Neurocomputing 173 (2016) 137–144

Contents lists available at ScienceDirect

Neurocomputing

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

Constrained discriminant neighborhood embedding for high dimensional data feature extraction

Bo Li [a,b,c,*], Lei Lei [a,b], Xiao-Ping Zhang [c,d]

[a] School of Computer Science and Technology, Wuhan University of Science and Technology, Wuhan, Hubei 430065, China
[b] Hubei Province Key Laboratory of Intelligent Information Processing and Real-time Industrial System, Wuhan, Hubei 430065, China
[c] Department of Electrical and Computer Engineering, Ryerson University, Toronto, Ontario, Canada M5B 2K3
[d] School of Electronics and Information Engineering, Tongji University, Shanghai, 201804, China

ARTICLE INFO

Article history:
Received 30 June 2014
Received in revised form
3 October 2014
Accepted 1 January 2015
Available online 1 September 2015

Keywords:
Dimensionality reduction
Feature extraction
Local uncorrelation
Neighborhood embedding

ABSTRACT

When handling pattern classification problem such as face recognition and digital handwriting identification, image data is always represented to high dimensional vectors, from which discriminant features are extracted using dimensionality reduction methods. So in this paper, we present a supervised manifold learning based dimensionality reduction method named constrained discriminant neighborhood embedding (CDNE). In the proposed CDNE, on one hand, the class information of samples is taken into account to construct both an inter-class graph and an intra-class graph, where neighborhood points in the intra-class graph are selected from those with the same class and any point in the inter-class graph should sample those labeled different classes as its neighborhood points. On the other hand, locally least linear reconstruction technique is also introduced to model an objective function under the local uncorrelation constraint to explore a discriminant subspace. Compared to some related and state-of-the-art dimensionality reduction methods such as discriminant neighborhood embedding (DNE), supervised locality discriminant manifold learning (SLDML), discriminant sparse neighborhood preserving embedding (DSNPE), local graph embedding based on maximum margin criterion (LGE/MMC), uncorrelated discriminant locality preserving projection (UDLPP) and locally uncorrelated discriminant projection (LUDP), the proposed CDNE has been validated to be efficient and feasible by experimental results on some benchmark face data sets including CMU PIE, ORL and FERET.

1. Introduction

Dimensionality reduction is indispensable in many applications such as bioinformatics, pattern recognition, machine learning and data mining because it can provide a way to overcome the curse of dimensionality [1,2]. Moreover, dimensionality reduction also contributes to extract discriminant features from the original data. When carrying out the task of classification, objects expect to be described in terms of a set of measurable features, where the selection and quality of the features representing different patterns have a considerable impact on the final performance. So dimensionality reduction or feature extraction is employed to derive new features from the original data to reduce the cost of feature measurement and improve the classification accuracy with high efficiency. Until to now, many current techniques involved linear and nonlinear transformations are applied to all kind of data classification [3-15, 51].

In the last decade, as one kind nonlinear method, manifold learning has attracted more and more attentions, where locally linear embedding (LLE) as well as its supervised extensions has been widely used due to its high efficiency for dimensionality reduction [16,17]. In the original LLE, manifold local geometry is explored using k nearest neighbors (KNN) criterion which is determined by the ascending Euclidean distances. In some cases, some points with different labels may have shorter Euclidean distances than those with the same class and then are selected as neighbors, which results in performance degradation for data classification. Under such circumstances, some researchers have reported their attempts to address the problem. When constructing KNN graph, class information is taken into account to adjust weights between KNN neighbors or to select the corresponding neighbors. The work was first presented by de Ridder et al., where only the Euclidean distances between those labeled different classes are simply enlarged by adding a constant [18]. Instead of

* Corresponding author.
E-mail address: liberol@126.com (B. Li).

http://dx.doi.org/10.1016/j.neucom.2015.01.099
0925-2312/

enlarging between-class distances, Wen et al. utilized a nonlinear function to shrink within-class distances [19], which shows similar effect on recognition performance. These methods either enlarge between-class distances or shrink within-class distances. Then Zhang brought forward an enhanced supervised LLE by reducing within-class distance and expanding between-class distance simultaneously [20]. Later, Zhang and Zhao introduced a probability-based distance to adjust the Euclidean distance between labeled and unlabeled points, respectively [21]. Combining to class information, on one hand, the above methods try to improve the performance of LLE by changing the weights between KNN graph nodes; on the other hand, they paid no attention to neighborhood points' selection. So Hui et al. [22] and Zhao et al. [23] strictly required that only points with the same class can be considered to be neighbors, which represents the compactness of the same class data. Conversely, how to characterize the apartness of different labeled data is discarded. In order to avoid the problem, some ideas based on feature line have been proposed [10,50]. In Ref. [10], Lu et al. presented a work based on nearest feature line (NFL), where both a within-class nearest feature line space and a between-class nearest feature line space are formulated. In the between-class nearest feature line space, the label of any point must be different to those of points consisting of its nearest feature line, on the contrary, in the within-class nearest feature line space, any point and its nearest feature lines points should be sampled from the same class data. Moreover, a discriminant multi-manifold learning (DMML) algorithm model an intrinsic graph representing intra-class compactness and a penalty graph characterizing inter-class separability based on which a difference between the locality preserving projections associated to the corresponding graphs is reasoned [3]. Similarly to Lu, Yang et al. [24], and Chen et al. [25] also addressed an inter-class graph, an intra-class graph and the corresponding scatters with the least reconstruction trick, by which a Fisher criterion can be modeled, by which a subspace can be explored with the maximum inter-class scatter and the minimum intra-class scatter, simultaneously. Meanwhile, with such Fisher criterion, only locality of the intra-class data instead of all samples is preserved and small sample size (SSS) problem always appears.

In addition, some other supervised LLE algorithms combing to LDA have also been boomed. Based on the projection distances of the preprocessed points in LDA subspace, Pang et al. selected k minimum-distance points as the neighbor set of each point and then used LLE [26], which can be viewed as the mode of LDA+LLE. Zhang et al. presented a unified framework of LLE and LDA [27,28]. This framework essentially equals to LLE+LDA. Pang et al. also brought forward a model which is linearly constructed by the objective function of LLE and LDA with some constraints [29]. Furthermore, a local Fisher embedding (LFE) was put forward by de Ridder et al. in 2004 [30], where local geometry and global class information were integrated to a Fisher form. Li et al. also proposed a supervised LLE algorithm named local linear discriminant embedding (LLDE) based on the fact that the embedding cost function is invariant to translation, where it is optimized by a modified LDA [31]. In these methods, LDA is introduced to LLE where discriminant features are expected to be extracted.

It is well-known that out-of-sample always exits in the original manifold learning methods besides LLE. Some researchers have introduced a linear transformation into LLE to overcome the problem, i.e. neighborhood preserving embedding (NPE) [32] and neighborhood preserving projection (NPP) [33]. Kokiopouloua et al. also proposed a similar work named orthogonal neighborhood preserving projection (ONPP) [34]. Later, Kokiopouloua et al. defined a repulsion graph to extract supervised features, where an objective function is constructed to minimize the weighted difference of the reconstruction errors to distances between any two points with different labels in low dimensional space [35]. Zhang et al. also designed an intra-class graph and attempted to approach a subspace with the minimum weighted difference of the reconstruction errors in the intra-class graph to distances between any two differently labeled points [36].

The methods mentioned above aim to explore a subspace spanned by the orthogonal vectors. However, in pattern recognition, statistical uncorrelation is a favorable property for discriminant features because statistical uncorrelated features contain the minimum redundancy and are theoretical superior to those under orthogonal constraint [37–42]. So in this paper, a new dimensionality reduction method is proposed for feature extraction from the original high dimensional data, which is titled constrained discriminant neighborhood embedding (CDNE). In the proposed algorithm, firstly, the class information is taken into account to construct an intra-class graph and an inter-class graph. Based on both graphs, the intra-class scatter and the inter-class scatter, characterizing the compactness of intra-class data and the saparatity of the inter-class data, can be modeled with locally least linear reconstruction technique. Moreover, a new locally statistical uncorrelation is constrained to explore discriminatively uncorrelated features. At last, a linear transformation is also introduced. Thus the proposed method expects to find a subspace spanned by locally statistical uncorrelated vectors that can maximize the difference between the inter-class scatter to the intra-class scatter.

However, some other LLE based linear approximate methods, such as ONPP, NPE and NPP, introduced a linear transformation to overcome out-of-sample problem and did not take into account class information when constructing KNN graph. In other words, for any point, ONPP, NPE and NPP just select those with k bottom Euclidean distances to it as its k nearest neighbors. However, both Euclidean distance and label information are all considered in the proposed CDNE, where k inter-class nearest neighbors and k intra-class nearest neighbors are chosen to model the inter-class graph and intra-class graph, respectively. Thus the corresponding inter-class scatter and the intra-class scatter are all involved in exploring the linear embeddings. In addition, some methods contribute to extract discriminant features under statistical uncorrelated constraint and most of them constrained their objective functions with statistical uncorrelation globally instead of locally, which fail to explore the local structure of data manifold. So Chen et al. presented a locally uncorrelated discriminant projection (LUDP) method [43], where a locally uncorrelation based on heat kernel was introduced and the corresponding parameter was involved. Different to LUDP, CDNE presents a new locally statistical uncorrelated constraint, which is related to the intra-class scatter based on locally linear reconstruction.

The rest of the paper is organized as follows: Section 2 simply describes LLE. In Section 3, the inter-class scatter and the intra-class scatter are defined with the locally least linear reconstruction technique, and then the local statistical uncorrelation is newly deduced following the principle addressing the proposed method. Experimental results on some benchmark face data sets are offered in Section 4 and the paper is finished with some conclusions in Section 5.

2. Review of LLE

The goal of LLE is to map the high dimensional data into a low dimensional manifold space, where neighborhoods including neighbors and their reconstruction weights will be kept unchanged. The outline of LLE can be summarized as follows.

Step 1: For each data point X_i, identify its k nearest neighbors by KNN criterion;