Neurocomputing 143 (2014) 208–221

Contents lists available at ScienceDirect

Neurocomputing

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

Max–min distance analysis by making a uniform distribution of class centers for dimensionality reduction

Guowan Shao *, Nong Sang

Science and Technology on Multi-spectral Information Processing Laboratory, School of Automation, Huazhong University of Science and Technology, Wuhan 430074, PR China

ARTICLE INFO

Article history:
Received 28 August 2013
Received in revised form
4 April 2014
Accepted 2 June 2014
Communicated by Xu Zhao
Available online 17 June 2014

Keywords:
Dimensionality reduction
Fisher's linear discriminant analysis
Max–min distance analysis

ABSTRACT

Max–min distance analysis (MMDA) for dimensionality reduction has been presented to guarantee class separation. However, class centers may be nonuniformly distributed and thus optimal classification accuracy may not be obtained. In this paper, we first give a novel method based on MMDA, called fractional-step max–min distance analysis (FMMDA), which relaxes max–min pairwise distances in fractional steps. The method can make a relatively uniform distribution of class centers and approximately maintain class separation. Then we present a more efficient method, called regularized max–min distance analysis (RMMDA), which achieves the same effect as FMMDA by integrating the Fisher criterion into MMDA. Moreover, we present the speedup and kernel versions of the methods to accelerate an optimization procedure and deal with the data distribution problem, respectively. Finally, we analyze the computational complexities of our methods. Empirical studies demonstrate that our methods can outperform or be comparable to some state-of-the-art discriminant analysis methods in terms of classification accuracy.

1. Introduction

Dimensionality reduction (DR) is a fundamental issue in many pattern recognition and machine learning applications. In unsupervised dimensionality reduction, the essential goal is to preserve significant properties of data in a low-dimensional subspace [1–7]. In supervised dimensionality reduction, class labels are available to influence the choice of a subspace [8–15]. Moreover, in semi-supervised dimensionality reduction, both labeled and unlabeled data are used to determine a desirable subspace [16–19].

As a classical technique for dimensionality reduction, Fisher's linear discriminant analysis (FLDA) [8,9] is fast and easy as an optimal subspace can be determined with simple matrix computations. However, the between-class scatter matrix implicitly limits the maximum dimensionality of the subspace since the rank of the matrix is deficient. Namely, the dimensionality of the subspace is smaller than or equal to $c-1$, wherein c is the class number. In general, a larger dimensional subspace may further contribute to classification accuracy. As expected, some approaches can increase the dimensionality of the subspace. Heteroscedastic LDA (HLDA) [20], wherein LDA is linear discriminant analysis, was presented to deal with heteroscedastic data by simultaneously considering differences in class means and covariances. For HLDA, the rank of the generalized between-class scatter matrix is in general larger than $c-1$, thus the algorithm can obtain a larger dimensional subspace. Subclass discriminant analysis (SDA) [21] was proposed to address the problem that the data distribution is unknown. Since in general the rank of the between-subclass scatter matrix increases, SDA can get a subspace whose dimensionality is larger than $c-1$.

In addition to the limitation of the dimensionality, FLDA suffers from another related problem that close classes tend to merge in the subspace. It can be explained as follows: The goal of FLDA is to maximize the sum of squared distances between class means when the Fisher criterion is decomposed into pairwise Fisher criteria under certain assumption [22]. In the eigenvalue decomposition (for FLDA the solution is given by eigenvalue decomposition), inappropriate weights are implicitly given to different class pairs. In other words, far apart class pairs are overemphasized and close class pairs are ignored to a great degree. Hence, FLDA does not achieve optimal classification accuracy. Lotlikar and Kothari [23] developed fractional-step LDA (F-LDA) by iteratively applying LDA, in order to improve classification performance. In each step, close class pairs are indirectly endowed with large weights by gradually shortening distances in certain direction between data points. Loog et al. [22] proposed the approximate pairwise accuracy criterion (aPAC) to improve the classification performance of FLDA. The criterion is based on pairwise Fisher criteria and each two-class Fisher criterion is emphasized by a weighting function. Although the function is derived by approximating the

* Corresponding author. Tel.: +86 18 9732 79902.
E-mail addresses: gwshao_ezhou@163.com, 821211624@qq.com (G. Shao).

http://dx.doi.org/10.1016/j.neucom.2014.06.005
0925-2312/

two-class Bayes accuracy, the various approximations may deteriorate classification accuracy. Tao et al. [24] gave the geometric mean-based subspace selection (GMSS) to deal with the class separation problem. The algorithm is motivated by the observation that FLDA can be viewed as maximization of average Kullback–Leibler divergences between classes under certain assumption. Although GMSS adaptively emphasizes effects of divergences of close class pairs, it still cannot guarantee the separation of the class pairs.

Moreover, Sugiyama [13] developed local Fisher discriminant analysis (LFDA) to deal with multimodal data, which integrates local structure information into the reformulation of FLDA. The method can not only better preserve local structure in a subspace, but can also increase separability between different classes. Hamsici et al. [25] presented a Bayes optimal method, Bayes optimality in linear discriminant analysis (BLDA), to improve classification performance. However, it complies with the Bayes criterion only for a one-dimensional subspace, and the greedy approach is adopted for other dimensions, thus resulting in large computational load.

Recently, Bian and Tao [26] defined a novel discriminative dimensionality reduction criterion, max–min distance analysis (MMDA). The criterion can guarantee separation of class pairs since the minimum pairwise distance is maximized in a subspace under the homoscedastic Gaussian assumption. However, it may not guarantee a uniform distribution of class centers and thus *many* class pairs may not be emphasized duly. For MMDA, the max–min pairwise distance absolutely determines the distribution of class centers. However, the criterion is clearly different from the uniform distribution. Thus, according to MMDA, class centers may be nonuniformly distributed and *many* relatively close class pairs may occur. It is to say that despite excellent separability, the class pairs are not properly emphasized. Therefore, MMDA may not provide optimal classification accuracy.

In this paper, we first present fractional-step max–min distance analysis (FMMDA) to obtain high classification accuracy. To this end, we relax max–min pairwise distances in fractional steps on the basis of MMDA. The goal of relaxing the distances is to get a relatively uniform distribution of class centers. Moreover, the fractional-step strategy aims to maintain excellent separability between class centers and the same effect can be obtained with a small relaxation in each step. Thus class centers are uniformly distributed to a large extent and separation of class pairs is approximately preserved. As a result, almost all class pairs are duly emphasized.

Although FMMDA achieves better classification performance, it performs an optimization in fractional steps. Thus, the computational cost may be high. We then give an efficient method, regularized max–min distance analysis (RMMDA), to overcome the drawback, and the method uses the Fisher criterion to regularize MMDA. We are primarily motivated by the observation that close class pairs under the max–min criterion may become more separable under the Fisher criterion. By adequate regularization, class centers can be distributed fairly uniformly and separation of class pairs is approximately preserved. As a result, RMMDA achieves high classification performance without an iterative procedure.

In addition, we give the speedup versions of FMMDA and RMMDA. The versions can greatly accelerate an optimization procedure in case the dimensionality is much larger than the sample size or the class number. Meanwhile, we give the kernel versions of the two methods to solve the problem that the data does not meet the homoscedastic Gaussian assumption. We evaluate the proposed methods on five types of data sets and experimental results show that the methods can perform better than or be comparable to some top-level discriminant analysis methods.

This paper is organized as follows. In Section 2, we briefly review FLDA and MMDA. In Section 3, we present our FMMDA method. The RMMDA method is introduced in Section 4. Then the speedup versions of our methods are presented in Section 5. We give the kernel versions of the methods in Section 6. Section 7 presents the computational complexity analysis. The experiments are given in Section 8. Finally, we give the conclusions and the discussion in Section 9.

2. Related work

As mentioned in Section 1, our methods are based on MMDA. On the other hand, RMMDA is a combination of FLDA and MMDA. Therefore, we provide a brief description of FLDA and MMDA below.

2.1. Fisher's linear discriminant analysis (FLDA)

It is well known that Fisher's linear discriminant analysis (FLDA) is a popular technique for supervised dimensionality reduction. Here, we first briefly review its definition and then reformulate it to better validate the proposed RMMDA method later.

Let there be given a data set including N samples $\mathbf{X}=[\mathbf{x}_i]_{i=1}^N$, where $\mathbf{x}_i \in \mathbb{R}^m$ is the ith labeled training sample in an m-dimensional space. The samples are from c classes and for class ω_k, $k=1,2,\ldots,c$, there are N_k samples in $\mathbf{X}$. Let μ_k denote the mean of class ω_k and let p_k be the prior probability of class ω_k. The between-class scatter matrix $\mathbf{S}_b$ and the within-class scatter matrix $\mathbf{S}_w$ are given by

$$\mathbf{S}_b=\sum_{k=1}^{c} p_k(\mu_k-\mu)(\mu_k-\mu)^T \quad \text{and}$$

$$\mathbf{S}_w=\frac{1}{N}\sum_{k=1}^{c}\sum_{l=1}^{N_k} p_k(\mathbf{x}_l^k-\mu_k)(\mathbf{x}_l^k-\mu_k)^T,$$

where μ is the mean of all the samples, i.e., $\mu=(1/N)\sum_{i=1}^{N}\mathbf{x}_i$, and $\mathbf{x}_l^k$ is the lth sample from class ω_k. Moreover, the mean μ_k is estimated as $\mu_k=(1/N_k)\sum_{l=1}^{N_k}\mathbf{x}_l^k$.

According to the definition of FLDA [9], it seeks a projection matrix $\mathbf{W}\in\mathbb{R}^{m\times d}$ which optimizes the following trace ratio problem:

$$\max_{\mathbf{W}}\frac{\mathrm{tr}(\mathbf{W}^T\mathbf{S}_b\mathbf{W})}{\mathrm{tr}(\mathbf{W}^T\mathbf{S}_w\mathbf{W})}. \tag{1}$$

Problem (1) does not have a close-form solution and it can be approximately simplified into the corresponding ratio trace problem:

$$\max_{\mathbf{W}}\mathrm{tr}((\mathbf{W}^T\mathbf{S}_w\mathbf{W})^{-1}(\mathbf{W}^T\mathbf{S}_b\mathbf{W})). \tag{2}$$

Problem (2) can be solved by the generalized eigenvalue decomposition method [27].

The between-class scatter matrix $\mathbf{S}_b$ can be reformulated as

$$\mathbf{S}_b=\sum_{k=1}^{c-1}\sum_{l=k+1}^{c} p_k p_l(\mu_k-\mu_l)(\mu_k-\mu_l)^T.$$

Accordingly, problem (2) can be transformed into the following form:

$$\max_{\mathbf{W}}\sum_{k=1}^{c-1}\sum_{l=k+1}^{c} p_k p_l\,\mathrm{tr}((\mathbf{W}^T\mathbf{S}_w\mathbf{W})^{-1}(\mathbf{W}^T\mathbf{S}_{kl}\mathbf{W})), \tag{3}$$

where $\mathbf{S}_{kl}=(\mu_k-\mu_l)(\mu_k-\mu_l)^T$. The form shows that the Fisher criterion is equivalent to a combination of two-class Fisher criteria [22].

Suppose that the given data set is obtained by sampling from c conditional Gaussian distributions with an equal covariance and different means. The distributions are denoted by $p(\mathbf{x}\mid\omega_k)=$