Neurocomputing 125 (2014) 119–124

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

A fuzzy support vector machine algorithm for classification based on a novel PIM fuzzy clustering method ☆

Zhenning Wu *, Huaguang Zhang, Jinhai Liu

School of Information Science and Engineering, Northeastern University, Shenyang, Liaoning 110004, China

ARTICLE INFO

Available online 27 February 2013

Keywords:
Fuzzy classifier
Clustering methods
Pattern classification
Statistical learning theory
Support vector machine

ABSTRACT

The support vector machine (SVM) has provided excellent performance and has been widely used in real-world classification problems. Fuzzy methods used on the SVM solve the problem that the SVM is sensitive to the outliers or noises in the training set. In this paper, a novel partition index maximization (PIM) clustering method is studied to get a more reasonable and robust fuzzy membership for fuzzy SVM (FSVM). First, we improve the PIM clustering algorithm to cluster each of the two classes from the training set to get proper data centers. Then an algorithm is given to modify the boundary of PIM and form a new training set with fuzzy membership degrees. Finally, we use FSVM to induce the final decision function to show classification results. All the results indicate that the performance of PIM-FSVM is excellent.

1. Introduction

The theory of support vector machine (SVM) which based on the idea of structural risk minimization, is an important classification technique and has drawn much attention on this topic in recent years [1–6]. With many applications, SVM has been shown to provide higher performance than traditional learning machines and has been introduced as powerful tool for solving classification problems and nonlinear function estimation problems in the fields of pattern recognition and machine learning.

The SVM first maps the input points into a high-dimensional feature space and then constructs a separating hyperplane that maximizes the margin between two classes in this space. Maximizing the margins between two classes can be described as a quadratic programming (QP) problem and can be solved from its dual problem by introducing Lagrangian multipliers. With no knowledge of the mapping, the SVM uses the dot product functions to find the optimal hyperplane, called kernels, in the high-dimensional feature space [4]. The solution of the optimal hyperplane is written as a combination of a few input points that are called support vectors.

Though SVM is good from the point of view of both theory and application, one of the main drawbacks of the standard SVM is that the training process of the SVM is sensitive to the outliers or noises in the training dataset due to overfitting [7]. A training data point may neither exactly belong to any of the two classes when the outliers or noises exist in many real-world classification problems. For example, a data point near the margin may belong to one class or just be a noise point. But these kinds of uncertainty points may be more important than others for making decision, which leads to the problem of overfitting. As fuzzy approaches are effective in solving uncertain problems, Lin and Wang [8] have proposed FSVM based on the standard SVM model for classification problems with outliers or noises. The key for the FSVM is how to get the fuzzy memberships of the training data, on which much work has been done. Lin and Wang also gave an approach to automatic set the fuzzy memberships. In this method, many parameters have to be optimized, which makes it very difficult to set the confident factor and the trashy factor automatically. Based on the distance between a sample and its class center in the high-dimensional feature space, a new fuzzy membership function has also been designed [9], which is a kernel extension of the formulation in [8]. Considering that the same training sample may belong to multiple classes, a bilateral weights-based FSVM model has been presented [10]. At present, this model faces two main difficulties: how to set fuzzy memberships and how to decrease computational complexity. Based on the fuzzy c-means (FCM) clustering in the original input space and the fuzzy IF-THEN rules, the ε-margin nonlinear classification model [11] have been proposed. Yang et al. [12] have proposed a kernel fuzzy c-means clustering based fuzzy support vector machine. The strategies of iteratively setting sample weights and ensemble learning have been adopted, where the strategy of iteratively setting sample weights is similar to the two-stage strategy of solving the linear system in [13]. At present, how to set the reasonable sample memberships in the field of pattern recognition and machine learning is still an important problem.

☆ This work was supported by the National Natural Science Foundation of China (50977008, 61034005, 61104021), National Basic Research Program of China (2009CB320601), Science and Technology Research Program of the Education Department of Liaoning Province (LT2010040), the National High Technology Research and Development Program of China (2012AA040104) and the Fundamental Research Funds for the Central Universities (N100304007).

* Corresponding author. Tel.: +86 2483687762; fax: +862 483671498.
E-mail address: wuzn2003@hotmail.com (Z. Wu).

0925-2312/$ - see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.neucom.2012.07.049

The goals of this paper are to propose a strategy of setting the reasonable memberships for training data and to design a robust FSVM algorithm for classification problems with outliers or noises. Clustering is an effective method in machine learning and pattern recognition. Wu et al. [14] have presented a kernel FCM (KFCM) clustering algorithm, Yang et al. [15] have proposed a Mercer kernel based deterministic annealing algorithm, Rhee et al. [16] have developed a kernel version of [17], and Kim et al. [18] have evaluated the performance of kernel fuzzy clustering algorithms. While many fuzzy methods were studied to set the fuzzy memberships for the training data set, they have a common drawback that almost no training data has a fuzzy membership of 1, which means that almost no data point is certainly belong to one class, which contradicts the fact in real world. To solve this problem, a cluster core concept has been recently proposed by Özdemir and Akarun [19] and Wu et al. [20]. Akarun proposed a partition index maximization (PIM) algorithm by adding the partition coefficient into the FCM objective function. Wu et al. proposed a fuzzy compactness and separation (FCS) clustering algorithm, whose objective function is modification of the validity index proposed by Fukuyama and Sugeno [21] and is generalization of the FCM objective function by adding a between-cluster variation. The other problem of the clustering method is that the sharp of the data set needs to be spherical, while in many real problems, which will not be satisfied. In this paper, based on the PIM clustering algorithm, we developed a new method to deal with nonsphere-type data sets.

To generalize a FSVM overcoming both the problems of nonsphere-type data set and almost no data point has the membership of 1, we propose a strategy of setting a set of reasonable memberships to design a robust FSVM algorithm for classification problems with outliers or noises.

The rest of the paper is organized as follows. The FSVM model for classification problems is briefly reviewed in Section 2. The PIM clustering algorithm and its validity are briefly reviewed in Section 3. In Section 4, the details of the PIM-FSVM algorithm are given. In Section 5, the computational complexity analysis of the PIM-FSVM algorithm is discussed. In Section 6 the experimental results and analysis are presented. Finally, we give the related conclusions in Section 7.

2. Fuzzy support vector machine model for classification problems

In this section we briefly review the basic theory of SVM for classification.

Suppose we are given a set of labeled training points $\{x_i, y_i, s_i\}_{i=1}^{l}$ for a binary classification problem, where $x_i \in R_n$ are the input data, $y_i \in \{-1, 1\}$ are the corresponding binary class labels, and $s_i \in [0,1]$ is the fuzzy membership degree of x_i belonging to y_i. The FSVM model for the binary classification problems can be described as a quadratic programming (QP) problem based on the inequality constraints as following [8]:

$$\min_{w,b,\varepsilon} J(w,b,\varepsilon) = \frac{1}{2} w^T w + \zeta \sum_{i=1}^{l} s_i \varepsilon_i, \tag{1}$$

s.t.

$$y_i[w^T \varphi(x_i) + b] \geq 1 - \varepsilon_i, \quad i = 1, \ldots l, \tag{2}$$

$$\varepsilon_i \geq 0, \quad i = 1, \ldots l, \tag{3}$$

where w is a normal vector of the hyperplane, b is a bias, $\varphi(x_i)$ is a nonlinear function which maps x_i to a high-dimensional feature space, s_i holds for misclassified examples, and C is a regularization constant that controls the tradeoff between the maximizing the classification margin and minimizing the cost of misclassification.

This quadratic-optimization problem can be solved by constructing a Lagrangian representation and transforming it into the following dual problem:

$$\max_{\alpha} \sum_{i=1}^{l} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{l} \alpha_i \alpha_j y_i y_j K(x_i, x_j), \tag{4}$$

s.t.

$$\sum_{i=1}^{l} \alpha_i y_i = 0, \tag{5}$$

$$0 \leq \alpha_i \leq \gamma s_i, \quad i = 1, \ldots l, \tag{6}$$

where α_i is a Lagrange multiplier with a value not equal to 0 when data point i is a support vector, and $K(x_i, x_j)$ is a kernel function.

We used Gaussian kernel function in this paper, which is shown as

$$K(x_i, x_j) = \exp\left(-\frac{1}{2\sigma^2} \|x_i - x_j\|^2\right). \tag{7}$$

Once the FSVM model has been solved, the class label of a testing x can be predicted as follows:

$$y(x) = sgn\left[\sum_{i=1}^{l} \alpha_i y_i K(x, x_i) + b\right]. \tag{8}$$

3. PIM fuzzy c-means clustering algorithm and its validity

Clustering is one of the most powerful tools in the data-mining process for discovering interesting distributions and patterns in the underlying data. Data centers are concerned in clustering methods. It assumes that the attributes of the data can form a center that data from different class may be on or near its own class attributes center. While FCM clusters data into two clusters, one data has two fuzzy memberships, one for positive class, the other for negative class. The fuzzy memberships are determined by the distance from this data point to each cluster centers. According to this, it is easy to see that almost no data has fuzzy membership of 1. Özdemir and Akarun [19] proposed a PIM algorithm for color quantization of images. We improved this algorithm by giving multi-boundary which is more reasonable. Our PIM objective function was constructed by minimizing the FCM objective function J_{PIM} as

$$J_{PIM}(\mu, a) = \sum_{i=1}^{C} \sum_{j=1}^{n} \mu_{ij}^m d(x_j, a_i) - \beta_i \sum_{i=1}^{C} \sum_{j=1}^{n} \mu_{ij}^m. \tag{9}$$

The cluster center update equation of PIM is equivalent to FCM,

$$a_i = \frac{\sum_{j=1}^{n} \mu_{ij}^m x_j}{\sum_{j=1}^{n} \mu_{ij}^m}, \quad i = 1, \ldots, C, \tag{10}$$

while the membership update equation is

$$\mu_{ij} = \frac{[d(x_j, a_i) - \beta_i]^{1/(1-m)}}{\sum_{i=1}^{C} [d(x_j, a_i) - \beta_i]^{1/(1-m)}}. \tag{11}$$

The detailed computational steps are as follows:

(1) Choose the cluster number C , the termination parameter ε and m.
(2) Choose the kernel function K and its parameters.
(3) Initialize the membership μ_{ij}, $j = 1, \ldots, n$, $i = 1, \ldots, C$.
(4) Compute the cluster center a_i with (10).
(5) Compute the objective function J_{PIM} with (9).
(6) Update μ_{ij} with (11) and a_i with (10).