Engineering Applications of Artificial Intelligence 75 (2018) 94–101

Contents lists available at ScienceDirect

Engineering Applications of Artificial Intelligence

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

A global manifold margin learning method for data feature extraction and classification

Bo Li [a,b,*], Wei Guo [a,b], Xiao-Long Zhang [a,b]

[a] School of Computer Science and Technology, Wuhan University of Science and Technology, Wuhan, 430065, PR China
[b] Hubei Province Key Laboratory of Intelligent Information Processing and Real-time Industrial System, Wuhan, 430065, PR China

ARTICLE INFO

Keywords:
Feature extraction
Supervised manifold learning
Manifold margin

ABSTRACT

This paper presents a global manifold margin learning approach for data feature extraction or dimensionality reduction, which is named locally linear representation manifold margin (LLRMM). Provided that points locating on one manifold are of the same class and those residing on the corresponding manifolds are varied labeled, LLRMM is desired to identify different manifolds, respectively. In the proposed LLRMM, it firstly constructs both a between-manifold graph and a within-manifold graph. In the between-manifold graph, for any point, its k nearest neighbors and itself must belong to different manifolds. However, any node and its neighborhood points should be on the same manifold in the within-manifold graph. Then we use the minimum locally linear representation trick to reconstruct any node with their corresponding k nearest neighbors in both graphs, from which a between-manifold graph scatter and a within-manifold graph scatter can be reasoned, followed by a novel global model of manifold margin. At last, a projection will be explored to map the original data into a low dimensional subspace with the maximum manifold margin. Experiments on some widely used face data sets including AR, CMU PIE, Yale, YaleB and LFW have been carried out, where the performance of the proposed LLRMM outperforms those of some other methods such as kernel principal component analysis (KPCA), non-parametric discriminant analysis (NDA), reconstructive discriminant analysis (RDA), discriminant multiple manifold learning (DMML) and large margin nearest neighbor (LMNN).

1. Introduction

For image pattern classification besides face recognition, it often rewrites the original data to high dimensional vectors. For instance, an appearance-based face image with size 80*80 can be transformed to a 6400-dimension vector. So it is required to extract discriminant features from the high dimensional vectors before making classification to them, which will contribute to improving recognition performance with low computational expense. Currently, researchers have reported many dimensionality reduction or feature extraction methods where both the linear and the nonlinear models are all involved (Wang et al., 2016; Yu et al., 2016; Sadatnejad and Ghidary, 2016; Motta et al., 2015). Moreover, they have been widely used in many applications with convincing performances (Zhang et al., 2016b, a; Sun et al., 2013; Zhang et al., 2016c).

As a traditional linear feature extraction method, principal component analysis (PCA) aims to locate a subspace where the covariance of all the original data can be maximized (Jolliffe, 2002). Meanwhile, it should be noted that no class information is taken into account in PCA, thus the supervised information do not play its role in the following feature extraction and classification (Yang and Zhang, 2008). However, another classical linear method, i.e. linear discriminant analysis (LDA), constructs an objective function by taking full consideration of data class labels (Kim et al., 2011; Martinez and Kak, 2001). In general, LDA projects the original data into a subspace with the maximum between-class apartness and the minimum within-class compactness.

The above mentioned methods concentrate on global linear structure of the original data and fail to dig nonlinear information hidden in them, which has been validated to be useful for dimensionality reduction (Tenenbaum et al., 2000; Roweis and Saul, 2000). Thus some nonlinear learning techniques are prevailing. As one kind of famous nonlinear models, neural networks have been attracting more and more attentions. In 1996, Huang (2004a, b) systematically concluded the theory of neural networks and their applications to pattern recognition. Moreover, neural networks have also been used to find polynomial roots (Huang, 2004a, b; Huang et al., 2005). On the basis of the traditional neural networks, radial basis probabilistic neural network (RBPNN) is constructed by a constructive hybrid strategy (Huang and Du, 2008). The

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

https://doi.org/10.1016/j.engappai.2018.08.004
Received 25 September 2017; Received in revised form 21 March 2018; Accepted 6 August 2018

proposed RBPNN has also been introduced for biometric identification (Shang et al., 2006; Zhao et al., 2004). However, neural networks optimize weights between nodes with so many iterations that much more computational cost will be paid when dealing with real world data.

In addition, kernel transformations, i.e. kernel principal component analysis (KPCA) (Scholkopf et al., 1998; Wen et al., 2012) and kernel Fisher discriminant analysis (KFDA) (Yang et al., 2004), are also presented to implicitly map observations into a space with high dimensionality, where they can be linearly classified. However, not local geometry but global structure information is approached from high-dimensional data using both KPCA and KFDA. Under such circumstances, manifold learning is put forward to explore manifold geometry hidden in the high dimensional data by locality learning.

It is well known that many manifold learning methods have been presented during last decade (Tenenbaum et al., 2000; Roweis and Saul, 2000; Saul and Roweis, 2003; Donoho and Grams, 2003; Belkin and Niyogi, 2003; Zhang and Zha, 2005; Weinberger and Saul, 2006; Lin and Zha, 2008). Among them, all the local patches on manifold are determined using k nearest neighbors or super-ball criterion, where linear tricks can be well performed to find the locality of manifold. Moreover, by taking advantage of data class information, some modifications are also made to them. For example, marginal Fisher analysis (MFA) (Xu et al., 2007; Yan et al., 2007) and discriminant multi-manifold learning (DMML) (Lu et al., 2013) construct two different graphs to represent the within-class compactness and the between-class separability, respectively. Additionally, by carrying out traditional LDA to all the local patches, local Fisher discriminant analysis (LFDA) (Sugiyama, 2006), non-parametric discriminant analysis (NDA) (Li et al., 2009) and reconstructive discriminant analysis (RDA) (Yang et al., 2008; Chen and Jin, 2012) maximize the trace ratio of the local inter-class graph scatter to the local intra-class graph scatter to find an optimal subspace. All the objective functions of these methods are under framework of trace ratio, which often incurs small sample size problem (Kim et al., 2011). To prevent the problem, locality sensitive discriminant analysis (LSDA) defines a local margin, which can be deduced to difference between the local inter-class graph scatter to the local intra-class graph scatter (Cai et al., 2007). However, it just pays more attentions to separateness between local patches. Another manifold learning based dimensionality reduction method titled local discriminant embedding (LDE) models two graphs based on data neighborhood and class relation, and then two graph Laplacians are used to find low-dimensional embeddings (Chen et al., 2005). Recently, a novel manifold method, i.e. t-distributed stochastic neighbor embedding (t-SNE), has also been focused on and high performances have been achieved by using it for feature extraction and pattern recognition (van der Maaten, 2014). However, both LDE and t-SNE show no concern on global manifold margin, which can characterize the total apartness of all the manifolds.

Thus how to globally measure all the manifolds' margin still needs further demonstration. In this paper, we will propose a globally defined manifold margin metric with locally linear representation strategy that can be introduced to measure apartness among all the manifolds. Based on the proposed manifold margin, a locally linear representation manifold margin (LLRMM) method will be put forward for multi-manifold identification. The main contributions of the proposed LLRMM are listed below:

(1) Although manifold margin is firstly proposed in the conference paper with a simply version (Li et al., 2015), in this paper, it is also presented to characterize the global apartness among different manifolds with a graphical illustration. Moreover, more details are offered either from the construction of three kind graphs and their corresponding scatters using the minimum linear representation trick or from theoretical derivations of the proposed manifold margin.

(2) Based on the proposed manifold margin, LLRMM is also put forward to extract discriminant features accompanying with its outline.

(3) Much more experiments on benchmark face data such as AR, CMU PIE, Yale, YaleB and LFW are carried out to obtain the statistics results including mean recognition rates and standard deviations, from which the performance of the proposed LLRMM can be validated.

The rest of the paper is organized as follows. The proposed algorithm is described and justified in Section 2. Section 3 presents the experimental results on face data accompanying with some discussions. At last, it draws some conclusions and makes some expectations for the future work in Section 4.

2. Locally linear representation manifold margin

2.1. Motivation

Recently, many supervised extensions have been made to LLE to deal with data classification problem. In these methods, some are proposed by combining LDA to LLE (Zhang et al., 2004, 2006; Pang et al., 2006; de Ridder et al., 2004; Li et al., 2008), some other take class information into account to direct the construction of local graph. However, when constructing k nearest neighbor graph, some distances between varied labeled nodes may be shorter than those between points sampled from the same class, which will lead to wrong neighborhood selection for discriminant analysis. In order to overcome the problem, some methods are put forward either by adjusting distances between nodes or by just selecting neighbors from nodes with the same class (de Ridder et al., 2003; Wen and Jiang, 2006; Zhang and Zhao, 2007; Zhao and Zhang, 2009; Hui and Wang, 2008; Zhao et al., 2005; Han et al., 2008). Assume that points locating on one manifold are of the same class and those residing on the corresponding manifolds are sampled from varied labeled data, these approaches are aiming to construct a k nearest neighbor graph to characterize the within-manifold data. However, they ignore to set up another k nearest neighbor graph which is composed of the between-manifold data. Thus both the within-manifold graph and the between-manifold graph will be constructed, by which a manifold margin metric can be globally proposed to quantify the apartness among different manifolds. At the same time, following the within-manifold graph and the between-manifold graph, a total-manifold graph will also be introduced to measure locality of all the samples without considering manifold label information. Compared to the original LLE (Lawrence, 2001), which is an unsupervised dimensionality reduction method, the proposed LLRMM takes manifold label information into account to construct a between-manifold graph and a within-manifold graph, respectively. In the between-manifold graph, any node and its k nearest neighbors must belonging to different manifolds. Thus the distances between multiple manifolds may exist in the between-manifold graph, which shows close relation to the expected global manifold margin.

Fig. 1 illustrates the proposed LLRMM method, where binary classification problem is involved. In Fig. 1, there are two differently labeled manifolds M1 and M2. For one point in a manifold M1, its four within-manifold nearest neighbors are selected to be composed of its within-manifold graph. Meanwhile, it also chooses other four nearest neighbors on another manifold to consist of its between-manifold graph. However, from the left sub-figure in Fig. 1, it can be found that two manifold data are mixed together and cannot be distinguished in the original high dimensional space. So in order to identify these two manifold data, it is expected to find a low dimensional subspace to maximize the manifold margin shown in right subfigure in Fig. 1.

But how to define the global manifold margin is still a problem. In the following, it will be reasoned from the within-manifold graph scatter and the between-manifold graph scatter using the minimum linear representation technique.

2.2. Locally linear representation weights

When constructing the within-manifold graph, both local geometry and manifold label information are all employed. On one hand, any node and its neighborhood should be on a manifold in the within-manifold graph. On the other hand, its local neighborhood should be composed