

# 正交化近邻关系保持的降维及分类算法

刘小明<sup>1),2)</sup> 尹建伟<sup>2)</sup> 冯志林<sup>3)</sup> 董金祥<sup>2)</sup>

<sup>1)</sup>(武汉科技大学计算机科学与技术学院,武汉 430081) <sup>2)</sup>(浙江大学计算机科学与技术学院,杭州 310027)

<sup>3)</sup>(浙江工业大学之江学院,杭州 310024)

**摘要** 针对近邻关系保持嵌入(NPE)算法易于受到降低后的维数影响,而且性能依赖于正确的维数估计的问题,提出了一种正交化的近邻关系保持的嵌入降维方法——ONPE。ONPE方法是使用数据点间的近邻关系来构造邻接图,假设每个数据点都能由其近邻点的线性组合表示,则可以通过提取数据点的局部几何信息,并在降维中保持提取的局部几何信息,迭代地计算正交基来得到数据的低维嵌入坐标。同时,在ONPE算法的基础上,利用局部几何信息,提出了一种在低维空间中使用标签传递(LNP)的分类算法——ONPC。其是假设高维空间中的局部近邻关系在降维后的空间中依然得到保持,并且数据点的类别可由近邻点的类别得到。在人工数据和人脸数据上的实验表明,该算法在减少维数依赖的同时,能有效提高NPE算法的分类性能。

**关键词** 流形学习 近邻保持嵌入 线性近邻传递算法

中图法分类号: TP391.41 文献标识码: A 文章编号: 1006-8961(2009)07-1319-08

## Orthogonal Neighborhood Preserving Embedding Based Dimension Reduction and Classification Method

LIU Xiao-ming<sup>1),2)</sup>, YIN Jian-wei<sup>2)</sup>, FENG Zhi-lin<sup>3)</sup>, DONG Jin-xiang<sup>2)</sup>

<sup>1)</sup>(College of Computer Science and Technology, Wuhan University of Science and Technology, Wuhan 430081)

<sup>2)</sup>(College of Computer Science and Technology, Zhejiang University, Hangzhou 310029)

<sup>3)</sup>(Zhijiang College, Zhejiang University of Technology, Hangzhou 310024)

**Abstract** To overcome the sensitivity to the dimensions of reduced space, and performance degradation with wrong dimension estimation of neighborhood preserving embedding (NPE) method, an orthogonal neighborhood preserving embedding (ONPE) method is proposed for manifold dimension reduction. ONPE uses neighborhood information to construct the adjacent graph, and assuming that each data point can be represented by linear combination of its neighbor points. ONPE then extracts local geometry information embedded in reconstruction weights, and obtains the low dimensional coordinates by iteratively computes the mutually orthogonal basis functions. Moreover, utilizing the local geometry during ONPE dimension reduction, a new classification method (ONPC) based on a label propagation method (LNP) is proposed. The reasonable assumption is that local neighbor information in high dimensional space is also preserved in reduced space, and the class label of a data point can be obtained through the class labels of its neighbors. Several experiments on artificial datasets and face database demonstrate the effectiveness of the algorithm.

**Keywords** manifold learning, neighborhood preserving embedding, linear neighborhood propagation

**基金项目:**国家自然科学基金项目(60703042);国家高技术研究发展计划“863”项目(2006AA01Z170,2007AA01Z124);浙江省自然科学基金项目(Y106045)

**收稿日期:**2007-06-13;**改回日期:**2008-02-22

**第一作者简介:**刘小明(1980 ~ ),男。2007年获浙江大学工学博士学位。主要研究方向为模式识别与图像处理。E-mail: lxmzju@gmail.com

**通讯作者:**尹建伟,E-mail: zjujyw@cs.zju.edu.cn

## 1 引言

维数约简的目的是为了将高维数据变换到低维空间中,同时尽量保留数据的内在结构,这一直是模式识别、数据挖掘和机器学习领域的重要课题。国内外学者提出了多种降维方法,可粗略分为 2 类:线性方法和非线性方法。主成分分析(PCA)与线性判别分析(LDA)是最常见的线性降维方法,其是假设数据样本分布在全局线性的流形上。实际上,许多高维数据分布在嵌入高维线性空间的低维非线性流形上。近来,许多学者致力于从高维数据空间中的数据样本中构造低维非线性流形的算法——流形学习算法,例如 Isomap<sup>[1]</sup>、局部线性嵌入(local linear embedding, LLE)<sup>[2-3]</sup>、局部切空间排列算法(local tangent space alignment, LTSA)<sup>[4-6]</sup>。虽然这些算法在一些基准人工合成数据和一些实际应用中取得了很好的效果,但因其是非线性的,故计算工作量很大,并且难以处理新样本。

近邻关系保持嵌入(neighborhood perserving embeddin, NPE)算法<sup>[7]</sup>是一种新的线性降维方法。与传统 PCA 保持全局线性结构不同,NPE 尽量保持局部线性结构,能更真实地反映数据的分布。但与 PCA 不同,NPE 得到的投影矩阵不是正交的。由于非正交性使其不能保持高维空间中的测度结构,因此不能用于估计高维数据的内蕴维数。事实上,对高维空间的内蕴维数的不正确估计会极大地降低 NPE 的性能。LPP(locality preserving projection)算法<sup>[8]</sup>也有相似的问题,文献[9]、[10]中提出了正交化的 LPP 算法。

本文提出了一种称为正交化的近邻关系保持嵌入(orthogonal neighborhood perserving embedding, ONPE)的新算法。ONPE 与 NPE 算法一样,能保持数据点间的近邻关系,但同时要求变换基函数是正交的,其不仅能保持原始高维数据空间中的测度结构,且性能对降维后的维数不敏感。同时,在 ONPE 算法的基础上,本文提出了使用局部近邻标签传递(local neighborhood labels propagation, LNP)在低维空间中对数据进行分类的算法(orthogonal neighborhood perserving classification, ONPC),其自然地扩展了 LNP 算法中的假设。ONPC 算法是假设高维空间中数据间的局部近邻关系在降维后的数据间依然存在,并且数据点的类别信息可以由近邻点的类别信

息得到,后者正是 LNP 算法<sup>[9]</sup>的假设。ONPC 算法与 LNP 算法<sup>[9]</sup>不同的是,LNP 算法是作用在高维数据空间中,其对高维的大数据集效率低,而 ONPC 算法是对经 ONPE 算法降维后的低维空间中的数据进行处理,其能更好地处理大数据集。

## 2 NPE 简介

NPE<sup>[7]</sup>算法的目标是在降维中保留局部流形结构,这里“局部结构”是指每个数据点能由其近邻点的线性组合表示。线性降维问题可表述为:给定  $\mathbf{R}^D$  中的数据点集  $\{\mathbf{x}_1, \dots, \mathbf{x}_n\}$ ,通过寻找变换矩阵  $\mathbf{A} \in \mathbf{R}^{D \times d}$  将这些点映射为  $\mathbf{R}^d$  中的数据点集  $\{\mathbf{y}_1, \dots, \mathbf{y}_n\}$  ( $d \ll D$ ),使点  $\mathbf{y}_i$  代表  $\mathbf{x}_i$ ,即  $\mathbf{y}_i = \mathbf{A}^T \mathbf{x}_i$  ( $i = 1, \dots, n$ )。NPE 算法要求在高维空间中可线性局部重构的数据点在降维后依然能以相同的权重重构<sup>[7]</sup>。

给定权重矩阵  $\mathbf{W}$ ,NPE 算法映射可通过求解如下问题得到:

$$\begin{aligned} \alpha_{\text{opt}} &= \arg \min_{\alpha} \sum_i (\alpha^T \mathbf{x}_i - \sum_j W_{i,j} \alpha^T \mathbf{x}_j)^2 \\ &= \arg \min_{\alpha} \alpha^T \mathbf{X} \mathbf{M} \mathbf{X}^T \alpha \end{aligned}$$

在  $\alpha^T \mathbf{X} \mathbf{X}^T \alpha = 1$  的限制下。这里  $\mathbf{X} = (\mathbf{x}_1, \dots, \mathbf{x}_n)$ ,  $\mathbf{M} = (\mathbf{I} - \mathbf{W})^T (\mathbf{I} - \mathbf{W})$ 。NPE 算法通过最小化下述代价函数来构造权重矩阵  $\mathbf{W}$ :

$$\min \sum_i \left\| \mathbf{x}_i - \sum_j W_{i,j} \mathbf{x}_j \right\|^2$$

限制为  $\sum_j W_{i,j} = 1, j = 1, 2, \dots, n$ ,此外,如果数据点  $\mathbf{x}_i$  与  $\mathbf{x}_j$  不是近邻点,则强制  $W_{i,j} = 0$ 。最后,NPE 算法的变换矩阵由如下与广义特征值问题的最小特征值所对应的特征向量集组成:

$$\mathbf{X} \mathbf{M} \mathbf{X}^T \alpha = \lambda \mathbf{X} \mathbf{X}^T \alpha$$

由于  $\mathbf{X} \mathbf{X}^T$  在 NPE 算法中经过基于奇异值分解(SVD)预处理可保证为非奇异,因此,NPE 算法的基函数可认为是与  $(\mathbf{X} \mathbf{X}^T)^{-1} \mathbf{X} \mathbf{M} \mathbf{X}^T$  的最小特征值集对应的特征向量。因  $(\mathbf{X} \mathbf{X}^T)^{-1} \mathbf{X} \mathbf{M} \mathbf{X}^T$  一般为非对称,故 NPE 算法得到的特征向量集为非正交向量集。

## 3 ONPE 与 ONPC 算法

在介绍 ONPC 算法前,首先介绍 ONPE 算法,ONPE 算法的理论分析将在第 4 节中介绍;然后,给出了一种基于 ONPE 降维和局部近邻标签传递的分

类算法——ONPC。

### 3.1 ONPE 算法

在许多问题中,例如人脸识别,通常数据的特征数  $D$  远大于样本的数目  $n$ ,此时矩阵  $XX^T \in \mathbf{R}^{D \times D}$  是奇异的。为克服此问题,可在不损失任何信息的条件下使用 PCA 对数据进行投影,使得矩阵  $XX^T$  为非奇异。

预处理(可选):将数据点  $\mathbf{x}_i$  投影到 PCA 子空间,并丢弃对应于特征值 0 的成分,同时使用  $\mathbf{A}_{\text{PCA}}$  表示这里 PCA 的变换矩阵。为描述方便,仍使用  $\mathbf{x}_i$  代表经过 PCA 投影后的数据。如果不进行预处理,则  $\mathbf{A}_{\text{PCA}}$  是单位变换矩阵。

(1) 构造邻接图 令  $\mathbf{H}$  表示有  $n$  个节点的图,第  $i$  个节点对应于数据点  $\mathbf{x}_i$ 。邻接图的构造有  $k$  近邻或者  $\varepsilon$  邻居<sup>[2,7]</sup> 两种方式。注意,如果数据点已经被分为不同的类,则可以依据类别信息构造邻接图,即在图  $\mathbf{H}$  中与 2 个数据对应的顶点间用一条边连接,当且仅当它们属于同一类。在下面的描述中,假设采用的是  $k$  近邻方法。

(2) 计算权重 这里的基本假设是每个数据点和其  $k$  近邻近似分布在局部线性流形上,并且每个数据点可由其  $k$  近邻的线性组合重构。令  $\mathbf{W} \in \mathbf{R}^{n \times n}$  表示权重矩阵,其中  $W_{i,j}$  表示节点  $i$  到节点  $j$  的边上的权重,如果节点  $i$  与节点  $j$  间无边,则权重为 0。权重矩阵可通过最小化以下重构误差函数得到:

$$\varepsilon(\mathbf{W}) = \sum_i \left\| \mathbf{x}_i - \sum_j W_{i,j} \mathbf{x}_j \right\|_2^2$$

限制为  $\sum_j W_{i,j} = 1, j = 1, 2, \dots, n$ 。注意重构误差可对各数据点分别最小化。对于数据点  $\mathbf{x}_i$ ,可将局部 Gram 矩阵定义为  $\mathbf{G}_i \in \mathbf{R}^{k \times k}$ ,其中元素值为

$$\mathbf{G}_{p,l} = (\mathbf{x}_i - \mathbf{x}_p)^T (\mathbf{x}_i - \mathbf{x}_l) \in \mathbf{R}^{k \times k}$$

其中,  $\mathbf{x}_p$  与  $\mathbf{x}_l$  为  $\mathbf{x}_i$  的近邻点。

易知,含约束的权重矩阵有如下解析解<sup>[9]</sup>:

$$\mathbf{w}_i = \frac{\sum_p \mathbf{G}_{i,p}^{-1}}{\sum_{p,l} \mathbf{G}_{p,l}^{-1}}$$

其中,  $\mathbf{w}_i$  是权重矩阵  $\mathbf{W}$  的第  $i$  列。

(3) 计算正交化近邻关系保持嵌入 令  $[\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \dots, \boldsymbol{\alpha}_k]$  是正交近邻关系嵌入,并定义

$$\mathbf{A}^{(k-1)} = [\boldsymbol{\alpha}_1, \dots, \boldsymbol{\alpha}_{k-1}],$$

$$\mathbf{S}^{(k-1)} = [\mathbf{A}^{(k-1)}]^T (\mathbf{X}\mathbf{X}^T)^{-1} \mathbf{A}^{(k-1)}$$

则正交的近邻关系保持矩阵  $[\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \dots, \boldsymbol{\alpha}_k]$  可进行

如下迭代计算:

① 计算  $\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_1$  是对应于矩阵  $(\mathbf{X}\mathbf{X}^T)^{-1} \mathbf{X}\mathbf{M}\mathbf{X}^T$  最小特征值的特征向量。

② 计算  $\boldsymbol{\alpha}_k, \boldsymbol{\alpha}_k$  是对应于如下矩阵  $\mathbf{J}^{(k)}$  最小特征值的特征向量:

$$\mathbf{J}^{(k)} = \{ \mathbf{I} - (\mathbf{X}\mathbf{X}^T)^{-1} \mathbf{A}^{(k-1)} [\mathbf{S}^{(k-1)}]^{-1} [\mathbf{A}^{(k-1)}]^T \} \cdot (\mathbf{X}\mathbf{X}^T)^{-1} \mathbf{X}\mathbf{M}\mathbf{X}^T$$

(4) ONPE 投影:令  $\mathbf{A}_{\text{ONPE}} = [\boldsymbol{\alpha}_1, \dots, \boldsymbol{\alpha}_d]$ , 则嵌入映射为

$$\begin{aligned} \mathbf{x} \rightarrow \mathbf{y} &= \mathbf{A}^T \mathbf{x} \\ \mathbf{A} &= \mathbf{A}_{\text{PCA}} \mathbf{A}_{\text{ONPE}} \end{aligned} \quad (1)$$

其中,  $\mathbf{y}$  是  $\mathbf{x}$  的  $d$  维表达,  $\mathbf{A}$  是变换矩阵。

### 3.2 ONPC 分类算法

文献[11]、[12]提出的基于局部近邻标签传递的半监督式分类算法,其中利用了局部几何信息,但由于其在原始高维空间中处理,因此对大数据集的高维数据的处理效率不高。

ONPC 算法使用 LNP 算法相似的想法,并进一步假设高维数据中的近邻几何关系在降维后的低维空间中依然得到保留,注意到,因这是 NPE 算法在降维过程中要保留的目标信息,故此假设自然成立。ONPC 与 LNP 算法不同之处是,ONPC 算法是在经过 ONPE 算法降维后的低维空间中进行分类,即 ONPC 算法要分类的数据集为  $\mathbf{Y} = \mathbf{A}^T \mathbf{X}, \mathbf{y}_i = \mathbf{A}^T \mathbf{x}_i, i = 1, \dots, n$ 。假设有  $c$  类,类别标签集为  $\mathbf{L} = \{1, 2, \dots, c\}$ 。令  $\mathbf{M}$  为非负矩阵集,  $\mathbf{M} \subset \mathbf{R}^{n \times c}$ 。任意矩阵  $\mathbf{F} = [\mathbf{f}_1^T, \mathbf{f}_2^T, \dots, \mathbf{f}_n^T]^T \in \mathbf{M}$  可对应于  $\mathbf{Y}$  的一个分类,将其  $\mathbf{y}_i$  分类为  $z_i = \operatorname{argmax}_{j \leq c} F_{i,j}$ , 这样  $\mathbf{F}$  可视为一个分类函数。初始时,令  $\mathbf{F}_0 = \mathbf{Z}$ , 其中,如果  $\mathbf{y}_i$  的类别为  $j$ , 则  $Z_{i,j} = 1 (1 \leq j \leq c)$ , 否则  $Z_{i,j} = 0$ 。对于未分类的数据点,赋值  $Z_{u,j} = 0$ 。ONPC 算法的主要过程如下(详见文献[11]、[12]):

输入:  $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_l, \mathbf{x}_{l+1}, \dots, \mathbf{x}_n\} \in \mathbf{R}^D$ , 其中  $\{\mathbf{x}_i\}_{i=1}^l$  已分类,  $\{\mathbf{x}_u\}_{u=l+1}^n$  未分类。初始分类矩阵  $\mathbf{Z}$ 。近邻数  $k$ , 低维空间的维数  $d$ , 常数  $\alpha (0 \leq \alpha \leq 1)$ , 缺省值 0.99。

输出:所有数据点的类别信息

(1) 使用 ONPE 算法降维,  $\mathbf{Y} = \mathbf{A}^T \mathbf{X}$ , 即可通过 ONPE 算法得到权重矩阵  $\mathbf{W}$ 。

(2) 构造传递矩阵  $\mathbf{P} = \mathbf{W}$ , 并循环  $\mathbf{F}_{t+1} = \alpha \mathbf{P}\mathbf{F}_t + (1 - \alpha)\mathbf{Z}$  直到收敛。可知由于在收敛时  $\tilde{\mathbf{F}} = \lim_{t \rightarrow \infty} \mathbf{F}_t = (1 - \alpha)(\mathbf{I} - \alpha\mathbf{W})^{-1}\mathbf{Z}$ , 因此  $\mathbf{F}$  可以直接计算得到。

(3) 令  $F^*$  是  $F_i$  的收敛极限, 输出数据点  $x_i(y_i)$  的类别标签  $z_i = \arg \max_{j \leq c} F_{i,j}^*$ .

## 4 ONPE 算法理论分析

定义: 令  $\alpha \in \mathbf{R}^D$  是一个投影映射, 则近邻关系保持函数  $f$  定义为

$$f(\alpha) = \frac{\alpha^T X M X^T \alpha}{\alpha^T X X^T \alpha}$$

注意, 近邻关系保持函数  $f(\alpha)$  可用于度量投影映射  $\alpha$  的近邻关系保持能力

$$\begin{aligned} f(\alpha) &= \frac{\alpha^T X M X^T \alpha}{\alpha^T X X^T \alpha} \\ &= \frac{\alpha^T X (I - W)^T (I - W) X^T \alpha}{\alpha^T X X^T \alpha} \\ &= \frac{[\alpha^T X (I - W)^T] [\alpha^T X (I - W)^T]^T}{(\alpha^T X) (\alpha^T X)^T} \\ &= \frac{[\alpha^T (X - XW^T)] [\alpha^T (X - XW^T)]^T}{(\alpha^T X) (\alpha^T X)^T} \end{aligned}$$

由此可知,  $f(\alpha)$  越小, 近邻信息保留得越多<sup>[3]</sup> (考虑将  $x$  映射到位于一条线上)。直接最小化此函数即可得到原始 NPE 算法, ONPE 算法在最小化近邻保持函数的同时, 还能确保基向量正交。

因此, ONPE 算法的目标函数是

$$\begin{aligned} \alpha_1 &= \arg \min_{\alpha} \frac{\alpha^T X M X^T \alpha}{\alpha^T X X^T \alpha} \\ \alpha_k &= \arg \min_{\alpha} \frac{\alpha^T X M X^T \alpha}{\alpha^T X X^T \alpha}, \text{ 其中} \\ \alpha_k^T \alpha_1 &= \alpha_k^T \alpha_2 = \cdots = \alpha_k^T \alpha_{k-1} = 0 \end{aligned}$$

容易验证,  $\alpha_1$  是与广义特征值问题  $X M X^T \alpha = \lambda X X^T \alpha$  最小特征值对应的特征向量。因  $X X^T$  非奇异, 故  $\alpha_1$  是矩阵  $(X X^T)^{-1} X M X^T$  最小特征值对应的特征向量。

接下来, 可以将最小化目标函数重新表达为

$$f(\alpha_k) = \frac{\alpha_k^T X M X^T \alpha_k}{\alpha_k^T X X^T \alpha_k} \quad (2)$$

其中,  $\alpha_k^T \alpha_1 = \alpha_k^T \alpha_2 = \cdots = \alpha_k^T \alpha_{k-1} = 0, \alpha_k^T X X^T \alpha_k = 1$ 。

注意到, 由于  $X X^T$  为正定矩阵, 它对任意投影映射  $\alpha$ , 总可以对其规范化, 并可在保持  $\alpha^T X M X^T \alpha$  与  $\alpha^T X X^T \alpha$  比值不变的条件下, 使  $\alpha^T X X^T \alpha = 1$ , 因此, 最小化问题等价于在约束  $\alpha^T X X^T \alpha = 1$  的条件下, 最小化  $\alpha^T X M X^T \alpha$  的问题。

为了得到第  $k$  个基向量, 可使用 Lagrange 乘数

法表达包含所有约束的上述目标函数, 即

$$C^{(k)} = \alpha_k^T X M X^T \alpha_k - \lambda (\alpha_k^T X X^T \alpha_k - 1) - \mu_1 \alpha_k^T \alpha_1 - \cdots - \mu_{k-1} \alpha_k^T \alpha_{k-1}$$

令  $C^{(k)}$  对  $\alpha_k$  的偏导为 0, 即可得到

$$\begin{aligned} \frac{\partial C^{(k)}}{\partial \alpha_k} &= 0 \\ \Rightarrow 2 X M X^T \alpha_k - 2 \lambda X X^T \alpha_k - \mu_1 \alpha_1 - \cdots - \mu_{k-1} \alpha_{k-1} &= 0 \end{aligned} \quad (3)$$

用  $\alpha_k^T$  乘以式(3)的左边, 即得到

$$\begin{aligned} 2 \alpha_k^T X M X^T \alpha_k - 2 \lambda \alpha_k^T X X^T \alpha_k &= 0 \\ \Rightarrow \lambda &= \frac{\alpha_k^T X M X^T \alpha_k}{\alpha_k^T X X^T \alpha_k} \end{aligned} \quad (4)$$

比较式(2)知,  $\lambda$  正是要最小化的表达式。

分别使用  $\alpha_1^T (X X^T)^{-1}, \cdots, \alpha_{k-1}^T (X X^T)^{-1}$  乘以等式(3)的左边, 即可得到以下  $k-1$  个等式:

$$\begin{aligned} \mu_1 \alpha_1^T (X X^T)^{-1} \alpha_1 + \cdots + \mu_{k-1} \alpha_1^T (X X^T)^{-1} \alpha_{k-1} \\ = 2 \alpha_1^T (X X^T)^{-1} X M X^T \alpha_k \\ \mu_1 \alpha_2^T (X X^T)^{-1} \alpha_1 + \cdots + \mu_{k-1} \alpha_2^T (X X^T)^{-1} \alpha_{k-1} \\ = 2 \alpha_2^T (X X^T)^{-1} X M X^T \alpha_k \\ \dots\dots \\ \mu_1 \alpha_{k-1}^T (X X^T)^{-1} \alpha_1 + \cdots + \mu_{k-1} \alpha_{k-1}^T (X X^T)^{-1} \alpha_{k-1} \\ = 2 \alpha_{k-1}^T (X X^T)^{-1} X M X^T \alpha_k \end{aligned}$$

为了简化记号, 可定义

$$\begin{aligned} \mu^{(k-1)} &= [\mu_1, \cdots, \mu_{k-1}]^T \\ A^{(k-1)} &= [\alpha_1, \cdots, \alpha_{k-1}] \\ S^{(k-1)} &= [S_{i,j}^{(k-1)}] = [A^{(k-1)}]^T (X X^T)^{-1} A^{(k-1)} \\ S_{i,j}^{(k-1)} &= \alpha_i^T (X X^T)^{-1} \alpha_j \end{aligned}$$

使用这些符号, 则前面的  $k-1$  个等式可用一个矩阵来表达, 即

$$\begin{aligned} S^{(k-1)} \mu^{(k-1)} &= 2 [A^{(k-1)}]^T (X X^T)^{-1} X M X^T \alpha_k \\ \text{或者,} \\ \mu^{(k-1)} &= 2 [S^{(k-1)}]^{-1} [A^{(k-1)}]^T (X X^T)^{-1} X M X^T \alpha_k \end{aligned} \quad (5)$$

使用  $(X X^T)^{-1}$  乘以式(3)的左边, 则可得到

$$\begin{aligned} 2 (X X^T)^{-1} X M X^T \alpha_k - 2 \lambda \alpha_k - \mu_1 (X X^T)^{-1} \alpha_1 - \cdots - \mu_{k-1} (X X^T)^{-1} \alpha_{k-1} &= 0 \end{aligned}$$

当使用矩阵记号, 则可表达为

$$2 (X X^T)^{-1} X M X^T \alpha_k - 2 \lambda \alpha_k - (X X^T)^{-1} A^{(k-1)} \mu^{(k-1)} = 0$$

使用式(5), 即可得到

$$\begin{aligned} \{I - (X X^T)^{-1} A^{(k-1)} [S^{(k-1)}]^{-1} [A^{(k-1)}]^T\} \cdot \\ (X X^T)^{-1} X M X^T \alpha_k = \lambda \alpha_k \end{aligned}$$

如式(4)所示,  $\lambda$  正是要最小化的目标,  $\alpha_k$  是与  $J^{(k)}$  最小化特征值对应的特征向量:

$$J^{(k)} = \{ \mathbf{I} - (\mathbf{X}\mathbf{X}^T)^{-1} \mathbf{A}^{(k-1)} [\mathbf{S}^{(k-1)}]^{-1} [\mathbf{A}^{(k-1)}]^T \} \cdot (\mathbf{X}\mathbf{X}^T)^{-1} \mathbf{X}\mathbf{M}\mathbf{X}^T$$

最后,即得到了优化的正交基向量集。ONPE 算法的正交基能保持原始高维空间中的度量结构。

注意在 NPE 算法<sup>[7]</sup>中,NPE 算法的基向量是与广义特征值问题  $\mathbf{X}\mathbf{M}\mathbf{X}^T \mathbf{b} = \lambda \mathbf{X}\mathbf{X}^T \mathbf{b}$  的  $k$  个最小特征值对应的特征向量集。这些特征向量满足如下约束:

$$\mathbf{b}_i^T \mathbf{X}\mathbf{X}^T \mathbf{b}_j = 0 \quad (i \neq j)$$

由此可知,NPE 变换不是正交变换,实际上,NPE 变换为  $\mathbf{X}\mathbf{X}^T$  正交变换。

## 5 实验结果

### 5.1 人工合成数据

本文使用 S-curve 和 swiss-roll 数据<sup>[13]</sup>来测试 ONPE 算法的降维效果。图 1 给出了将 S-curve 和 swiss-roll 数据集降到 2 维的投影,实验中两个方法的近邻数参数都取为  $k = 8$ 。注意到两种方法都能保持局部近邻关系(由颜色的平滑渐变可知),进一步,ONPE 算法还能保持局部与全局的几何结构信息,在降维后的空间中,ONPE 算法还能体现数据在降维前高维空间中的分布信息,而 NPE 算法因为不能保留高维空间中的测度关系,所以在降维后的空间中扭曲了近邻结构信息。

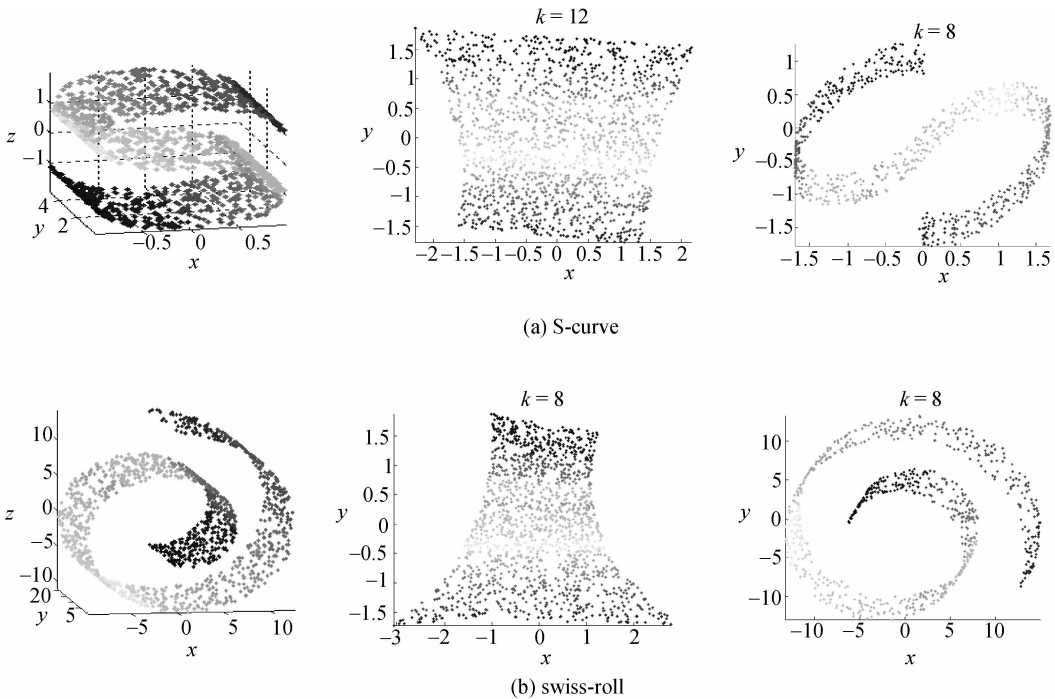


图 1 NPE 与 ONPE 算法降维的结果

Fig. 1 Results of applying NPE and ONPE on dataset

### 5.2 人脸表示和识别

人脸识别是计算机视觉中的一个重要问题,现在已经有许多方法被提出,其中 Eigenface (PCA) 和 Fisherface (PCA + LDA) 是两种最常用的基于表象的人脸识别方法。一幅人脸图像可以表达为图像空间中的一个点,大小为  $m \times n$  的图像可表示为  $m \times n$  维图像空间的一个点。但由于光照,表情,姿态等的变化,图像空间可能并不适合人脸的表达和识别。

本文在 ORL 和 Yale 人脸数据集<sup>[14]</sup>上进行了实验。在 ORL 数据集中,共有 40 个人,每人 10 幅灰度图像,包含了不同光照和不同表情(睁眼/闭眼,笑/不笑)和不同人脸细节(戴眼镜/不戴眼镜),图像大小为  $92 \times 112$  pixels,示例图见图 2。Yale 数据集包含 15 个人的 165 幅灰度图像,每人 11 幅,包含了不同光照,不同表情(普通,笑,悲伤,惊讶等)的图像。



图 2 ORL 数据集中的一个人的示例人脸图像

Fig. 2 Sample face images from the ORL database.

实验时,首先在 ORL 数据集上进行测试,为了与 NPE 算法进行比较,还采用文献[7]中的预处理过程对数据进行了预处理,并使用经过预处理后的数据集<sup>[14]</sup>进行识别。经过预处理后的图像为  $32 \times 32$  大小的灰度图像,这样每幅图像就可以表示为图像空间中 1 024 维的向量。如文献[7],其对于 Eigenface, Fisherface 和 NPEface,都是使用  $k = 1$  的 KNN 方法进行识别,而对于 ONPEfaces,则使用本文提出的 ONPC 方法进行识别。

第 3 节说明了如何得到正交化的近邻关系保持子空间,而且使用训练数据集中的人脸图像就可以得到相应的人脸子空间,这样人脸子空间中的任意一幅图像就可以由基向量的线性组合表示。类似 Eigenface 和 Fisherface 方法,可以将基向量表示为特征图像。识别时,使用预处理过后的 Yale 人脸数据集<sup>[14]</sup>作为训练集,图 3 显示了与 Eigenface, Fisherface, NPEface 和 ONPEface 方法对应的前 5 个基向量,由图 3 可见, NPEface 与 ONPEface 与 Fisherface 方法的向量较相似。

实验时,对 ORL 数据集<sup>[14]</sup>中的每个人,先选择  $l(l = 3, 4, 5)$  幅图像作为训练集,其余图像作为测试集;然后对每个选定的  $l$  幅图像,对训练集和测试集进行 10 次随机划分,并记录其识别率。一般地,识别率随着降维后的人脸子空间的维数变化。图 4 显示了 Eigenface, Fisherface, NPE 和 ONPE 方法在 ORL 数据集上的识别率与降维后子空间维数  $d$  的关系。其中采用的基准方法是在原始 1 024 维图像空间上直接使用  $k = 1$  的 KNN 方法。表 1 列出了 ORL 数据集上各种方法在不同训练样本数下,取得的最高识别率及对应的维数  $d$ 。注意,这里对于 Eigenface, Fisherface, NPE 和 ONPE 方法,子空间的最高维数分别为  $n-1, c-1, n$  和  $n$ 。

可见,ONPE 方法在几乎所有的  $d$  维数据上的识别率都超过其他方法。虽然 Fisherface 方法在  $d$

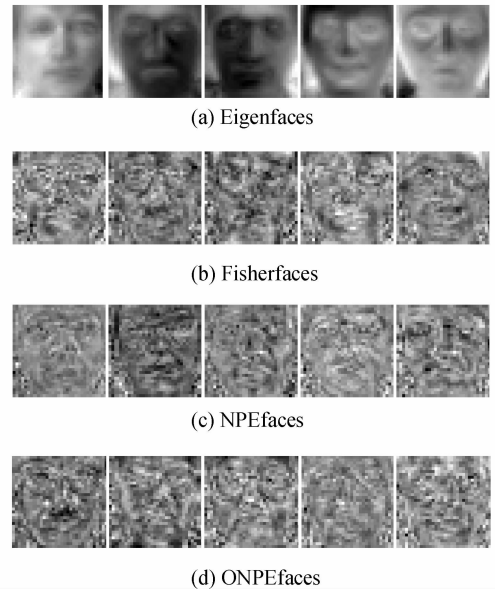


图 3 Yale 数据集上的前 5 个特征向量

Fig. 3 The first 5 basic vectors on Yale face database

表 1 ORL 数据集上的识别率

Tab. 1 Recongition accuracy on the ORL database

方法	不同训练样本数的识别率(%)		
	3	4	5
Baseline	81.3(1 024)	82.3(1 024)	87.2(1 024)
Eigenfaces	81.3(106)	81.3(118)	86.2(58)
Fisherfaces	86.7(38)	91.4(39)	92.5(39)
NPEfaces	87.2(48)	91.6(42)	94.3(46)
ONPEfaces	91.7(42)	94.2(34)	97.8(44)

注:括号中的数据是维数

较小时能取得较好的识别效果,但其性能随着  $d$  的增加提高并不大。从表 1 可看出,ONPE 取得最好识别率时的维数  $d$  很小,并且随着训练样本数  $l$  的增加,性能显著提高。Fisherface 方法随着训练样本的加大,其性能与 ONPE 方法较接近。NPE, ONPE 和 Fisherface 方法取得较好识别率时的维数  $d$  比 Eigenface 方法小很多。同时可看出,ONPE 方法在取得与 NPE 方法相似的识别率时的维数  $d$  比 NPE 方法小很多。注意这里在 NPE 方法中是使用 KNN 方法进行分类,而 ONPE 方法则使用了 ONPC 方法进行分类。

同时还在 Yale 数据集<sup>[14]</sup>上进行了相似的实验,实验时,对其中每个人,首先选  $l(l = 3, 4, 5)$  幅图像作为训练样本,其余图像作为测试集,然后采用 10 次随机划分方法进行测试。各种方法的识别率与降维后子空间维数  $d$  的关系如图 5 所示。由图 5 可看出,ONPE 方法的识别率高于所有其他方法。

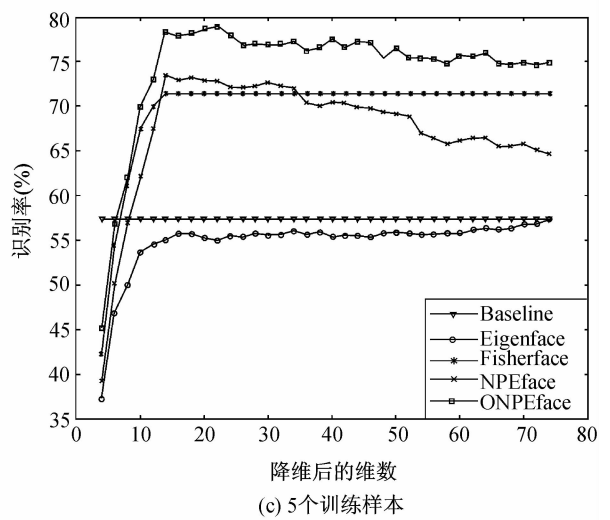
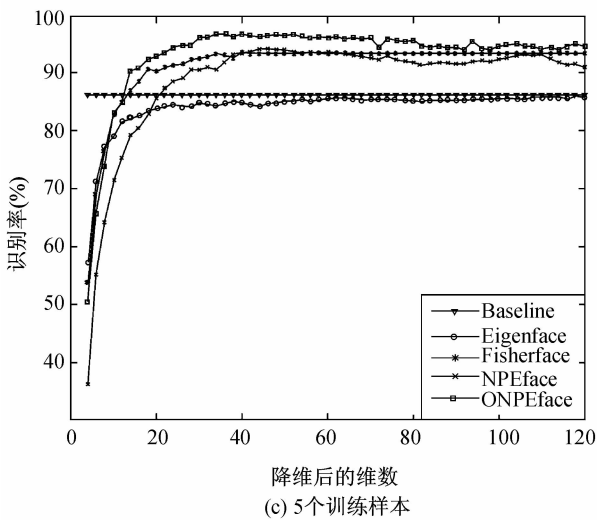
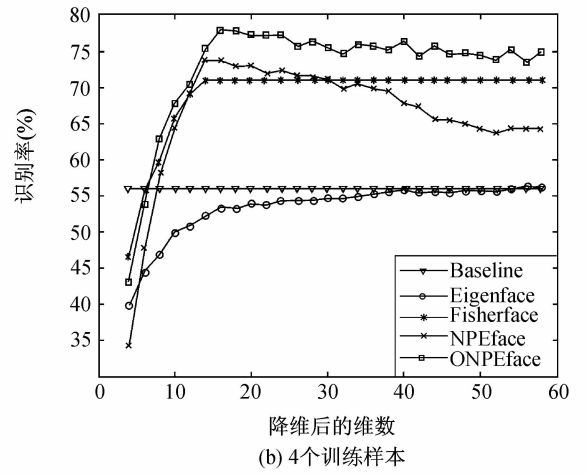
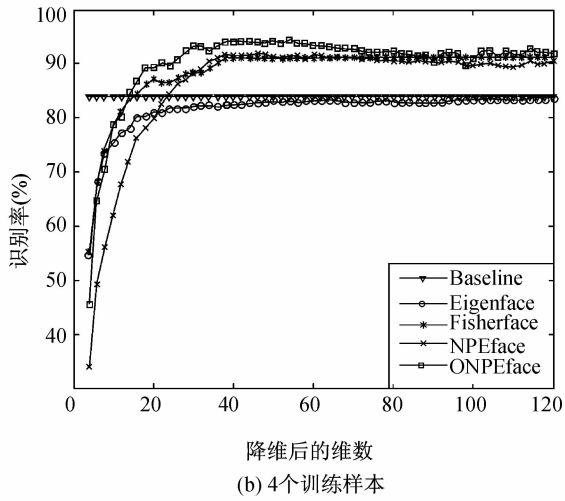
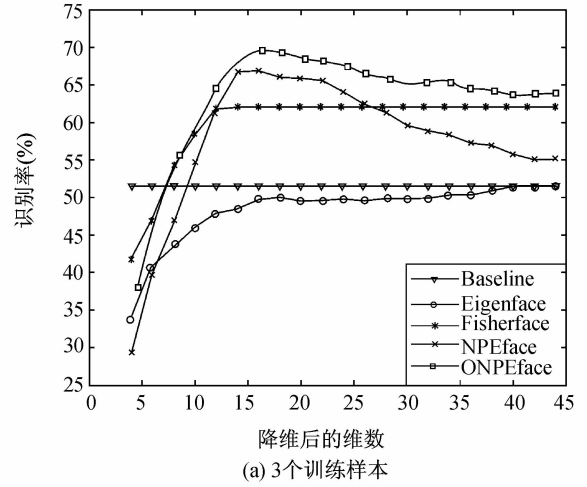
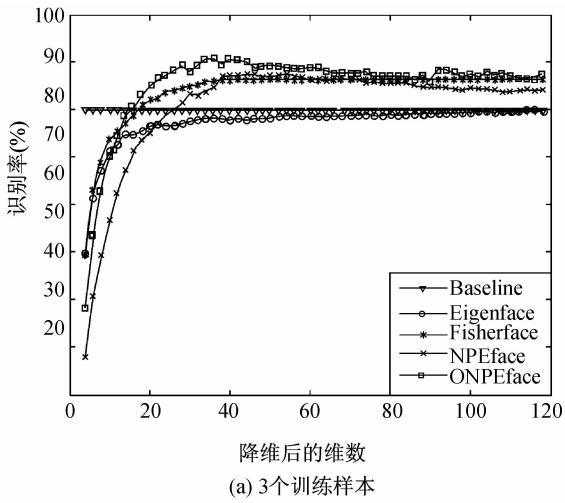


图 4 ORL 数据集上各方法在不同子空间下的识别率

Fig. 4 Recognition rate vs dimension of reduced space on the ORL database

图 5 Yale 数据集上各方法在不同子空间下的识别率

Fig. 5 Recognition rate vs dimension of reduced space on the Yale database

## 6 结 论

本文提出了一种新的流形学习方法——正交化的近邻关系保持嵌入方法。ONPE 方法同时具有 PCA 变换的正交性和 NPE 方法的近邻保持性。理论分析和实验结果说明, ONPE 方法在具有 NPE 方法的近邻关系保持能力的同时, 还克服了 NPE 方法对维数的依赖性。同时, 在 ONPE 映射的基础上, 结合局部标签传递方法, 提出了一种在低维空间中利用近邻几何信息进行分类的方法——ONPC。实验结果表明, ONPC 方法的分类性能优于使用  $k$  近邻分类的 NPE 方法。在 ONPC 方法中, 虽然使用了 LNP 方法利用局部几何信息, 但没有显式地最大化类间距离, 其已证明能进一步提高分类性能。因此如何在利用局部几何信息的同时, 能最大化类间距离, 同时不失去 ONPC 方法的简单性, 这是下一步工作的重点。

### 参考文献 (References)

- 1 Tenenbaum J B, Silva V, Langford J C. A global geometric framework for nonlinear dimensionality reduction [J]. *Science*, 2000, **290**(12): 2319-2323.
- 2 Roweis S T, Saul L K. Nonlinear dimensionality reduction by locally linear embedding [J]. *Science*, 2000, **290**(12): 2323-2326.
- 3 Belkin M, Niyogi P. Laplacian eigenmaps for dimensionality reduction and data representation [J]. *Neural Computation*, 2003, **15**(6): 1373-1396.
- 4 Yin J W, Liu X M, Feng Z L, *et al.* A local tangent space alignment based transductive classification algorithm [A]. In: *Lecture Notes in Computer Science: Artificial Neural Networks in Pattern Recognition [C]*, Ulm, Germany; Springer, 2006: 93-106.
- 5 Liu X M, Yin J W, Feng Z L, *et al.* Incremental manifold learning via tangent space alignment [A]. In: *Lecture Notes in Computer Science: Artificial Neural Networks in Pattern Recognition [C]*, Ulm Germany; Springer, 2006: 107-121.
- 6 Zhang Z, Zha H. Principal manifolds and nonlinear dimensionality reduction via tangent space alignment [J]. *SIAM Journal on Scientific Computing*, 2006, **26**(1): 313-338.
- 7 He X, Cai D, Yan S C, *et al.* Neighborhood preserving embedding [A]. In: *Proceedings of the Tenth IEEE International Conference on Computer Vision [C]*, Beijing, China, 2005: 1208-1213.
- 8 He X, Niyogi P. Locality preserving projections [A]. In: *Vancouver and Whistler (Eds), Advances in Neural Information Processing Systems [C]*, Cambridge, MA, USA; MIT press, 2003.
- 9 Cai D, He X. Orthogonal locality preserving indexing [A]. In: *Proceedings of the 28th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval [C]*, Salvador, Brazil, 2005: 3-10.
- 10 Cai D, He X, Han J, *et al.* Orthogonal laplacian faces for face recognition [J]. *IEEE Transactions on Image Processing*, 2006, **15**(11): 3609-3614.
- 11 Wang F, Zhang C. Label propagation through linear neighborhoods [A]. In: *Proceedings of the 23rd International Conference on Machine Learning [C]*, Pittsburgh, Penn., USA, 2006: 985-992.
- 12 Wang F, Wang J, Zhang C, *et al.* Semi-supervised classification using linear neighborhood propagation [A]. In: *Proceedings IEEE CVPR [C]*, New York, USA, 2006: 160-167.
- 13 Saul L K, Roweis S T. Think globally, fit locally: unsupervised learning of low dimensional manifolds [J]. *Journal of Machine Learning Research*, 2004, **4**(2): 119-155.
- 14 ORL & yale [EB/OL]. <http://www.ews.uiuc.edu/dengcai2/Data/data.html>