Physica A 509 (2018) 422–438

Contents lists available at ScienceDirect

Physica A

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

A biased least squares support vector machine based on Mahalanobis distance for PU learning

Ting Ke*, Hui Lv, Mingjing Sun, Lidong Zhang

Department of Mathematics, College of Science, Tianjin University of Science and Technology, Tianjin, 300457, China

HIGHLIGHTS

- We propose a new Mahalanobis distance-based least squares support vector machines (MD-BLSSVM) classifier, in which two Mahalanobis distances are constructed according to the covariance matrices of two class data for PU learning.
- Excellent kernel technique can be introduced to solve the linear non-separable problem in a reproducing kernel Hilbert space after making certain linear transformation ingeniously.
- MD-BLSSVM not only possesses faster learning speed, but also obtains better generalization than BLSSVMs and other methods.

ARTICLE INFO

Article history:
Received 29 January 2018
Received in revised form 11 May 2018
Available online xxxx

Keywords:
Positive and unlabeled learning
Least squares support vector machine
Mahalanobis distance
Regularization

ABSTRACT

In many domains, the presence of both positive and negative examples is not satisfied and only one class of examples is available. This special case of binary classification is called as PU (positive and unlabeled) learning in short. At present, many classification algorithms have been introduced for PU learning, such as BLSSVM, BSVM and so on. However, all of these classical approaches were measured by Euclidean distance, which did not take into account the correlative information of each class and the fluctuation of various attributions. In order to reflect this information, we propose a new Mahalanobis distance-based least squares support vector machines (MD-BLSSVM) classifier, in which two Mahalanobis distances are constructed according to the covariance matrices of two class data for optimizing the hyper-planes. Actually, MD-BLSSVM has a special case of BLSSVMs when the covariance matrices are degenerated to the identity matrix. The merits of MD-BLSSVM are (1) Mahalanobis distance of two classes can measure more suitable distance with certain weights on attributions; (2) Excellent kernel technique can be introduced in a reproducing kernel Hilbert space after making certain linear transformation ingeniously; (3) A solution is obtained simply by solving the system of linear equations. In all, MD-BLSSVM is appropriate to many real problems, especially for the case that the distribution and correlation of two classes of data are obviously different. The experimental results on several artificial and benchmark datasets indicate that MD-BLSSVM not only possess faster learning speed, but also obtains better generalization than BLSSVMs and other methods.

* Corresponding author.
E-mail address: kk.ting@163.com (T. Ke).

https://doi.org/10.1016/j.physa.2018.05.128
0378-4371/

1. Introduction

Training binary classifiers on positive and unlabeled data is referred to as PU learning [1]. PU learning has no training negative data because negative training data is either hard to obtain or even not available at all in some domains. Fig. 1 shows the distribution of PU data in two-dimensional space. In Fig. 1, each example has two features, $[x]_1$ and $[x]_2$. "Red $+$" represents positive data and "blue circle" represents unlabeled data, where "blue $\oplus$" and "blue $\ominus$" describe positive and negative data in unlabeled examples respectively. How to identify the label of unlabeled data (blue circle) or any new data point effectively is our goal in this paper.

To cope with this setting, a theoretical study of probably approximately correct (PAC) learning from positive and unlabeled examples was first done in [2]. The study concentrated on the computational complexity of learning and showed that function classes learnable under the statistical queries model. Recently, learning from positive examples was also studied theoretically in [3] within a Bayesian framework where the distribution of functions and examples were assumed known. Roughly speaking, most research works in this area are divided into three types of methods.

The first type is the two-step strategy, which selects possible negative or positive examples from unlabeled examples, and then builds classifiers using positive examples and negative examples. The popular used techniques for extracting possible negative or positive examples included spy, Rocchio and 1-DNF [4]. After extracting possible negative or positive examples, standard machine learning methods such as Naïve Bayes and SVM [5] are used to train classifiers. At present, the two-step strategy is a widely used method, such as S-EM [6], ROC-SVM [7], 1-DNF, PNLH [8], LCLC [9], Pulce [10,11] etc. Pulce took advantage of the intrinsic relationships between attribute values and exceeded the independence assumption made by Naïve Bayes. In fact, leverages on the statistical properties of the data to learn a distance metric employed during the classification task.

The second type of methods is the probability estimation approach [12,13]. In [12], it took each unlabeled example as both positive and negative example with weights pre-computed by an additional classifier for protein record identification. Luigi Cerulo et al. made use of the approach in [14] to reconstruct gene regulatory networks without negative examples.

The third type of methods is a one-step method. It includes one-class SVM [15,16], BSVM [17], WL [18,19], BLSSVM [20] LUHC [21]. One class SVM is effective only for the case where the number of positive examples is large enough to capture the characteristics of the positive class, and their performance would be rather poor when this number is very small [22]. BSVM is built by giving appropriate weights to the positive examples and unlabeled examples which are regarded as negative examples with noise, respectively. Experimental results indicated that the performance is better than most of two-step strategies. WL used Logistic Regression technique after weighting the negative class. BLSSVM utilized all training data to learn a biased least squares SVM classifier. It reduces to be more stable and effective than BSVM. Moreover, the time complexity of BLSSVM is lower than that of BSVM, where BLSSVM only needs to solve linear equations and BSVM is a quadratic programming. LUHC proposed a Laplacian unit hyper-plane classifier which adding a manifold regularizer to make the predicted labels and the initial labels sufficiently close on the labeled points.

As an excellent state-of-the-art tools for classical binary classifier in machine learning [23,24], support vector machines (SVMs) has already outperformed most of other learning algorithms. One of the important reasons for the successfulness of SVMs is the employment of kernel technique. However, the kernels used in SVMs are based on Euclidean distance. That is, SVMs assume data points are more likely distributed within a hyper-spherical region. However, data points in two classes are more likely distributed within two different hyper-ellipsoidal regions. For this more general case, Mahalanobis distance, which was introduced by P. C. Mahalanobis in 1936 [25], takes into account the correlations of the dataset and is scale-invariant, is a better choice [26]. Many efforts were made toward classifying data based on Mahalanobis distance in a reproducing kernel Hilbert space [27–31]. Correspondingly, as one of the most effective technique for PU learning, BLSSVM and BSVM are still measured by Euclidean distance. This is inappropriate for certain data distributions.

Therefore, a more suitable distance measure for different distribution of data points for PU learning is introduced in this paper. We present a Mahalanobis distance-based biased least squares support vector machine (MD-BLSSVM) classifier for PU learning. In MD-BLSSVM, a pair of Mahalanobis distance-based inner products in a reproducing kernel Hilbert space is first introduced according to the empirical covariance matrices of the two classes of data. In fact, MD-BLSSVM has a special case of BLSSVM when the covariance matrices of two classes of data in a reproducing kernel Hilbert space are both degenerated to the identity matrix. Compared to other Mahalanobis distance-based methods, the idea in this paper is not only simpler and more natural, but also considers the Mahalanobis distance-based kernels for the two classes of data respectively. MD-BLSSVM successfully inherits the merit of BLSSVM, i.e., fast learning speed and good adaptability and robustness for PU learning. In addition, MD-BLSSVM effectively combines the covariance matrix information of two classes of data in the prediction stage. As a simple illustration on the MD-BLSSVM significance, Fig. 2 shows the learning result of the linear MD-BLSSVM. It can be found that the linear MD-BLSSVM obtains the much better separating hyper-plane than the linear BLSSVM. Computational comparisons on BLSSVM, MD-BLSSVM in terms of generalization performance and learning speed have been made on several artificial and benchmark datasets, indicating MD-BLSSVM is not only fast, but also shows comparable generalization.

The rest of this paper is organized as follows. Section 2, briefly dwells on distance measures and previous PU algorithm, such as BLSSVM. The proposed MD-BLSSVM is presented in Section 3. Then in Section 4, experiments on both synthetic and real datasets are reported. Finally, we conclude this paper in Section 5.