Available online at www.sciencedirect.com
ScienceDirect
Pattern Recognition Letters 28 (2007) 2088–2093
Pattern Recognition Letters
www.elsevier.com/locate/patrec

Reducing the number of sub-classifiers for pairwise multi-category support vector machines

Wang Ye *, Huang Shang-Teng

Department of Computer Science and Engineering, Shanghai JiaoTong University, China

Received 4 January 2005; received in revised form 1 June 2007
Available online 25 July 2007

Communicated by L. Bottou

Abstract

Among the SVM-based methods for multi-category classification, "1-a-r", pairwise and DAGSVM are most widely used. The deficiency of "1-a-r" is long training time and unclassifiable region; the deficiency of pairwise and DAGSVM is the redundancy of sub-classifiers. We propose an uncertainty sampling-based multi-category SVM in this paper. In the new method, some necessary sub-classifiers instead of all $N \times (N-1)/2$ sub-classifiers are selected to be trained and the uncertainty sampling strategy is used to decide which samples should be selected in each training round. This uncertainty sampling-based method is proved to be accurate and efficient by experimental results on the benchmark data.

Keywords: SVM; Multi-category classification; Pairwise; Uncertainty sampling

1. Introduction

The Support Vector Machine (SVM) is a powerful technique for classification. It classifies positive and negative samples by searching a hyperplane with the largest margin between them, so that better generalization performance and fewer training errors can be obtained. In this paper, we will discuss SVM for multi-category classification, which means the number of the categories is more than two.

Generally, the binary (two-category) SVM can be extended to multi-category case in two ways. The first way is considering all categories in one optimization problem. According to this way, a multi-category problem is formulated into one optimization equation, but there are too many parameters to adjust, so it is inefficient. The second way is constructing several binary sub-classifiers. In this way, multi-category problems are treated as a series of binary sub-problems, and many methods are developed based on this idea. Compared with the first, the second way is more widely used.

Although many methods of the second way are available, when the number of the categories or the size of each category is quite large, these methods are faced with a common problem, that is, it takes a very long time for all binary SVM sub-classifiers being trained. Targeted on this, we propose an uncertainty sampling-based multi-category SVM (abbreviated as US_MSVM) in this paper. Faster than "1-a-r" and pairwise, the new method has similar average accuracy with them. In each round of US_MSVM, samples of the two most indistinguishable categories are selected for the next training round. After a training round, the probabilities of positive samples (PPS) are used to decide which two categories are most indistinguishable.

The remaining of this paper is organized as follows: In Section 2, we briefly review the current research situation of multi-category SVM. The main idea of uncertainty sampling strategy will be introduced in Section 3. The new method, US_MSVM, is presented and analyzed in Section

* Corresponding author. Tel./fax: +86 21 5448 5310.
E-mail address: wangye34@126.com (W. Ye).

0167-8655/$ - see front matter © 2007 Elsevier B.V. All rights reserved.
doi:10.1016/j.patrec.2007.06.020

4. Experimental results of the performance comparison between the new method and pairwise classifier on the benchmark data are shown in Section 5. Conclusions are drawn in Section 6.

2. Multi-category SVM

The basic form of SVM is presented to solve the problem of two-category linearly separable cases (Vapnik, 1995). By using kernel functions and slack variables, SVM can be extended to solve problems of nonlinearly cases and non-separable cases. A multi-category problem can be converted into a series of two-category sub-problems. "1-a-r" (Bennett, 1999), pairwise ("1-a-1") (Kreßel, 1999), Decision Directed Acyclic Graph (DDAG) (Platt et al., 2000) and Adaptive Directed Acyclic Graph (ADAG) (Kijsirikul and Ussivakul, 2002) are all based on this idea.

The "1-a-r" method is used to combine N binary sub-classifiers, where N is the number of the categories. In the ith round of the training phase, samples of the ith category are labeled positive, and all others are labeled negative. The advantage of "1-a-r" is simple architecture and high testing speed, but it costs long time for training and the unclassifiable region is quite large (Shigeo, 2003).

The pairwise method is used to combine $N\times(N-1)/2$ binary sub-classifiers and each sub-classifier is trained on samples of two out of N categories. In the testing phase, the Max Wins algorithm is adopted, that is, the final result is the category gets more supports. According to Shigeo (2003), Abe and Inoue (2002), The pairwise classifier costs less training time and has smaller unclassifiable region than "1-a-r".

To solve the unclassifiable region problem in "1-a-r" and pairwise, Platt proposed the DDAG, which is a special pairwise classifier. The training phase of DDAG is the same as the pairwise method by solving $N\times(N-1)/2$ binary SVMs. In the testing phase, these SVMs are arranged in an N-layer DAG. Excluding impossible categories step by step, The DDAG labels a sample with its most possible category label at the bottom of the DAG, as is shown in Fig. 1.

Kijsirikul and Ussivakul proposed a tournament-based classifier: ADAG. The training phase of ADAG and DDAG are the same. In each testing round of the ADAG, the number of the candidate categories reduces by half. The final label is given after the last decision is made, as is shown in Fig. 1b. Pontil and Verri (1998) proposed another version of the ADAG.

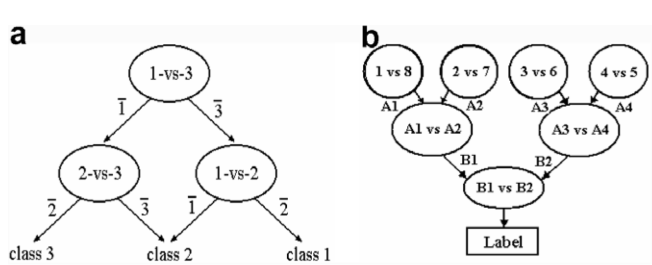

Fig. 1. The testing process of (a) DDAG and (b) ADAG.

Compared with the training time, the testing time of the Multi-category SVM can be ignored generally (Shigeo, 2003). As the training time as concerned, the complexity of pairwise is $2^{\gamma-1}cN^2-\gamma m^{\gamma}$ and the complexity of "1-a-r" is cNm^{γ} (Shigeo, 2003). Here, N is the number of categories and m is the number of all training samples and c is a constant. γ is equal to 2, when decomposition method is used to solve SVM (Shigeo, 2003). Clearly, the complexity of pairwise is lower than that of "1-a-r".

3. Uncertainty sampling

Before introducing our new method, we will review the uncertainty sampling strategy (Lewis and Gale, 1994) firstly. The uncertainty sampling strategy is an important sampling selecting strategy used in active learning. Active learning (Simon and Lea, 1974; Winston, 1975) is an efficient supervised learning algorithm that actively selects "helpful" samples to learn, instead of learning from the original training set passively. The uncertainty sampling strategy is used to select the "helpful" samples by measuring their uncertainty to the current classifier.

A typical active learning framework is described in (Tong, 2001). In active learning, the whole data are divided into labeled samples X and unlabeled samples U. There is also a learner l and a deciding module q. The learner l is trained on the labeled samples X and the module q is used to decide which samples of U should be selected and labeled, and should be added into X. The updated X will be used to train l in the next step. According to the difference mechanism of deciding modules, active learning methods can be divided into two groups: uncertainty sampling and query by committee (QBC) Seung et al. (1992).

The main idea of uncertainty sampling is that a classifier will benefit more from being trained on samples, which it is more uncertain to current classifier. Uncertainty sampling requires a probabilistic classifier that assigns to unlabeled samples each possible label with a certain probability. The unlabeled samples with most uncertainty are selected and labeled, and then are added into X. Various methods for measuring uncertainty have been proposed Lewis and Gale (1994), Iyengar et al. (2000). Query by committee is another group of active learning methods. It is based on the disagreement among a committee of classifiers.

Active learning is effective on saving labeled data and has been applied to various fields, such as natural language parsing, spoken language understanding, feature selection and text classification. In our method, we will use uncertainty sampling as a sample selecting strategy to decide which two categories are most indistinguishable.

4. Uncertainty sampling-based MSVM

As reviewed in Section 2, the sub-classifiers of all $N\times(N-1)/2$ pairs should be trained. Are these sub-classifiers