Pattern Recognition 43 (2010) 3468–3482

Contents lists available at ScienceDirect

Pattern Recognition

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

An online core vector machine with adaptive MEB adjustment

Di Wang [a], Bo Zhang [a,*], Peng Zhang [a], Hong Qiao [b]

[a] LSEC and Institute of Applied Mathematics, AMSS, Chinese Academy of Sciences, Beijing 100190, China
[b] Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China

ARTICLE INFO

Article history:
Received 28 September 2009
Received in revised form
18 April 2010
Accepted 15 May 2010

Keywords:
Minimum enclosing ball
Online classifier
Core vector machine
Support vector machine
Machine learning

ABSTRACT

Support vector machine (SVM) is a widely used classification technique. However, it is difficult to use SVMs to deal with very large data sets efficiently. Although decomposed SVMs (DSVMs) and core vector machines (CVMs) have been proposed to overcome this difficulty, they cannot be applied to online classification (or classification with learning ability) because, when new coming samples are misclassified, the classifier has to be adjusted based on the new coming misclassified samples and all the training samples. The purpose of this paper is to address this issue by proposing an online CVM classifier with adaptive minimum-enclosing-ball (MEB) adjustment, called online CVMs (OCVMs). The OCVM algorithm has two features: (1) many training samples are permanently deleted during the training process, which would not influence the final trained classifier; (2) with a limited number of selected samples obtained in the training step, the adjustment of the classifier can be made online based on new coming misclassified samples. Experiments on both synthetic and real-world data have shown the validity and effectiveness of the OCVM algorithm.

1. Introduction

So far, SVMs have been widely used in many real-world applications due to their good performance. These applications include face recognition [1,2], gene expression data clustering [3], pedestrian detection [4], handwriting recognition [5], as well as the classification tasks of text [6,7], fingerprint [8] and texture [9]. The main advantages of the SVMs can be summarized as follows [10]:

- a compromise between minimizing empirical risk and preventing overfitting is taken by implementing the structural risk minimization;
- the process of computing the classification hyperplane involves a convex quadratic optimization problem which can be solved efficiently and has a global solution;
- the obtained classifier is completely determined by the support vectors and the type of kernel functions which are used for training.

Despite the above advantages, SVMs have also the following two main disadvantages which limit their application in real-time pattern recognition problems:

- the convex quadratic optimization problem arising in SVMs is a large-scale problem for very large data sets, so it is difficult for SVMs to deal with very large data effectively;
- SVMs handle training samples in a batch mode; when a new training sample arrives, the whole training process has to be implemented once again to adjust the classifier; thus, it is not practical for SVMs to be used for online learning.

Recently, many algorithms have been proposed to address the fast computation issue of large-scale SVMs (see [18,19] for a good literature survey on this). Among these algorithms are the decomposed SVMs (DSVMs) [11–15] and core vector machines (CVMs) [16,17]. Their main ideas can be briefly summarized as follows.

- DSVMs essentially repeats two operations until some optimality condition is satisfied: one is to select a working set and the other is to minimize the original objective function of the quadratic programming (QP) problem arising in SVMs by updating the variables only associated with the working set. The key step in DSVMs is how to select a suitable working set at each iteration.
- CVMs reformulate SVMs as the minimum enclosing ball (MEB) problems in computational geometry. Then an approximate optimal solution to the original optimization problem can be obtained by utilizing efficient approximate MEB approaches. Reported experimental results on very large data sets have shown that classification results obtained by CVMs are as

* Corresponding author. Tel.: +86 10 6265 1358.
E-mail addresses: wangdi@amss.ac.cn (D. Wang), b.zhang@amt.ac.cn (B. Zhang), zhangpeng@amss.ac.cn (P. Zhang), hong.qiao@ia.ac.cn (H. Qiao).

0031-3203/$ - see front matter
doi:10.1016/j.patcog.2010.05.020

accurate as those obtained by SVMs, while the computation speed of the former is much faster than that of the latter since the computational complexity of the MEB problem is independent of the dimension and number of data samples.

DSVMs and CVMs have been successfully applied to solve many large-scale classification problems. However, the online learning issue of the DSVM and CVM classifiers is still not addressed. In these two algorithms, data are processed in a batch mode. When a new training sample arrives, the whole training process should be implemented once again to adjust the classifier. Thus, online adjustment of the classifier is impossible.

Online learning ability of a classifier is very important in real-time pattern recognition systems such as pedestrian detection system and aircraft visual navigation system. In such systems, data are input in a consecutive sequence. The classifier needs to be adjusted online with misclassified samples to achieve more accurate classification results. Recently, several successful approaches have been proposed to address the online learning issue of SVMs [21–33] which will be reviewed in the next section. However, very few work is concerned with deleting training samples effectively without influencing the final trained classifier. Efficient samples deletion is very important for online classification. For very large data sets, a very large amount of training samples will be used in order to get a good trained classifier, so if no samples are deleted then online adjustment of the classifier would be very difficult since all training samples will be needed to re-train the classifier.

In this paper, we propose an online CVM classification algorithm with adaptive MEB adjustment, based on an efficient redundant samples deletion technique. An advantage of our approach over the existing ones is that our online CVM algorithm, called OCVM, can be applied to deal with very large data sets efficiently, as shown by the experimental results on both synthetic and real-world data sets. Our OCVM consists of the following two steps:

- Off-line samples deletion: an upper bound is given of the distance between the center of the approximate MEB at each iteration and the accurate MEB of all the training samples and then used to identify training samples which definitely lie in the final computed MEB; such data samples are permanently deleted from the set of training samples to accelerate the speed of MEB computation.
- Online classifier adjustment: the selected training samples after the off-line samples deletion step, together with new coming misclassified samples, are used to compute the new classifier coefficients; then online updating of the classifier can be achieved since only very limited training samples are maintained in this process due to the efficient samples deletion.

Experimental results on both synthetic and real-world data sets have been presented to illustrate the validity and effectiveness of the proposed method.

The rest of the paper is organized as follows. A literature review on the existing online learning algorithms for SVMs is presented in Section 2. In Section 3, the classical CVM algorithm is briefly reviewed. The redundant samples deletion algorithm is described in Section 4, and the new OCVM algorithm is presented in Section 5. In Section 6, experiments are conducted on both synthetic and real-world data to illustrate the validity and effectiveness of the proposed method. Some concluding remarks are given in Section 7.

2. Literature review on existing online SVM algorithms

In this section, we briefly review the recent progress on the online learning issue of SVMs. Cheng and Shih [20] proposed an incremental training algorithm of SVMs by using active query. A subset of training samples is first selected by the K-Means clustering to compute the initial separating hyperplane. At each iteration, the classifier is updated based on the training samples selected by active query. The iteration stops until there are no unused informative training samples. Syed et al. [21] proposed a so-called support vector incremental algorithm where the SVM classifier is re-trained based on both the new samples and the support vectors from the trained SVM. Then the SVM classifier can be incrementally updated. Peng et al. [22] designed a questing process in which the machine quests the interrogative instance by the distance between the point and the hyperplane. Through this process, some samples are selected from the incremental training set, which are used, together with the support vectors from previous steps, to update the classifier. Bordes et al. [23] proposed a fast algorithm, called LaRank, for efficiently solving multi-class SVMs. Different from the traditional methods which rely on the full gradient, the LaRank algorithm computes and updates the gradient in a random pattern. Thus, the computational complexity can be greatly reduced, and online learning can be achieved.

In recent years, the stochastic gradient descent (SGD) method has been introduced to online learning of SVMs [24–27]. To solve the primal problem of SVMs, gradient-based methods compute the gradient using all the training samples, which is very expensive for very large data sets. Alternatively, SGD-based methods compute the gradient with respect to only one randomly selected sample. Therefore, these methods run significantly faster than the gradient descent methods if the number of training samples is very large, so online adjustment of the SVM classifier with newly arrived samples can be achieved. Although both the SGD-based methods and our proposed OCVM algorithm only utilize part of the training samples in online learning, they are quite different in motivation and algorithm design. Our OCVM algorithm aims to delete most redundant samples safely and permanently so that the adjustment of the classifier only involves a limited number of training samples. Moreover, our OCVM solves the dual problem using the sequential minimal optimization (SMO) method.

On the other hand, a series of online learning algorithms for SVMs have been proposed, which are mainly based on analyzing the change of the Karush–Kuhn–Tucker (KKT) conditions while updating the classifier. Note that the KKT conditions are the optimality criteria for the solution of the SVM quadratic programming problems. Cauwenberghs and Poggio [28] analyzed the change of the KKT conditions when one data point is added in or removed from the training set. Then a so-called bookkeeping step is used to compute the new coefficients of the classifier to achieve online updating. Agarwal et al. [29] introduced the concept of S-span and constructed a mechanism to remove/add support vectors. Through this mechanism, a new sample is first determined whether to be a support vector or not. If the new sample is a support vector, then a judgement has to be made to see if the support vector should be removed. The coefficients of the classifier are then adjusted by the method of [28]. Davy et al. [30] adopted a similar idea as in [28] to compute the change of the support vector decision function and designed an online novelty detection algorithm as well as an online abnormality detection algorithm for online updating of the classifier. Based on the traditional least square SVM model, Awad et al. [31] modified the SVM for multi-class classification. They utilized the change of the KKT conditions, in the case when a new sample is added, to