Neurocomputing ∎ (∎∎∎∎) ∎∎∎–∎∎∎

Contents lists available at ScienceDirect

Neurocomputing

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

A dual-layer supervised Mahalanobis kernel for the classification of hyperspectral images

Li Li, Chao Sun*, Lianlei Lin, Junbao Li, Shouda Jiang

Department of Automatic Test and Control, Harbin Institute of Technology, Harbin, China

ARTICLE INFO

Article history:
Received 5 January 2016
Received in revised form
31 May 2016
Accepted 12 June 2016
Communicated by Jun Yu

Keywords:
Hyperspectral image classification
Mahalanobis kernel
Supervised learning
SVM

ABSTRACT

To address the drawback of traditional Mahalanobis distance metric learning (DML) methods that learn the matrix without considering the weights of each class, in this paper, a novel dual-layer supervised Mahalanobis kernel is proposed for the classification of hyperspectral images. By modifying the traditional unsupervised Mahalanobis kernel, a supervised Mahalanobis matrix that can include more relativity information of different types of real materials in hyperspectral images is learned to obtain a new kernel. The proposed Mahalanobis matrix is obtained in two steps. In step one, we learn the first traditional Mahalanobis matrix with all samples to map the raw data. In step two, based on the data mapped by the first matrix, we pick several hard-to-identify classes from all the classes and learn the second Mahalanobis matrix using only these data. Finally, by combining these two matrices, we construct a new form of the Mahalanobis kernel. Simulation experiments are conducted on three real hyperspectral data sets. We use SVM as the kernel-based classifier to classify the dimensionally reduced data and compare with several methods from various aspects. The results show that the proposed methods perform better than other unsupervised or single-layer DML methods in classifying the hard-to-identify classes, especially under an extreme condition.

1. Introduction

Hyperspectral images provide a precise representation of the earth's surface with high geometrical precision and a high level of thematic detail [1]. In practice, hyperspectral image classification is a problem wherein we have to process high-dimensional data with insufficient a priori information. This problem can also result in the curse of dimensionality, the so-called Hughes phenomenon, and can produce overfitting estimations [2]. Therefore, in general, the major tasks of hyperspectral image (HSI) classification are twofold: extracting essential features from enormous bands to reduce the dimension and designing suitable classifiers to obtain significant classification accuracies.

In the past decade, kernel-based classifiers have attracted increased attention because of their excellent performance when addressing high-dimensional data. Among them, the support vector machine (SVM) classifier is the most representative such classifier [3–5]. Many studies in the field of remote sensing have addressed the fact that SVM-based classifiers provide superior performance in hyperspectral data classification [6–10] compared with other popular classifiers such as the k-nearest neighbors (KNN) classifier and artificial neural networks (ANNs).

To improve the generalization capability of kernel-based learning machines (such as SVMs and SRCs [11]), different forms of optimized-kernel methods have been proposed. One form consists of constructing multi-kernel learning (MKL) methods [12–21]. However, the solution for identifying the weight of each kernel is always obtained by solving an optimization problem, such as a semi-definite program (SDP) [12] or semi-infinite linear program (SILP) problem [22], which are excessively computationally complex.

Another shortcoming is that they do not account for the statistical regularities of the specific classification task and are consequently susceptible to noise or irrelevant spectral features.

The other form is to construct a kernel that can increase the spatial resolution near the separating boundary surface; a representative family of such methods is data-dependent kernels [23,24], and another family of methods consists of adopting distance metric learning (DML) methods. In recent years, there has been a growing body of work in the field of metric learning on the development of adaptive similarity measures that learn the relevances of features with respect to a given classification task [25–27]. Substituting an unweighted measure with a learned similarity measure offers a straightforward means of improving the classification accuracy with similarity-based classifiers. In addition, the DML algorithms have already been extended with kernel tricks

* Corresponding author.
E-mail addresses: hitsc@163.com (C. Sun), linlianlei@hit.edu.cn (L. Lin).

http://dx.doi.org/10.1016/j.neucom.2016.06.039
0925-2312/© 2016 Published by Elsevier B.V.

[28–34]. Among these metric learning methods, the Mahalanobis metric learning method has been found to provide the best performance [35–39,31,40–42]. Its expression is given by Eq. (1), and the goal is to compute a linear projection matrix M. Applying A to each sample pair x_i,x_j induces a Mahalanobis distance parametrized by the PSD matrix $M = AA^T$:

$$d_M(x_i, x_j) = \sqrt{(A^T x_i - A^T x_j)^T (A^T x_i - A^T x_j)} = \sqrt{(x_i - x_j)^T A A^T (x_i - x_j)} \\ = \sqrt{(x_i - x_j)^T M (x_i - x_j)} \quad (1)$$

Beyond the full-rank Mahalanobis distances, increasingly more works propose techniques to learn low-rank Mahalanobis distances. A low-rank Mahalanobis metric learning method is used to project samples from the original-dimensional feature space into a lower dimensional feature space. Low-rank matrices are particularly well-suited for hyperspectral image classification problems because of the high dimensionality of hyperspectral data and the limited amounts of training data. Moreover, the reduced dimensionality reduces the computational time and memory requirements [43,44]. In [45], Gomez and Camps-Valls proposed the Mahalanobis kernel (MK) in the formulation of the SVM and applied it to the field of hyperspectral classification. In addition, in [46], Bue introduced several typical Mahalanobis methods into the SVM and evaluated the resulting performance.

However, a drawback of the normal Mahalanobis kernel is that such methods simply apply the same weight to all the samples to learn the Mahalanobis matrix and cannot learn a matrix to increase the distance to concentrate on the partial features of several special classes.

Recently, through various experiments, we discovered some phenomena that can be observed in different hyperspectral images. Specifically, in a given hyperspectral image, when we remove one or several labels, another class's classification accuracy *can greatly increase*, whereas the accuracy of the remaining classes *is hardly affected*. This means that, for a certain hyperspectral image, there are always several classes whose clusters are quite close to some classes but sufficiently far from other classes, which makes it such that the samples in these classes can have one or several incorrect labels easily attached to them but not to the remaining samples.

Thus, in this paper, considering this property of hyperspectral images, by analyzing the drawback of the current Mahalanobis kernel, we propose a novel dual-layer supervised Mahalanobis DML method to learn a new form of the Mahalanobis matrix M and to construct a new Mahalanobis kernel. In contrast to current methods, and going beyond the method proposed in [40], where the author proposed a weighted-NCA method in which a weighted matrix is generated to tune the weights of different classes but in which a quantization method is not proposed, using an indiscriminate Mahalanobis DML on the first layer, we extract more precise relevancies of features between different classes and determine the potentially hard-to-identify classes. We then re-cluster these data by learning a second Mahalanobis matrix. Considering computational efficiency, we adopt an LDA-based Mahalanobis DML algorithm to compute the transformation matrix. Numerical experiments are conducted on several real hyperspectral data sets. To evaluate the performance of our proposed methods, we compare our methods with some single- and state-of-the-art multi-kernel methods based on various aspects. The experiments show that, with the proposed methods, we can increase the partial difference between the special classes better than we can under the traditional Mahalanobis kernel while simultaneously having little negative effect on the remaining classes.

The remainder of this paper is organized as follows. In Section 3, we introduce the proposed methods in detail. Numerical experimental results on three different real-world hyperspectral data sets are presented in Section 3. Finally, some conclusions are drawn from the experiments, and suggestions for future work are given in Section 4.

2. Algorithm

2.1. Mahalanobis kernel for SVM

Given a labeled training data set $\{(x_1, y_1), \dots, (x_n, y_n)\}$, where $x_i \in \mathbb{R}^N$ and $y_i \in \{-1, +1\}$, and a nonlinear mapping $\phi(.)$, usually to a higher (possibly infinite) dimensional Hilbert space, $\phi: \mathbb{R}^N \to \mathcal{H}$, the SVM method solves

$$\min_{w,\zeta_i,b} \left\{ \frac{1}{2} \| w \|^2 + C \sum_i \zeta_i \right\} \quad (2)$$

with the constraint

$$y_i(\langle \phi(x_i, w) \rangle + b) \ge 1 - \zeta_i \quad \forall\, i = 1, \dots, n \\ \zeta_i \ge 0 \quad (3)$$

where w and b define a linear classifier in the feature space. The nonlinear mapping function ϕ is applied in accordance with Cover's theorem [47], which guarantees that the transformed samples are more likely to be linearly separable in the resulting feature space. The regularization parameter C controls the generalization capabilities of the classifier and must be selected by the user, and ζ_i are positive slack variables enabling permitted errors to be addressed.

It is worth noting that all ϕ mappings used in the SVM learning occur in the form of inner products in $\mathcal{H}$, which allows us to define a kernel function $K(x_i, x_j) = \langle \phi(x_i), \phi(y_i) \rangle$. Then, a nonlinear SVM can be constructed using only the kernel function without having to consider the mapping ϕ explicitly [48]. Finally, the decision function implemented by the classifier for any test vector x_* is given by $f(x_* = \mathrm{sgn}(\sum_{i=1}^{n} y_i \alpha_i K(x_i, x_j) + b))$, where b can be easily computed from the α_i that are neither 0 nor C.

Gomez and Camps-Valls [45] introduced the Mahalanobis kernel in the formulation of the SVM, which is based on an RBF kernel $K(x_i, x_j) = \exp(-1/2\sigma^2(\| x_i - x_j \|^2))$ and is defined as

$$K(x_i, x_j) = \exp\left(-\frac{1}{2\sigma^2}(x_i - x_j)^T M^{-1} (x_i - x_j) \right) \quad (4)$$

where M is the estimated Mahalanobis distance metric matrix computed using the available training data. Note that this constitutes a nonlinear generalization of any Mahalanobis distance metric through the use of the kernel method's framework.

2.2. Mahalanobis DML algorithm

2.2.1. Related work

Suppose we have two data points $x_1, x_2 \in \mathbb{R}^n$, and their Mahalanobis distance can be defined as:

$$d_A(x_1, x_2) = \sqrt{(x_1 - x_2)^T A (x_1 - x_2)} \quad (5)$$

where $A \in \mathbb{R}^{n \times n}$ is a positive semi-definite matrix which can be decomposed into $A = WW^T$ by performing an eigenvalue decomposition(EVD). So we can solve this problem by learning the matrix W and have the following equation:

$$d_A(x_1, x_2) = \sqrt{(x_1 - x_2)^T W W^T (x_1 - x_2)} \quad (6)$$

Following Xiang et al. [49], we solve the optimization problem