Pattern Recognition 67 (2017) 32–46

Contents lists available at ScienceDirect

Pattern Recognition

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

Weighted linear loss multiple birth support vector machine based on information granulation for multi-class classification

Shifei Ding[a,b,*], Xiekai Zhang[a], Yuexuan An[a], Yu Xue[c]

[a] School of Computer Science and Technology, China University of Mining and Technology, Xuzhou 221116, China
[b] Key Laboratory of Intelligent Information Processing, Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100190, China
[c] School of Computer and Software, Nanjing University of Information Science & Technology, Nanjing 210044, China

ARTICLE INFO

Article history:
Received 23 November 2015
Revised 5 September 2016
Accepted 3 February 2017
Available online 4 February 2017

Keywords:
Multi-class classification
Twin support vector machine
Multiple birth support vector machine
Granular computing

ABSTRACT

Recently proposed weighted linear loss twin support vector machine (WLTSVM) is an efficient algorithm for binary classification. However, the performance of multiple WLTSVM classifier needs improvement since it uses the strategy 'one-versus-rest' with high computational complexity. This paper presents a weighted linear loss multiple birth support vector machine based on information granulation (WLMSVM) to enhance the performance of multiple WLTSVM. Inspired by granular computing, WLMSVM divides the data into several granules and builds a set of sub-classifiers in the mixed granules. By introducing the weighted linear loss, the proposed approach only needs to solve simple linear equations. Moreover, since WLMSVM uses the strategy "all-versus-one" which is the key idea of multiple birth support vector machine, the overall computational complexity of WLMSVM is lower than that of multiple WLTSVM. The effectiveness of the proposed approach is demonstrated by experimental results on artificial datasets and benchmark datasets.

1. Introduction

Standard support vector machine (SVM) [1], which is based on the statistical learning theory [2] and the Vapnik–Chervonenkis (VC) dimension, classifies 2-category points by assigning them to one of two disjoint half spaces. SVM has drawn extensive attention of scholars [3–8] and has been applied to many fields successfully [9–13]. Twin support vector machine (TWSVM) [14], as an excellent extension of SVM, generates two nonparallel hyperplanes such that each plane is close to one of the two classes and as far as possible from the other class. TWSVM assigns a new sample to one of the classes depending on which hyperplane the new sample is closer to. An illustrative diagram of the thought of TWSVM in 2-dimensional space is shown in Fig. 1. TWSVM solves two small-scale quadratic programming problems (QPPs), whereas SVM solves one single QPP with a large number of constraints. Because of the strategy, TWSVM is almost four times faster than standard SVM [15]. In the last several years, TWSVM has been studied extensively and greatly generalized [16-27]. Recently, Shao et al. [28] proposed a novel extension of TWSVM, called weighted linear loss twin support vector machine (WLTSVM). Different from TWSVM, WLTSVM solves linear equations. The two systems of linear equations in WLTSVM for binary classification can be solved efficiently by using the well-known conjugate gradient algorithm, resulting in the ability to deal with large-scale datasets without any extra external optimizers. Many pattern recognition problems in real world are multi-class classification problems [29–38]. WLSTVM has also been extended to multi-class classification problems. However, multiple WLTSVM uses the strategy "one-versus-rest" with high computational complexity. Multiple WLTSVM builds a binary WLTSVM classifier for each class. Each binary WLTSVM in multiple WLTSVM is constructed by considering samples in one of the classes as positive samples and the rest as negative samples and training them. Multiple WLTSVM does not keep the advantages of WLTSVM that has high performance and low computational complexity. Multiple birth support vector machine (MBSVM) [39] is another novel extension of TWSVM with high performance. MBSVM uses the strategy "all-versus-one". The strategy "all-versus-one" considers one of the classes as negative class and all the rest classes as positive class in turn to generate a serious of binary sub-classifiers to solve the multi-class classification problem. However, MBSVM needs to deal with a serious of QPPs.

Several multi-class TWSVMs have been proposed. The strategies that can be used to extend binary TWSVMs to multi-class TWSVMs include: one-versus-rest, one-versus-one, one-versus-one-versus-rest, binary tree, rest-versus-one, directed acyclic grape (DAG). The

* Corresponding author.
E-mail address: dingsf@cumt.edu.cn (S. Ding).

http://dx.doi.org/10.1016/j.patcog.2017.02.011
0031-3203/

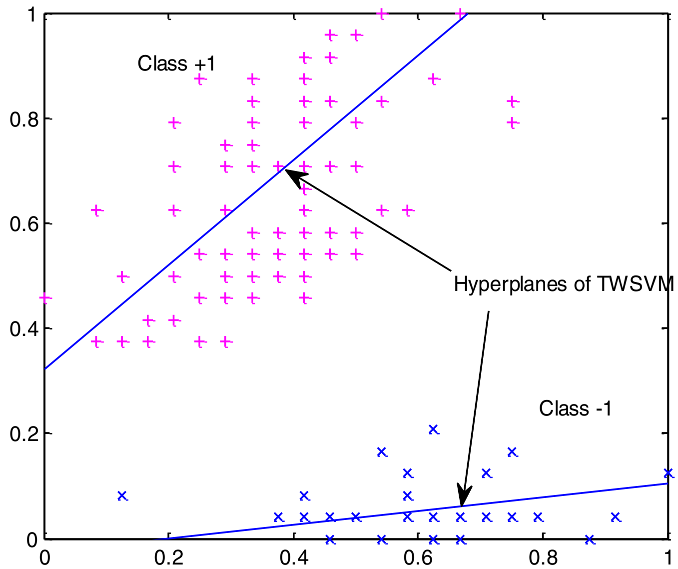

Fig. 1. Illustrative diagram of TWSVM.

strategy one-versus-rest is easy to understand and implement. However, the complexity of one-versus-rest based methods is high. In general, one-versus-one based multi-class TWSVMs and DAG based multi-class TWSVMs always get better classification accuracies than other methods. However, they need to build a large number of sub-classifiers. When the number of classes is big, they are complex systems. The complexity of one-versus-one-rest based methods is higher than one-versus-rest based methods. The binary TWSVMs in rest-versus-one based methods take one class as the negative class and the rest classes as the positive class, so the numbers of constrains are small. The complexity related to the number of constrains directly. Compared with other approaches, the advantage of rest-versus-one based methods is the lower complexity. In this paper, we employ the rest versus one strategy to reduce the time complexity.

Granular computing [40–42], covering all the research about theories, methods, techniques and tools of granulation, is a powerful method to handle large scale information. The essence of granular computing is to find an approximate solution, which is simple and low-cost, to replace the exact solution through using inaccurate and large scale information to achieve the tractability, robustness, low cost and better describing the real world of intelligent systems or intelligent control. The combination of granular computing with statistical learning theory is becoming a hotspot. Many effective granular SVM (GSVM) models for binary classification have been developed [41]. Wang et al. [44] proposed a GSVM model based on mixed measure; Ding et al. [45] proposed a fast fuzzy support vector machine based on information granulation; Cheng et al. [46] proposed a dynamic GSVM. However, combination of granular computing with extensions of TWSVM for multi-class classification is still an unsolved research problem.

This paper proposes a new classifier for multi-class classification, called weighted linear loss multiple birth support vector machine based on information granulation (WLMSVM) to enhance the performance of multiple WLTSVM. The proposed algorithm works as follows. Firstly, it splits the whole feature space into a set of information granules depending on the training data and signs the information granules into "pure granule" or "mixed granule" depending on the labels of training samples in them. A pure granule is a subspace in which there are only samples with same class label. A mixed granule is a subspace in which two or more classes present. Then, WLMSVM builds one multi-class sub-classifier in each mixed granule. Different from multiple WLTSVM, our approach uses the strategy "all-versus-one" which is the key idea of MBSVM. In a given mixed granule, the sub-classifier of WLMSVM generates one hyperplane for each class by solving a TWSVM-style QPP which considers the samples in one class as negative samples and all other samples in the mixed granule as positive samples. The last step is to predict the label of an unlabeled sample. Compared with other methods for multi-class classification, the proposed approach has three advantages. 1) By introducing the strategy "all-versus-one", our approach as a whole has low computational complexity. Especially when the number of classes is large, WLMSVM can always work faster than most of the other methods; 2) WLMSVM keeps the advantage of the multiple WLTSVM classifier. WLMSVM uses a weighted linear loss instead of hinge loss. The use of weighted linear loss leads to that WLMSVM only needs to solve several systems of linear equations; 3) Granular computing technique frees each sub-classifier to focus on the local information of data in the granules.

The rest of this article is organized as follows. In the next section, a brief review to TWSVM, WLTSVM, MBSVM and granular computing is provided. In Section 3, we introduce the proposed weighted linear loss multiple birth support vector machine based on information granulation for multi-class classification in detail. Experimental results are given in Section 4. In the last section, concluding remarks and further research to be developed are presented.

2. Related works

2.1. Twin support vector machine

Assume a binary classification problem with m samples in the n-dimensional real space R^n. The set of training data points is represented by $T=\{(x_i,\ y_i)\mid i=1,2,3,\ldots,m\}$, where x_i is input sample and $y_i \epsilon \{+1,-1\}$ is corresponding output. Let $m_1\times n$ matrix A denote the samples belonging to class $+1$ and $m_2\times n$ matrix B denote the samples belonging to class -1. Each row of A is a sample belonging to class $+1$, and each row of B is a sample belonging to class -1.

For a linear binary classification problem, TWSVM generates two nonparallel separating hyperplanes in R^n.

$$x^T w_2 + b_2 = 0 \quad \text{and} \quad x^T w_1 + b_1 = 0 \tag{1}$$

where w_1 and w_2 are the normal vectors of the separating hyperplanes, and b_1 and b_2 are the bias vectors. Each hyperplane of TWSVM is close to samples of one class and as far as possible from the rest samples. TWSVM obtains the hyperplanes by solving two small QPPs with constrains responding to only one of the classes.

The model of TWSVM is written as follows [14]:

$$\begin{aligned} \min \quad & \frac{1}{2}\|Aw_1+e_1b_1\|^2 + c_1 {e_2}^T\xi \\ \text{s.t} \quad & -(Bw_1+e_2b_1)+\xi \ge e_2, \quad \xi \ge 0 \end{aligned} \tag{2}$$

$$\begin{aligned} \min \quad & \frac{1}{2}\|Bw_2+e_2b_2\|^2 + c_2 {e_1}^T\eta \\ \text{s.t} \quad & (Aw_2+e_1b_2)+\eta \ge e_1, \quad \eta \ge 0 \end{aligned} \tag{3}$$

where c_1 and $c_2 > 0$ are parameters, e_1 and e_2 are column vectors of ones of appropriate dimensions and ξ and η are error variables.

By introducing formula of Hinge loss function [25]:

$$H_s(z) = max(0, s - z) \tag{4}$$

The QPPs (2) and (3) can be rewritten as:

$$\min \quad \frac{1}{2}(Aw_1+e_1b_1)^T(Aw_1+e_1b_1)+c_1e_2^T H_{e_1}(-Bw_1 - e_2b_1) \tag{5}$$