

改进型低秩线性回归新方法

于传波 聂仁灿 周冬明 黄帆 丁婷婷

(云南大学信息学院 昆明 650500)

摘要 对于遮挡、光照等影响因素,低秩线性回归模型具有很好的鲁棒性。LRRR(Low Rank Ridge Regression)以及 DENLR(Discriminative Elastic-net Regularized Linear Regression)通过正则化系数矩阵在一定程度上减少了 LRLR(Low Rank Linear Regression)产生的过拟合现象。但其没有考虑子空间数据的错误逼近,投影矩阵不能准确地将数据映射到目标空间。鉴于此,提出了一种运算更快、更具判别性的低秩线性回归分类新方法。首先,将 0-1 构成的矩阵作为线性回归的目标值;其次,利用核范数作为低秩约束的凸近似;然后,通过正则化各类别之间的距离矩阵和模型输出矩阵来降低过拟合,同时可以增强投影子空间的判别性;再次,利用增广拉格朗日乘子(Augmented Lagrangian Multiplier, ALM)优化目标函数;最后,在子空间中利用最近邻分类器进行分类。在 AR、FERET 人脸数据库、Stanford 40 Actions、Caltech-UCSD Bird 以及 Oxford 102 Flowers 数据库上进行相关算法的对比实验,结果表明所提算法是有效的。

关键词 低秩,线性回归,正则化,分类

中图法分类号 TP391.41 文献标识码 A

Novel Method of Improved Low Rank Linear Regression

YU Chuan-bo NIE Ren-can ZHOU Dong-ming HUANG Fan DING Ting-ting

(School of Information Science and Engineering, Yunnan University, Kunming 650500, China)

Abstract The low rank linear regression model has good robustness for the influence of occlusion and illumination and so on. To a certain extent, the overfitting phenomenon of LRLR (Low Rank Linear Regression) is reduced in LRRR (Low Rank Ridge Regression) and DENLR (Discriminative Elastic-Net Regularized Linear Regression) by regularization coefficient matrix. Because the error approximation of data in subspace is ill-considered, the data are hardly mapped to the target space accurately via projection matrix. This paper proposed has low rank linear regression classification method which has a faster computing speed and is more discriminative. Firstly, the 0-1 constitutive matrix is regarded as the target value of the linear regression. Secondly, the kernel norm is used as the convex approximation of low rank constraints. Thirdly, all kinds of the distance matrix and the model output matrix are regularized to reduce overfitting phenomenon, at the same time it can enhance the spatial discriminant of projection subspace. Then, the augmented Lagrange multiplier (ALM) is used to optimize the objective function. Finally, the nearest neighbor classifier is used for classification in subspace. We compared the related algorithms on AR, FERET face database, Stanford 40 Actions database, Caltech-UCSD Birds database and Oxford 102 Flowers database. The experimental results show that the proposed algorithm is effective.

Keywords Low rank, Linear regression, Regularization, Classification

1 引言

回归模型方法是数据挖掘和机器学习领域的重要方法之一,被广泛用于解决计算机视觉问题。回归模型旨在通过拟合方法建立线性关系来得到投影矩阵从而观测数据,并对数据进行分类。

线性回归模型有两大受制约的因素:1)分析高维数据的能力不足。样本和标签不易采集,除此之外,通常会遇到 curse-of-dimensionality 问题^[1]使得线性回归模型不能很好地

分析高维数据,但可以使用降维算法来降低特征的维度,从而有效地解决上述问题。常用的降维算法有 Linear Discriminant Analysis (LDA)^[2], Locality Preserving Projection (LPP)^[3]等。2)线性回归模型没有强调不同类别线性预测之间的关联,而且鲁棒性问题、判别性问题也没有得到更有效的解决。

最小二乘回归(Least Square Regression, LSR)公式简单、求解容易,在计算机视觉和模式识别中应用广泛^[4]。在最小二乘回归的基础上研究者相继提出了权重最小线性回归

本文受国家自然科学基金(61365001,61463052)资助。

于传波(1991-),男,硕士生,主要研究方向为模式识别、图像处理, E-mail: 18213018302@163.com;聂仁灿(1982-),男,博士,副教授,主要研究方向为图像处理、机器学习等;周冬明(1963-),男,博士,教授,主要研究方向为神经网络、优化计算、信号处理技术等, E-mail: zhoum@ynu.edu.cn(通信作者)。

(weighted LSR)^[5]、非负最小二乘 (Nonnegative Least Squares, NLS)^[6]等方法。此外,在回归模型的基础上也有学者提出了增强判别性的最小二乘回归,例如判别性最小二乘回归(Discriminative Least Square Regression, DLSR)^[7]。文献[8]陈述了稀疏表示的问题在最小值收缩问题上与最小二乘也有很大的关联性。稀疏表示分类器(Sparse Representation based Classification, SRC)^[9]是最早应用于人脸识别的稀疏表示方法,测试图像由所有训练图像线性表示,根据每一类的残差进行分类,将测试样本的标签归为最小残差那一类别。SRC采用了正则化导致计算复杂度较高。协同表示的分类器(Collaborative Representation based Classification method, CRC)指出, SRC 识别效果好的原因不在于 ℓ_1 范数,而是样本的协同表示^[10]。CRC 用 ℓ_2 范数替代 ℓ_1 范数,在不降低识别率的前提下大大提高了运算的速度。

低秩方法可以有效地恢复图像,减少图像的噪声和阴影,并使算法的运算速度大大提高。基于低秩最小化可以更好地了解底层数据的关联结构。鲁棒性主成分分析(Robust Principal Component Analysis, RPCA)^[11]是低秩方法的一个重要应用,可以将图像分解成“干净”图像和稀疏噪声图像。由于低秩独特的特征,低秩回归模型相继被提出^[12-13]。基于传统的低秩线性回归模型,文献[12]提出了低秩岭回归 LRR 和稀疏低秩回归(Sparse Low-Rank Regression, SLRR)。文献[13]提出了判别性弹性网正则化线性回归 DENLR 和边缘弹性网正则化线性回归(Marginalized Elastic-Net Regularized Linear Regression, MENLR)。相比于传统的低秩线性回归模型, LRR, SLRR, DENLR 和 MENLR 都加入了正则化模型输出矩阵,有效地降低了过拟合的现象。但是上述几种方法并没有考虑到测试样本在投影子空间会出现错误逼近的情况,本文针对该问题在低秩线性回归模型的基础上结合样本类别信息提出了判别性低秩线性回归模型。所提方法可以有效地降低过拟合带来的错误分类。除此之外,由于在目标函数中加入了类与类之间的判别信息,可以有效地减少测试样本在子空间的错误逼近。本文采用增广拉格朗日乘子 ALM 法来优化函数,将样本数据通过投影矩阵转移到投影子空间,用简单的最近邻分类器进行分类。

2 线性回归和低秩线性回归模型

假设有 K 类训练样本,其中每类训练样本数为 N_k ,总训练样本数为 $N = N_1 + N_2 + \dots + N_K$ 。 $\mathbf{x}_{i,j} \in R^{d \times 1}$ 表示第 i 类的第 j 个样本数据,且维度为 d 的列向量。第 i 类所有训练样本由 $\mathbf{X}_i = [\mathbf{x}_{i,1}, \mathbf{x}_{i,2}, \dots, \mathbf{x}_{i,N_i}] \in R^{d \times N_i}$ 表示,则总样本 $\mathbf{X} = [\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_K] \in R^{d \times N}$ 。

2.1 线性回归模型

传统的用于分类的线性回归模型是通过解决式(1)问题,将训练样本的特征尽可能地逼近目标矩阵:

$$\min_D \| \mathbf{Y} - \mathbf{X}^T \mathbf{D} \|_F^2 \tag{1}$$

其中, $\mathbf{Y} \in R^{N \times K}$ 是表示类别的目标矩阵(比如, $Y_{i,j} = 1$ 表示的是第 i 个数据属于第 j 类,且该数据属于其他类别的用 0 表示), $\mathbf{D} \in R^{d \times K}$ 表示投影矩阵(模型输出系数矩阵)。将测试样本通过投影矩阵 $\mathbf{x}^T \mathbf{D}$ 转移到子空间进行判别预测。

2.2 低秩线性回归模型

低秩线性回归 LRLR^[12]是标准线性回归模型的改进模

型。LRLR 在传统线性回归模型的基础上增加了对投影矩阵秩的限制,投影矩阵的秩 $s < \min(N, K)$,可以有效地去除样本数据中无用的信息(噪声、光照阴影、重复类似样本)。LRLR 目标函数如下:

$$\min_D \| \mathbf{Y} - \mathbf{X}^T \mathbf{D} \|_F^2, \text{ s. t. } \text{rank}(\mathbf{D}) < s \tag{2}$$

假设 $\mathbf{D} = \mathbf{A}\mathbf{B}$ ($\mathbf{A} \in R^{d \times s}, \mathbf{B} \in R^{s \times K}$),那么式(2)可以转化为:

$$\min_D \| \mathbf{Y} - \mathbf{X}^T \mathbf{A}\mathbf{B} \|_F^2 \tag{3}$$

式(2)由于引入了低秩的限制,因此是一个非凸、非光滑的目标函数,文献[14]指出低秩限制条件可以通过核范数形式替代。式(2)还可以改写成:

$$\min_D \| \mathbf{Y} - \mathbf{X}^T \mathbf{D} \|_F^2 + \lambda \| \mathbf{D} \|_* \tag{4}$$

其中正则化参数 $\lambda (\lambda > 0)$ 起到使解更加稳定以及平衡矩阵 \mathbf{D} 的秩和线性回归目标矩阵关系的作用。

与 LRLR 相比, Low-rank ridge regression (LRRR) 在 LRLR 的基础上加入了对投影矩阵 \mathbf{D} 的正则化,从而有效地缓解了 LRLR 中的过拟合现象。LRRR 目标函数如下:

$$\min_D \| \mathbf{Y} - \mathbf{X}^T \mathbf{A}\mathbf{B} \|_F^2 + \lambda \| \mathbf{A}\mathbf{B} \|_F^2 \tag{5}$$

理论上 LRRR 模型也可以写成:

$$\min_D \| \mathbf{Y} - \mathbf{X}^T \mathbf{D} \|_F^2 + \lambda_1 \| \mathbf{D} \|_* + \lambda_2 \| \mathbf{D} \|_F^2 \tag{6}$$

从式(6)可以看出, LRRR 在一定程度上与文献[13]所提出的广义上的弹性网络正则化线性回归模型是一致的。

3 判别性低秩线性回归

3.1 判别性低秩线性回归目标函数的建立

虽然 LRRR, SLRR, DENLR 和 MENLR 等方法在一定程度上减小了过拟合现象带来的影响,但是其只是针对投影矩阵(模型输出系数矩阵)的正则化。而且,上述几种方法都只考虑样本点之间的关联性,忽视了样本类与类之间的整体性,从而导致待测样本在投影子空间极可能错误逼近。从图 1 可以更直观地看出测试样本投影到子空间后,得到的数据为 $[0 \ 0 \ 0 \ 0.2 \ 0.4 \ 0.4]$ 。我们很难判断出测试样本属于第 5 类还是第 6 类,这就导致测试样本很难选择正确的类别。针对以上不足,本文提出了判别性低秩线性回归。其目标函数如下:

$$\min_D \| \mathbf{Y} - \mathbf{X}^T \mathbf{D} \|_F^2 + \alpha \| \mathbf{D} \|_F^2 - \beta \sum_{i=1}^K \sum_{j=1}^K \| \mathbf{X}_i^T \mathbf{D} - \mathbf{X}_j^T \mathbf{D} \|_F^2 \tag{7}$$

由于式(7)是一个 NP-hard 问题,我们用和函数替代低秩限制条件,得到:

$$\min_D \| \mathbf{Y} - \mathbf{X}^T \mathbf{D} \|_F^2 + \lambda_1 \| \mathbf{D} \|_* + \frac{\lambda_2}{2} \| \mathbf{D} \|_F^2 - \frac{\lambda_3}{2} \sum_{i=1}^K \sum_{j=1}^K \| \mathbf{X}_i^T \mathbf{D} - \mathbf{X}_j^T \mathbf{D} \|_F^2 \tag{8}$$

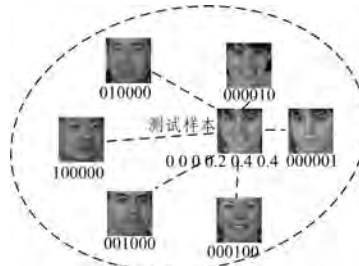


图 1 待测样本错误逼近

将投影矩阵 \mathbf{D} 进行奇异值分解(Singular Value Decom-

position, SVD), 可以理解投影的本质。 $\mathbf{D} = \mathbf{U}\Sigma\mathbf{V}^T = \sum_{i=1}^s \mathbf{u}_i \alpha_i \mathbf{v}_i^T$, 其中 \mathbf{u}_i 和 \mathbf{v}_i 分别表示奇异值分解的左右向量, α_i 表示奇异值。数据投影到子空间可以表示为: $\mathbf{x}^T \mathbf{D} = \sum_{i=1}^m \alpha_i (\mathbf{x}^T \mathbf{u}_i) \mathbf{v}_i^T$, 其中 m 为奇异值选择的个数。因此, 可以将目标子空间看作权值的一种线性组合。在文献[13]中已经证明了 $\|\mathbf{D}\|_* = \min_{\mathbf{D}=\mathbf{AB}} \|\mathbf{A}\|_F \|\mathbf{B}\|_F = \min_{\mathbf{D}=\mathbf{AB}} \frac{1}{2} (\|\mathbf{A}\|_F^2 + \|\mathbf{B}\|_F^2)$, 因此式(7)可以改写成:

$$\min_{\mathbf{D}} \|\mathbf{Y} - \mathbf{X}^T \mathbf{D}\|_F^2 + \frac{\lambda_1}{2} (\|\mathbf{A}\|_F^2 + \|\mathbf{B}\|_F^2) + \frac{\lambda_2}{2} \|\mathbf{D}\|_F^2 - \frac{\lambda_3}{2} \sum_{i=1}^K \sum_{j=1}^K \|\mathbf{X}_i^T \mathbf{D} - \mathbf{X}_j^T \mathbf{D}\|_F^2 \quad (9)$$

s. t. $\mathbf{D} = \mathbf{AB}$

3.2 目标函数的优化

因为式(9)附有低秩 $\mathbf{D} = \mathbf{AB}$ 的限制, 因此属于非凸、非光滑问题。我们利用增广拉格朗日乘子 ALM^[15] 来迭代优化目标函数。式(9)的拉格朗日乘子函数可以表示为:

$$\min_{\mathbf{D}, \mathbf{A}, \mathbf{B}, \mathbf{M}} \|\mathbf{Y} - \mathbf{X}^T \mathbf{D}\|_F^2 + \frac{\lambda_1}{2} (\|\mathbf{A}\|_F^2 + \|\mathbf{B}\|_F^2) + \frac{\lambda_2}{2} \|\mathbf{D}\|_F^2 - \frac{\lambda_3}{2} \sum_{i=1}^K \sum_{j=1}^K \|\mathbf{X}_i^T \mathbf{D} - \mathbf{X}_j^T \mathbf{D}\|_F^2 + \langle \mathbf{M}, \mathbf{D} - \mathbf{AB} \rangle + \frac{\lambda_4}{2} \|\mathbf{D} - \mathbf{AB}\|_F^2 \quad (10)$$

其中, $\langle \mathbf{M}, \mathbf{D} - \mathbf{AB} \rangle = \text{tr}(\mathbf{M}^T (\mathbf{D} - \mathbf{AB}))$, \mathbf{M} 表示拉格朗日乘子。以上惩罚参数都是选取大于 0 的值。

更新矩阵 \mathbf{A} 。假设除矩阵 \mathbf{A} 外的其他变量固定不变, 式(9)变为典型的正则化最小二乘问题:

$$\begin{aligned} f(\mathbf{A}) &= \arg \min_{\mathbf{A}} \frac{\lambda_1}{2} \|\mathbf{A}\|_F^2 + \langle \mathbf{M}, \mathbf{D} - \mathbf{AB} \rangle + \frac{\lambda_4}{2} \|\mathbf{D} - \mathbf{AB}\|_F^2 \\ &= \arg \min_{\mathbf{A}} \frac{\lambda_1}{2} \|\mathbf{A}\|_F^2 + \frac{\lambda_4}{2} \|\mathbf{D} - \mathbf{AB} + \frac{\mathbf{M}}{\lambda_4}\|_F^2 \end{aligned} \quad (11)$$

通过求导, 可得到优化后的矩阵 \mathbf{A} 。

$$\frac{\partial f(\mathbf{A})}{\partial \mathbf{A}} = \lambda_1 \mathbf{A} + \lambda_4 (\mathbf{D} - \mathbf{AB} + \frac{\mathbf{M}}{\lambda_4}) (-\mathbf{B}^T) = 0$$

即 $\mathbf{A}_{\text{new}} = (\mathbf{M} + \lambda_4 \mathbf{D}) \mathbf{B}^T (\lambda_1 \mathbf{I} + \lambda_4 \mathbf{B} \mathbf{B}^T)^{-1}$, $\mathbf{I} \in \mathbb{R}^{s \times s}$ 是单位矩阵。

更新矩阵 \mathbf{B} 。假设除矩阵 \mathbf{B} 外的其余变量固定不变, 式(9)可写成:

$$\begin{aligned} f(\mathbf{B}) &= \arg \min_{\mathbf{B}} \frac{\lambda_1}{2} \|\mathbf{B}\|_F^2 + \langle \mathbf{M}, \mathbf{D} - \mathbf{AB} \rangle + \frac{\lambda_4}{2} \|\mathbf{D} - \mathbf{AB}\|_F^2 \\ &= \arg \min_{\mathbf{B}} \frac{\lambda_1}{2} \|\mathbf{B}\|_F^2 + \frac{\lambda_4}{2} \|\mathbf{D} - \mathbf{AB} + \frac{\mathbf{M}}{\lambda_4}\|_F^2 \end{aligned} \quad (12)$$

对 $f(\mathbf{B})$ 求导:

$$\frac{\partial f(\mathbf{B})}{\partial \mathbf{B}} = \lambda_1 \mathbf{B} + \lambda_4 (\mathbf{D} - \mathbf{AB} + \frac{\mathbf{M}}{\lambda_4}) (-\mathbf{A}^T) = 0$$

更新后 $\mathbf{B}_{\text{new}} = (\lambda_1 \mathbf{I} + \lambda_4 \mathbf{A} \mathbf{A}^T)^{-1} \mathbf{A}^T (\lambda_4 \mathbf{D} + \mathbf{M})$ 。

更新矩阵 \mathbf{D} 。假设除矩阵 \mathbf{D} 以外的其余变量均保持不变, 式(9)可变为:

$$\begin{aligned} f(\mathbf{D}) &= \arg \min_{\mathbf{D}} \|\mathbf{Y} - \mathbf{X}^T \mathbf{D}\|_F^2 + \frac{\lambda_2}{2} \|\mathbf{D}\|_F^2 + \frac{\lambda_3}{2} \sum_{i=1}^K \sum_{j=1}^K (\mathbf{X}_i^T \mathbf{D} - \mathbf{X}_j^T \mathbf{D})^2 + \langle \mathbf{M}, \mathbf{D} - \mathbf{AB} \rangle + \frac{\lambda_4}{2} \|\mathbf{D} - \mathbf{AB}\|_F^2 \\ &= \arg \min_{\mathbf{D}} \|\mathbf{Y} - \mathbf{X}^T \mathbf{D}\|_F^2 + \frac{\lambda_2}{2} \|\mathbf{D}\|_F^2 + \frac{\lambda_3}{2} \sum_{i=1}^K \sum_{j=1}^K (\mathbf{X}_i^T \mathbf{D} - \mathbf{X}_j^T \mathbf{D})^2 + \frac{\lambda_4}{2} \|\mathbf{D} - \mathbf{AB} - \frac{\mathbf{M}}{\lambda_4}\|_F^2 \end{aligned} \quad (13)$$

对 $f(\mathbf{D})$ 求导:

$$\frac{\partial f(\mathbf{D})}{\partial \mathbf{D}} = 2(\mathbf{X} \mathbf{X}^T \mathbf{D} - \mathbf{X} \mathbf{Y}) + \lambda_2 \mathbf{D} + \lambda_3 \sum_{i=1}^K \sum_{j=1}^K (\mathbf{X}_i - \mathbf{X}_j) (\mathbf{X}_i - \mathbf{X}_j)^T + \lambda_4 (\mathbf{D} - \mathbf{AB} + \frac{\mathbf{M}}{\lambda_4}) = 0$$

更新后得到:

$$\mathbf{D}_{\text{new}} = (2\mathbf{X} \mathbf{X}^T + \lambda_2 \mathbf{I} + \lambda_3 \sum_{i=1}^K \sum_{j=1}^K (\mathbf{X}_i - \mathbf{X}_j) (\mathbf{X}_i - \mathbf{X}_j)^T)^{-1} \times (\mathbf{X} \mathbf{Y} + \lambda_4 \mathbf{AB} - \mathbf{M})$$

算法的流程如图 2 所示。

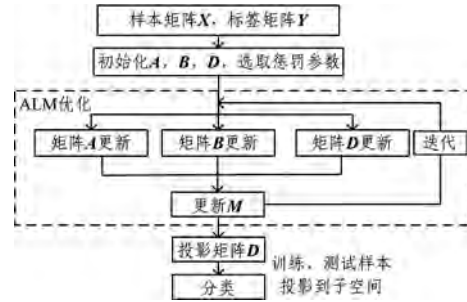


图 2 算法流程图

判别性低秩线性回归分类算法对某一测试样本 y 的分类步骤如算法 1 所示。

算法 1

输入: 训练样本 $\mathbf{X} \in \mathbb{R}^{d \times N}$ 及对应的标签矩阵 $\mathbf{Y} \in \mathbb{R}^{N \times K}$

输出: 优化后的投影矩阵 $\mathbf{D}(\mathbf{D}_{\text{new}})$

初始化: $\mathbf{D} \in \mathbb{R}^{d \times K}$, $\mathbf{A} \in \mathbb{R}^{d \times s}$, $\mathbf{B} \in \mathbb{R}^{s \times K}$, $\mathbf{M} \in \mathbb{R}^{d \times K}$, $\lambda_1 > 0, \lambda_2 > 0, \lambda_3 > 0, \lambda_4 > 0$

LSR 初始化 \mathbf{D}

for $i = 1$: iterations

根据公式 $\mathbf{A}_{\text{new}} = (\mathbf{M} + \lambda_4 \mathbf{D}) \mathbf{B}^T (\lambda_1 \mathbf{I} + \lambda_4 \mathbf{B} \mathbf{B}^T)^{-1}$ 更新矩阵 \mathbf{A}

根据公式 $\mathbf{B}_{\text{new}} = (\lambda_1 \mathbf{I} + \lambda_4 \mathbf{A} \mathbf{A}^T)^{-1} \mathbf{A}^T (\lambda_4 \mathbf{D} + \mathbf{M})$ 更新矩阵 \mathbf{B}

根据公式 $\mathbf{D}_{\text{new}} = (2\mathbf{X} \mathbf{X}^T + \lambda_2 \mathbf{I} + \lambda_3 \sum_{i=1}^K \sum_{j=1}^K (\mathbf{X}_i - \mathbf{X}_j) (\mathbf{X}_i - \mathbf{X}_j)^T)^{-1} (\mathbf{X} \mathbf{Y} + \lambda_4 \mathbf{AB} - \mathbf{M})$ 更新矩阵 \mathbf{D}

更新 \mathbf{M} : $\mathbf{M} = \mathbf{M} + \lambda_4 (\mathbf{D} - \mathbf{AB})$

end

将待测试样本映射到子空间 $\mathbf{y}^T \mathbf{D}$, 在子空间中通过最近邻分类器进行分类

4 实验仿真

为了验证本文算法的有效性, 将其与 LRC^[16], OMP^[17], Homotopy^[18], TPTSR^[19], CRC^[10], RPCA^[20], DENLR^[13], LRLR^[12], LRRR^[12], SLRR^[12] 共 10 种相关的模式识别方法在 AR^[21], FERET^[22] 人脸数据库, Stanford 40 Actions 数据库^[23], Caltech-UCSD Birds 数据库^[24] 以及 Oxford 102 Flowers 数据库^[25] 上进行算法的仿真比较, 其中 OMP, Homotopy, TPTSR, SPCA 几种方法用 SCR 做分类。实验中数据均先用 PCA^[25] 做处理, 保存 98% 的能量; 低秩相关算法实验中低秩的选择为 $s = K - 1$; 每次实验重复运行 10 次, 最终识别率为 10 次的均值。本文实验的运行环境为 Inter(R) Core(TM) i7-4500U CPU @ 1.80GHz 2.40GHz, 8.00GB 内存, Windows 8 操作系统, 软件是 MATLAB R2014a。实验数据库的部分样本图片如图 3 所示。



图3 数据库部分样本图片

4.1 AR人脸数据库

AR人脸数据共有126个类别。每人有26张人脸图片，其中有14张图片包含光照、表情，有12张为遮挡图片(围巾、墨镜遮挡)。本文选取了50个男性和50个女性的所有图片，共计100个类别，图片像素调整为 32×32 。实验中从每一类分别随机选取4,6,8,10张图片作为训练样本，剩余的作为测试样本，实验参数为： $\lambda_1 = 0.01, \lambda_2 = 0.01, \lambda_3 = 0.01, \lambda_4 = 10^{-8}$ 。实验结果如表1所列。

表1 AR数据库对比实验

算法	4Train	6Train	8Train	10Train
LRLR	0.8070	0.8333	0.8664	0.8893
LRRR	0.8007	0.8843	0.9289	0.9598
DENLR	0.8674	0.9171	0.9379	0.9491
CRC	0.7587	0.8467	0.8979	0.9328
SLRR	0.8201	0.9144	0.9450	0.9656
LRC	0.3793	0.5425	0.6698	0.7608
OMP	0.7495	0.8500	0.9133	0.9325
Homotopy	0.6545	0.7525	0.8294	0.8738
TPTSR	0.7673	0.8420	0.9161	0.9475
RPCA	0.7802	0.8693	0.9098	0.9373
本文算法	0.8476	0.9272	0.9516	0.9692

为了清楚地观察低秩的大小对实验结果的影响，以AR26数据库6Train为例，给出基于低秩线性回归方法的识别率曲线，如图4所示。从图中可以清楚地看出，随着低秩大小 s 的改变，本文方法与其他几种低秩线性回归方法相比识别率是最高的。

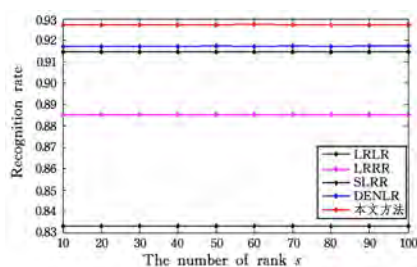


图4 低秩大小对线性回归的影响

4.2 FERET人脸数据库

FERET数据库是人脸识别领域应用最广泛的人脸数据库之一。其包含200人，每人7张共计1400张多姿态、光照的灰度人脸图像。本文选取的图片像素大小为 80×80 。实验中从每一类随机选取5张和6张作为训练样本，剩余的用于测试。实验参数为： $\lambda_1 = 0.1, \lambda_2 = 0.1, \lambda_3 = 0.005, \lambda_4 = 10^{-8}$ 。实验结果如表2所列。

表2 FERET数据库对比实验结果

算法	5Train	6Train	算法	5Train	6Train
LRLR	0.2291	0.2588	OMP	0.6725	0.7293
LRRR	0.7455	0.8193	Homotopy	0.6224	0.7430
DENLR	0.7800	0.7915	TPTSR	0.7040	0.7430
CRC	0.6159	0.6553	RPCA	0.6091	0.6457
SLRR	0.7775	0.8172	本文算法	0.7925	0.8290
LRC	0.4940	0.5490			

4.3 Stanford 40 Actions 数据库

Stanford 40 Actions数据库包含40种动作(刷牙、扫地、读书等)，本文选取每一个动作样本100张，共计4000张作为实验数据集。每张图片采用VGG-verydeep-19^[26]提取了VGG19特征，作为数据向量。实验中从每类分别选取了30,40,50张图片作为训练集，将剩余样本作为测试集。实验参数如下： $\lambda_1 = 0.1, \lambda_2 = 0.1, \lambda_3 = 0.0001, \lambda_4 = 10^{-8}$ 。平均实验结果如表3所列。

表3 Stanford 40 Actions数据库对比实验结果

算法	30Train	40Train	50Train
LRLR	0.6314	0.6483	0.6795
LRRR	0.7600	0.7750	0.7850
DENLR	0.7650	0.7800	0.8020
CRC	0.7646	0.7750	0.7895
SLRR	0.7564	0.7713	0.8031
LRC	0.7018	0.7263	0.7520
OMP	0.7360	0.7538	0.7795
Homotopy	0.7421	0.7517	0.7711
TPTSR	0.7546	0.7746	0.7925
RPCA	0.7389	0.7483	0.7605
本文算法	0.7736	0.7904	0.8110

4.4 Caltech-UCSD Birds 数据库

Caltech-UCSD Birds数据库是由加利福尼亚理工学院采集的鸟类数据库，数据库包含了200种鸟类的11788张图片。由于鸟类之间有高度的相似性，因此该数据库是模式识别领域中极具挑战性的。本文实验选取了每种鸟类的40张，共计4000张图片作为数据集，并用VGG-verydeep-19提取了VGG19特征，作为数据向量。实验分别选取每种鸟类的5,10,15张图片作为训练样本，剩余的用于测试。实验参数如下： $\lambda_1 = 0.1, \lambda_2 = 0.1, \lambda_3 = 0.0001, \lambda_4 = 10^{-8}$ 。平均实验结果如表4所列。

表4 Caltech-UCSD Birds数据库对比实验结果

算法	5Train	10Train	15Train
LRLR	0.7149	0.7477	0.7660
LRRR	0.7094	0.7710	0.8032
DENLR	0.7066	0.7643	0.8028
CRC	0.6677	0.7113	0.7508
SLRR	0.7083	0.7705	0.8056
LRC	0.6480	0.7077	0.7344
OMP	0.6414	0.7227	0.7772
Homotopy	0.6874	0.7403	0.7620
TPTSR	0.6809	0.7533	0.7852
RPCA	0.6883	0.7440	0.7816
本文算法	0.7203	0.7750	0.8119

4.5 Oxford 102 Flowers 数据库

Oxford 102 Flowers 数据库是图像识别领域的重要数据库之一,其包含了 102 种花,共计 8189 张图片,每种花至少 40 张图片。其中每种花具有不同的大小、光照、样式。本文实验选取每种花 40 张图片作为数据集,共计 4080 张图片;并用 VGG-verydeep-19 提取了 VGG19 特征,作为数据向量。实验分别选取每种鸟类的 5, 10, 15 张图片作为训练,剩余的用于测试。实验参数如下: $\lambda_1 = 0.05, \lambda_2 = 0.05, \lambda_3 = 0.0001, \lambda_4 = 10^{-8}$ 。为了测试各算法的运行时间,以 Oxford 102 Flowers 数据库为例在识别率后面用括号将运行时间表示出来,平均实验结果如表 5 所列。从表 5 中的实验结果来看,本文提出的算法不仅在识别率上优于对比算法,而且在运行时间上也是最快的。

表 5 Oxford 102 Flowers 数据库对比实验结果

算法	5Train(时间/s)	10Train(时间/s)	15Train(时间/s)
LRLR	0.8252(9.575)	0.8330(15.96)	0.9114(51.10)
LRRR	0.8641(5.339)	0.9291(15.59)	0.9537(46.44)
DENLR	0.8703(3.704)	0.9317(5.568)	0.9533(12.04)
CRC	0.8633(49.59)	0.9134(327.9)	0.9211(365.9)
SLRR	0.8779(7.791)	0.9265(40.53)	0.9545(194.5)
LRC	0.8148(52.64)	0.8833(213.4)	0.9165(292.9)
OMP	0.8044(70.96)	0.8859(209.4)	0.9043(357.2)
Homopoty	0.8560(316.9)	0.9120(612.3)	0.9290(882.7)
TPTSR	0.8535(75.21)	0.9222(384.4)	0.9454(816.7)
RPCA	0.8571(118.7)	0.9268(271.4)	0.9435(453.8)
本文算法	0.8868(2.640)	0.9431(4.046)	0.9608(7.171)

4.6 实验分析

从上述 5 个数据库的实验结果来看,在 AR 数据库上 4Train 时本文方法的识别率低于 DENLR,不过本文方法依然比另外 9 种方法的识别率要高。本文方法除了在 AR 数据库 4Train 时的识别率不是最佳的,在其他的 14 次实验中整体上识别率对比的 10 类方法都更高。为了便于观测最佳实验结果,我们在实验结果中用黑体表示该实验中的最佳识别率。由于引入了类别之间的判别信息,因此算法不仅可以更好地缓解过拟合现象,而且能够降低迭代次数,减少运行时间,实验中一般迭代 2~3 次就能达到优化的效果;值得注意的是,低秩线性回归方法运行的时间总体上快于基于稀疏表示的分类方法。这是由于稀疏表示需要将每一个待测试的样本由所有训练样本线性表示求得稀疏表示的系数值,而低秩线性回归方法类似于求一种低维的子空间,通过将数据投影到子空间进行分类识别。从表 5 的各类方法的运行时间来看,本文方法的运行速度最快,不仅远快于基于稀疏表示的分类方法,而且与其他几类基于低秩线性回归的方法相比,运行的速度也快了将近 1 倍及以上。从 5 个数据库的实验结果来看,参数选择的变化不是很大,从侧面反映出本文方法对参数选择不是很敏感。

结束语 线性回归方法在模式识别领域的效果显著。低秩线性回归在传统线性回归的基础上加以改进,具有很好的鲁棒性,可以有效地减少噪声、阴影等干扰因素对识别率的影响。文献[12]解释了低秩线性回归与 LDA 性能相似。但是由于训练样本获取困难,很难搜集到足够的样本,因此会产生过拟合现象,对未知样本的归类有很大的误判性。LRRR, SLRR, DENLR 等方法通过对投影矩阵的正则化能有效地缓解过拟合现象,但是并没有对投影到子空间的数据加以处理,

因此很容易引起样本类别之间的错误逼近。本文在低秩线性回归模型的基础上增加了类别之间的判别性,使得判别性、鲁棒性一定程度上都得到了提高,不仅有效地削弱了样本错误逼近的可能性,而且更有效地减少了过拟合现象,降低了 ALM 优化的次数,使运行时间大幅度缩短。

参考文献

- [1] DONOHO D. High-dimensional data analysis: The curses and blessings of dimensionality [C] // AMS Math Challenges Lecture. 2000: 1-32.
- [2] YANG J, FRANGI A F, YANG J Y, et al. KPCA Plus LDA: a complete kernel Fisher discriminant frame work for feature extraction and recognition [J]. IEEE Trans. Pattern Anal. Mach. Intell., 2005, 27(2): 230-244.
- [3] HE X F, YAN S C, HU Y X. Face recognition using Laplacian faces [J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2005, 27(3): 328-340.
- [4] TORRE F. A least-squares framework for component analysis [J]. IEEE Trans. Pattern Anal. Mach. Intell., 2012, 24(6), 1041-1055.
- [5] STRUTZ T. Data Fitting and Uncertainty: A Practical Introduction to Weighted Least Squares and Beyond [M]. Wiesbaden, Germany: Vieweg and Teubner, 2010.
- [6] LI Y, NGOM A. Nonnegative least-squares methods for the classification of high-dimensional biological data [J]. IEEE/ACM Transactions on Computational Biology & Bioinformatics, 2013, 10(2): 447-456.
- [7] XIANG S, NIE F, MENG G, et al. Discriminative least squares regression for multiclass classification and feature selection [J]. IEEE Trans. Neural Netw. Learn. Syst., 2012, 23(11): 1738-1754.
- [8] ZHANG Z, XU Y, YANG Y, et al. A survey of sparse representation: Algorithms and applications [J]. IEEE Access, 2015, 3(1): 490-530.
- [9] WRIGHT J, YANG A Y, GANESH A, et al. Robust face recognition via sparse representation [J]. IEEE Trans. Pattern Anal. Mach. Intell., 2009, 31(2): 210-227.
- [10] ZHANG L, YANG M, FENG X. Sparse representation or collaborative representation: Which helps face recognition? [J]. International Conference on Computer Vision, 2012, 2011(5): 471-478.
- [11] CANDÈS E J, LI X, MA Y, et al. Robust principal component analysis? [J]. ACM, 2011, 58(3): 11.
- [12] CAI X, DING C, NIE F, et al. On the equivalent of low-rank linear regressions and linear discriminant analysis based regressions [C] // Proc. 19th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining, 2013: 1124-1132.
- [13] ZHANG Z, LAI Z H, XU Y. Discriminative Elastic-Net Regularized Linear Regression [J]. IEEE Transactions on Image Processing A Publication of the IEEE Signal Processing Society, 2017, 26(3): 1466-1481.
- [14] WRIGHT J, PENG Y G, MA Y. Robust principal component analysis: exact recovery of corrupted low-rank matrices by convex optimization [J]. International Conference on Neural Information Processing Systems, 2009, 58(3): 2080-2088.
- [15] LIN Z, CHEN M, MA Y. The augmented lagrange multiplier

- method for exact recovery of corrupted low-rank matrices[OL]. <https://arxiv.org/abs/1009.5055>.
- [16] NASEEM I, TOGNERI R, BENNAMOUN M. Linear regression for face recognition [J]. *IEEE Trans. Pattern Anal. Mach. Intell.*, 2010, 32(11): 2106-2112.
- [17] TROPP J A, GILBERT A C. Signal recovery from random measurements via orthogonal matching pursuit [J]. *IEEE Transactions on Information Theory*, 2007, 53(12): 4655-4666.
- [18] OSBORNE M R, PRESNELL B, TURLACH B A. A new approach to variable selection in least squares problems [J]. *IMA Journal of Numerical Analysis*, 2000, 20(3): 389-403.
- [19] XU Y, ZHANG D, YANG J. A two-phase test sample sparse representation method for use with face recognition [J]. *IEEE Transactions on Circuits and Systems for Video Technology*, 2011, 21(9): 1255-1262.
- [20] MARTINEZ A, BENAVENTE R. The AR Face Database[R]. Purdue University, Computer Vision Center, West Lafayette, 1998.
- [21] PHILLIPS P J. The facial recognition technology (FERET) database[OL]. http://www.itl.nist.gov/iad/humanid/feret/feret_master.html.
- [22] YAO B, JIANG X, KHOSLA A, et al. Human action recognition by learning bases of action attributes and parts [C]// *International Conference on Computer Vision (ICCV)*. 2011: 1331-1338.
- [23] WAH C, BRANSON S, WELINDER P, et al. The caltech-ucsd birds-200-2011 dataset[OL]. <http://www.vision.caltech.edu/visipedia/CUB-200-2011.html>.
- [24] NILSBACK M E, ZISSERMAN A. Automated flower classification over a large number of classes [C]// *Sixth Indian Conference on Computer Vision, Graphics & Image Processing*, 2008 (ICVGIP'08). 2008: 722-729.
- [25] ABDI H, WILLIAMS L J. Principal component analysis[J]. *Wiley Interdiscip. Rev. Comput. Stat.*, 2010, 2(4): 433-459.
- [26] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition [C]// *Proceedings of the International Conference on Learning Representations (ICLR)*. 2014.

(上接第 142 页)

参 考 文 献

- [1] BARNES D E, YAFFE K. The projected effect of risk factor reduction on Alzheimer's disease prevalence [J]. *Alzheimers & Dementia the Journal of the Alzheimers Association*, 2011, 10(9): 819.
- [2] CHANG C C, LIN C J. LIBSVM: A library for support vector machines [J]. *ACM*, 2011, 2(3): 1-27.
- [3] CHAN T H, JIA K, GAO S, et al. PCANet: A Simple Deep Learning Baseline for Image Classification? [J]. *IEEE Transactions on Image Processing*, 2015, 24(12): 5017-5032.
- [4] CLARK C M, DAVATZIKOS C, BORTHAKUR A, et al. Biomarkers for Early Detection of Alzheimer Pathology [J]. *Neurosignals*, 2007, 16(1): 11-18.
- [5] GUPTA A, MAIDA A S, AYHAN M. Natural Image Bases to Represent Neuroimaging Data [C]// *International Conference on Machine Learning*. 2013: 987-994.
- [6] CHOI H, JIN K H. Predicting Cognitive Decline with Deep Learning of Brain Metabolism and Amyloid Imaging [J]. *Behavioral Brain Research*, 2017, 344: 103-109.
- [7] HOSSEINI A S L, GIMEL'FARB G, ELBAZ A. Alzheimer's Disease Diagnostics by a Deeply Supervised Adaptable 3D Convolutional Network [J]. *Front Biosci*, 2018, 23: 584-596.
- [8] JANOUŠOVÁ E, VOUNOU M, WOLZ R, et al. Biomarker discovery for sparse classification of brain images in Alzheimer's disease [J]. *Neurology Neurosurgery Neurosciences*, 2012, 2012(2): 1-11.
- [9] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. ImageNet Classification with Deep Convolutional Neural Networks [J]. *Advances in Neural Information Processing Systems*, 2012, 25(2): 2012.
- [10] MA J, SHERIDAN R P, LIAW A, et al. Deep Neural Nets as a Method for Quantitative Structure-Activity Relationships [J]. *Journal of Chemical Information & Modeling*, 2015, 55(2): 263-274.
- [11] MRICron [OL]. <http://www.nitrc.org/projects/mricron>.
- [12] BATMANGHELICH N, TASKAR B, DAVATZIKOS C. A general and unifying framework for feature construction, in image-based pattern classification [C]// *Information Processing in Medical Imaging*. Springer, 2009: 423-434.
- [13] PAYAN A, MONTANA G. Predicting Alzheimer's disease: a neuroimaging study with 3D convolutional neural networks [J]. *arXiv:1502.02506v1*.
- [14] SAINATH T N, MOHAMED A R, KINGSBURY B, et al. Deep convolutional neural networks for LVCSR [C]// *IEEE International Conference on Acoustics*. 2013: 8614-8618.
- [15] LIU S Q, LIU S D, CAI W D, et al. Early diagnosis of Alzheimer's disease with deep learning [C]// *2014 IEEE 11th International Symposium on Biomedical Imaging (ISBI)*. IEEE, 2014: 1015-1018.
- [16] TONG T, WOLZ R, GAO Q, et al. Multiple instance learning for classification of dementia in brain MRI [J]. *Medical Image Analysis*, 2013, 16(Pt 2): 599-606.
- [17] YANG W L, LUI R L M, GAO J H, et al. Independent component analysis-based classification of Alzheimer's disease MRI data [J]. *Journal of Alzheimer's Disease*, 2011, 24(4): 775-783.
- [18] WOLZ R, JULKUNEN V, KOIKKALAINEN J, et al. Multi-method analysis of MRI images in early diagnostics of Alzheimer's disease [J]. *Plos One*, 2011, 6(10): e25446.
- [19] YIANNPOULOU K G, PAPAGEORGIU S G. Current and future treatments for Alzheimer's disease [J]. *Therapeutic Advances in Neurological Disorders*, 2013, 6(1): 19-33.