Available online at www.sciencedirect.com

SCIENCE DIRECT®

Pattern Recognition 39 (2006) 342–350

PATTERN RECOGNITION
THE JOURNAL OF THE PATTERN RECOGNITION SOCIETY

www.elsevier.com/locate/patcog

An adaptive error penalization method for training an efficient and generalized SVM

Yiqiang Zhan[a, b, c, *], Dinggang Shen[a, c]

[a] Section of Biomedical Image Analysis, Department of Radiology, University of Pennsylvania, Philadelphia, PA, USA
[b] Department of Computer Science, The Johns Hopkins University, Baltimore, MD, USA
[c] Center for Computer-Integrated Surgical Systems and Technology, The Johns Hopkins University, Baltimore, MD, USA

Received 7 June 2005; received in revised form 14 September 2005

Abstract

A novel training method has been proposed for increasing efficiency and generalization of support vector machine (SVM). The efficiency of SVM in classification is directly determined by the number of the support vectors used, which is often huge in the complicated classification problem in order to represent a highly convoluted separation hypersurface for better nonlinear classification. However, the separation hypersurface of SVM might be unnecessarily over-convoluted around extreme outliers, as these outliers can easily dominate the objective function of SVM. This situation eventually affects the efficiency and generalization of SVM in classifying unseen testing samples. To avoid this problem, we propose a novel objective function for SVM, i.e., an adaptive penalty term is designed to suppress the effects of extreme outliers, thus simplifying the separation hypersurface and increasing the classification efficiency. Since maximization of the margin distance of hypersurface is no longer dominated by those extreme outliers, our generated SVM tends to have a wider margin, i.e., better generalization ability. Importantly, as our designed objective function can be reformulated as a dual problem, similar to that of standard SVM, any existing SVM training algorithm can be borrowed for the training of our proposed SVM. The performances of our method have been extensively tested on the UCI machine learning repository, as well as a real clinical problem, i.e., tissue classification in prostate ultrasound images. Experimental results show that our method is able to simultaneously increase the classification efficiency and the generalization ability of the SVM.

Keywords: Support vector machine; Training method; Computational efficiency; Generalization ability

1. Introduction

Support vector machine (SVM) is a new generation of learning systems based on statistical learning theory [1]. Considering a two-class classification problem with m labeled training samples, $\{(\vec{x}_i, y_i) | \vec{x}_i \in R^n, y_i \in \{-1, 1\}, i = 1 \cdots m\}$, SVM aims to generate a hypersurface that has maximum margin to separate two classes. The classification of a testing sample is accomplished by calculating its distance to the hypersurface:

$$d(\vec{x}) = \sum_{i=1}^{m} \alpha_i y_i K\left(\vec{x}_i, \vec{x}\right) + b, \tag{1}$$

where α_i and b are the parameters determined by SVM's learning algorithm, and $K(\vec{x}_i, \vec{x})$ is the kernel function. Samples $\vec{x}_i$ with nonzero parameters α_i are called "support vectors".

Since its generation in 1995 [1], SVM has drawn considerable attentions in various research areas [3–8] due to its striking properties as described next. First, based on the idea of structural risk minimization [1], SVM can achieve high generalization ability by minimizing the Vapnik–Chervonenkis

* Corresponding author. Department of Computer Science, The Johns Hopkins University, Baltimore, MD, USA. Tel.: +1 215 662 7362.
E-mail addresses: yzhan@cs.jhu.edu (Y. Zhan), Dinggang.Shen@uphs.upenn.edu (D. Shen).

0031-3203/$30.00 © 2005 Pattern Recognition Society. Published by Elsevier Ltd. All rights reserved.
doi:10.1016/j.patcog.2005.09.008

dimension. Second, by using the kernel trick [2], the samples are implicitly mapped to a higher dimensional space. Therefore, SVM can generate a convoluted hypersurface to nonlinearly separate different classes. Finally, the training procedure of SVM can be eventually formulated as a constraint quadratic optimization problem, which has a unique global minimum.

However, although SVM shows superior classification ability in pattern recognition problems, it usually needs a huge number of support vectors to parameterize the separation hypersurface, particularly when confronting large data classification problems. Since the calculation of the decision function with many nonzero parameters α_i in Eq. (1) is very time consuming, SVM exhibits substantially slower classification speed compared to the neural network [9]. This disadvantage unavoidably limits the capability of SVM in the applications that require a massive number of classifications [5] or real-time classification [10].

In this paper, we propose a novel training method to increase the efficiency as well as the generalization ability of SVM. We notice that the extreme outliers in the training set usually make the separation hypersurface unnecessarily over-convoluted, thus affecting both the efficiency and the generalization of SVM. This problem is actually resulted from the domination of the extreme outliers over the objective function of the standard soft-margin SVM [12]. To overcome this problem, we reformulate the objective function by designing an adaptive penalty term to suppress the effects of extreme outliers in objective function, thereby simplifying the separation surface and increasing the generalization ability of SVM. Importantly, we find that our designed objective function can be reformulated as a quadratic optimization problem with adaptive constraints, which is similar to the dual problem of the standard soft-margin SVM. Therefore, any existing SVM training method can be borrowed for training our proposed SVM.

The remainder of this paper is organized as following. In Section 2, we will first analyze the problem in details. Then, the reformulated objective function with an adaptive penalty term to outliers is presented. The training method for the reformulated SVM will also be provided. Section 3 will present the experimental results of our method on the UCI machine learning repository, as well as a real clinical problem, i.e., tissue classification in a set of prostate ultrasound images. This paper concludes in Section 4.

2. Methods

2.1. Problem description

As indicated in Eq. (1), the computational cost of SVM is determined by the number of the support vectors, i.e. training samples with nonzero parameters α_i. According to their relative positions to the separation hypersurface,

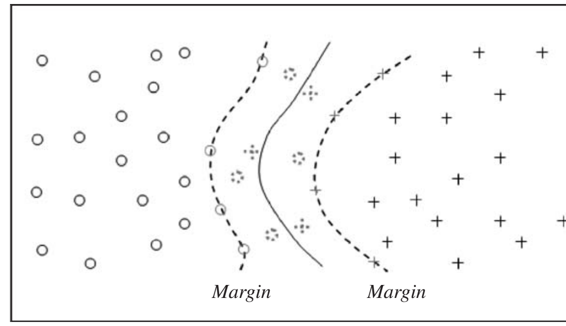

Fig. 1. Schematic explanation of the separation hypersurface (solid curves), margins (dashed curves), and support vectors of SVM (grey circles/crosses). The positive and the negative training samples are indicated by circles and crosses, respectively.

support vectors can be categorized into two types. The first type of support vectors are the training samples that exactly locate on the margins of the separation hypersurface, i.e., $\mathrm{d}(\vec{x}_i) = \pm 1$, such as solid grey circles/crosses shown in Fig. 1. The second type of support vectors are the training samples that locate beyond their corresponding margins, i.e., $y_i\, \mathrm{d}(\vec{x}_i) < 1$, such as dashed grey circles/crosses shown in Fig. 1. For a SVM, the second type of support vectors are regarded as misclassified samples, although some of them still locate at the correct side of the hypersurface.

SVM usually has a huge number of support vectors, when the distributions of the positive and the negative training samples from a large dataset highly overlap with each other. This unfavorable situation is resulted from two reasons: (1) a large number of the first-type support vectors are needed to construct a highly convoluted hypersurface, in order to separate two classes; (2) even the highly convoluted separation hypersurface has been constructed, a lot of confounding samples will be misclassified, thus selected as the second type of support vectors.

Some support vectors might be redundant to parameterize the separation hypersurface. Based on this hypothesis, researchers have proposed efficient SVM training methods [11,13]. Compared to Ref. [13], the method proposed by Osuna and Girosi in Ref. [11] is more feasible, and it offered a principle for controlling the accuracy of approximation. This method approximates the separation hypersurface with a subset of the support vectors by using a Support vector regression machine (SVRM). If the separation hypersurface is relatively simple, Osuan's method is quite effective to reduce the number of support vectors without system degradation. However, in many large dataset classification problems, SVM usually generates a locally over-convoluted separation hypersurface, which is difficult to be parameterized by a small number of support vectors as Osuan's method did. Therefore, in order to further decrease the number of support vectors, it is necessary to simplify the hypersurface without losing its classification ability.