Neurocomputing 161 (2015) 47–55

Contents lists available at ScienceDirect

Neurocomputing

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

Distance-based human action recognition using optimized class representations

Alexandros Iosifidis*, Anastasios Tefas, Ioannis Pitas

Department of Informatics, Aristotle University of Thessaloniki, Box 451, 54124 Thessaloniki, Greece

ARTICLE INFO

Article history:
Received 17 May 2014
Received in revised form
15 September 2014
Accepted 31 October 2014
Available online 16 March 2015

Keywords:
Distance-based classification
Optimized class representations

ABSTRACT

We study distance-based classification of human actions and introduce a new metric learning approach based on logistic discrimination for the determination of a low-dimensional feature space of increased discrimination power. We argue that for effective distance-based classification, both the optimal projection space and the optimal class representation should be determined. We qualitatively and quantitatively illustrate the superiority of the proposed approach to metric learning approaches employing the class mean for class representation. We also introduce extensions of the proposed metric learning approach to allow for richer class representations and to operate in arbitrary-dimensional Hilbert spaces for non-linear feature extraction and classification. Experimental results denote that the performance of the proposed distance-based classification schemes is comparable (or even better) to that of Support Vector Machine classifier (in both the linear and kernel cases) which is currently the standard choice for human action recognition.

1. Introduction

In this paper we focus on the problem of distance-based, multi-class classification of human actions and specifically on the Nearest Class Centroid (NCC) classification scheme that has been employed in many Computer Vision tasks, including image and action classification [22,21,35,26,5]. The success of NCC critically depends on the adopted distance function, which is usually learned by applying a learning process exploiting training samples. We follow this line of work and cast our classifier learning problem as the one of learning a low-rank Mahalanobis distance which is shared across all classes. Such a Mahalanobis distance can be used in order to map the samples to a low-dimensional feature space of increased discrimination power, where classification is performed by employing the minimal Euclidean distance from the class representation.

Typically, NCC classification schemes employ the class mean vector for class representation, assuming that the classes forming the classification problem follow unimodal probability distributions having the same covariance structure. However, this is a strong assumption, which is difficult to be met in real classification problems. Consider the example illustrated in Fig. 1. Fig. 1a illustrates two classes formed by 2D data following different probability distributions and having different covariance structures. Fig. 1b illustrates the projection space obtained by applying Linear Discriminant Analysis (LDA) [3] on the 2D data forming the two classes. As can be seen, LDA fails to determine a useful for classification subspace, since the two classes are mapped to the same region resulting to a classification rate equal to 46.45%. On the other hand, logistic discrimination is able to merely overcome these issues and increases class discrimination in the projection space, as illustrated in Fig. 1c, leading to a classification rate equal to 94.17%. Finally, logistic discrimination employing the class vectors denoted by triangles in Fig. 1a, d for class representation is able to perfectly discriminate the two classes in the projection space, leading to a classification rate equal to 100%. As can be seen in this, rather simple, example, critical role on the performance of NCC classifier plays, not only the adopted distance function, but also the adopted class representation.

In this paper, we propose a new metric learning algorithm based on multi-class logistic discrimination, where a sample is enforced to be closer to its class representation than to any other class representation in the projection space. The proposed algorithm determines both the optimal projection matrix and the optimal class representation that can be, subsequently, used for classification. In order to distinguish our approach from the NCC classifier, it is referred to as the Nearest Class Vector (NCV) classifier hereafter. In order to overcome the unimodality assumption that is inherently set by all the NCC, including the proposed NCV, classifiers, we introduce an extension, namely Nearest Subclass Vector (NSV) classifier, which exploits multiple representations per class. Finally, since kernel methods have been found to be very effective in many Computer

* Corresponding author.
E-mail addresses: aiosif@aiia.csd.auth.gr (A. Iosifidis), tefas@aiia.csd.auth.gr (A. Tefas), pitas@aiia.csd.auth.gr (I. Pitas).

http://dx.doi.org/10.1016/j.neucom.2014.10.088
0925-2312/

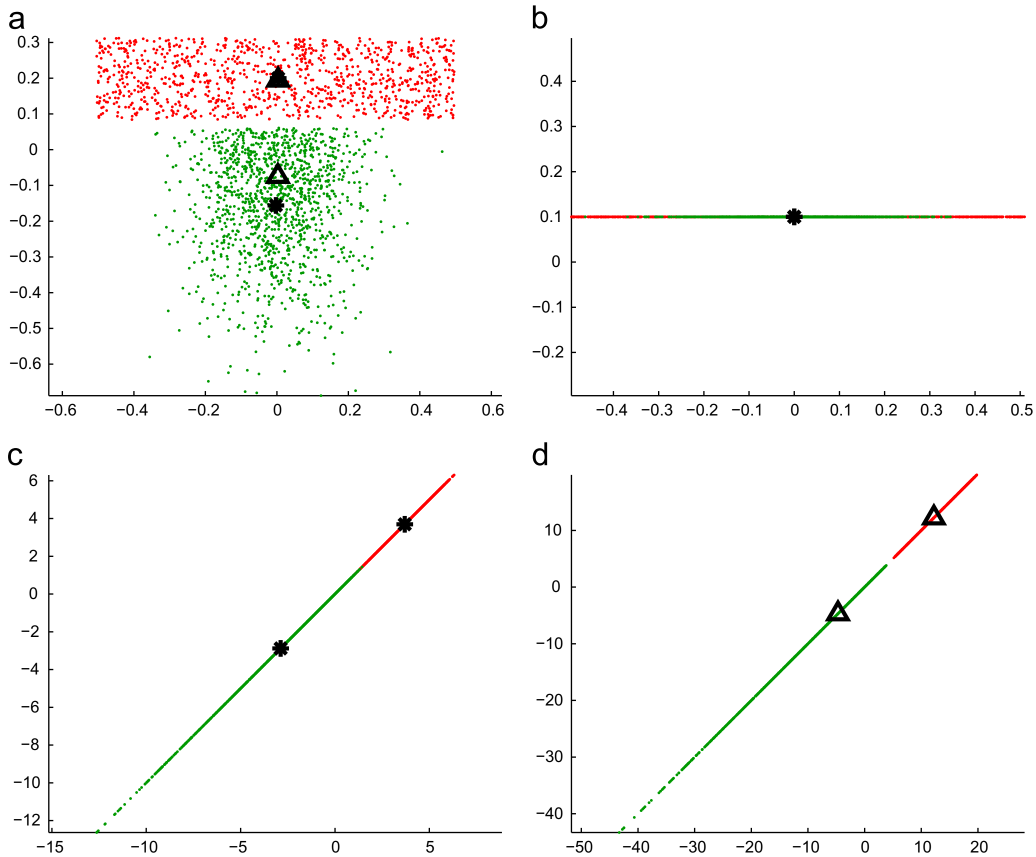

Fig. 1. (a) Two 2D classes having different probability class distributions and different covariance structures, (b) projection space of LDA, (c) projection space of logistic discrimination employing the class mean vectors (asterisks) for class representation and (d) projection space of the proposed logistic discrimination scheme employing the optimal class vectors (triangles).

Vision tasks, including human action recognition [31,30,29,18], we extend both the NCV and the NSV classifiers in order to determine an optimal data projection matrix and optimal class representations in arbitrary-dimensional Hilbert spaces [24].

We apply the proposed classification schemes on the Hollywood2 [19], Olympic Sports [23] and the, recently introduced, ASLAN [6] datasets. As baseline approaches, we use the state-of-the-art methods proposed in [30,6]: on the ASLAN dataset we employ a set of 12 similarity values calculated for histogram similarity measure between pairs of videos, represented by using the Bag of Words (BoW) model for HOG, HOF and HNF descriptors evaluated on STIP video locations [17]. This video pair similarity representation is employed for classification using a linear Support Vector Machine (SVM) classifier. We employ this baseline to evaluate the performance of the linear version of the proposed NCV and NSV classification schemes. For the remaining datasets we employ the BoW-based video representation by using HOG, HOF, MBH and Trajectory descriptors evaluated on the trajectories of densely sampled interest points [30]. Classification is performed by employing a kernel SVM classifier and the χ^2 kernel. We employ this baseline to evaluate the performance of the kernel version of the proposed NCV and NSV classification schemes.

The rest of the paper is structured as follows. We first discuss a selection of works related to this paper in Section 2. We describe the proposed metric learning algorithm for NCV classification in Section 3. Extensions towards two directions, in order to exploit multiple representations per class and in order to operate in arbitrary-dimensional Hilbert spaces, are presented in Sections 3.1 and 3.2, respectively. Experimental results on human action recognition are presented in Section 4. Finally, conclusions are drawn in Section 5.

2. Related work

Closely related to the proposed NCV classifier is the LDA algorithm and its variances [32,14]. LDA determines an optimal discriminant subspace by adopting the between-class to within-class scatter ratio. LDA assumes unimodal class probability distributions having the same covariance structure and employs the mean class vectors for class representation. As has been discussed above, these are two strong assumptions that are difficult to be met in several real-world classification problems. A variant of LDA that tries to determine the optimal class representation for LDA-based data projection (in the linear case) is proposed in [8]. This idea has also been extended for nonlinear data projection in [11]. Our approach differs significantly in that: (i) we employ multi-class logistic discrimination for the determination of the data projection matrix and the optimal class representation. As it will be shown in Section 4.4, the adoption of logistic discrimination leads to increased performance compared to the criterion used in [8]. (ii) We extend the proposed NCV classification in order to exploit multiple representations per class.

Other works related to this paper include LESS [27], Taxonomy Embedding [34], the Sift-bag kernel [38], NCC classifier [22] and sample-to-class metric learning [33]. The LESS model [27] is used to learn a diagonal scaling matrix for the modification of the Euclidean distance by scaling the data dimensions and includes an l_1 penalty term in order to perform feature selection. Taxonomy Embedding [34] exploits a hierarchical cost function in order to map the samples to a lower dimensional feature space where each class is represented by the class mean vector. The Sift-bag kernel [38] determines a lower dimensional feature space that is orthogonal to the subspace with the maximal within-class variance, which is