J. Vis. Commun. Image R. 25 (2014) 1676–1685

Contents lists available at ScienceDirect

J. Vis. Commun. Image R.

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

Robust locality preserving projection based on maximum correntropy criterion

Fujin Zhong [a,b], Defang Li [a], Jiashu Zhang [a,*]

[a] Sichuan Province Key Lab of Signal and Information Processing, Southwest Jiaotong University, Chengdu 610031, PR China
[b] School of Computer & Information Engineering, Yibin University, Yibin 644000, PR China

ARTICLE INFO

Article history:
Received 12 October 2013
Accepted 14 August 2014
Available online 26 August 2014

Keywords:
Locality preserving projections
Outliers
Robustness
Half-quadratic optimization
Correntropy
Maximum correntropy criterion
Optimization
Dimensionality reduction

ABSTRACT

Conventional local preserving projection (LPP) is sensitive to outliers because its objective function is based on the L2-norm distance criterion and suffers from the small sample size (SSS) problem. To improve the robustness of LPP against outliers, LPP-L1 uses L1-norm distance metric. However, LPP-L1 does not work ideally when there are larger outliers. We propose a more robust version of LPP, called LPP-MCC, which formulates the objective problem based on maximum correntropy criterion (MCC). The objective problem is efficiently solved via a half-quadratic optimization procedure and the complicated non-linear optimization procedure can thereby be reduced to a simple quadratic optimization at each iteration. Moreover, LPP-MCC avoids the SSS problem because the generalized eigenvalues computation is not involved in the optimization procedure. The experimental results on both synthetic and real-world databases demonstrate that the proposed method can outperform LPP and LPP-L1 when there are large outliers in the training data.

1. Introduction

There are many dimensionality reduction methods used to reduce the number of input variables to simplify data analysis problems, which play an important role in machine learning, information retrieval, pattern recognition and so on. Linear methods such as principal component analysis (PCA) [1], linear discriminant analysis (LDA) [2] and $(2D)^2$PCALDA [3] have demonstrated excellent performance in many fields. These methods are dimensionality reduction algorithms 'thinking globally' and can successfully discover low dimensional manifold on the premise of Gaussian data. At the same time, they are easy to implement and their optimizations are well understood and efficient. However, these methods are inadequate for embedding the nonlinear manifolds and cannot preserve the local structure of data.

A number of nonlinear dimensionality reduction techniques have been developed to address the aforementioned problem. Some representative nonlinear manifold learning methods include Isomap [4], locally linear embedding (LLE) [5], Laplacian Eigenmaps [6] and Hessian Eigenmaps [7–9]. These methods preserve local properties of the given data by constructing a graph representation of the data points and have demonstrated good performances on some databases. However, many nonlinear manifold learning methods yield maps that are defined only on the training data samples and how to evaluate the maps on the novel test data samples is unclear [10]. For dealing with this problem, some linear versions of these methods are proposed such as neighborhood preserving projection (NPP) [11], neighborhood preserving embedding (NPE) [12] and locality preserving projection (LPP) [10,13]. LPP is a linear approximation of Laplacian Eigenmaps and provides a way to the projection of the novel test data samples. Therefore, LPP not only have the locality preserving property but also is a linear technique. However, LPP is sensitive to outliers because its objective function is based on the L2-norm distance criterion and suffers from the small sample size (SSS) problem. VDE [14] is proposed to overcome the SSS problem of LPP by adopting the maximum margin criterion. Moreover, more manifold learning algorithms are proposed in recent years [15–21].

Recently, many researchers focus on improving the robustness of dimensionality reduction methods. One of the efforts is searching for more robust distance metric because it is well known that L2-norm-based distance criterion is sensitive to outliers for the square operation. A number of L1-norm-based dimensionality reduction methods are proposed to alleviate the negative effect of outliers [22–27]. These methods, such as PCA-L1 [23], LDA-L1 [26] and LPP-L1 [27], adopt a greedy strategy to learn a set of projection vectors for optimizing a L1-norm-based objective function

* Corresponding author.
E-mail addresses: fujin-zhong@163.com (F. Zhong), ldf125@home.swjtu.edu.cn (D. Li), jszhang@home.swjtu.edu.cn (J. Zhang).

http://dx.doi.org/10.1016/j.jvcir.2014.08.004
1047-3203/

and have demonstrated better robustness to outliers than the corresponding L2-norm-based versions. He et al. adopted correntropy to measure the construction error and proposed a robust PCA based on maximum correntropy criterion (termed as PCA-MCC) [28]. PCA-MCC uses a half-quadratic optimization algorithm to compute the correntropy objective function and has outperformed robust rotational-invariant PCAs based on L1-norm [28]. On the other hand, some recent works [29,30] present that one key effort is looking for efficient optimization algorithms in robust learning methods.

In this paper, we adopt correntropy to measure the similarity between all pairs of data points in the feature space for improving the robustness of LPP to outliers. Since the objective function of the proposed method is based on maximum correntropy criterion (MCC) which is a useful measurement to handle nonzero mean and non-Gaussian noise with large outliers, we denote the new LPP method as LPP-MCC. The correntropy-based objective function of LPP-MCC can be optimized efficiently by half-quadratic optimization framework in an iterative manner, so the complex optimization problem can be solved by a standard optimization method. Therefore, the idea of LPP-MCC can be easily generalized to other graph embedding algorithms. It is worthwhile to highlight three important advantages of LPP-MCC as follows: (1) LPP-MCC is more robust to outliers than the conventional LPPs based on L2-norm or L1-norm. (2) The optimal solutions are obtained via half-quadratic optimization framework which can be achieved by a simple iterative standard optimization method. (3) It avoids the small sample size problem that LPP often encounters because of the generalized eigenvalues problem.

The remainder of this paper is organized as follows. In Section 2, we briefly give a quick review of LPP-L2 and LPP-L1. In Section 3, we propose a robust LPP method based on maximum correntropy criterion and present its optimization procedure using a half-quadratic technique. In Section 4, the experimental results are shown to demonstrate the robustness of the proposed method to outliers. Finally, the paper is summarized in Section 5.

2. LPP and LPP-L1

For convenience, we present in Table 1 the important notations used in this paper and their descriptions.

Let $\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_N] \in \mathbf{R}^{m\times N}$ be the given training samples, where N is the number of samples and $\mathbf{x}_i$ denotes an m-dimensional column vector. LPP aims to find an optimal projection matrix $\mathbf{W} = [\mathbf{w}_1, \ldots, \mathbf{w}_n] \in \mathbf{R}^{m\times n}$ $(n < m)$ whose columns $\{\mathbf{w_k}\}(k = 1, \cdots, n)$ constitute the base of the n-dimension subspace. Projecting the sample $\mathbf{x}_i$ onto $\mathbf{W}$ yields an n-dimension vector $\mathbf{y}_i$, i.e. $\mathbf{y}_i = \mathbf{W}^T\mathbf{x}_i$, where $\mathbf{y}_i$ is called the feature of $\mathbf{x}_i$ in the n-dimension subspace. The optimal projection vector $\mathbf{w} \in \mathbf{R}^{m\times 1}$ can be gained by solving the following constrained optimization problem as:

$$\begin{aligned}&\mathbf{w} = \arg\min_{\mathbf{w}} J_{L2}(\mathbf{w}) = \arg\min_{\mathbf{w}} \sum_{i,j}^{N} (\mathbf{w}^T\mathbf{x}_i - \mathbf{w}^T\mathbf{x}_j)^2 s_{ij} \\ &\text{s.t. } \mathbf{w}^T\mathbf{X}\mathbf{D}(\mathbf{w}^T\mathbf{X})^T = 1\end{aligned} \tag{1}$$

where s_{ij} is the similarity measure between $\mathbf{x}_i$ and $\mathbf{x}_j$, and all s_{ij} constitute the similarity matrix $\mathbf{S}$. $\mathbf{D}$ is a diagonal matrix and its entries are $D_{ii} = \sum_j s_{ij}$. A possible way of defining s_{ij} is heat kernel [31]:

$$s_{ij} = \begin{cases} exp(-\|\mathbf{x}_i - \mathbf{x}_j\|^2/t), & \text{if } \mathbf{x}_i \text{ is among } k \text{ nearest neighbors of } \mathbf{x}_j \\ & \text{or } \mathbf{x}_j \text{ is among } k \text{ nearest neighbors of } \mathbf{x}_i, \\ 0 & \text{otherwise.} \end{cases} \tag{2}$$

Minimizing the objective function of (1) is an attempt to ensure that, if $\mathbf{x}_i$ and $\mathbf{x}_j$ are close, then $\mathbf{y}_i$ and $\mathbf{y}_j$ are close as well [10]. In other words, for large similarity between $\mathbf{x}_i$ and $\mathbf{x}_j$, the distance between $\mathbf{y}_i$ and $\mathbf{y}_j$ should be so small that the objective function is minimized [27,32]. The optimization of (1) can be reduced to the following generalized eigenvalues problem [10,13]:

$$\mathbf{XLX}^T\mathbf{w} = \lambda\mathbf{XDX}^T\mathbf{w} \tag{3}$$

where $\mathbf{L}$ is the Laplacian matrix which is formed by subtracting $\mathbf{S}$ from $\mathbf{D}$, i.e. $\mathbf{L} = \mathbf{D} - \mathbf{S}$. The projection vectors $\{\mathbf{w_k}\}(k = 1, \cdots, n)$ are given by the minimum eigenvalues solutions. However, if $\mathbf{XDX}^T$ is singular, the solution of LPP is unstable. That it is caused by the so called small sample size problem [27].

Because the L2-norm-based LPP is sensitive to outliers, L1-norm is applied in (1) to substitute the L2-norm [27]. Thus, the objective function of LPP-L1 is formulated as the following:

$$\begin{aligned}&\mathbf{w} = \arg\min_{\mathbf{w}} J_{L1}(\mathbf{w}) = \arg\min_{\mathbf{w}} \sum_{i,j}^{N} |\mathbf{w}^T\mathbf{x}_i - \mathbf{w}^T\mathbf{x}_j| s_{ij} \\ &\text{s.t. } \mathbf{w}^T\mathbf{X}\mathbf{D}(\mathbf{w}^T\mathbf{X})^T = 1.\end{aligned} \tag{4}$$

However, the optimization of (4) is very difficult because it contains the absolute value operation, which is nonlinear. In [27], the solution of LPP-L1 is approximated by maximizing a new objective problem as:

$$\begin{aligned}&\mathbf{w} = \arg\max_{\mathbf{w}} J_{L1}(\mathbf{w}) = \arg\max_{\mathbf{w}} \sum_{i,j}^{N} |\mathbf{w}^T\mathbf{x}_i - \mathbf{w}^T\mathbf{x}_j| (1 - s_{ij}) \\ &\text{s.t. } \mathbf{w}^T\mathbf{w} = 1.\end{aligned} \tag{5}$$

The optimization procedure of (5) is similar to that of PCA-L1 [24]. When the first projection vector is extracted, the procedure can easily be extended to learn more projection vectors by applying the same optimization procedure greedily to the remainder of the projected samples [25].

Table 1
Notations used in this paper and their descriptions.

Notation	Description
N	Sample size
m	Sample dimensions
$\mathbf{X}$	Sample data matrix
$\mathbf{x}_i$	The i-th sample
$\mathbf{y}_i$	The feature of $\mathbf{x}_i$
$\mathbf{D}$	Diagonal matrix and $D_{ii} = \sum_j s_{ij}$
$\mathbf{L}$	Laplacian matrix and $\mathbf{L} = \mathbf{D} - \mathbf{S}$
p_{ij}	The auxiliary variable
$\mathbf{P}$	The auxiliary variable matrix
$\lvert\cdot\rvert$	Absolute value
n	Subspace dimensions
$\mathbf{W}$	Optimal projection matrix
$\mathbf{w}_k$	The k-th column vector of $\mathbf{W}$
s_{ij}	Similarity measure between $\mathbf{x}_i$ and $\mathbf{x}_j$
$\mathbf{S}$	Similarity matrix
$k_\sigma(\cdot)$	Kernel function that satisfies Mercer's theory
$E(\cdot)$	Mathematical expectation
A, B	Two random variables
$g(x)$	Gaussian kernel
$\lVert\cdot\rVert_2$	L2 norm

3. Robust LPP based on MMC

Although LPP-L1 can alleviate the negative effect of outliers to some extent, it is not sufficient for handling nonzero mean and non-Gaussian noise with large outliers. Liu et al. have made enough theoretical analysis to indicate that MCC is a robust measurement to handle nonzero mean and non-Gaussian noise with large outliers [33]. In addition, He et al. have validated the robustness of MCC by developing PCA-MCC [28] and MCC-based face recognition [34]. Considering the relation of PCA and LPP, this paper