

局部敏感非负矩阵分解

姜 伟^{1,2} 杨炳儒¹ 隋海峰¹

(北京科技大学信息工程学院 北京 100083)¹ (辽宁师范大学数学学院 大连 116029)²

摘 要 非负矩阵分解是一种新的基于部分学习的矩阵分解方法,反映了人类思维中局部构成整体的概念。算法只将非负矩阵近似地分解成两个非负矩阵的积,忽略了数据几何结构和判别信息。提出了一个局部敏感非负矩阵分解降维算法来克服这一缺点。该算法既保持了数据非负性,又保持了数据的几何结构和判别信息。构造了一个有效的乘积更新算法并且在理论上证明了算法的收敛性。ORL 和 Yale 人脸数据库实验表明该算法性能超过许多已存在的方法。

关键词 非负矩阵分解,局部敏感分析,判别信息,几何结构

中图法分类号 TP181 **文献标识码** A

Local Sensitive Nonnegative Matrix Factorization

JIANG Wei^{1,2} YANG Bing-ru¹ SUI Hai-feng¹

(School of Information Engineering, University of Science and Technology Beijing, Beijing 100083, China)¹

(School of Mathematics, Liaoning Normal University, Dalian 116029, China)²

Abstract Non-negative matrix factorization (NMF) is a new matrix decomposition method based on the part of the study, which has a reflection of human thinking partial constitute the overall concept. It only find two nonnegative matrices whose product can approximate the nonnegative data matrix without considering the geometric structure and the discriminative information in the data. We presented a local sensitive nonnegative matrix factorization for dimensionality to overcome the disadvantage, which preserves not only the nonnegativity but also the geometric structure and discriminative information of the data. An efficient multiplicative updating procedure was produced, and its convergence was guaranteed theoretically. Experiments on ORL and Yale face recognition databases demonstrate that proposed method outperforms many existing dimensionality reduction methods.

Keywords Nonnegative matrix factorization, Local sensitive analysis, Discriminative information, Geometric structure

在数据挖掘、机器学习、模式识别、计算机视觉和图像处理的研究过程中,经常要用到子空间学习算法对高维数据进行分解,子空间学习算法的思想是把高维空间中松散分布的数据通过变换压缩到低维子空间中,使样本在低维子空间中的分布更紧凑,以便于分类。已有的矩阵分解方法有 PCA(主成分分析)、ICA(独立成分分析)、SVD(奇异值分解)、VQ(矢量量化)、FA(因子分析)等,这些方法通常是在一定的限制下对数据进行线性变换或分解,不同的方法施加在其上的限制是不同的,这些方法在矩阵进行分解的时候允许分解的对象和结果为负。与上面所提出的算法不同,非负矩阵分解(Nonnegative Matrix Factorization, NMF)是一种新的变换方法,能使分解的对象和分解的结果均为非负值。1999 年科学杂志《Nature》刊登了两位科学家 D. D. Lee 和 H. S. Seung^[1]对数学中非负矩阵的研究成果,对分解的因子矩阵施加不同的约束,可以得到不同的变形, Li 等人^[2]在约束原始基向量的数目尽可能少并要求不同的基向量尽可能地接近正交化的

基础上提出了将局部非负矩阵分解(LNMF)。根据稀疏编码原则, Hoyer^[3]提出了一种非负稀疏编码(NNSC), 使其分解后的系数具有较好的稀疏特性。针对 NNSC 并不能从复杂情况下学习基于部分的表示, Liu^[4]等提出了将 NMF 与稀疏编码相结合而形成的 SNMF。Wang 等^[5]提出判别非负矩阵分解,在非负约束的基础上增加了最大化类间散度和最小化类内约束限制。David Guillam 等人^[6]提出了加权 NMF, Ding 等人^[7]将非负矩阵分解成 3 个矩阵的乘积。

许多现实世界中的数据都是嵌入在高维欧氏空间中的非线性低维流形上^[8]的。然而 NMF 算法没有考虑数据的内蕴几何结构,它是假设数据在全局欧氏空间而非流形上,这就大大地限制了当数据位于非线性流形 NMF 算法时的使用。与传统的线性降维算法相比,流形学习算法能够揭示数据内在的几何结构,寻找高维数据在低维空间中的紧致嵌入。Cai 等^[9]提出了图正则化 NMF(GNMF)算法,这个算法假定样本分布在低维流形上,通过构建所有样本的近邻图来估计数据

到稿日期:2010-01-25 返修日期:2010-04-10 本文受国家自然科学基金项目(60875029)资助。

姜 伟(1969—),男,博士生,副教授,主要研究方向为数据挖掘、流形学习、模式识别, E-mail: swxxjw@yahoo.com.cn; 杨炳儒(1943—),男,教授,博士生导师,主要研究方向为知识工程、数据挖掘和人工智能; 隋海峰(1976—),男,博士生,主要研究方向为知识工程、数据挖掘和人工智能。

空间的几何结构,然后将其作为正则化项添加到 NMF 的目标函数,GNMF 算法没有考虑对分类问题非常重要的类别信息,只使用未标记样本,浪费了标记样本中所提供的有用信息,在本文中提出了局部敏感非负矩阵分解方法,新算法既考虑了带标签数据又考虑了不带标签数据。

1 标准非负矩阵分解算法

假设一个非负数据矩阵 $X=[x_1, \dots, x_n] \in R_+^{m \times n}$, X 的每一列都是一个数据点, NMF 的目的是找到两个非负矩阵 $U=[u_{ij}] \in R_+^{m \times k}$ 和 $V=[v_{ij}] \in R_+^{n \times k}$, 其中 $k < m$, 也就是最小化下面的目标函数:

$$J_{\text{NMF}} = \|X - UV^T\|_F^2 \quad s. t. U \geq 0, V \geq 0 \quad (1)$$

$\|\cdot\|_F$ 是 Frobenius 范数, 目标函数 J_{NMF} 是非凸函数, 所以不能找到其全局最优解, 相应的乘积迭代规则为:

$$u_{ij}^{t+1} \leftarrow u_{ij}^t \frac{(XV)_{ij}}{(UV^TV)_{ij}} \quad (2)$$

$$v_{ij}^{t+1} \leftarrow v_{ij}^t \frac{(X^TU)_{ij}}{(VU^TU)_{ij}} \quad (3)$$

2 局部敏感非负矩阵分解

尽管非负矩阵分解能够得到有关物体基于部分的信息, 然而非负性约束并没有明确考虑数据的几何结构及其判别信息。为此首先引入经典的流形学习算法 LPP^[10] 和 LSD^[11] 算法。

2.1 LPP 算法

设 W 为相似度矩阵, LPP 定义的目标函数如下:

$$\min \sum_{ij} (y_i - y_j)^2 W_{ij}$$

$$W_{ij} = \begin{cases} 1, & \text{if } x_i \in N_k(x_j) \text{ or } x_j \in N_k(x_i) \\ 0, & \text{otherwise} \end{cases}$$

$N_k(x_i) = \{x_1^i, x_2^i, \dots, x_k^i\}$ 表示 x_i 的 k 个最近邻。

假设 a 为投影向量, 则有 $y^T = a^T X$, 经过简单变换之后, 目标函数变为:

$$\min a^T X L X^T a \quad s. t. a^T X D X^T a = 1 \quad (4)$$

利用拉格朗日乘数法求解下面的特征值问题:

$$X L X^T a = \lambda X D X^T a \quad (5)$$

最小化目标函数的解就是上述广义特征值问题最小的特征值对应的特征向量。

2.2 LSD 算法

LPP 算法是一种非监督流形学习算法, 用于分类问题时效果不一定好, 当两类样本靠得很近, 由于 LPP 只考虑局部保持的特性而没有考虑类别信息, 将会导致两类样本混淆在一起。局部敏感判别分析使用类内邻域关系图 G_w 和类间邻域关系图 G_b 对数据流形的几何结构和判别结构建模。设 $l(x_i)$ 是 x_i 的类标签, 对于每一数据点 x_i , 集合 $N(x_i)$ 分成两个互不相交的子集分别为 $N_b(x_i)$ 和 $N_w(x_i)$, $N_w(x_i)$ 表示 x_i 的 k_1 个具有相同类别的最近邻点的集合, $N_b(x_i)$ 表示 x_i 的 k_2 个具有不相同类别的最近邻点的集合。定义:

$$N_w(x_i) = \{x_j^i \mid l(x_j^i) = l(x_i), 1 \leq j \leq k_1\}$$

$$N_b(x_i) = \{x_j^i \mid l(x_j^i) \neq l(x_i), 1 \leq j \leq k_2\}$$

式中, $N_b(x_i) \cap N_w(x_i) = \Phi$, 并且 $N_b(x_i) \cup N_w(x_i) = N(x_i)$ 。 G_w 与 G_b 的权矩阵分别为 W_w 和 W_b 。定义其中的元素分别为:

$$W_{w,ij} = \begin{cases} 1, & \text{if } x_i \in N_w(x_j) \text{ or } x_j \in N_w(x_i) \\ 0, & \text{otherwise} \end{cases} \quad (6)$$

$$W_{b,ij} = \begin{cases} 1, & \text{if } x_i \in N_b(x_j) \text{ or } x_j \in N_b(x_i) \\ 0, & \text{otherwise} \end{cases} \quad (7)$$

定义映射 $y = (y_1, y_2, \dots, y_n)^T$ 使 G_w 中的点映射后尽可能地聚集在一起, 而 G_b 映射后的点尽可能地远离, 为此, 优化下面的目标函数:

$$\min \sum_{ij} (y_i - y_j)^2 W_{w,ij} \quad (8)$$

$$\max \sum_{ij} (y_i - y_j)^2 W_{b,ij} \quad (9)$$

设 a 是一个投影向量, 目标函数式(8)、式(9)经过简单的数学推导得到:

$$\begin{aligned} \frac{1}{2} \sum_{ij} (y_i - y_j)^2 W_{w,ij} &= a^T X D_w X^T a - a^T X W_w X^T a \\ &= a^T X L_w X^T a \end{aligned} \quad (10)$$

$$\frac{1}{2} \sum_{ij} (y_i - y_j)^2 W_{b,ij} = a^T X D_b X^T a - a^T X W_b X^T a \quad (11)$$

对于式(11), 给出如下的约束条件:

$$y^T D_b y = 1 \Rightarrow a^T X D_b X^T a = 1 \quad (12)$$

这样, 目标函数式(9)变为如下形式:

$$\max_a a^T X W_b X^T a \quad (13)$$

等价于:

$$\min_a a^T X W_b X^T a \quad (14)$$

目标函数(8)能被重写为如下形式:

$$\min_a a^T X L_w X^T a \quad (15)$$

最优化问题可以写成如下形式:

$$\arg \min_a a^T X (\alpha L_w + (1-\alpha) W_b) X^T a \quad (16)$$

为了使得与非负矩阵表示法的一致, 这里取 $v^T = a^T X$, 式(16)可以表示为:

$$\arg \min_v v^T (\alpha L_w + (1-\alpha) W_b) v \quad (17)$$

2.3 局部敏感非负矩阵分解

算法既考虑矩阵分解结果为非负, 又考虑当数据取样于嵌入在高维空间低维流形数据的几何结构和判别结构时, 得到如下的目标函数:

$$\begin{aligned} L_{\text{NEW}} &= \|X - UV^T\|_F^2 + \lambda \text{Tr}[V^T (\alpha L_w + (1-\alpha) W_b) V] \\ s. t. U &\geq 0, V \geq 0 \end{aligned} \quad (18)$$

经过简单的推导, 我们用矩阵的迹表示为:

$$\begin{aligned} L_{\text{NEW}} &= \text{Tr}((X - UV^T)(X - UV^T)^T) + \lambda \text{Tr}(V^T (\alpha L_w + (1-\alpha) W_b) V) \\ &= \text{Tr}(X X^T) - 2 \text{Tr}(X V U^T) + \text{Tr}(U V^T V U^T) + \lambda \text{Tr} \\ &\quad (V^T (\alpha L_w + (1-\alpha) W_b) V) \\ s. t. U &\geq 0, V \geq 0 \end{aligned} \quad (19)$$

令

$$\begin{aligned} L &= \alpha L_w + (1-\alpha) W_b = \alpha (D_w - W_w) + (1-\alpha) W_b \\ &= (\alpha D_w + W_b) - \alpha (W_w + W_b) \\ &= (1 \quad 1) \begin{pmatrix} \alpha D_w + W_b & 0 \\ 0 & -\alpha (W_w + W_b) \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix} \\ &= L_1 - L_2 \end{aligned}$$

在这里

$$L_1 = (1 \quad 1) \begin{pmatrix} \alpha D_w + W_b & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

$$L_2 = (1 \quad 1) \begin{pmatrix} 0 & 0 \\ 0 & \alpha (W_w + W_b) \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

则式(19)为:

$$L_{NEW} = \text{Tr}(XX^T) - 2\text{Tr}(XVU^T) + \text{Tr}(UV^TVU^T) + \lambda \text{Tr}(V^TLV) \quad (20)$$

$$s. t. U \geq 0, V \geq 0$$

利用 Lagrange 定理有:

$$J_{NEW} = \text{Tr}(XX^T) - 2\text{Tr}(XVU^T) + \text{Tr}(UV^TVU^T) + \lambda \text{Tr}(V^TLV) + \text{Tr}(\Psi U^T) + \text{Tr}(\Phi V^T)$$

J_{NEW} 分别对 U 和 V 求偏微分得:

$$\frac{\partial J_{NEW}}{\partial U} = -2XV + 2UV^TV + \Psi \quad (21)$$

$$\frac{\partial L_{NEW}}{\partial V} = -2X^TU + 2VU^TU + 2\lambda LV + \Phi \quad (22)$$

使用 KKT 条件 $\phi_{ij}u_{ij} = 0$ 和 $\phi_{ij}v_{ij} = 0$, 得到如下关于 u_{ij} 和 v_{ij} 的等式:

$$u_{ij}^{(t+1)} \leftarrow u_{ij}^{(t)} \frac{(XV)_{ij}}{(UV^TV)_{ij}} \quad (23)$$

$$v_{ij}^{(t+1)} \leftarrow v_{ij}^{(t)} \frac{(X^TU + \lambda L_2 V)_{ij}}{(VU^TU + \lambda L_1 V)_{ij}} \quad (24)$$

2.4 算法收敛性分析

在这一部分, 我们研究算法的收敛性, 首先引入辅助函数的定义。

定义 1 设 $G(v, v')$ 为 $F(v)$ 辅助函数, 满足:

$$G(v, v') \geq F(v), G(v, v) = F(v) \quad (25)$$

引理 1 如果 $v^{(t+1)} = \argmin G(v, v')$, 则 F 是一个非增函数。

证明:

$$F(v^{(t+1)}) \leq G(v^{(t+1)}, v^{(t)}) \leq G(v^{(t)}, v^{(t)}) = F(v^{(t)})$$

因为只有当 $v^{(t)}$ 为 $G(v, v^{(t)})$ 的局部极小点时, $F(v^{(t+1)}) = F(v^{(t)})$ 。如果目标函数 F 的导数存在且在 $v^{(t)}$ 的一个极小邻域内连续, 则这也意味着微分 $\nabla F(v^{(t)}) = 0$ 。因此若每次取 $v^{(t+1)} = \argmin G(v, v^{(t)})$ 进行迭代, 最终会收敛到一个局部极小点 $v_{\min} = \argmin_h F(v)$:

$$F(v_{\min}) \leq \dots \leq F(v^{(t+1)}) \leq F(v^{(t)}) \leq \dots \leq F(v^{(1)}) \leq F(v^{(0)})$$

对于 V 中任何元素 $v_{\mathcal{A}}$, 我们用 $F_{\mathcal{A}}$ 表示 L_{NEW} 中仅与 $v_{\mathcal{A}}$ 有关的部分, 很容易得到:

$$F'_{\mathcal{A}} = \left(\frac{\partial L_{NEW}}{\partial V} \right)_{\mathcal{A}} = \begin{pmatrix} -2X^TU + \\ 2VU^TU + 2\lambda LV \end{pmatrix}_{\mathcal{A}} \quad (26)$$

$$F''_{\mathcal{A}} = 2(U^TU)_{\mathcal{A}\mathcal{A}} + 2\lambda L_{\mathcal{A}\mathcal{A}} \quad (27)$$

引理 2

$$G(v, v_{\mathcal{A}}^{(t)}) = F_{\mathcal{A}}(v_{\mathcal{A}}^{(t)}) + F'_{\mathcal{A}}(v_{\mathcal{A}}^{(t)})(v - v_{\mathcal{A}}^{(t)}) + \frac{(VU^TU)_{\mathcal{A}} + \lambda(L_1 V)_{\mathcal{A}}}{v_{\mathcal{A}}^{(t)}}(v - v_{\mathcal{A}}^{(t)})^2 \quad (28)$$

是 $F_{\mathcal{A}}$ 的辅助函数。

由于 $G(v, v) = F_{\mathcal{A}}(v)$ 是显而易见的, 我们只须证明 $G(v, v_{\mathcal{A}}^{(t)}) \geq F_{\mathcal{A}}(v)$, 为了证明这个问题, 我们比较 $F_{\mathcal{A}}(v)$ 的泰勒展开式:

$$F_{\mathcal{A}}(v) = F_{\mathcal{A}}(v_{\mathcal{A}}^{(t)}) + F'_{\mathcal{A}}(v_{\mathcal{A}}^{(t)})(v - v_{\mathcal{A}}^{(t)}) + [U^TU]_{\mathcal{A}} + \lambda L_{\mathcal{A}}](v - v_{\mathcal{A}}^{(t)})^2 \quad (29)$$

与式(29)比较后得到 $G(v, v_{\mathcal{A}}^{(t)}) \geq F_{\mathcal{A}}(v)$ 等价于:

$$\frac{(VU^TU)_{\mathcal{A}} + \lambda(L_1 V)_{\mathcal{A}}}{v_{\mathcal{A}}^{(t)}} \geq (U^TU)_{\mathcal{A}} + \lambda L_{\mathcal{A}} \quad (30)$$

我们有:

$$(VU^TU)_{\mathcal{A}} = \sum_{l=1}^k v_{\mathcal{A}}^{(l)} (U^TU)_{\mathcal{A}} \geq v_{\mathcal{A}}^{(t)} (U^TU)_{\mathcal{A}} \quad (31)$$

并且

$$(\lambda L_1 V)_{\mathcal{A}} = \lambda \sum_{l=1}^k (L_1)_{\mathcal{A}} v_{\mathcal{A}}^{(l)} \geq (L_1)_{\mathcal{A}} v_{\mathcal{A}}^{(t)} \geq (L_1 - L_2)_{\mathcal{A}} v_{\mathcal{A}}^{(t)} = L_{\mathcal{A}\mathcal{A}} v_{\mathcal{A}}^{(t)}$$

引理 3 通过最小化辅助函数 $G(v, v_{\mathcal{A}}^{(t)})$ 可以得到式(24), $v_{\mathcal{A}}^{(t)}$ 是第 t 次循环的解。

证明: 为了得到最优解, 辅助函数 $G(v, v_{\mathcal{A}}^{(t)})$ 对 $v_{\mathcal{A}}$ 进行微分有:

$$\frac{\partial G(v, v_{\mathcal{A}}^{(t)})}{\partial v} = F'_{\mathcal{A}}(v_{\mathcal{A}}^{(t)}) + \frac{2(VU^TU + \lambda L_1 V)_{\mathcal{A}}}{v_{\mathcal{A}}^{(t)}}(v - v_{\mathcal{A}}^{(t)}) = 0 \quad (32)$$

因此我们能够得到 V 的更新规则为:

$$\begin{aligned} & (-2X^TU + 2VU^TU + 2\lambda LV)_{\mathcal{A}} + \\ & \frac{2(VU^TU + \lambda L_1 V)_{\mathcal{A}}}{v_{\mathcal{A}}^{(t)}}(v - v_{\mathcal{A}}^{(t)}) = 0 \\ & -X^TU + VU^TU + \lambda LV + \frac{VU^TU + \lambda L_1 V}{v_{\mathcal{A}}^{(t)}}(v - v_{\mathcal{A}}^{(t)}) = 0 \\ & -X^TU + VU^TU + \lambda LV + \frac{VU^TU + \lambda L_1 V}{v_{\mathcal{A}}^{(t)}} \\ & v - VU^TU - \lambda LV_1 = 0 \\ & -X^TU + \lambda LV + \frac{VU^TU + \lambda L_1 V}{v_{\mathcal{A}}^{(t)}}v - \lambda LV_1 = 0 \\ & -X^TU - \lambda L_2 V + \frac{VU^TU + \lambda L_1 V}{v_{\mathcal{A}}^{(t)}}v = 0 \\ & \frac{VU^TU + \lambda L_1 V}{v_{\mathcal{A}}^{(t)}}v = X^TU + \lambda L_2 V \\ & v = v_{\mathcal{A}}^{(t)} \frac{X^TU + \lambda L_2 V}{VU^TU + \lambda L_1 V} \end{aligned}$$

所以更新规则为:

$$v_{\mathcal{A}}^{(t+1)} = v_{\mathcal{A}}^{(t)} \frac{X^TU + \lambda L_2 V}{VU^TU + \lambda L_1 V}$$

同理可得 $u_{\mathcal{A}}^{(t+1)}$ 的更新规则为:

$$u_{ij}^{(t+1)} = u_{ij}^{(t)} \frac{(XV)_{ij}}{(UV^TV)_{ij}}$$

3 实验

为了验证所提出的局部敏感非负矩阵分解算法的有效性, 实验比较了该算法与几种经典的子空间学习算法, 比较的指标是降维后低维投影在分类问题上的效果好坏。这些对比的算法包括基线方法(Baseline)(直接在原始输入空间上使用最近邻法)、主成分分析(PCA)、线性判别分析(LDA)、保局投影算法(LPP)、边际费舍尔分析(MFA)、标准非负矩阵分解NMF、局部非负矩阵分解(LNMF)。实验采用了两个比较著名的人脸数据库, 分别是 ORL 人脸库和 Yale 人脸库。在实验中, 所有人脸图像都经过了标准化处理, 即首先标定出人脸图像中两眼的位置并在该位置将图像对齐, 然后剪切出图像中的面部区域并将其缩放成统一的 32×32 大小。图像最终按列堆叠成 1024 维的向量。在实验中, 我们首先利用这些算法将数据降低至低维空间, 然后用简单的最近邻分类器在低维特征空间对其进行分类。表 1 给出了不同算法在 ORL 数据库上的识别率; 表 2 给出了不同算法在 Yale 数据库上的识别率。

表1 不同算法在 ORL 数据库上的识别率

	2 Train	3 Train	4 Train
Baselines	66.8%(1024)	77%(1024)	81.7%(1024)
Eigenfaces	66.8%(78)	77%(119)	81.7%(159)
Fisherfaces	71.1%(22)	83.2%(39)	89.5%(39)
MFA	71.6%(37)	84.1%(39)	89.6%(39)
LPP	76.1%(39)	86.6%(39)	90.42%(39)
NMF	70.87%(97)	78.98%(81)	84.48%(95)
LNMF	71.73%(178)	81.09%(168)	86.31%(195)
LSDA	76.7%(39)	85%(39)	90.5%(39)
LSDNMF	78.7%(198)	87.9%(89)	93.5%(83)

表2 不同算法在 Yale 数据库上的识别率

	2 Train	3 Train	4 Train
Baselines	43.4%(1024)	49.4%(1024)	56.2%(1024)
Eigenfaces	43.4%(29)	49.4%(44)	56.2%(58)
Fisherfaces	47.2%(9)	64.9%(14)	72.9%(14)
MFA	47.7%(10)	65.7%(14)	74.1%(14)
LPP	46.5%(14)	69.5%(14)	74.6%(14)
NMF	44.11%(112)	49.00%(195)	52.19%(164)
LNMF	44.00%(157)	48.83%(198)	53.57%(14)
LSDA	56.5%(14)	68.5%(14)	74.4%(14)
LSDNMF	58.7%(108)	70.5%(128)	78.35%(146)

数据集

ORL 人脸库包括 40 个人的 400 幅 112×92 大小的人脸灰度图像,其中有些图像是拍摄于不同时期的,人的脸部表情和脸部细节有着不同程度的变化;比如,笑或不笑,眼睛或睁或闭,戴或不戴眼镜。人脸姿态也有相当程度的变化,深度旋转和平面旋转可达 20° ,人脸的尺度也有多达 10% 的变化。Yale 人脸库中包括了 15 个人的 165 幅灰度人脸图像,每个人由 11 幅 92×112 的照片构成,这些图片涵盖了光照(包括左、中、右 3 个方向)、面部表情(包括正常、高兴、难过、困乏、惊讶以及眨眼)和遮挡(戴/不戴眼镜)等多种因素的变化。为了节省时间和内存,图像被预处理成为 32×32 维的形式,对图像的灰度不做任何处理。图 1 显示了 ORL, Yale 人脸数据库中的部分人脸样本。



图1 顶行是 ORL 数据集,底行是 Yale 数据集

为了避免小样本问题,我们首先采用 PCA 算法对数据进行降维。对于 LPP 算法,所保留的 PCA 主分量个数为 $n-1$;而对于 LDA 和 MFA 算法,所保留的 PCA 主分量的个数为 $n-N_c$,这里 n 是训练数的个数, N_c 是类数。对于非负矩阵 NMF 和 LNMF,参数 k 为 $n \times m / (n+m)$,我们所提算法的近

邻参数 $k_1=2, k_2=2$ 。

结束语 实验表明,与传统的线性降维方法相比,基于局部敏感判别非负矩阵分解的学习方法通常能够得到更好的识别效果。PCA, LDA, NMF, LNMF 降维仅仅考虑了线性的欧氏结构而没有考虑全局的非线性属性。LPP, MFA, SDA 算法采用保持数据的局部几何结构,能够在一定程度上刻画数据的非线性结构,因而在实验中能够得到较优的识别性能。本文所提出的 SDNMF 算法在实验中表现最好,这主要是因为 SDNMF 同时考虑了数据的局部几何结构、判别信息以及分解结果的非负性,从而得到了最优的识别性能。

参考文献

- [1] Lee D D, Seung H S. Learning the parts of objects by non-negative matrix factorization[J]. Nature, 1999, 401(6755): 788-791
- [2] Li S Z, Hou X W, Zhang H J, et al. Learning spatially localized, parts-based representation[C]// Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. 2001, 1: 207-212
- [3] Hoyer P O. Non-negative sparse coding[A]// Proc. IEEE Workshop on Neural Netw. Signal Process [C]. Martigny, Valais, Switzerland: IEEE, 2002: 557-565
- [4] Liu W X, Zheng N N, Lu X F. Non-negative matrix factorization for visual coding[C]// Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing. 2003, 3: 293-296
- [5] Wang Yuan, Jia Yunde, Hu Changbo, et al. Fisher non-negative matrix factorization for learning local features[C]// ACCV. 2004
- [6] Guillemet D, Vitria J, Schiele B. Introducing a weighted non-negative matrix factorization for image classification[J]. Pattern Recognition Letters, 2003, 24(14): 2447-2454
- [7] Ding C, Li T, Peng W, et al. Orthogonal nonnegative matrix t-factorizations for clustering[C]// Proceedings of the 12th ACM SIGKDD International Conference. 2006
- [8] Roweis S T, Saul L K. Nonlinear dimensionality reduction by locally linear embedding[J]. Science, 2000, 290(5500): 2323-2326
- [9] Cai Deng, He Xiaofei, Wu Xiaoyun, et al. Non-negative matrix factorization on manifold[C]// ICDM. 2008
- [10] He Xiao-fei, Yan Shui-cheng. Face recognition using laplacian-faces[J]. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2005, 27(3): 328-340
- [11] Cai Deng, He Xiao-fei. Locality sensitive discriminant analysis [C]// International Joint Conference on Artificial Intelligence. 2007: 708-713

(上接第 200 页)

- [10] Fan Shi-qing, Zhang Wen-xiu, Wei Xu. Fuzzy Inference based on fuzzy concept lattice[J]. Fuzzy sets and systems, 2006, 157: 3177-3187
- [11] Shao M W, Liu M, Zhang W X. Set approximations in fuzzy formal concept analysis[J]. Fuzzy Sets and Systems, 2007, 158 (23): 2627-2640
- [12] Belohlavek R, Vychodil V. Reducing the size of fuzzy concept lattices by hedges[C]// FUZZ-IEEE 2005, The IEEE Interna-

tional Conference on Fuzzy Systems. Reno, Nevada, USA, 2005: 663-668

- [13] Belohlavek R, Sklenar V. Formal concept analysis constrained by attribute-dependency formulas[J]. Lecture Notes in Artificial Intelligence, 2005, 3403: 176-191
- [14] Zhang Wen-xiu, Ma Jian-min, Fan Shi-qing. Variable threshold concept lattice[J]. Information Sciences, 2007, 177 (22): 4883-4892