Far East J. Math. Sci. (FJMS) **22**(1) (2006), 17-24

RBF NETWORKS BASED ON STRUCTURAL RISK FUNCTION

ZHENGWEI SHEN, HONG-LIANG ZHENG

and

DE-SHAN SUN

(Received August 18, 2005)

Submitted by K. K. Azad

Abstract

Support vector machines, based on statistical learning theory, are a kind of novel machine learning algorithm. Due to their outstanding learning performance, support vector machines have become a hot spot for research in the field of international machine learning, and have been developed for solving classification and regression problems. In this paper, we apply the structural risk function to radial basis function (RBF) networks, and discuss the relationship between support vector regression model and RBF networks. Simulation experiments reveal that this algorithm can improve the generalization ability of RBF networks.

1. Introduction

Neural networks have wide application in many areas [6]. BP networks and RBF networks are the most common network models. BP networks change weights by error propagation, and have relatively slow

---

2000 Mathematics Subject Classification: 68-XX.

Keywords and phrases: support vector machines, RBF networks, regression, structure risk function.

This work is supported by LEC (2004C068).

convergence rate and local minima. RBF networks are three-layer feed-forward neural networks. The action function of hidden layer units is a radial basis function. The key of radial basis function is center selection which can be obtained by assemble and learning methods.

In recent years, support vector machines learning algorithm [1, 2, 4, 5], introduced by Vapnik and others according to statistical learning theory, has drawn broad attention in international academic circle. An important character of the algorithm is that it uses the structural risk minimization principle instead of the traditional risk minimization principle to improve the generalization ability of learning machine. In addition, because support vector machines algorithm is a convex optimal control problem, local optimal solutions must be global optimal solutions. Support vector machines have been utilized to solve classification and regression problems. In this paper, the structural risk function is used in RBF networks so that the generalization ability of RBF networks can be improved; furthermore, the relationship between support vector regression model and RBF networks is discussed. This paper also gives an example of chaos time sequence forecast, forecasted by the RBF networks based on the structural risk function and the general RBF networks respectively and the result indicates that the generalization ability of the former is improved remarkably.

2. RBF Networks based on Support Vector Regression Algorithm

First we introduce RBF networks, and then apply the support vector regression algorithm to RBF networks learning.

2.1. RBF networks

The structure of RBF networks is similar to that of multilayer feed-forward networks. They are three-layer feed-forward networks. The input weights of hidden layer units are set to be 1. Only weights input from hidden layer units to output units can be transformed. The action function of hidden layer units is a radial basis function. Suppose $\{x_i, y_i\}_{i=1}^{M}$ is a training set, where M is the number of training data, $x_i \in R^m$ are input data and $y_i \in R$ are output data. Then RBF network structure can be expressed as

$$f(x) = \sum_{i=1}^{N} w_k g_k(x) + B,$$

where $g_k(x)$ is a radial basis function, $g_k(x) = \exp\left(-\frac{\| x - \mu_k \|^2}{\sigma_k}\right)$, μ_k is the center, σ_k is the standard deviation, N is the number of radial basis functions. Center selection of RBF networks is crucial. There are three selection methods:

(1) Select center by experience. We can select several centers according to the distribution of the samples.

(2) Select center by assemble method. We can first divide the sample into some units and then let the center of each unit be the centers of RBF networks.

(3) Select center by sample learning.

After selecting parameters of RBF networks the weight can be calculated using the method of least squares.

2.2. RBF networks based on structure risk function

Firstly we select all samples as a network center with the same width σ, and let

$$W = (w_1, w_2, \ldots, w_M)^T,$$

$$G(x) = (g_1(x), g_2(x), \ldots, g_M(x))^T.$$

Then RBF networks can be showed

$$f(x) = W^T G(x) + B. \tag{1}$$

Introduce structure risk into the networks, according to the structure risk minimization principle, then optimal problem is a minimal function

$$\frac{1}{2}\| W \|^2 + C \sum_{i=1}^{M} (\xi_i + \xi_i^*). \tag{2}$$

The conditions are

$$f(x_i) - y_i \le \xi_i^* + \varepsilon, \quad i = 1, \dots, M,$$

$$y_i - f(x_i) \le \xi_i + \varepsilon, \quad i = 1, \dots, M,$$

$$\xi_i, \xi_i^* \ge 0, \quad i = 1, \dots, M.$$

The first term of the formula (2) makes the function more smoother and improves the generalization ability. The second term reduces the errors while the constant C compromises them, ε is a positive constant. When the difference between $f(x_i)$ and y_i is smaller than it, then the error is neglected. On the contrary, the error is $|f(x_i) - y_i| - \varepsilon$. This is a convex quadratic optimal problem, take Lagrange function

$$\begin{aligned} &L(W, B, \xi, \xi^*, \alpha, \alpha^*, \gamma, \gamma^*) \\ &= \frac{1}{2}\|W\|^2 + C\sum_{i=1}^{M}(\xi_i + \xi_i^*) - \sum_{i=1}^{M}\alpha_i[\xi_i + \varepsilon - y_i + f(x_i)] \\ &\quad - \sum_{i=1}^{M}\alpha_i^*[\xi_i^* + \varepsilon + y_i - f(x_i)] - \sum_{i=1}^{M}(\xi_i\gamma_i + \xi_i^*\gamma_i^*), \end{aligned} \tag{3}$$

where $\alpha_i, \alpha_i^* \ge 0$, $\gamma_i, \gamma_i^* \ge 0$, $i = 1, \dots, M$. The minimax of function L satisfies the conditions:

$$\frac{\partial}{\partial W}L = 0, \quad \frac{\partial}{\partial B}L = 0, \quad \frac{\partial}{\partial \xi_i}L = 0, \quad \frac{\partial}{\partial \xi_i^*}L = 0.$$

Then we get

$$\sum_{i=1}^{M}(\alpha_i - \alpha_i^*) = 0, \tag{4}$$

$$W = \sum_{i=1}^{M}(\alpha_i - \alpha_i^*)G(x_i), \tag{5}$$

$$C - \alpha_i - \gamma_i = 0, \quad i = 1, \dots, M, \tag{6}$$

$$C - \alpha_i^* - \gamma_i^* = 0, \quad i = 1, \dots, M. \tag{7}$$

In terms of conditions (4), (5), (6) and (7), we can obtain the dual form of the optimal problem from (3), the maximal function

$$W(\alpha, \alpha_i^*) = -\frac{1}{2}\sum_{j=1}^{M}(\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*)\langle G(x_i), G(x_j)\rangle$$

$$+\sum_{i=1}^{M}(\alpha_i - \alpha_i^*)y_i - \sum_{i=1}^{M}(\alpha_i + \alpha_i^*)\varepsilon.$$

The constraints are

$$\sum_{i=1}^{M}(\alpha_i - \alpha_i^*) = 0, \quad 0 \le \alpha_i, \alpha_i^* \le C, \quad i = 1, \dots, M,$$

where $\langle\, , \rangle$ shows the vector inner product. This is also a quadratic optimal problem, W can be obtained by (5), B can be obtained by the next formula.

$$B = y_j - \varepsilon - \sum_{i=1}^{M}(\alpha_i - \alpha_i^*)\langle G(x_i), G(x_j)\rangle, \quad \alpha_j \in (0, C).$$

$$B = y_j + \varepsilon - \sum_{i=1}^{M}(\alpha_i - \alpha_i^*)\langle G(x_i), G(x_j)\rangle, \quad \alpha_j^* \in (0, C).$$

The final solution of the network is

$$f(x) = \sum_{i=1}^{M}(\alpha_i - \alpha_i^*)\langle G(x_i), G(x)\rangle + B.$$

According to the character of support vector regression algorithm, in generally, most of $\alpha_i - \alpha_i^* = 0$. Those samples with nonzero are called *support vectors*, which are regarded as the center of radial basis function. Hence, we can determine the center of radial basis function by introducing the structural risk function.

3. The Relationship between Support Vector Regression Model and RBF Networks

Applying the structural risk function into RBF networks learning, we obtain the center of the radial basis function and the network weight directly. In fact, the ideas of constructing both support vector regression algorithm and RBF networks are identical, so that the structural risk function can be applied to RBF networks learning. They both map the original problems to a high-dimensional characteristic space, then conduct linear algorithm on the space. The key of RBF networks is the center determination of the radial basis function, which is obtained by assemble method or learning method in order to simplify the networks. The key of support vector regression is to obtain support vectors. Then we describe the model by the kernel expansion of support vector in order to simplify the model.

The algorithm in this article does not involve kernel function, which is extremely important in support vector regression and in avoiding dimension problems. If we find an appropriate kernel function $K(x, y)$, which satisfies $K(x, y) = \langle G(x), G(y) \rangle$, then the RBF network above is a standard support vector regression model when replacing $\langle G(x), G(y) \rangle$ by $K(x, y)$. Here the selection of kernel function is a key point. It has been proved that there exists a kernel function. At present, there are no explicit guiding methods concerning the selection problem of kernel function. We always select the kernel function from those satisfying the Mercer condition. RBF networks have a close relationship with support vector regression model, and kernel function is the linking bridge.

Actually, the linkage between support vector machine and RBF networks is of multi-aspect. Regarding classification problems, document [3] shows a hybrid algorithm. This algorithm first uses support vector machine classification algorithm to work out the support vector, which is considered as the network center, then trains the RBF networks independently to obtain the network weights. The algorithm also has a good effect.

4. Example

Let us forecast Lorenz chaos time sequence using the method introduced in this paper, and then compare the result with that of the general RBF networks forecast. Lorenz time sequence is generated from this differential equation

$$\frac{dx}{dt} = 10(y - x), \quad \frac{dy}{dt} = (28 - z)x - y, \quad \frac{dz}{dt} = xy - \frac{8}{3}z.$$

Let the initial value be [0.0031, 0.1928, 0.4208]. Take 0.1 as the length of step of t. Collect a one-dimensional time sequence. Take 300 data starting from 100th data. The first 280 data form the training samples and the rest 20 data form testing samples. For forecasting the sequence, we need to reconstruct the phase space and take 3 as the embedding dimension. In addition, in order to improve quality of the training samples, we select samples which are close to the testing samples as the training set for every forecasting, and the select standard is those samples whose Euclidean distance is less than $\delta(> 0)$. Here we set $\delta = 10$. Meanwhile let every training sample be the center of the radial basis function, with same standard deviation $\sigma = 10.5$, and parameters $C = 100000$, $\varepsilon = 0.001$.

Take the mean square error as the testing index:

$$(MSE) = \sqrt{\frac{1}{K}\sum_{i=1}^{K}(X_i - \hat{X}_i)^2},$$

where X_i is the real value, $\hat{X}_i$ is the forecasted value, K is the number of the testing samples. The forecasted result calculated by this algorithm is $MSE = 0.0913$. Using the same data and parameters, regarding all the testing samples as centers, and working out the weights of RBF networks by the method of least squares, we can also get the forecasted result, which is $MSE = 0.6893$. Additionally, if we forecast the same data by the RBF networks tool box of Matlab 6.1, the result will be $MSE = 1.6059$.

This example implies that the generalization ability of the RBF networks based on the structural risk function exceeds that of the traditional RBF networks greatly.

5. Conclusions

Since the global optimal solutions can be found using the support vector machine algorithm, the algorithm has many advantages in other problems and has successful applications in many fields. This paper applies the structural risk function to the RBF networks. From the example of the chaos time sequence forecasting, we see that the method improves the generalization ability of RBF networks. Furthermore, this paper shows the relationship between support vector regression model and RBF networks, and indicates that the ideas of constructing both of them are identical.

References

[1] C. J. C. Burges, A tutorial on support vector machines for pattern recognition, Data Mining and Knowledge Discovery 2(2) (1998), 121-167.

[2] C. Cortes and V. Vapnic, Support-vector networks, Machine Learning 20(3) (1995), 273-297.

[3] B. Scholkopf et al., Comparing support vector machines with Gaussian kernels to radial basis function classifiers, IEEE Trans. Signal Process. 45(11) (1997), 2758-2765.

[4] A. J. Smola and B. Scholkopf, A tutorial on support vector regression, Neuro COLT TR NC-TR-98-030, Royal Holloway College University of London, UK, 1998.

[5] V. N. Vapnik, Statistical Learning Theory, Wiley, New York, 1998.

[6] N. Y. Zhang and P. F. Yan, Neural Networks and Fuzzy Control, Tsinghua University Press, Beijing, 1998.

Zhengwei Shen and De-Shan Sun
Mathematics School
Liaoning Normal University
Dalian 116029, P. R. China
e-mail: szw1229@163.com

Hong-Liang Zheng
School of Computer and Information Technology
Liaoning Normal University
Dalian 116029, P. R. China