Information Sciences 429 (2018) 164–176

Contents lists available at ScienceDirect

Information Sciences

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

# Locality-regularized linear regression discriminant analysis for feature extraction

Pu Huang[a,b,*], Tao Li[a,b], Zhenqiu Shu[c], Guangwei Gao[d,e], Geng Yang[a,b], Chengshan Qian[f]

[a] School of Computer Science and Technology, Nanjing University of Posts and Telecommunications, Nanjing, 210023, China
[b] Jiangsu Key Laboratory of Big Data Security and Intelligent Processing, Nanjing University of Posts and Telecommunications, Nanjing, 210023, China
[c] School of Computer Engineering, Jiangsu University of Technology, Changzhou, 231001, China
[d] Institute of Advanced Technology, Nanjing University of Posts and Telecommunications, Nanjing, 210023, China
[e] Fujian Provincial Key Laboratory of Information Processing and Intelligent Control, Minjiang University, Fuzhou, 350121, China
[f] School of Computer & Software, Nanjing University of Information Science & Technology, Nanjing, 210044, China

A R T I C L E I N F O

*Article history:*
Received 8 September 2016
Revised 20 October 2017
Accepted 1 November 2017

*Keywords:*
Dimensionality reduction
Feature extraction
Face recognition
LLRC
Discriminant analysis

A B S T R A C T

Locality-regularized linear regression classification (LLRC) is an effective classifier that shows great potential for face recognition. However, the original feature space cannot guarantee the classification efficiency of LLRC. To alleviate this problem, we propose a novel dimensionality reduction method called locality-regularized linear regression discriminant analysis (LLRDA) for feature extraction. The proposed LLRDA is developed according to the decision rule of LLRC and seeks to generate a subspace that is discriminant for LLRC. Specifically, the intra-class and inter-class local reconstruction scatters are first defined to characterize the compactness and separability of samples, respectively. Then, the objective function for LLRDA is derived by maximizing the inter-class local reconstruction scatter and simultaneously minimizing the intra-class local reconstruction scatter. Extensive experimental results on CMU PIE, ORL, FERET, and Yale-B face databases validate the effectiveness of our proposed method.



## 1. Introduction

Feature extraction (or dimensionality reduction) has become an effective tool to aid understanding of the underlying structure of data in many real-world applications such as face recognition, image retrieval [37], fingerprint authentication [47], bioinformatics [45], and data mining [24]. The goal of feature extraction is to reduce the redundant, irrelevant, and noise information while preserving the intrinsic information contained in the data.

Over the past few decades, numerous dimensionality reduction (DR) methods have been developed for feature extraction. Principal component analysis (PCA) [33] and linear discriminant analysis (LDA) [1] are the two most representative DR algorithms. PCA aims to find a set of projection axes such that the data's variance is maximized after the data projection. PCA is an unsupervised approach as it ignores the useful prior class information. In contrast, LDA is a supervised algorithm. LDA

* Corresponding author at: School of Computer Science and Technology, Nanjing University of Posts and Telecommunications, Nanjing, 210023, China.
*E-mail address:* huangpu@njupt.edu.cn (P. Huang).

https://doi.org/10.1016/j.ins.2017.11.001

searches for a set of projection axes such that the samples from the same class are in close proximity and the samples from different classes are separated, simultaneously. LDA takes advantage of the class information to discover the discriminant structure of the data; thus, LDA is usually more effective than PCA for supervised pattern recognition tasks. However, LDA still has some limitations. For example, LDA usually suffers from small sample size (SSS) problems [29] (which lead to the singularity of the within-class scatter matrix), and LDA can only extract a maximum of $C - 1$ features (where $C$ is the number of sample classes).

In recent years, inspired by manifold learning, a family of locality-based DR algorithms has been proposed. Locality preserving projections (LPP) [12] is one of the most popular locality-based DR methods. Unlike PCA and LDA, which can only discover the global Euclidean structure of the data, LPP is able to determine the intrinsic manifold structure of the data. LPP seeks to find a set of projection directions such that the neighbourhood structure of the data is preserved after projection. To exploit the class information to improve the classification performance, some supervised versions of LPP have been proposed, such as class-wise locality preserving projection (CLPP) [23], supervised optimal locality preserving projection (SOLPP) [36], and local similarity preserving projections (LSPP) [16]. In addition, motivated by LDA and LLP, many discriminant locality-based DR algorithms [3,13–15,17,18,21,25,26,31,34,38,39,42,46] have been developed. These methods combine the ideas behind LDA and LPP, and aim to discover the local and discriminant structures of the data. Yan et al. proposed a general DR framework called graph embedding [39], which provides a powerful platform on which various DR algorithms can be developed. It has also been shown that all the aforementioned methods can be interpreted by the graph embedding framework.

After feature extraction, data classification is another important step in pattern recognition. Once the distinctive features are obtained using the DR algorithms, a suitable classifier can be chosen to classify these features. So far, many useful classification methods [5,7–11,22,35,41,49] have been developed, such as nearest neighbour classifier (NNC) [5], nearest feature line (NFL) classifier [22], and minimum distance classifier (MDC) [7]. Recently, representation-based classification methods have attracted considerable research attention and show great potential for pattern classification. Sparse representation-based classification (SRC) [20] is one of the most representative representation-based classification methods. SRC first codes a testing sample as a sparse linear combination of the training samples, and then classifies the testing sample by evaluating which class leads to the minimum reconstruction error. The SRC method has shown promising and robust results for face recognition as it can handle errors due to occlusion and corruption effectively. However, SRC is usually time-consuming owing to computation of the reconstruction coefficients by solving a $l_1$-norm minimization problem. Zhang et al. [48] analysed the working mechanism of SRC and pointed out that it is the collaborative representation (CR), and not the $l_1$-norm sparsity, that makes SRC powerful in terms of face recognition. To this end, they proposed a CR-based classification (CRC) method and demonstrated that CRC achieves highly competitive classification results with significantly less complexity than SRC. Motivated by this promising property of CRC, Yang et al. developed a novel DR algorithm called collaborative representation-based projection (CRP) [43], to preserve the collaborative representation relationships between samples. In addition to SRC and CRC, linear regression classification (LRC) [19] is a well-known representation-based classifier. LRC codes the testing sample as a linear combination of class-specific samples, and the decision is determined in favour of the class with the minimal reconstruction error. Compared to SRC, LRC spends less time in estimating the reconstruction coefficients using the least-squares method.

As mentioned above, both feature extraction and data classification are of extreme importance for pattern recognition. However, most DR algorithms (for feature extraction) were designed independently of the decision rules of the classifiers, and the classifier is generally selected heuristically (e.g., based on experience) at the stage of classification. The subspaces learned by DR algorithms generally have different characteristics that are invisible to the classifiers. Consequently, the classifiers selected after feature extraction cannot effectively make use of the characteristics of the learned subspace. As a result, the DR algorithms may not work well with the selected classifier, which often degrades the pattern recognition performance. To better connect the DR algorithm to one specific classifier, some researchers consider learning the DR algorithm according to the decision rule of the classifier. For example, Yang et al. proposed local mean-based nearest neighbour discriminant analysis (LM-NNDA) [44] and sparse representation classifier steered discriminative projection (SRC-DP) [40] according to the decision rules of local mean based nearest neighbour (LM-NN) classifier [27] and SRC, respectively. Inspired by Yang's work, Chen et al. used the decision rule of LRC to develop a novel DR algorithm named reconstructive discriminant analysis (RDA) [4]. These studies demonstrate remarkable results and validate the feasibility and effectiveness of using the classification rules to design various algorithms.

More recently, Brown et al. proposed a locality-regularized linear regression classification (LLRC) [2] method using a manifold learning procedure to expand on conventional LRC and increase the accuracy. However, the original feature space cannot guarantee the efficiency of LLRC. Motivated by the aforementioned methods [4,40,44], this paper proposes a novel DR algorithm, termed locality-regularized linear regression discriminant analysis (LLRDA), to generate an efficient feature subspace that is optimum for LLRC using the decision rule of LLRC. Concretely, the inter-class and the intra-class local reconstruction scatters are first constructed to characterize the separability and compactness of the samples, respectively; then, the feature extraction criterion is derived by maximizing the ratio between them. Since the decision rule of LLRC is used to steer the design of LLRDA, LLRDA is able to perform well with LLRC in pattern recognition.

The remainder of this paper is organized as follows. Section 2 briefly reviews LRC and LLRC. Section 3 details the proposed LLRDA method. Section 4 discusses the relationship between LLRDA and other techniques such as LRC and RDA.