



An adaptive support vector regression based on a new sequence of unified orthogonal polynomials

Jinwei Zhao^{a,*}, Guirong Yan^b, Boqin Feng^a, Wentao Mao^c, Junqing Bai^b

^a Department of Computer Science, School of Electronic and Information Engineering, Xi'an Jiaotong University, Xi'an 710049, China

^b State Key Laboratory of Strength and Vibration of Mechanical Structures, Xi'an Jiaotong University, Xi'an 710049, China

^c College of Computer and Information Technology, Henan Normal University, Xinxiang 453007, China

ARTICLE INFO

Article history:

Received 19 August 2010

Received in revised form

1 August 2012

Accepted 2 September 2012

Available online 18 September 2012

Keywords:

Chebyshev polynomials

Kernel function

Adaptable measures

Small sample

Generalization ability

ABSTRACT

In practical engineering, small-scale data sets are usually sparse and contaminated by noise. In this paper, we propose a new sequence of orthogonal polynomials varying with their coefficient, unified Chebyshev polynomials (UCP), which has two important properties, namely, orthogonality and adaptivity. Based on these new polynomials, a new kernel function, the unified Chebyshev kernel (UCK), is constructed, which has been proven to be a valid SVM kernel. To find the optimal polynomial coefficient and the optimal kernel, we propose an adaptive algorithm based on the evaluation criterion for adaptive ability of UCK. To evaluate the performance of the new method, we applied it to learning some benchmark data sets for regression, and compared it with other three algorithms. The experiment results show that the proposed adaptive algorithm has excellent generalization performance and prediction accuracy, and does not cost more time compared with other SVMs. Therefore, this method is suitable for practical engineering application.

© 2012 Elsevier Ltd. All rights reserved.

1. Introduction

Support vector regression (SVR), a support vector machine (SVM) for regression, has been widely applied to the fields of machinery fault diagnostic technique, dynamics environmental forecasting, and earthquake prediction. Based on VC dimension and structural risk minimization [1], it can solve some practical problems such as sparsity, nonlinearity, high dimension, etc.

However, in practical applications, the training data sets in some important fields, such as telemetry data of rockets and missiles and data of human experimentation of vaccine, are sparse, of small size and contaminated by noise. This may decrease the generalization ability and the prediction accuracy of the algorithm. One way to solve this problem is to improve algorithm structure. Recently, several new SVM learning algorithms have been proposed for more powerful generalization ability. Least squares support vector machine (LSSVM) [2,3], for example, curtails the training time obviously due to the availability of equality instead of inequality constraints during the modeling process [4]. However, it cannot solve the sparsity problems and may increase the predictive time due to the large number of support vectors [5]. Other algorithms, such as proximal support vector machine (PSVM) [6], fuzzy support vector machine (FSVM) [7], possibility support vector machine (PSVM) [8] and semi-supervised support vector machine [9], all have certain shortcomings, which limit their

application in practical engineering projects. Another idea is to find an excellent kernel function or hybrid kernel function for the algorithm, and optimize it so as to get higher generalization ability. The most common kernel functions include Gauss kernel $K(\mathbf{x}, \mathbf{z}) = \exp(-\|\mathbf{x} - \mathbf{z}\|^2 / 2\sigma^2)$ [1], polynomial kernel $K(\mathbf{x}, \mathbf{z}) = ((\langle \mathbf{x}, \mathbf{z} \rangle + 1) / \beta)^n$ [1], sigmoid $K(\mathbf{x}, \mathbf{z}) = \tanh(a \langle \mathbf{x}, \mathbf{z} \rangle + r)$ [1], wavelet kernel $K(\mathbf{x}, \mathbf{z}) = \prod_{j=1}^m (\cos(1.75(x_j - z_j)/a) \exp(-(\|x_j - z_j\|^2 / 2a^2)))$ [10], where σ , β , a and r are the kernel parameters. These kernel yield an inner product of two given vectors in a high dimensional feature space where all input data can be linearly separated, without the need of an appropriate transformation function $\Phi(\bullet)$ [1]. However, the adaptive ability of any single kernel function is limited, resulting in good prediction accuracy on some data sets while poor accuracy on others. In addition, with the increase of the number of kernel parameters in the hybrid kernel function, the parameter optimization and weight selection requires a high-efficiency mechanism, for instance, boosting algorithm [11,12], semi-definite programming (SDP) [13], quadratically constrained quadratic program (QCQP) [14], semi-infinite linear program (SILP) [15,16], hyperkernels [17], simple multiple kernel learning (SimpleMKL) method [18] and the kernel target alignment [19–21]. Actually, the task is quite complex and may reduce the efficiency of the algorithm.

In this paper, we propose a kernel function based on a sequence of orthogonal polynomials varying with their coefficient, and design an adaptive algorithm for optimizing the parameters of the kernel function. Our aim is to improve the generalization ability of the algorithm and avoiding the complicated parameter optimization and weight selection in the learning process as well.

* Corresponding author. Tel.: +86 29 82660978.
E-mail address: zjw76888@gmail.com (J. Zhao).

The issue we are concerned about first is to find a new sequence of orthogonal polynomials. As is known to all, Chebyshev polynomials, a sequence of orthogonal polynomials, have been commonly applied in many fields [22,23], and have attracted wide attention in the field of machine learning due to their orthogonality. Ye et al. based on the first-kind Chebyshev polynomials, constructed an orthogonal Chebyshev kernel [22]. Sedat et al. further proposed a generalized Chebyshev kernel [23,24] which extended the orthogonal Chebyshev kernel from single variable input to vector input, and modified Chebyshev kernel [25] which replaced the weighting function with an exponential function to better capture the non-linearity along the decision surface. Compared with the traditional kernels, these kernels endue the algorithm with better generalization ability and prediction accuracy, and enable it to create less support vectors. However, the key problem in small sample regression is essentially how the kernel function and machine learning algorithm correctly determine the distribution characteristics of the small sample data sets. If the kernels derived from only the first-kind Chebyshev polynomial kernel are used, the limited adaptive ability of the kernels may result in poor prediction accuracy on some data sets. Considering this, we propose a new sequence of unified polynomials based on both the first- and second-kind Chebyshev polynomials, the unified Chebyshev polynomials (UCP), which may vary with their coefficient. Once the polynomial coefficient changes, the mathematical formulation of the polynomials will change correspondingly. By means of this property, the kernel function constructed by UCP can be applied to different kinds of data sets. The next issue to consider is how to change the polynomial coefficient so as to apply the kernel to different data sets and obtain higher generalization performance. To realize this, an evaluation criterion for adaptive capability of the proposed kernel and an adaptive algorithm based on Riemannian manifold are proposed.

The remaining part of this paper is organized as follows. Section 2 gives a brief description of Chebyshev polynomials kernels and the proposed unified polynomials and presents the UCP kernel and proves the validity of it. In Section 3, we propose two adaptive measures and their calculation formula, and construct the evaluation criterion for the adaptive capability of the kernel. Based on the proposed kernel and the constructed evaluation criterion, an adaptive algorithm is designed. In Section 4, the role of the polynomial coefficient in the proposed kernel is evaluated. Then the proposed kernel and the adaptive algorithm are tested on the benchmark data sets. The results of the experiments are discussed in Section 5, while conclusion and possible future work are reported in Section 6. The detailed proofs of the orthogonality of the proposed polynomials and some proposed theorems are provided in the Appendix.

2. Construction of the unified Chebyshev kernel (UCK)

Many studies on machine learning have found the existence of redundant attributes of sample sets. Li and Liu once studied the distribution of three types of Iris flower [26]. They chose three attributes of the flower, namely, petal width, petal length and sepal width, to obtain the data set. It was found that the flowers could not be distinguished when all the three attributes were concerned about, but they were easily separated from one another when one of the attributes was excluded. This result indicated that the redundant attributes of the flowers were not propitious for classification. Since SVM solves the regression by classification method, the problem of redundant attributes should be taken into consideration. As is known to all, in SVM, the kernel function maps the sample points in the sample space into the feature space, where the problem of redundant attributes exists and leads to the decline in the generalization performance and “over-fitting” of SVM. The sample points in the feature space are linearly separated by

increasing the dimension. However, we know very little about the dimension and cannot control it. In the paper, we intend to find a valid kernel function that can construct a feature space with lowest attribute redundancy so as to improve the separability of small samples, and control the dimension efficiently.

Firstly, two kinds of Chebyshev polynomials and two kinds of common Chebyshev kernels will be introduced and discussed in the following part.

2.1. Chebyshev polynomials kernel

The Chebyshev polynomials, named after Pafnuty Chebyshev, are a sequence of orthogonal polynomials which are easily defined recursively. The Chebyshev polynomials of the first and second kinds are denoted by T_n and U_n , respectively, which can be deduced directly from Eqs. (1) and (2) as follows:

$$(1-x^2)y'' - xy' + n^2y = 0, \tag{1}$$

$$(1-x^2)y'' - 3xy' + n(n+2)y = 0, \tag{2}$$

where $x \in [-1, 1]$, $y \in \mathbb{R}$ and $n = 0, 1, 2, 3, \dots$

If $x = \cos(\theta)$, the Chebyshev polynomials of the first kind are described by the recurrence relation

$$T_n(\cos(\theta)) = \cos(n\theta), \quad n = 0, 1, 2, 3, \dots \tag{3}$$

and those of the second kind are described by

$$U_n(\cos(\theta)) = \frac{\sin((n+1)\theta)}{\sin(\theta)}, \quad n = 0, 1, 2, 3, \dots \tag{4}$$

Since the conventional generating function for T_n is $(1-xt)/(1-2xt+t^2)$, we can obtain

$$\frac{1-xt}{1-2xt+t^2} = \sum_{n=0}^{\infty} T_n(x)t^n, \quad |x| \leq 1, |t| \leq 1. \tag{5}$$

While, as the conventional generating function for U_n is $1/(1-2xt+t^2)$, we can obtain

$$\frac{1}{1-2xt+t^2} = \sum_{n=0}^{\infty} U_n(x)t^n, \quad |x| \leq 1, |t| \leq 1. \tag{6}$$

Both T_n and U_n form two sequences of polynomials, which are orthogonal with respect to the weights $1/\sqrt{1-x^2}$ and $\sqrt{1-x^2}$, respectively. On the interval $(-1, 1)$, i.e., we have

$$\int_{-1}^1 \frac{T_m(x)T_n(x)}{\sqrt{1-x^2}} dx = \begin{cases} 0 & (m \neq n), \\ \frac{\pi}{2} & (m = n \neq 0), \\ \pi & (m = n = 0) \end{cases} \tag{7}$$

and

$$\int_{-1}^1 \sqrt{1-x^2} U_m(x)U_n(x) dx = \begin{cases} \frac{\pi}{2} & (m = n \neq 0), \\ 0 & ((m \neq n) \text{ or } (m = n = 0)). \end{cases} \tag{8}$$

The Chebyshev polynomials have the best uniform proximity and orthogonality which guarantee the minimum redundant attributes in feature space [22–25]. This makes it possible for the kernel to represent the sample space concisely, and reduces the complexity of the kernel matrix and the computational complexity of SVM. The kernel derived from the first-kind Chebyshev polynomials has been applied in machine learning, while that from the second-kind has not been involved in any application. Here we construct two Chebyshev kernels based respectively on the two Chebyshev polynomials, $K(x, z) = \sum_{i=0}^n T_i(x)T_i^T(z)/\sqrt{M-\langle x, z \rangle}$ and $K(x, z) = \sqrt{M-\langle x, z \rangle} \sum_{i=0}^n U_i(x)U_i^T(z)$, where M denotes dimension. To compare the regression results of the two kinds of kernels, we use them in the approximation of the same training data sets, of which the distribution is not easily recognized (Fig. 1). To confirm the result, this is done on another data set again. The experiments yield very different results, indicating that

متن کامل مقاله

دریافت فوری ←

ISIArticles

مرجع مقالات تخصصی ایران

- ✓ امکان دانلود نسخه تمام متن مقالات انگلیسی
- ✓ امکان دانلود نسخه ترجمه شده مقالات
- ✓ پذیرش سفارش ترجمه تخصصی
- ✓ امکان جستجو در آرشیو جامعی از صدها موضوع و هزاران مقاله
- ✓ امکان دانلود رایگان ۲ صفحه اول هر مقاله
- ✓ امکان پرداخت اینترنتی با کلیه کارت های عضو شتاب
- ✓ دانلود فوری مقاله پس از پرداخت آنلاین
- ✓ پشتیبانی کامل خرید با بهره مندی از سیستم هوشمند رهگیری سفارشات