a set of N arbitrary (generally nonlinear) functions, known as radial basis functions, and $\|\square\|$ denotes a norm that is usually Euclidean. The known data points $x_i \in R^{m_0}$, $i = 1, \cdots, N$ are taken to be the centers of radial basis functions. Inserting the interpolation conditions of Eq.(1) in (2). We obtain the following equation for the unknown weights of the expansion $\{w_i\}$:

$$\Phi \mathbf{w} = \mathbf{d} \tag{3}$$

where $\Phi = \left\{\varphi_{ji}\left(\left\| x_j - x_i \right\|\right) \mid i = 1, \cdots, N, j = 1, \cdots, N\right\}$, $\mathbf{d} = [d_1, d_2, \cdots, d_N]$, $\mathbf{w} = [w_1, w_2, \cdots, w_N]$.
Assuming that Φ is nonsingular and therefore that the inverse matrix Φ^{-1} exists, we may go on to solve Eq. (3) for the weight vector w as shown by $\mathbf{w} = \Phi^{-1}\mathbf{d}$ .In order to be sure that the interpolation matrix Φ is nonsingular, the following Micchelli's theorem [16] is given:

Let $\{x_i\}_{i=1}^{N}$ be a set of distinct points in R^m. Then the N-by-N interpolation matrix Φ , whose *ji*-th element is $\varphi_{ji} = \varphi\left(\left\| x_j - x_i \right\|\right)$, is nonsingular.

There is a large class of radial basis functions that is covered by the theorem, it concludes that the function $\varphi(x)=\sin(x)/x$ is of particular interest in the study of RBF networks. What is even more remarkable is that a radial basis function $\varphi(\mathrm{x})$ can be used to approximate input-output mapping with greater accuracy and with good generalization to new data. So we have $\mathbf{w}=\Phi^{-1}\mathbf{d}=\mathbf{d}$, here $\Phi=I$.

3 SRBF neural networks

This section introduces a new family of reformulated RBF neural networks constructed by Shannon radial basis functions, which are obtained from Shannon sampling theorem. As we know, the unfit centers in conventional function approximation methods of RBF neural networks have a large effect on approximation error. From this viewpoint, the paper presents the following SRBF neural networks with some fixed location of its centers, avoiding to the error for function approximation bringing from the unfit centers.

3.1 Shannon Radial Basis Functions

In $L^2(R)$: $\left\{f_a(t),\int|f_a(t)|^2dt<+\infty\right\}$, let $F(\omega)$ denote the Fourier transform of the function $f_a(t)$, $\forall\ f_a(t)\in L^2(R)$, $\exists\ B>0$ to make $F(\omega)=0,|\omega|\le B$, so that $f_a(t)$ has compact support, namely, $f_a(t)$ is B frequency truncation. When sampling period $T\le\pi/B$, $g(t)=\sin(\pi t/T)/(\pi t/T)$ can be regarded as an ideal low pass filter, the waveform of function $g(t)$ is shown in Fig 1:

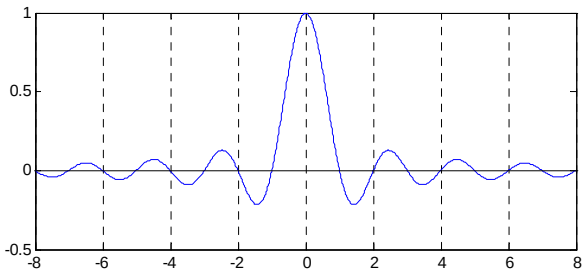

Fig. 1. The waveform of function g(t)

Input of $g(t)$ is $\hat{f}_a(t)=\sum_{n=-\infty}^{\infty}f_a(t)\delta(t-nT)$, output is $y_a(t)$, so that

$$y_a(t)=\hat{f}_a(t)*g(t)=\int_{-\infty}^{\infty}\hat{f}_a(t)g(t-\tau)d\tau=\sum_{n=-\infty}^{\infty}f_a(nT)g(t-nT)$$
$$=\sum_{n=-\infty}^{\infty}f_a(nT)\frac{\sin(\pi(t-nT)/T)}{\pi(t-nT)/T}$$

If it satisfies Shannon sampling theorem, $y_a(t)=f_a(t)$, then we have

$$f_a(t)=\sum_{n=-\infty}^{\infty}f_a(t)\frac{\sin(\pi(t-nT)/T)}{\pi(t-nT)/T}. \tag{4}$$

Under the condition that approximation rate and accuracy can meet our desired expectation, for arbitrary B, supposing $\sup F(\omega) \in [-B, B]$. Consequently, SRBF network output in Eq.(1) is exact at the sampling points, so that

$$\hat{f}_a(t) = \sum_{k=1}^{N} f_a(kT) \frac{\sin(\pi(t-kT)/T)}{\pi(t-kT)/T}$$

Let $w_k = f(kT)$, $\phi_k(x) = \frac{\sin(\pi(t-kT)/T)}{\pi(t-kT)/T}$, We may simplify the Eq.(2) as follows:

$$\hat{f}_a(t) = \sum_{k=1}^{N} w_k \phi_k(t) \tag{5}$$

The expansion of the approximation function $\hat{f}_a(t)$ given in Eq.(5) in terms of the function $\phi_k(t)$ suggests that the network structure as a method for its implementation is a kind of three-layer forward structure in which hidden layer is constituted by the Shannon radial basis function.

The input layer broadcasts the coordinates of the input vector to each of the units in the hidden layer. Each unit in the hidden layer then produces an activation based on the associated SRBF. Finally, each unit in the output layer computes a linear combination of the activations of the hidden units. How a SRBF network reacts to a given input stimulus is completely determined by the activation functions associated with the hidden units and the weights associated with the links between the hidden layer and the output layer.

3.2 Bounds on approximation Error of SRBF neural networks

There are two main aspects as follow that lead to error: the first one is $T > \pi/B$, the second is that $f_a(t)$ has compact support in frequency domain，but does certainly not have compact support in the time domain. In practice, this unavoidable error is caused by truncation of SRBF $\phi_k(t)$. Generalization error is defined by

$$e(t) = \hat{f}_a(t) - f_a(t)$$

Generalization error for SRBF networks is a tradeoff between approximation and estimation errors. The mapping realized by SRBF depends on the sampling $\{f(nT) \mid n = 0,1,\cdots,N\}$ and the sampling time T. Then, the generalization error produced by the network is bounded by

$$\begin{aligned} |e(t)| &= \left| \sum_{n=-\infty}^{-1} f_a(nT) \frac{\sin(\pi(t-nT)/T)}{\pi(t-nT)/T} + \sum_{m=N+1}^{\infty} f_a(nT) \frac{\sin(\pi(t-mT)/T)}{\pi(t-mT)/T} \right| \\ &< K \left(\left| \sum_{l}^{-1} f_a(nT) \frac{\sin(\pi(t-nT)/T)}{\pi(t-nT)/T} \right| + \left| \sum_{n=N+1}^{m} f_a(nT) \frac{\sin(\pi(t-nT)/T)}{\pi(t-nT)/T} \right| \right) \end{aligned} \tag{6}$$

where, $K = \sup\{f_a(nT) \mid n = 0,1,\cdots,N\}$, $l \in Z$, $l < -1$; $N+1 < m$, N is the number of samples. From the bound of Eq.(6), we may make the following deductions: the error $e(t)$ will be reduced with increasing numbers of samples. In other words, to reduce the error, the number of samples should be increased. In addition, the errors

at the boundary are larger than in middle. The generalization error converges to zero only if the number of hidden units verge on infinity. For a given size, N, of training samples, the optimum number of hidden units is the minimum value yielding the error desired.

3.3 Generalized SRBF neural networks

When the samples are acquired by nonuniform sampling, coefficient w_k of the network can be acquired by the learning strategy of traditional RBF neural networks. Assume that the training sample $\{(x_i, d_i)\}_{i=1}^{N}$ is obtained by randomly sampling from $f(x)$. Eq. (5) is a linear equation, so coefficient w_k can be attained by solving linear equations. When the dimensionality of the input space is low, the simple LMS algorithm is good enough to fulfill the task of optimization. But the key issues are how we can construct these approximating networks and how many hidden-layer units and their centers are chosen for approximating specific functions within some specified error. To solve this problem, this paper presents the following method to determine the optimal number and the locations of centers. This strategy prevents the training set from participating in the formation of the radial basis function centers.

For an unknown function to be approximated, the number of centers of the radial basis functions is determined according to an unsupervised procedure relying on frequency domain of approximated function. Let $f(x): S \to R$ denote the function to be approximated, where $S \subset R^n$ is a closed bounded region. More specifically, if the real number a and b denote the limits of x, respectively, then $S = \{x \in R \mid a \le x \le b\}$.

It is assumed that values of $f(x)$ are available at discrete points within S. Let r denote the number of centers. So that the SRBF network includes r terms with the k th term consisting of SRBF $\phi_k(x)$ centered about point $x^k = k(b-a)/r$ and scaled by weight w_k. The constraint condition $r \ge (b-a)B/\pi$ can guarantee satisfaction with the sampling theorem. Then the network output can be expressed formally as $f(x) = \sum_{k=1}^{r} w_k \phi_k(x)$.

The algorithm for computing the coefficient w_k is as follows:

1. Training sample: Input signal vector = $x(k)$, Desired response = $d(k)$
2. Selected parameter η
3. Initialization: Set $w(k) = d(i)$, $i = \arg\min_i \left| x^k - x(i) \right|$
4. Computation. For $n = 1, 2, \cdots, N$, to compute

$$e(k) = d(k) - w^T(k)x(k)$$
$$w(k+1) = w(k) + \eta x(k) e(k)$$

The initialization method of the algorithm can accelerate convergence significantly. These parameters were kept fixed during the learning process to avoid substantial fluctuations of the error. The training of all neural-network models utilized in experimental study was terminated according to the stopping criterion described below: each adaptation cycle was followed by the calculation of the average error on

the testing set over the previous. Parameters and weights are initialized appropriately, and then tuned and adjusted via the presented method to improve the performance of the network.

4 Computer experiment

An approximated function is

$$y = 3\sin(4\pi x) + 2\cos(6\pi x) + 4 \qquad x \in (0,\ 3.0)$$

The sampling points acquired by nonuniformly sampling are $\{x_i \in R \mid i = 1,\cdots,128\}$ for the training set, and $\{\tilde{x}_i \in R \mid i = 1,\cdots,128\}$ for the test set. There are 128 patterns in the training and the test sets respectively. The training set is plused by $0.2\mathrm{N}(0,1)$ normal noise.

In this case, frequency truncation $B = 6\pi = 18.84$, then the sampling period $T = \pi/B = 0.167$, so that the number of initial centers $r = 16$ by $r \geq (b-a)B/\pi$, the l th initial location centers are $x^l = 3k/16$, $k = 1,2,\cdots,16$. The initial structure of network is 1-16-1. The approximation accuracy of RBF networks is evaluated by the formula:

$$\varepsilon = \sqrt{\frac{1}{2N}\sum_{x=1}^{N}\left|e(x)\right|^2}$$

where N is the member of samples, $N = 128$, and $e(x) = y - y^*$, y is output of network and y^* is the desired output. Through the algorithm for computing the coefficient w_k , the error is $\varepsilon = 0.248$ for training set and $\varepsilon = 0.271$ for the test set. In order to improve the approximation accuracy, the number of final centers selected is 32, the error is then $\varepsilon = 0.137$ for the training set and $\varepsilon = 0.113$ for the test set. The approximation result by 1-32-1 neural networks is shown in Fig. 2.

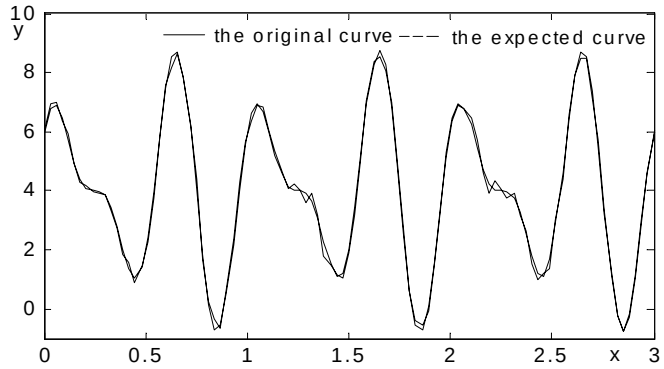

Fig. 2. The approximation result by 1-32-1 neural networks

To solve the relatively complex approximation problems, multilayer perceptrons are not as effective as the proposed SRBF networks. Compared with the results in [17], SRBF networks have better properties in terms of accuracy and training time. For the network proposed, the approximation accuracy for test data can be guaranteed by the Shannon sampling theorem and the least trainning time can be seen from the learning process that does not need to adjust the location of centers of RBF, as is necessary in [17].

5 Conclusions

In order to effectively solve the approximation problems using RBF networks, a technique for approximation has been presented. The corresponding learning algorithm is equivalent to finding a suitable sample that provides a best fit to the training data. Correspondingly, generalization is equivalent to the use of these samples to interpolate the test data. Such a viewpoint is the motivation behind the method of radial-basis functions that it draws upon research work on traditional strict interpolation in $L^2(R)$ space.

References

[1] Nicolaos B. Karayiannis, Mary M. Randolph-Gips. The Construction and Training of Reformulated Radial Basis Function Neural Networks. IEEE Transactions on Neural Networks, 4: 835-844, 2003.

[2] Chun-guo Wu, Yan-chun Liang, Center Selection for RBF Neural Network in Prediction of Nonlinear Time Series. Proceedings of the Second International Conference on Machine Learning and Cybernetics, 1355-1359, 2003.

[3] Hong-Ruiwang, Hong-Binwang, A New Algorithm of Selection the Radial Basis Function Networks Center. Proceedings of the First International Conference on Machine Learning and Cybernetics, Beijing, 1801-1804, 2002.

[4] J.B.Gomm,D.L, Yu, Selecting Radial Basis Function Network Centers with Recursive Orthogonal Least Squares Training, IEEE Transactions on Neural Networks, 2: 306-314, 2000.

[5] W.Kaminski, P.Strumollo, Kernel, Orthonormalization in Radial Basis Function Neural Networks. IEEE Transactions on Neural Nnetworks, 6:1177-1183, 1997.

[6] A.Sherstinsky, R,W. Picard, On the Efficiency of The Orthogonal Least Squares Training Method for Radial Basis Function Networks, IEEE Transactions on Neural Networks, 1: 195-200, 1996.

[7] S.Chen,C.F.N. Cowan,P.M.Grant, Orthogonal Least Squares Learning Algorithm for Radial Basis Function Networks. IEEE Transactions on Neural Networks, 2: 302-309, 1991.

[8] J.Platt, A Resource Allocating Network for Function Interpolation. Neural computation, 2: 213-215, 1991.

[9] L.Yingwei, N. Sundarajan, P.Saratchandran, Performance Evaluation of a Sequential Minimal Radial Basis Function(RBF) Network Learning Algorithm. IEEE Transactions on Neural Networks, 2: 308-318, 1998.

[10] B.A. Whitcheas, T.D. Choate, Evolving Space-Filling Curves to Distribute Radial Basis Functions Over an Input Space. IEEE Transactions on Neural Networks, 1: 15-23, 1994.

[11] B.A. Whitcheas, T.D. Choate, Cooperative Genetic Evolution of Radial Basis Function Centers and Widths for Time Series Prediction, IEEE Transactions on Neural Networks, 4: 869-880, 1996.

[12] E.S. Chan, S.Chen, B.Mulgrew, Gradient Radial Basis Function Networks for Nonlinear and Nonstationary Time Series Prediction. IEEE Transactions on Neural Networks, 1: 190-194, 1996.

[13] N.B. Karayiannis and G.W. Mi, Growing Radial Basis Neural Networks: Merging Supervised and Unsupervised Learning with Network Growth Techniques, IEEE Transactions on Neural Networks, 6: 1492-1506, 1997.

[14] T.Poggio,F.Girosi, Networks for Approximation and Learning. Proceedings of IEEE, 9: 1481-1495, 1990.

[15] Powell,M.D. Radial Basis Function Approximation to Polynomials. Numerical Analysis Proceedings, 223-241, 1987.
[16] Micchelli,C.A. Interpolation of Scattered Data: Distance Matrices and Conditionally Positive Definite Functions. Constructive Approximation, 1: 11-22, 1986.
[17] Gao Daqi, Ji Yan. Frequency Domain Analysis Based RBF Networks and Their Applications Function Approximations. IEEE International Joint Conference on Neural Networks Proceedings. 2111-2116, 2004.

Shenmin Song works as a associate professor at the Center for Control Theory and Guidance Technology at the School of Aerospace, Harbin Institute of Technology. He received his PhD degree in Control Theory and Engineering from the Institute of Technology. His main research interests include robust control for linear and non-linear uncertain systems, soft computation and control systems design.

Zhigang Yu is a PhD students in the Center for Control Theory and Guidance Technology at the School of Aerospace, Harbin Institute of Technology. He received his MSc. in Automatic Control from the Northeast Forest University in 2003. His current research interests are Soft Computing and Intelligent Control.

Xinglin Chen is a professor at the department of control science and engineering at the school of aerospace, Harbin Institute of technology and the Director of the Center for experiment of control science and engineering. His main research interests include control systems analysis and design, soft computation and robot automatic control.