Neurocomputing 129 (2014) 289–297

Contents lists available at ScienceDirect

Neurocomputing

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

Semi-supervised discriminative common vector method for computer vision applications

Hakan Cevikalp *

Electrical and Electronics Engineering Department of Eskisehir Osmangazi University, Machine Learning and Computer Vision Laboratory, Meselik 26480 Eskisehir, Turkey

A R T I C L E I N F O

Article history:
Received 13 April 2013
Received in revised form
15 August 2013
Accepted 2 September 2013
Communicated by M. Wang
Available online 30 October 2013

Keywords:
Semi-supervised distance learning
Classification
Clustering
Image retrieval
Similarity constraints
Discriminative common vector

A B S T R A C T

We introduce a new algorithm for distance metric learning which uses pairwise similarity (equivalence) and dissimilarity constraints. The method is adapted to the high-dimensional feature spaces that occur in many computer vision applications. It first projects the data onto the subspace orthogonal to the linear span of the difference vectors of similar sample pairs. Similar samples thus have identical projections, i.e., the distance between the two elements of each similar sample pair becomes zero in the projected space. In the projected space we find a linear embedding that maximizes the scatter of the dissimilar sample pairs. This corresponds to a pseudo-metric characterized by a positive semi-definite matrix in the original input space. We also kernelize the method and show that this allows it to handle cases with low-dimensional input spaces and large numbers of similarity constraints. Despite the method's simplicity, experiments on synthetic problems and on real-world image retrieval, visual object classification, gender classification and image segmentation ones demonstrate its effectiveness, yielding significant improvements over the existing distance metric learning methods.

1. Introduction

In a wide range of computer vision problems, including object classification, segmentation and image and video retrieval, the performance depends critically on the similarity metric used to compare examples, so it is important to develop effective methods for learning distance metrics for such applications [1–5]. Measuring distances is comparatively simple when the features used are hand-chosen to be independent and highly relevant, but in computer vision applications with modern feature sets there are typically a great many features, many of which are either highly correlated with other features or irrelevant for the task being considered. This happens because at present, despite their redundancy, large and comparatively generic modern feature sets typically give better performance than smaller hand-chosen ones, and because vision problems are often somewhat open with the most relevant features depending on the exact problem and dataset being considered. (For example, when organizing image collections, it is possible to group images in many different ways, based on objects that they contain, natural versus built, outdoors versus indoors, *etc.*) When there are irrelevant and/or correlated features, similarity judgments based on Euclidean feature space distances often give unacceptable results, so it is necessary to learn more discriminative distance metrics.

In the simplest forms of distance learning, explicit class labels are supplied for the training samples, thus establishing a global notion of the similarity that is to be learned. However, there are many applications in which explicit labels are not available, either because the underlying problem does not involve classes or involves only poorly defined ones, or owing to the high cost of supplying a full labeling. In such cases, side information in the form of categorical similar/dissimilar judgments linking pairs of examples may still be available at a reasonable cost. For example in surveillance applications such as [4], objects (*e.g.* faces) extracted at roughly the same location in successive video frames can be assumed to represent the same individual, whereas ones extracted at different locations in the same frame must represent different individuals. In some applications such as relevance-feedback based image retrieval [6] or interactive semi-supervised image segmentation [2], such similarity judgments are actually the most natural form of supervision.

This paper focuses on distance metric learning from similarity judgments of this kind. Our strategy is to handle the similarity constraints first by projecting the data to a lower-dimensional subspace in which each similar pair becomes an identical pair, and then to address the dissimilarity constraints by finding a linear embedding that maximizes the distances between the projected dissimilar pairs. There are several advantages of this procedure: Projection onto the null space is the optimal linear projection in

* Tel.: +90 5423501529.
E-mail address: hakan.cevikalp@gmail.com

0925-2312/$ - see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.neucom.2013.09.029

the sense that it preserves the variance along the orthogonal directions to the projection direction, hence the original distance measure is best preserved. Moreover, as the experiments show, the resulting method is particularly suitable for computer vision problems based on modern high-dimensional feature sets since one does not need to approximate complex distance model parameters.

2. Related work

In recent years there has been a growing interest in methods for learning distance metrics due to their broad applications. Some of these approaches find the desired distance function directly, while others find embeddings in which the Euclidean distance serves as the new distance function. The two problems are equivalent and we will present them interchangeably here. We only discuss methods based on similarity judgments: ones that require explicit class labels [7–10] will not be considered here. A more comprehensive survey of distance metric learning techniques can be found in [11].

Similarity judgment based distance learning methods modify their input distances to accommodate the given pairwise constraints, and at present most of them focus on learning linear Mahalanobis-like distances parameterized by positive-definite or semi-definite matrices. Xing et al. [12] used a convex programming formulation under equivalence constraints to learn a full-rank Mahalanobis metric. The metric is learned via an iterative procedure that involves projection and eigendecomposition in each step. Tsang and Kwok [13] formulated the problem as a quadratic optimization one. They also extend their method to the nonlinear case using the kernel trick. Shalew-Shwartz et al. [14] proposed a sophisticated online distance metric learning algorithm that uses side information. The method incorporates the large margin concept, and the distance metric is modified based on two successive projections involving an eigendecomposition. Davis et al. [15] proposed an information-theoretic approach to learn a Mahalanobis distance function. They formulated the metric learning problem as that of minimizing the differential entropy between two multivariate Gaussians under equivalence constraints on the distance function. Yang et al. [11] proposed a Bayesian framework that estimates a posterior distribution for the distance metric from the pairwise constraints. All of the above algorithms attempt to learn full-rank distance metrics. This makes them less suitable for high-dimensional computer vision problems, in which it is usually more effective to learn lower-rank distance metrics or embeddings. To this end, Cevikalp and Paredes [2] introduce a low-rank distance metric learning algorithm based on sigmoid functions. A similar weakly supervised method was introduced in [16]. A semi-supervised low-rank Mahalanobis distance learning algorithm for high-dimensional spaces using log-determinant matrix divergence was introduced in [17]. More recently, a sparse (low-rank) metric learning method using Nesterov's smooth optimization has been proposed for high-dimensional data [18]. Unlike other methods, the sparse metric learning algorithm uses relative comparisons (given in terms of triplets) instead of pairwise equivalence constraints and the authors showed that it outperforms competing methods. However, they reported results on relatively small-dimensional datasets selected from UCI repository rather than on challenging high-dimensional real-world datasets. Another sparse metric learning method using alternating linearization optimization has been proposed in [19].

Graph-based methods that incorporate pairwise side information by modifying the weights of the graph have also been proposed [20–23]. Their major limitation is that they assume that local nearest-neighbor samples typically have the same class label (*c.f.* local neighborhood-based nonlinear dimensionality reduction methods in which each class is modeled as a manifold that is locally close to linear). This is only true if the classes are sampled densely relative to the inter-class spacing, which is hard to achieve with feasible training set sizes in high dimensional problems with difficult-to-distinguish classes. As a result, the graph-based approaches tend to perform poorly in practical vision problems because the constraints that they assume become too noisy. To alleviate this problem, some authors [24,6] use multiple graphs which operate on different feature sets. Then, they learn more reliable distance metrics by fusing those graphs with different techniques.

A method that is more closely related to ours is relevant components analysis (RCA) [25]. It searches for an embedding that assigns large weights to the most relevant dimensions and lower weights to less relevant ones, where relevance is estimated using the pairwise similarity constraints. RCA does not incorporate dissimilarity constraints and it is restricted to learning linear transformations in the original input space. Tsang et al. [26] improved RCA and kernelized it. Hoi et al. [5] proposed discriminative components analysis (DCA), a method that allows dissimilarity constraints to be incorporated into RCA and Kernel RCA. Our approach is similar in spirit to DCA, but it overcomes a serious drawback of DCA (see Section 2.3).

Finally, there are some hybrid methods that unify clustering and metric learning into a common framework based on side information [27,28]. Among these, [28] is worth mentioning because it projects onto the null space of the similarity constraints as we do.

2.1. Metric learning under side constraints

Before presenting our method in its linear and kernelized forms, we summarize the setting for distance metric learning under side constraints and sketch the RCA and DCA approaches.

Let $\mathbf{x}_i \in \mathbb{R}^d$, $i = 1, \ldots, N$, denote the samples of the training set. We are given a set of side constraints in the form of similar and dissimilar pairs and we aim to find a pseudo-metric that reflects the underlying relationships imposed by them. We focus on pseudo-metrics of the form

$$d_{\mathbf{A}}(\mathbf{x}_i, \mathbf{x}_j) = \|\mathbf{x}_i - \mathbf{x}_j\|_{\mathbf{A}} = \sqrt{(\mathbf{x}_i - \mathbf{x}_j)^\top \mathbf{A} (\mathbf{x}_i - \mathbf{x}_j)}, \tag{1}$$

where $\mathbf{A} \geq \mathbf{0}$ is a symmetric positive semi-definite matrix. Equivalently, if $q = \text{Rank}(\mathbf{A}) \leq d$, $\mathbf{A}$ can be written in the form $\mathbf{A} = \mathbf{W}\mathbf{W}^\top$ where $\mathbf{W}$ is a full-rank rectangular matrix of size $d \times q$, so that

$$\|\mathbf{x}_i - \mathbf{x}_j\|^2_{\mathbf{A}} = \|\mathbf{W}^\top \mathbf{x}_i - \mathbf{W}^\top \mathbf{x}_j\|^2, \tag{2}$$

i.e. distances between points under the metric $\mathbf{A}$ are equivalent to Euclidean distances on their linear projections by $\mathbf{W}^\top$.

2.2. Relevant component analysis (RCA)

The basic strategy of RCA is to identify irrelevant dimensions and reduce their effects by assigning lower weights to them. RCA does not exploit dissimilarity information. Similarity information is provided in the form of "chunklets": groups of two or more data samples that are considered "similar" (*e.g.* that belong to the same class). Assume that we are given C chunklets with chunklet c containing n_c patterns $\{\mathbf{x}_{c,1}, \ldots, \mathbf{x}_{c,n_c}\}$. RCA centers each chunklet then finds their combined covariance matrix:

$$\mathbf{S} = \frac{1}{n} \sum_{c=1}^{C} \sum_{i=1}^{n_c} (\mathbf{x}_{c,i} - \boldsymbol{\mu}_c)(\mathbf{x}_{c,i} - \boldsymbol{\mu}_c)^\top, \tag{3}$$