

An New Algorithm for Modeling Regression Curve

JiSheng Hao, LeRong Ma and WenDong Wang

College of Computer Science, Yanan University, Shanxi Yanan, China
716000, Email: yadxhjs1963@163.com

Abstract: A new algorithm for modeling regression curve is put forward in the paper, it combines the B-spline network with improved support vector regression. Our experimental results on simulated data demonstrate that it is feasible and effective.

Keywords: Support vector machines; support vector regression; B-spline network; regression curve;

1 introduction

Support vector machines (SVM) is a new method of machine learning from statistical learning theory, which is a fresh tool of solving machine learning problem in virtue of optimization. It was first put forward by Vapnik in the 1990s [1]. In recent years, it have made breakthrough progress in its theoretical research, applied research and algorithm implementation that become a good way of overcoming traditional difficulty for dimension of the disaster and overfitting [2]. SVM uses structural risk minimization principle that has the promotion of good performance because it is considered the fitting and complexity for training samples. SVM can solve pattern classification and regression problems.

B-spline network is the lattice three-tier structure of associative memory networks [3][5], its structure is illustrated in Figure.1. B-spline function in latent layer is used as the basic function that is defined in input space of lattice. For a random input, a few B-spline functions in latent layer is activated and the network output is a linear combination of these active basic function. Since the support set of the basic function is limited, the network has the following features:

a) The knowledge in the network is locally stored without global distributed storing, and learning is local. Therefore the learning from a part in input space isn't influence the learning results in other part.

b) The learning algorithm converges fast, and the network is convenient for real time online applications.

c) The network has the good expressing capacity for fuzzy knowledge.

Thereby this kind network becomes more and more important, and widely used in control, modeling and pattern recognition, and other fields [5]

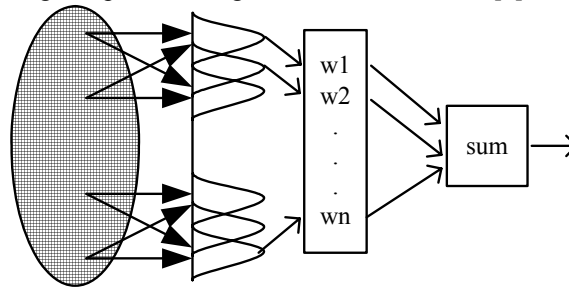


Fig.1. B-spline network structure

In the paper, a new algorithm for Modeling Regression Curve is put forward. Firstly, we obtain the support vectors from input sample set by an improved support vectors regression algorithm [2][4][6]. Then, we train the B-spline network with these support vectors as training sample set. Thus, we can quickly create the model of regression curve with this method because the B-spline network is characterized by short training period, fast convergence. Moreover, the regression curve that is created is smoothing curve as the basis functions for B-spline network are continuous functions.

2 ϵ —Support Vectors Regression Algorithm

The basic problem of Regression is to find $f \in F$, where F is function set, such that

$$R(f) = \int l(y - f(\bar{x})) dP(\bar{x}, y), \quad \bar{x} \in R^n, y \in R$$

minimizes, where $l(\bullet)$ is loss function that denotes the deviation between y and $f(\bar{x})$ which always is defined to be

$$l(\bullet) = |y - f(\bar{x})|^m$$

where m is a positive integer, and $P(x, y)$ is a Probability distribution function. Since we don't prior know $P(x, y)$, we can't calculate the $R(f)$ directly. According to structural risk minimization principle we have

$$R(f) \leq R_{emp} + R_{gen}$$

where $R_{emp} = \frac{1}{l} \sum_{i=1}^l I(y_i - f(\bar{x}_i))$ is empirical risk, and R_{gen} is a measure of the complexity for $f(\bar{x})$. Hence, $R_{emp} + R_{gen}$ can be a upper bound of $R(f)$.

The core of \mathcal{E} —Support Vectors Regression Algorithm for solving the regression problem is the following:

Given a sample set $\{(x_i, y_i) | (x_i, y_i) \in R^n \times R, i=1, 2, \dots, l\}$ which probability distribution function is $P(x, y)$, assumed that $F = \{f | f(x) = \bar{\omega}^T \Phi(x) + b, \bar{\omega} \in R^n\}$ is a regression function set.

Introduced structural risk function

$$R_{reg} = \frac{1}{2} \|\bar{\omega}\|^2 + C \cdot R_{emp}^\varepsilon[f] \quad (1)$$

where $\|\bar{\omega}\|^2$ is the complexity of $f(\bar{x})$, and c is punishment coefficient, $R_{emp}^\varepsilon[f]$ is the empirical risk which role is a tradeoff between the model complexity and the empirical risk.

Introduced ε -insensitive loss function

$$|y - f(\bar{x})|_\varepsilon = \begin{cases} 0 & |y - f(\bar{x})| \leq \varepsilon \\ |y - f(\bar{x})| - \varepsilon & other \end{cases} \quad (2)$$

Its meaning don't punish the item which deviation is less than ε , thus can increase the robustness of regression.

According to (2) we can define $R_{emp}^\varepsilon[f] = \frac{1}{l} \sum_{i=1}^l |y_i - f(\bar{x}_i)|_\varepsilon$.

In order to make the mean risk minimize we not only control training error but also control the model complexity, thus can improve the generalized ability of the model. Hence, (1) minimized is the core idea of statistic learning theory.

The above regression problem is equal to

$$\min \frac{1}{2} \bar{\omega}^T \bar{\omega} + C \sum_{i=1}^l (\zeta_i + \zeta_i^*) \quad (3)$$

$$s.t. \quad y_i - \bar{\omega}^T x_i - b \leq \varepsilon + \zeta_i \quad ; \quad \bar{\omega}^T x_i + b - y_i \leq \varepsilon + \zeta_i^* \quad ;$$

$$\zeta_i, \zeta_i^* \geq 0, i=1, 2, \dots, l$$

Where ζ_i, ζ_i^* are introduced relaxation variables which intent are made(3)solution exist.

We can obtain the following dual problem by Lagrangian and dual theorem:

$$\min_{\alpha, \alpha^*} \left\{ \frac{1}{2} [\bar{\alpha}, (\bar{\alpha}^*)^T] \begin{bmatrix} \bar{Q} & -\bar{Q} \\ -\bar{Q} & \bar{Q} \end{bmatrix} \begin{bmatrix} \bar{\alpha} \\ \bar{\alpha}^* \end{bmatrix} + [\varepsilon I^T + \bar{y}^T, \varepsilon I^T - \bar{y}^T] \begin{bmatrix} \bar{\alpha} \\ \bar{\alpha}^* \end{bmatrix} \right\} \quad (4)$$

$$s.t. \begin{bmatrix} I^T & -I^T \end{bmatrix} \begin{bmatrix} \bar{\alpha} \\ \bar{\alpha}^* \end{bmatrix} = 0, \alpha_i, \alpha_i^* \in [0, C]$$

where $Q_{i,j} = \Phi^T(x_i)\Phi(x_j)$; $I = [1, 1, \Lambda, 1]^T$; $\bar{\alpha}, \bar{\alpha}^*$ are Lagrange multiplier.

We can get the value of $\bar{\alpha}$ by solving this quadratic programming problem, and obtain

$$\bar{\omega} = \sum_{i=1}^l (\alpha_i - \alpha_i^*) \Phi(x_i) \quad (5)$$

With *KKT* condition we can calculate b as the following

$$\begin{cases} b = y_j - \varepsilon - \sum_{i=1}^l (\alpha_i - \alpha_i^*) K(x_j, x_i); \alpha_i, \alpha_i^* \in [0, C] \\ b = y_j + \varepsilon - \sum_{i=1}^l (\alpha_i - \alpha_i^*) K(x_j, x_i); \alpha_i, \alpha_i^* \in [0, C] \end{cases} \quad (6)$$

so we can obtain

$$f(\bar{x}) = \sum_{i=1}^l (\alpha_i - \alpha_i^*) K(\bar{x}, x_i) + b \quad (7)$$

Where $K(\bar{x}, x_i) = \Phi^T(\bar{x})\Phi(x_i)$ is the kernel function that satisfy Mercer's condition. Without knowing the Φ this function achieves the nonlinear transform between the input space and feature Space, it is an important feature for SVM.

The common kernel functions are the following:

Polynomial function: $K(\bar{x}_i, \bar{x}_j) = ((\bar{x}_i \bullet \bar{x}_j) + c)^d$

Gaussian radial basis kernel function:

$$K(\bar{x}_i, \bar{x}_j) = \exp(-\|\bar{x}_i - \bar{x}_j\|^2 / \sigma^2)^d$$

Sigmoid kernel function: $K(\bar{x}_i, \bar{x}_j) = \tanh(c_1(\bar{x}_i \bullet \bar{x}_j) + c_2)$

3 The improved support vectors regression machines

In [6] the author put forwards an improved support vectors regression machine, which modifies (3) the following:

$$\min\left(\frac{1}{2}\bar{\omega}^T \bar{\omega} + b^2\right) + C \sum_{i=1}^l (\zeta_i + \zeta_i^{*2}) \quad (8)$$

$$s.t. \quad y_i - \bar{\omega}^T \bar{x}_i - b \leq \varepsilon + \zeta_i \quad ; \quad \bar{\omega}^T \bar{x}_i + b - y_i \leq \varepsilon + \zeta_i^* \quad ;$$

$$\zeta_i, \zeta_i^* \geq 0, i = 1, 2, \Lambda, l$$

With the similar approach like the second section we can obtain the regression function for the improved support vectors regression machines

$$f(\bar{x}) = \sum_{i=1}^l (\alpha_i - \alpha_i^*) (K(\bar{x}, \bar{x}_i) + 1) \quad (9)$$

Obviously, this improved support vectors regression machines simplify the constraints of optimization, which have no equality constraints and the constraints of rectangular and there only have nonnegative constraints [6]. In addition, it has a more concise dual expression than standard SVM because there is no parameter b in (9).

4 The Algorithm of Regression Curve based the improved support vectors regression machines and B-spline network

Given a training sample set $\{(x_i, y_i) | (x_i, y_i) \in R^n \times R, i = 1, 2, \Lambda, l\}$, and assumed that the sample points are independence with the same probability distribution $P(\bar{x}, y)$ in $R^n \times R$, and also given ε -insensitive loss function (2), thus the regression problem is to find a $f(\bar{x})$ such that

$$R(f) = \int l(y - f(\bar{x})) dP(\bar{x}, y)$$

minimizes. The Algorithm of Regression Curve based the improved support vectors regression machines and the B-spline network is the following:

By selecting the proper kernel function and C、 ε in (1) we can obtain the support vectors for input sample points with the support vectors regression algorithm in the third section.

We train the B-spline network with the get support vectors.

Since the basis functions for the B-spline network are continuous functions and this network can approximate a random functional with arbitrary precision, we can create the model of smoothing regression curve.

Let the training set $T = \{(x_i, y_i) | i = 1, 2, \Lambda, 65\}$ from $y = f(x) = \sin x$ that has the noise. Namely, $x_1, x_2, \Lambda, x_{65}$ are the points that are *subject to* uniform distribution in $[-3.2, 3.2]$, and $y_i = \sin x_i + \zeta_i, i = 1, 2, \Lambda, 65$,

where the noise ξ_i is subject to normal distribution, and $E\xi_i = 0, E\xi_i^2 = \sigma^2, \sigma = 0.5$. With the algorithm that we put forward in the paper we finally create the model of regression curve as the Figure 2.

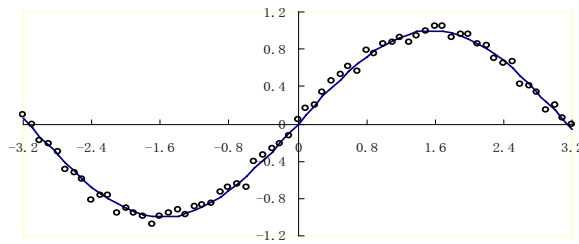


Fig.2 The regression curve model

6 Conclusion

The Algorithm of Regression Curve based the improved support vectors regression machines and B-spline network is put forward in this paper. With this algorithm we can create the model of smoothing regression curve. The algorithm is very feasible and effective that can be showed the above simulation.

References

- [1] Vapnik V. Statistical learning theory[M]. New York: John Wiley& Sons, 1998.
- [2] Deng NaiYang, Tian YingJie. A New Method in Data Mining—Support Vector Machine [M]. Science Press, 2004. 6, 224-273.
- [3] Moody J. Fastlearning in multi-resolution hierarchies[J]. Advances in Neural information Processing System, vol.1, 1989:29-39.
- [4] Scholkopf B, Smola A J. Learning with Kernels Support Vector Machines, Regularization, Optimization, and Beyond[M]. The MIT Press, 2002.
- [5] Martin Brown, Chris Harris. Neurofuzzy adaptive modeling and control [M]. Prentice Hall International (UK) Limited, 1994: 89-100.
- [6] Zhang HaoRan, Learning Algorithm for a New Regression SVM[J]. Journal of Test and Measurement Technology, Vol.20 No.2 2006