Pattern Recognition 47 (2014) 1709–1720

Contents lists available at ScienceDirect

Pattern Recognition

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

Linear reconstruction measure steered nearest neighbor classification framework

Jian Zhang [a,b], Jian Yang [a,*]

[a] Department of Computer Science, Nanjing University of Science and Technology, Nanjing 210094, PR China
[b] Department of Computer Engineering, Huaihai Institute of Technology, Lian yungang 222000, PR China

ARTICLE INFO

Article history:
Received 6 September 2012
Received in revised form
22 August 2013
Accepted 16 October 2013
Available online 4 November 2013

Keywords:
Linear reconstruction measure (LRM)
Pattern classification
Classifier
Match learning
Face recognition

ABSTRACT

The linear reconstruction measure (LRM), which determines the nearest neighbors of the query sample in all known training samples by sorting the minimum L_2-norm error linear reconstruction coefficients, is introduced in this paper. The intuitive interpretation and mathematical proofs are presented to reveal the efficient working mechanism of LRM. Through analyzing the physical meaning of coefficients and regularization items, we find that LRM provides more useful information and advantages than the conventional similarity measure model which calculates the distance between two entities (i.e. conventional point-to-point, C-PtP). Inspired by the advantages of LRM, the linear reconstruction measure steered nearest neighbor classification framework (LRM-NNCF) is designed with eight classifiers according to different decision rules and models of LRM. Evaluation on several face databases and the experimental results demonstrate that these proposed classifiers can achieve greater performance than the C-PtP based 1-NNs and competitive recognition accuracy and robustness compared with the state-of-the-art classifiers.

1. Introduction

The study of similarity measure is fundamentally important in pattern recognition and computer vision [1–7,23,24]. In these areas, the function of the similarity measure is generally to determine the neighbor relationship among samples. Conventionally, the precise distance between two samples is calculated with various defined similarity (distance) functions, such as Euclidean distance, Manhattan distance, correlation coefficient, Chi-squared distance and so on. Then the neighbor relationship is established according to the sequence of these distances. This paper denotes process of finding the neighbor relationships above as the conventional point-to-point measure model (C-PtP).

It has a long history for the application of the C-PtP model in the classifier design. Since Cover and Hart laid the theoretical foundation of 1-NN and showed that the error rate of the 1-NN classifier, independent of the similarity (distance) functions used, is bounded above by twice the Bayes error rate in large sample cases in 1967 [7], many successful variants, such as LM-NN, NNL, NNP, NFL, NFS [8–19], were generalized to get better results. Among these classifiers, the C-PtP model plays a critical role which first calculates the distance between each training sample and the query sample respectively and then classifies the query sample into the class of the nearest of the training samples (prototypes).

Although the C-PtP is so simple to be easily achieved and has an ideal performance, it has many inherent drawbacks for classifier design. First, the C-PtP just involves the distance information between two entities with ignorance of other useful information including the training sample distribution information, the class member information, etc. This feature makes the C-PtP based classifiers difficult to deal with some special problems encountered in the classification, such as how to correctly classify the class-edge samples. Second, the C-PtP model with precise calculated distance is not robust enough for the noise and occlusion in practical applications.

In order to avoid these drawbacks, some approaches were proposed by incorporating some useful information into the C-PtP for improving the classification ability and robustness. These approaches can be roughly divided into two categories: (1) Combining or fusing feature extraction (weighting) with C-PtP. The well-known Eigenface and Fisherface are generally viewed as two typical methods with combining strategy, which first use the PCA or LDA to extract more representation or discriminative features and then perform the similarity measure with C-PtP [20]. In fact, all feature extraction (weighting) plus C-PtP methods can be treated as the combining strategy. In [21], Hastie and Tibshirani proposed a fusion strategy coined discriminant adaptive nearest neighbor (DANN) algorithm with a local linear discriminant analysis and the C-PtP. Following the idea underlying the

* Corresponding author.
E-mail addresses: zhangjianhhit@gmail.com (J. Zhang), csjyang@mail.njust.edu.cn (J. Yang).

0031-3203/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.patcog.2013.10.018

DANN algorithm, Domenicone et al. presented the ADAMENN algorithm [22]. Yang et al., in [23], advocated a new fusion strategy of connecting the discriminant analysis and the C-PtP; (2) Distance learning. The distance learning refers to the acquisition of a good metric with high classification accuracy among raw samples (prototypes). Mahalanobis distance [24] and whitened cosine distance [1,6,25], which usually learns some information from the covariance matrix of the known samples, are two typical similarity measures based on distant learning with C-PtP. In [3,26,27], several distance learning methods were also proposed to achieve good performance with C-PtP.

Different from above works focused on C-PtP, this paper takes our mind into another fresh model, namely linear reconstruction measure model (LRM), which takes the linear regression process as a tool for achieving the function of similarity measure in classifier design. Specifically, the LRM represents the query sample as a linear combination of all the training samples and determines the nearest neighbors of the query sample by sorting the minimum L_2-norm error linear reconstruction coefficients.

Historically, the LRM has been widely used for measuring the strength of relationship between samples [55,56]. Given a variable y and a number of variables $x_1, x_2, \ldots, x_n$ probably related to y, the LRM will quantify the strength of the relationship between y and x_i to assess the identity and degree of x_i related to y with the linear regression coefficients. For example, widely used in the function Magnetic resonance imaging (fMRI) analysis [57,58], the researches take the linear regression coefficients between the observation signal and some known signals to determine the exact known single possessing the stronger or weaker relationship with the observation signal.

Recently, the LRM shows great potential for the classifier design. An important relevant work was first presented in [28], where the sparse representation based classifier (SRC) is proposed and the nearest neighbors are obtained by a linear sparse reconstruction process. In contrast to the SRC with the L_1-optimizer, Shi et al. advocated the orthonormal L_2-norm method of estimating the linear regression coefficients with QR and determining the nearest neighbors with non-sparse L_2-optimizer in [29]. Zhang et al. proposed the collaborative representation classifier with regularized least square (CRC_RLS) in [30]. In CRC_RLS, the linear reconstruction model established in all training samples and solved by L_2-optimizer is also introduced to reconstruct the query sample. Unlike SRC and orthonormal L_2-norm method, the decision rule of CRC_RLS relies on the ratio of the reconstruction errors and the coefficients in each class. Xu et al. proposed a two-phase test sample sparse representation method (TPTSR) for face recognition in [53] and further presented a classification method named representation-contribution-based classification procedure (RCBCP) in [54]. TPTSR employed twice L_2-optimizer to approximately alternative L_1-optimizer and obtained a competed performance with smaller complexity. RCBCP directly used the construction contribution of single training sample to determine the nearest neighbors of the query sample.

In spite of the impressive performance and widespread concerns in statistics and pattern recognition, the theoretical foundation of the LRM is still not clear. What caused the linear reconstruction process to determine the true nearest neighbors of the query sample? How to obtain the nearest neighbors with the linear reconstruction process? Where does the striking discriminative power of this similarity measure model come from? These essential problems remain open.

In this paper, we intend to solve these problems and build a unified theoretical foundation for the impressive LRM model. First, we present a mathematical proof to explain the reasons and mechanism of obtaining the true same class neighbors for the query sample with coefficients. Second, the physical meaning of generalization items was explored. Finally, we perform the comparison and analysis between the LRM model and the C-PtP model for classification problem.

To verify our proofs and analysis, we further design the linear reconstruction measure based nearest neighbor classifier framework (LRM-NNCF) with the combination of a series of classifiers according to different decision rules and algorithms of LRM. While the proposed classification framework is generally of broad interest in pattern recognition, this paper will focus on the face recognition problem, an active research area in computer vision and pattern recognition driven by the broad applications [44]. We perform experiments on some benchmark face databases to demonstrate the performance of these classifiers. Experimental results show more desirable performance from LRM based classifiers than the C-PtP based ones.

It should be clarified that our proposed classifier framework is quite different from the linear subspace projection (or statistical) approaches using the linear regression, such as LDA, Rosenblatt Perceptron Machine, and SVM. These methods often build an explicit hyper-plane by the learning process with the linear regression and all training samples, and then make the decision according to the learned hyper-plane. However, our approaches still fall in the category of nearest neighbor based classifier (NN) without learning process. Another noted issue is that the proposed LRM_NNCF is independent of the feature representation or extraction. In other words, these approaches can work well in other projection subspaces.

The remainder of this paper is organized as follows. Section 2 presents the general and the regularized version of LRM. Section 3 reveals the physical meaning of the coefficients, the regularization items and the advantages of LRM over the C-PtP. Section 4 develops the linear reconstruction measure based nearest neighbor classification framework (LCM-NNCF). Section 5 conducts extensive experiments to verify the validity of our approaches. Section 6 offers our conclusions and future work.

2. Linear reconstruction measure

2.1. Basic ideas and general model of LRM

Given a query sample $\mathbf{y} \in R^n$ and a dictionary matrix $\mathbf{A} = [\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_N] \in R^{n \times N}$ formed by the known samples, where N is the number of training samples. For $\forall i, j, \mathbf{a}_i \neq \mathbf{a}_j$ when $i \neq j$. Our aim is to find the true nearest neighbors (defined as the ones which have the same class labels with the query sample) of $\mathbf{y}$ in $\mathbf{A}$.

Traditional algorithms accurately calculate each similarity indicator $d(\mathbf{y}, \mathbf{a}_i)$ between $\mathbf{y}$ and $\mathbf{a}_i$ respectively with a variety of defined similarity (distance) functions, such as Euclidean distance, Manhattan distance, correlation coefficient, Chi-squared distance, etc. Then the nearest neighbors of $\mathbf{y}$ were obtained by sorting them.

This paper formulates the problem as a general linear reconstruction process with minimum L_2-norm loss expressed as

$$\min \ \|\mathbf{y} - \mathbf{A}\mathbf{w}\|_2^2 \tag{1}$$

where $w = (w_1, w_2, \ldots, w_N)$ denotes the linear representation coefficients, and $\mathbf{y}$ and each $\mathbf{a}_i$ should be normalized. Then we determine the nearest neighbors by sorting the indicators $\|w_i^*\|$ which were obtained by solving Eq. (1). We refer to this fresh similarity measure method for searching nearest neighbors as linear reconstruction measure (LRM) with Eq. (1) as the general model of LRM.

2.2. Regularized LRM

As mentioned in Section 2.1, LRM can be achieved by solving a linear regression equation with minimum error, i.e. Eq. (1). However, Eq. (1) is often practically be fitted with minimizing a penalized version of the coefficients in some norms (regularization