

一种基于最大相关熵和局部约束的协同表示分类器

余沁茹, 卢桂馥

安徽工程大学计算机与信息学院, 安徽 芜湖 241000

摘要: 提出一种基于最大相关熵和局部约束的协同表示分类器 (CRC/MCCLC), 该分类器能同时利用最大相关熵和局部信息。一方面, 通过利用最大相关熵准则, CRC/MCCLC 不仅在异常值处理上比 L_1 范数鲁棒性更高, 还可以使用半二次优化技术进行更有效的计算; 另一方面, CRC/MCCLC 通过使用局部信息得到近似稀疏表示, 以此从训练样本中获得更多的判别信息。在 ORL、Yale 以及 AR 人脸数据集等图像数据集上的实验结果验证了 CRC/MCCLC 方法的有效性。

关键词: 人脸识别; 协同表示; 稀疏表示; 最大相关熵; 局部约束

中图分类号: TP931.4

文献标识码: A

doi: 10.11959/j.issn.2096-6652.202133

Collaborative representation based classifier with maximum correntropy criterion and locality constraint

YU Qinru, LU Guifu

School of Computer and Information, Anhui Polytechnic University, Wuhu 241000, China

Abstract: A method which utilizes maximum correntropy criterion and locality information called collaborative representation based classifier with maximum correntropy criterion and locality constraint (CRC/MCCLC) was proposed. On the one hand, CRC/MCCLC was not only more robust to outliers than L_1 norm but also could be computed efficiently using half-quadratic optimization technique because of the use of maximum correntropy criterion. On the other hand, CRC/MCCLC could obtain more discriminative information from the training samples and could lead to an approximately sparse representation because of the use of locality information. Extensive experimental results on some image databases demonstrate that CRC/MCCLC can achieve the state-of-the-art performance on these image databases.

Key words: face recognition, collaborative representation, sparse representation, maximum correntropy criterion, locality constraint

1 引言

人脸识别因其巨大的应用潜力吸引了计算机视觉、机器学习和模式识别领域的大量研究人员, 在过去的二十年中诞生了许多算法^[1-8]。一些众所周知的人脸识别方法, 例如基于主成分分析 (principal component analysis, PCA) 的 Eigenfaces^[9]、基于线性

判别分析 (linear discriminant analysis, LDA) 的 Fisherfaces^[10]、基于独立成分分析 (independent component analysis, ICA) 的 Independentfaces^[11]和基于局部保留投影 (locality preserving projection, LPP) 的 Laplacianfaces^[12], 已经获得了较好的成果, 然而这些方法对噪声和异常值很敏感。

Li S Z 等人^[13]提出了一种最近特征线 (nearest

收稿日期: 2021-07-08; 修回日期: 2021-07-19

通信作者: 卢桂馥, luguifu_jsj@163.com

基金项目: 国家自然科学基金资助项目 (No.61976005, No.61772277), 安徽省自然科学基金资助项目 (No.1908085MF183)

Foundation Items: The National Natural Science Foundation of China (No.61976005, No.61772277), The Natural Science Foundation of Anhui Province (No.1908085MF183)

feature line, NFL) 分类方法, 该方法使用线性方式对属于同一类的每对训练样本进行外推和内插。NFL 可以泛化样本的代表性能力, 并已被广泛应用于许多模式识别任务中^[14]。沿着这个研究方向, Chien J T 等人^[15]提出了一种最近特征空间(nearest feature space, NFS) 分类方法。NFS 中构建每个类的特征空间的方式与 NFL 中构建特征线的方式类似。

近年来, 研究人员们在假设测试图像位于由同一类图像组成的线性子空间的条件下, 提出了一些基于线性表示的新型分类算法^[16-22]。Wright J 等人^[16]提出了一种基于稀疏表示的分类器 SRC (sparse representation classifier), SRC 在人脸识别方面取得了较好的结果。SRC 使用所有训练样本来制定一个过完备字典, 然后通过 L_1 最小化问题在训练样本上对探测图像进行稀疏编码。SRC 得到的大部分组合系数很小, 接近于零, 只有小部分组合系数具有显著值, 因此可以使用非零系数来确定探测图像的标签。

尽管 SRC 在鲁棒人脸识别问题上非常成功, 但对于存在连续遮挡的人脸图片, 其鲁棒性仍然不够高^[23], 且计算复杂度非常高。为了解决这个问题, He R 等人^[22]提出了一种基于相关熵的稀疏表示(correntropy based sparse representation, CESR) 方法。相关熵^[24-25]与 M 估计(M-estimation) 密切相关, 可以有效地处理噪声和大的异常值^[26-27]。实验结果表明, 对于人脸识别中的围巾遮挡问题, CESR 的识别率高于 SRC。

然而, 参考文献[28]认为 SRC 中用于约束的 L_1 范数可以用 L_2 范数代替, 这是因为真实人脸图像不支持 SRC 的稀疏假设。同样, Zhang L 等人^[21]认为协同表示的方法可以获得比使用 L_1 范数的 SRC 更好的结果。因此, Zhang L 等人^[21]提出了一种基于协同表示的分类方法 CRC (collaborative representation based classification), CRC 使用 L_2 范数而获得了解析解, 因此其展现出比 SRC 更优秀的性能。

参考文献[17]提出了另一种同样使用 L_2 范数的有效的基于线性回归的分类方法 LRC (linear regression-based classification)。与 CRC 相比, LRC 仅使用来自特定类别的图像获取表示系数。Ren C X 等人^[19]提出了一种基于 $L_{2,1}$ 范数的分类方法, 称为基于旋转不变范数的回归分类 (rotational in-

variant norm based regression for classification, RRC) 方法, $L_{2,1}$ 范数是旋转不变的, 且比 L_2 范数对噪声和异常值的处理更具鲁棒性。Ren C X 等人^[19]还为 RRC 开发了一种高效的优化算法, 这与 Nie F P 等人^[29]提出的优化算法非常相似。

流形学习的成功表明局部几何信息对构建学习算法非常重要^[12,30-38]。参考文献[38-39]认为局部性比稀疏性更重要, 这是因为局部性总会导致稀疏性, 但反过来不一定成立。在局部约束下, L_2 范数也可以产生近似稀疏的表示结果。

利用相关熵和局部约束的优点, 笔者提出了一种基于最大相关熵和局部约束的协同表示分类器 (collaborative representation based classifier with maximum correntropy criterion and locality information, CRC/MCCLC)。对于异常值, CRC/MCCLC 比 L_1 范数的鲁棒性更高, 且由于使用了最大相关熵准则, CRC/MCCLC 还可以使用半二次优化技术进行更有效的计算。此外, CRC/MCCLC 可以从训练样本中获得更多的判别信息, 并且由于使用了局部信息, CRC/MCCLC 可以获得近似稀疏的表示结果。本文还提出了一种用于实现 CRC/MCCLC 的有效优化算法。在一些图像数据集上的大量实验结果证明了 CRC/MCCLC 的有效性。

本文的主要贡献如下。

(1) 本文将相关熵和流形学习集成到一个无缝模型中, 因此本文提出的方法不仅可以有效地处理非高斯噪声和异常值, 还可以获得近似稀疏的表示结果。

(2) 笔者分析了 CRC/MCCLC 的优化问题。基于半二次优化技术, 笔者提出了一种求解 CRC/MCCLC 的有效程序。

本文其余部分安排如下: 第 2 节简要回顾了 SRC、CRC 和相关熵; 第 3 节提出了 CRC/MCCLC 方法, 并为其优化问题提出了一种有效的基于半二次优化的算法; 第 4 节介绍本文所做的实验; 第 5 节总结全文。

2 局部约束与熵

2.1 SRC 与 CRC

给定一个数据矩阵 $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n\} = [\mathbf{X}_1, \dots, \mathbf{X}_c] \in R^{d \times n}$, 其中 $\mathbf{x}_i \in R^d$ ($i=1, 2, \dots, n$), 为 d 维空间中的第 i 个训练样本, $\mathbf{X}_i \in R^{d \times n_i}$ ($i=1, 2, \dots, c$) 为

第 i 类训练样本集合且 $\sum_{i=1}^c n_i = n$ 。测试样本 $\mathbf{y} \in R^d$ 属于第 i 类, 可以近似表示为第 i 类训练样本的线性组合:

$$\mathbf{y} \approx \mathbf{X}_i \boldsymbol{\beta}_i, \quad i=1,2,\dots,c \quad (1)$$

其中, $\boldsymbol{\beta}_i \in R^n$ 为向量系数。 $\mathbf{y}$ 也可以近似表示为所有训练样本的线性组合:

$$\mathbf{y} \approx \mathbf{X} \boldsymbol{\beta} \quad (2)$$

其中, $\boldsymbol{\beta} \in R^n$ 为向量系数。

SRC^[16]试图用式(3)求解稀疏解:

$$\min_{\boldsymbol{\beta}} \|\boldsymbol{\beta}\|_0, \text{ s.t. } \mathbf{y} = \mathbf{X} \boldsymbol{\beta} \quad (3)$$

其中, $\|\cdot\|_0$ 表示 L_0 范数, 即求 $\boldsymbol{\beta}$ 中非零项的个数。

由于对式(3)的求解存在 NP 难题, 因此可以通过将 L_0 最小化问题转为 L_1 最小化问题进行求解, 如式(4)所示:

$$\min_{\boldsymbol{\beta}} \|\boldsymbol{\beta}\|_1, \text{ s.t. } \mathbf{y} = \mathbf{X} \boldsymbol{\beta} \quad (4)$$

其中, $\|\cdot\|_1$ 表示 L_1 范数, $\|\boldsymbol{\beta}\|_1 = \sum_{i=1}^n |\beta_i|$ 。在许多实际应用中, 人脸图像经常存在一些遮挡和损坏问题。为了解决这个问题, 模型被修改为:

$$\min_{\boldsymbol{\beta}} \|\boldsymbol{\beta}\|_1, \text{ s.t. } \|\mathbf{y} - \mathbf{X} \boldsymbol{\beta}\|_2 \leq \varepsilon \quad (5)$$

其中, $\varepsilon > 0$ 为容错量。该模型可以有效地解决遮挡和损坏问题。

通常, 式(5)可以被重写为具有适当参数 $\lambda \geq 0$ 的等效无约束形式:

$$\min_{\boldsymbol{\beta}} \|\mathbf{y} - \mathbf{X} \boldsymbol{\beta}\|_2^2 + \lambda \|\boldsymbol{\beta}\|_1 \quad (6)$$

式(6)被称为套索问题。在接下来的部分中, 笔者将式(6)作为 SRC 的目标函数。

CRC^[21]仅用 L_2 范数代替式(6)中的 L_1 范数就可以得到较高的识别率。CRC 的目标函数为:

$$\min_{\boldsymbol{\beta}} \|\mathbf{y} - \mathbf{X} \boldsymbol{\beta}\|_2^2 + \lambda \|\boldsymbol{\beta}\|_2^2 \quad (7)$$

CRC 存在解析解, 如 $\boldsymbol{\beta} = (\mathbf{X}^\top \mathbf{X} + \lambda \mathbf{I})^{-1} \mathbf{X}^\top \mathbf{y}$, 其中 $\mathbf{I}$ 是一个 $n \times n$ 的单位矩阵。从运算速度来说, CRC 比 SRC 快了许多。

2.2 熵

熵^[24-25]被定义为两个任意变量 A 和 B 之间的局部相似性度量:

$$V_\sigma(A, B) = E[k_\sigma(A - B)] \quad (8)$$

其中, $E[\cdot]$ 是数学期望, $k_\sigma(\cdot)$ 表示核函数。通常只能获得数量有限的数数据 $\{(a_j, b_j)\}_{j=1}^n$ 。由于不知道 A 和 B 的联合概率密度函数, 可以得到相关熵的样本估计为:

$$\hat{V}_{n,\sigma}(A, B) = \frac{1}{n} \sum_{j=1}^n k_\sigma(a_j - b_j) \quad (9)$$

其中, k_σ 是高斯核函数 $g(x) = \exp\left(-\frac{x^2}{2\sigma^2}\right)$ 。可以将式(9)重写为:

$$\hat{V}_{n,\sigma}(A, B) = \frac{1}{n} \sum_{j=1}^n g(a_j - b_j) \quad (10)$$

在参考文献[24]中, 最大相关熵准则(maximum correntropy criterion, MCC)被定义为误差的最大相关熵。

3 基于最大相关熵和局部约束的协同表示分类器

3.1 算法模型

显然, $\|\mathbf{y} - \mathbf{X} \boldsymbol{\beta}\|_2^2$ 可以被重写为:

$$\|\mathbf{y} - \mathbf{X} \boldsymbol{\beta}\|_2^2 = \sum_{j=1}^d \left(y_j - \sum_{i=1}^n x_{ij} \beta_i \right)^2 \quad (11)$$

其中, x_{ij} 表示第 i 个训练样本的第 j 项, y_j 表示 $\mathbf{y}$ 的第 j 项。

将式(11)代入式(7), 再结合式(10), 同时给出具有局部约束 $\mathbf{W}$ 的系数 $\boldsymbol{\beta}$, 则 CRC/MCCLC 可定义为下式:

$$J_{\text{CRC/MCCLC}} = \max_{\boldsymbol{\beta}} \sum_{j=1}^d g\left(y_j - \sum_{i=1}^n x_{ij} \beta_i\right) - \lambda \|\mathbf{W} \boldsymbol{\beta}\|_2^2 \quad (12)$$

其中, $\mathbf{W} = \text{diag}([\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_n])$, $\text{diag}(\cdot)$ 是将向量转换为对角矩阵的运算符, $\mathbf{w}_i (i=1, 2, \dots, n)$ 是 $\mathbf{x}_i$ 和 $\mathbf{y}$ 之间的间距:

$$\mathbf{w}_i = \frac{\exp\left(\|\mathbf{y} - \mathbf{x}_i\|^2 / \sigma^2\right)}{\max\left(\exp\left(\|\mathbf{y} - \mathbf{x}_i\|^2 / \sigma^2\right)\right)}, i=1, 2, \dots, n \quad (13)$$

式(12)中的局部约束 $\mathbf{W}$ 有双重解释: 当 $\mathbf{y}$ 和 $\mathbf{x}_i$ 之间距离较大时, 系数应该很小, 从而使式(12)最大化; 当 $\mathbf{y}$ 和 $\mathbf{x}_i$ 之间的距离较小时, 系数应较大。

3.2 CRC/MCCLC 的目标函数的求解算法

由于 CRC/MCCLC 的求解是一个非线性优化

问题, 很难直接进行优化。因此, 笔者设计了一种基于半二次优化技术求解式 (12) 的算法^[27]。

定理 1^[22,27] 存在 $g(x)$ 的凸共轭函数 φ 使得:

$$g(x) = \max_{p'} \left(p' \frac{\|x\|^2}{\sigma^2} - \varphi(p') \right) \quad (14)$$

其中, $p' \in R$ 是标量, 对于固定的 x , 最大值为 $p' = -g(x)$ 。

将式 (14) 代入式 (12), 得到 CRC/MCCLC 的增强目标函数:

$$\hat{J}_{\text{CRC/MCCLC}} = \max_{\beta, p} \sum_{j=1}^d \left(p_j \left(y_j - \sum_{i=1}^n x_{ij} \beta_i \right)^2 - \varphi(p_j) \right) - \lambda \|\mathbf{W}\beta\|_2^2 \quad (15)$$

其中, $\mathbf{p} = [p_1, p_2, \dots, p_d]^\top$ 由半二次优化技术引入, 是辅助变量。显然, 对于固定的 β 有:

$$J_{\text{CRC/MCCLC}}(\beta) = \max_p \hat{J}_{\text{CRC/MCCLC}}(\beta, p) \quad (16)$$

$$\max_p J_{\text{CRC/MCCLC}}(\beta) = \max_{\beta, p} \hat{J}_{\text{CRC/MCCLC}}(\beta, p) \quad (17)$$

因此, 可以通过最大化 $\hat{J}_{\text{CRC/MCCLC}}(\beta, p)$ 来获得 $J_{\text{CRC/MCCLC}}(\beta)$ 的最大值。

以交替最大化的方式计算局部最大化值:

$$\begin{aligned} p_j^{t+1} &= -g \left(y_j - \sum_{i=1}^n x_{ij} \beta_i^t \right) \quad (18) \\ \beta^{t+1} &= \arg \max_{\beta} (\mathbf{y} - \mathbf{X}\beta)^\top \text{diag}(\mathbf{p})(\mathbf{y} - \mathbf{X}\beta) - \lambda \|\mathbf{W}\beta\|_2^2 \quad (19) \end{aligned}$$

其中, t 表示迭代次数。式 (19) 可被重写为:

$$\begin{aligned} &\arg \max_{\beta} \beta^\top \mathbf{X}^\top \text{diag}(\mathbf{p}) \mathbf{X} \beta - \mathbf{y}^\top \text{diag}(\mathbf{p}) \mathbf{X} \beta - \\ &\beta^\top \mathbf{X}^\top \text{diag}(\mathbf{p}) \mathbf{y} - \lambda \beta^\top \mathbf{W}^\top \mathbf{W} \beta \quad (20) \end{aligned}$$

令 $\tilde{\mathbf{X}} = \text{diag}(\sqrt{-\mathbf{p}^{t+1}}) \mathbf{X}$ 且 $\tilde{\mathbf{y}} = \text{diag}(\sqrt{-\mathbf{p}^{t+1}}) \mathbf{y}$, 有:

$$\arg \min_{\beta} \beta^\top \tilde{\mathbf{X}}^\top \tilde{\mathbf{X}} \beta - \tilde{\mathbf{y}}^\top \tilde{\mathbf{X}} \beta - \beta^\top \tilde{\mathbf{X}}^\top \tilde{\mathbf{y}} + \lambda \beta^\top \mathbf{W}^\top \mathbf{W} \beta \quad (21)$$

式 (21) 存在解析解:

$$\beta = (\tilde{\mathbf{X}}^\top \tilde{\mathbf{X}} + \lambda \mathbf{W}^\top \mathbf{W})^{-1} \tilde{\mathbf{X}}^\top \tilde{\mathbf{y}} \quad (22)$$

根据以上分析, 式 (15) 的优化过程可总结为算法 1。

算法 1 CRC/MCCLC 目标函数的求解算法

输入 数据矩阵 $\mathbf{X}$, 测试样本 $\mathbf{y}$, 正则化参数 $\lambda > 0$, 核函数 $\sigma > 0$, $\mathbf{p}^1 = [-1, 1, \dots, 1] \in R^d$, 一个较小的值 ε

输出 β

步骤 1: 令 $t=1$ 并通过式 (13) 计算 $\mathbf{W}$;

步骤 2: 计算 $\tilde{\mathbf{X}} = \text{diag}(\sqrt{-\mathbf{p}^1}) \mathbf{X}$ 与 $\tilde{\mathbf{y}} = \text{diag}(\sqrt{-\mathbf{p}^1}) \mathbf{y}$;

步骤 3: 通过式 (19) 计算 β^t ;

步骤 4: 如果 $\|\beta^t - \beta^{t-1}\| < \varepsilon$, 停止计算; 否则通过式 (18) 更新辅助向量 $\mathbf{p}^{t+1}$, 且 $t=t+1$, 回到步骤 2。

由以上算法可以看出, CRC/MCCLC 生成的序列 $\{\hat{J}_{\text{CRC/MCCLC}}(\beta, p), t=1, 2, \dots\}$ 最终会收敛。

3.3 CRC/MCCLC 的分类过程

通常, 计算分类器需要将测试样本 $\mathbf{y}$ 分类到现有类中。理想情况下, 测试样本 $\mathbf{y}$ 应该通过与 $\mathbf{y}$ 同类的训练样本进行最佳重构获得。

因此, 笔者将测试样本 $\mathbf{y}$ 进行如下分类: 对于每个类 $i(i=1, 2, \dots, c)$, 令 $\delta_i: R^n \rightarrow R^n$ 为一个用于选择所属类系数的函数, 如 $\delta_i(\beta) \in R^n$ ($i=1, 2, \dots, c$) 是一个列向量, 其是 β 中与类 i 有关的项。由此, 测试样本 $\mathbf{y}$ 可以由来自 i 的类样本重构为:

$$\hat{\mathbf{y}}_i = \mathbf{X}_i \delta_i(\beta) \quad (23)$$

因此, 测试样本 $\mathbf{y}$ 可以被归类到 $\mathbf{y}$ 和 $\hat{\mathbf{y}}_i$ 之间具有最大非线性相似度的第 i 类, 如:

$$\begin{aligned} \max_i r_i(\mathbf{y}) &= g(\|\mathbf{y} - \hat{\mathbf{y}}_i\|_2) = g(\|\mathbf{y} - \mathbf{X}_i \delta_i(\beta)\|_2) = \\ &\exp \left(-\frac{\|\mathbf{y} - \mathbf{X}_i \delta_i(\beta)\|_2^2}{2\sigma^2} \right) \quad (24) \end{aligned}$$

由于 $g(x)$ 是一个单调递减函数, $\mathbf{y}$ 和 $\hat{\mathbf{y}}_i$ 之间的最大非线性相似度相当于 $\mathbf{y}$ 和 $\hat{\mathbf{y}}_i$ 之间的最小欧氏距离。那么有:

$$\min_i r_i(\mathbf{y}) = \|\mathbf{y} - \hat{\mathbf{y}}_i\|_2 = \|\mathbf{y} - \mathbf{X}_i \delta_i(\beta)\|_2 \quad (25)$$

根据以上分析, CRC/MCCLC 的分类过程可总结为算法 2。

算法 2 CRC/MCCLC 的分类过程

输入 数据矩阵 $\mathbf{X} = [\mathbf{X}_1, \dots, \mathbf{X}_c] \in R^{d \times n}$, 测试样本 $\mathbf{y}$, 算法 1 中获得的系数向量 β

输出 $\text{identity}(\mathbf{y})$

步骤 1: 利用算法 1 计算系数向量 β ;

步骤 2: 计算残差 $r_i(\mathbf{y}) = \|\mathbf{y} - \hat{\mathbf{y}}_i\|_2 = \|\mathbf{y} - \mathbf{X}_i \delta_i(\beta)\|_2$;

($i=1,2,\dots,c$)

步骤 3: $\text{identity}(\mathbf{y}) = \arg \max_i r_i(\mathbf{y})$ 。

3.4 算法复杂度

在算法 1 的步骤 1 中, 计算 $\mathbf{W}$ 时的复杂度为 $O(dn)$; 在步骤 2 中, 计算 $\tilde{\mathbf{X}} = \text{diag}(\sqrt{-\mathbf{p}^t})\mathbf{X}$ 与 $\tilde{\mathbf{y}} = \text{diag}(\sqrt{-\mathbf{p}^t})\mathbf{y}$ 时的复杂度为 $O(dn)$ 和 $O(d)$; 在步骤 3 中, 计算 β 时的复杂度为 $O(dn^2)$ 。由此可以得出, CRC/MCCLC 的算法复杂度为 $O(t(dn^2 + dn + d))$, 其中 t 为迭代次数。

4 实验与结果

本节在人脸识别方面将本文提出的 CRC/MCCLC 与几种相关的先进算法进行比较, 如 SRC^[16]、CRC^[21]、LRC^[17] 以及 CESR^[22]。SRC 是使用 l1-magic 实现的, 是一个公开可用的数据包。对于 CRC/MCCLC 中的 σ 和 λ , 笔者尝试从 $\{0.1, 0.5, 1, 5, 10, 50\}$ 与 $\{0.001, 0.002, 0.003, 0.004, 0.005, 0.006, 0.007, 0.008, 0.009, 0.010\}$ 中分别取不同的值进行测试。实验是在移动双核 Intel Pentium (1.8 GHz) 处理器、4G RAM 的 acer 计算机上进行的, 实验所用操作系统为 Windows7, 编程环境为 MATLAB2008。

4.1 ORL 人脸数据集上的实验

ORL 人脸数据集中共有 400 张人脸图像 (共 40 人, 每人 10 个样本)。某些人脸图像是在不同时间拍摄的, 灯光、面部动作 (睁眼/闭眼、微笑/不微笑) 和面部细节 (戴眼镜/不戴眼镜) 都不同。ORL 人脸数据集中某人的一些示例图像如图 1 所示。

实验选取每个人的 $i(i=4,5)$ 个样本进行训练, 其余样本用于测试。对于每个给定的 i , 执行 10 次随机选择样本操作并训练样本, 然后计算平均识别率和标准偏差。实验中, PCA 被用于降低每个图像的

维数。不同算法在 ORL 人脸数据集上的平均识别率和标准偏差见表 1。从表 1 可以看出, 本文提出的算法 CRC/MCCLC 较原有算法在误差允许范围内有明显的效果提升。



图 1 ORL 人脸数据集中某人的一些示例图像

4.2 Yale 人脸数据集上的实验

Yale 人脸数据集包含 15 个人的 165 张灰度图像, 每个人有 11 张图像。这些图像展示了光照条件、面部表情 (正常、快乐、悲伤、困倦、惊讶和眨眼) 的变化。在本文的实验中, Yale 人脸数据集中每张图像的尺寸都被手动裁剪并调整为 32×32 。Yale 人脸数据集中某人的一些示例图像如图 2 所示。

实验同样选取每个人的 $i(i=4,5)$ 个样本进行训练, 其余样本用于测试。对于每个给定的 i , 笔者执行 10 次随机选择样本操作并训练样本, 然后计算平均识别率和标准偏差, 结果见表 2。其中, LRC 的性能较差, SRC、CRC 与 CRC/MCCLC 的性能较好, CRC/MCCLC 的性能比 SRC 和 CRC 更好。



图 2 Yale 人脸数据集中某人的一些示例图像

4.3 AR 人脸数据集上的实验

AR 人脸数据集包含 126 个人的不同面部表情、

表 1 不同算法在 ORL 人脸数据集上的平均识别率和标准偏差

样本大小	SRC	CRC	LRC	CESR	CRC/MCCLC
4	91.8%±1.0%	91.5%±1.6%	88.9%±1.3%	90.9%±1.1%	92.8%±1.2%
5	95.1%±1.5%	94.5%±1.7%	92.0%±1.6%	92.8%±1.7%	95.5%±1.8%

表 2 不同算法在 Yale 人脸数据集上的平均识别率和标准偏差

样本大小	SRC	CRC	LRC	CESR	CRC/MCCLC
4	74.5%±3.2%	74.0%±3.6%	59.3%±2.7%	65.4%±2.1%	76.2%±3.8%
5	81.1%±3.6%	80.3%±3.3%	65.9%±3.3%	69.7%±3.6%	82.1%±3.0%

光照条件和遮挡情况下的 4 000 多张灰色人脸图像。在本文的实验中,所有图像的尺寸都被手动裁剪并调整为 50×40 。

笔者在 AR 人脸数据集上进行了两个实验。在第一个实验中,笔者使用了 AR 人脸数据集的子集,其中包含来自 120 个人的 1 680 张图像。对于每个人,AR 人脸数据集中会话 1 中的 7 张图像用于训练,而会话 2 中的 7 张图像用于测试。AR 人脸数据集中某人的一些示例图像如图 3 所示。PCA 被用于将每张图像的维数降低到 300。不同算法在 AR 人脸数据集上不同聚类量的平均识别率见表 3。

在第二个实验中,笔者使用了 AR 人脸数据集的其他子集,包含来自 120 个人的 1 440 张图像。其中,960 张无遮挡图像(每个人约 8 张图像)用于训练,其余戴墨镜和围巾的图像(如图 4 所示)用于测试。PCA 被用于将每张图像的维数降低到 300。不同算法在 AR 人脸数据集上不同聚类量的墨镜/围巾遮挡识别率分别见表 4 和表 5。

总体看来,除 LRC 外的几种算法有较好的实验表现,这些算法在 AR 数据集以及其墨镜遮挡数据集

上展现出的性能不分伯仲,但 CRC/MCCLC 在围巾遮挡数据集上的表现明显优于其他算法。



图3 AR人脸数据集中某人的一些示例图像



图4 AR人脸数据集中存在墨镜与围巾遮挡的图像

4.4 参数敏感性

核大小 σ 和正则化参数 λ 都是重要参数,经过良好调整的 σ 和 λ 可以提高 CRC/MCCLC 的性能。在本节中,笔者将研究 σ 和 λ 对 CRC/MCCLC 的识别性能的影响,实验的数据集合为 $\{0.1, 0.5, 1, 5, 10, 50\}$ 和 $\{0.001, 0.002, 0.003, 0.004, 0.005, 0.006, 0.007,$

表3 不同算法在 AR 人脸数据集上不同聚类量的平均识别率

算法	50	100	150	200	250	300
SRC	69.4%	73.3%	73.8%	73.9%	74.3%	72.0%
CRC	68.8%	72.2%	75.1%	75.0%	75.2%	73.8%
LRC	63.6%	65.8%	67.0%	68.1%	68.0%	68.0%
CESR	70.2%	72.3%	73.8%	74.4%	74.4%	74.4%
CRC/MCCLC	72.8%	75.1%	75.7%	74.5%	74.4%	73.8%

表4 不同算法在 AR 人脸数据集上不同聚类量的墨镜遮挡识别率

算法	50	100	150	200	250	300
SRC	93.3%	95.4%	97.9%	96.2%	95.4%	95.4%
CRC	88.3%	92.1%	95.4%	94.6%	94.6%	95.4%
LRC	94.2%	96.1%	96.7%	96.7%	96.7%	96.2%
CESR	93.8%	97.5%	97.9%	98.1%	98.1%	98.1%
CRC/MCCLC	97.2%	97.5%	98.1%	98.3%	98.1%	97.9%

表5 不同算法在 AR 人脸数据集上不同聚类量的围巾遮挡识别率

算法	50	100	150	200	250	300
SRC	43.7%	60.8%	71.1%	70.0%	73.3%	73.4%
CRC	34.2%	50.0%	58.3%	62.9%	65.8%	69.6%
LRC	24.2%	28.3%	31.7%	31.7%	31.7%	31.7%
CESR	42.1%	48.0%	58.8%	62.8%	65.8%	67.8%
CRC/MCCLC	45.1%	62.3%	70.7%	72.4%	73.6%	74.2%

0.008, 0.009, 0.010}。在 ORL 人脸数据集上, σ 和 λ 对 CRC/MCCLC 的识别性能的影响如图 5 所示(实验时每个人都有 4 个样本用于训练)。

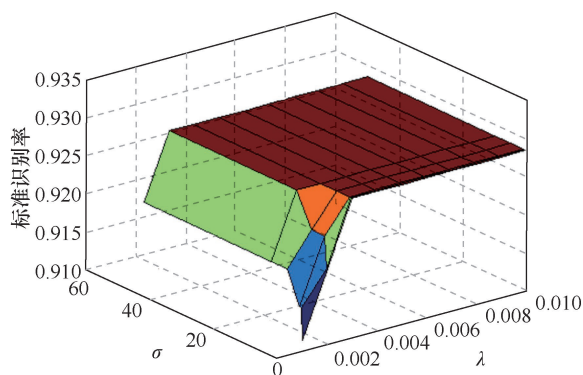


图5 σ 和 λ 对 CRC/MCCLC 的识别性能的影响

4.5 讨论

从实验结果可以得出以下结论。

(1) 与其他方法相比, 本文提出的 CRC/MCCLC 在大多数实验中可以达到最高的平均识别率。然而, CRC/MCCLC 在所有实验中并不总是显著优于 SRC、CRC、LRC 和 CESR 的。

(2) 从第 4.4 节的实验可以看出, CRC/MCCLC 的识别性能取决于 σ 和 λ 。然而, 当 σ 和 λ 分别从集合 {0.1, 0.5, 1, 5, 10, 50} 和 {0.001, 0.002, 0.003, 0.004, 0.005, 0.006, 0.007, 0.008, 0.009, 0.010} 中取值时, CRC/MCCLC 的性能相对稳定。在实践中, 研究人员可以使用交叉验证方法进行选择, 这是因为可以获得最高识别率的最佳 σ 和 λ 以及最高识别率取决于测试数据, 并且在实际问题中事先未知。

5 结束语

本文提出了一种新的基于最大相关熵和局部约束的协同表示分类器 CRC/MCCLC。借助最大相关熵准则的优势, CRC/MCCLC 对异常值的鲁棒性比 L_1 范数更高, 而且可以使用半二次优化技术进行有效的计算。同时, CRC/MCCLC 对局部信息的使用使得其可以从训练样本中获得更多的判别信息及近似稀疏表示。这些方法的运用使得 CRC/MCCLC 可以获得较好的性能。

参考文献:

[1] GU B, SHENG V S. A robust regularization path algorithm for v-support vector classification[J]. IEEE Transactions on Neural Networks and Learning Systems, 2017, 28(5): 1241-1248.

[2] GU B, SUN X M, SHENG V S. Structural minimax probability machine[J]. IEEE Transactions on Neural Networks and Learning Systems, 2017, 28(7): 1646-1656.

[3] LIU Y F, CHEN J H. Unsupervised face frontalization for pose-invariant face recognition[J]. Image and Vision Computing, 2021, 106: 104093.

[4] TANG L L, LU H F, PANG Z, et al. A distance weighted linear regression classifier based on optimized distance calculating approach for face recognition[J]. Multimedia Tools and Applications, 2019, 78(22): 32485-32501.

[5] WEN X Z, SHAO L, XUE Y, et al. A rapid learning algorithm for vehicle classification[J]. Information Sciences, 2015, 295: 395-406.

[6] ZHOU Z L, WANG Y L, WU Q M J, et al. Effective and efficient global context verification for image copy detection[J]. IEEE Transactions on Information Forensics and Security, 2017, 12(1): 48-63.

[7] 沈宇, 韩金朋, 李灵犀, 等. 游戏智能中的 AI: 从多角色博弈到平行博弈[J]. 智能科学与技术学报, 2020, 2(3): 205-213.

SHEN Y, HAN J P, LI L X, et al. AI in game intelligence-from multi-role game to parallel game[J]. Chinese Journal of Intelligent Science and Technology, 2020, 2(3): 205-213.

[8] 韩冬, 李其花, 蔡巍, 等. 人工智能在医学影像中的研究与应用[J]. 大数据, 2019, 5(1): 39-67.

HAN D, LI Q H, CAI W, et al. Research and application of artificial intelligence in medical imaging[J]. Big Data Research, 2019, 5(1): 39-67.

[9] TURK M, PENTLAND A. Eigenfaces for recognition[J]. Journal of Cognitive Neuroscience, 1991, 3(1): 71-86.

[10] BELHUMEUR P N, HESPAHNA J P, KRIEGMAN D J. Eigenfaces vs. Fisherfaces: recognition using class specific linear projection[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1997, 19(7): 711-720.

[11] KIM J, CHOI J, YI J, et al. Effective representation using ICA for face recognition robust to local distortion and partial occlusion[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2005, 27(12): 1977-1981.

[12] HE X F, YAN S C, HU Y X, et al. Face recognition using Laplacianfaces[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2005, 27(3): 328-340.

[13] LI S Z, LU J W. Face recognition using the nearest feature line method[J]. IEEE Transactions on Neural Networks, 1999, 10(2): 439-443.

[14] LI S Z, CHAN K L, WANG C L. Performance evaluation of the nearest feature line method in image classification and retrieval[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2000, 22(11): 1335-1339.

[15] CHIEN J T, WU C C. Discriminant waveletfaces and nearest feature classifiers for face recognition[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2002, 24(12): 1644-1649.

[16] WRIGHT J, YANG A Y, GANESH A, et al. Robust face recognition via sparse representation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2009, 31(2): 210-227.

[17] NASEEM I, TOGNERI R, BENNAMOUN M. Linear regression for face recognition[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2010, 32(11): 2106-2112.

[18] NASEEM I, TOGNERI R, BENNAMOUN M. Robust regression for

- face recognition[J]. Pattern Recognition, 2012, 45(1): 104-118.
- [19] REN C X, DAI D Q, YAN H. Robust classification using $L_{2,1}$ -norm based regression model[J]. Pattern Recognition, 2012, 45(7): 2708-2718.
- [20] SHEN F M, TANG Z M, XU J S. Locality constrained representation based classification with spatial pyramid patches[J]. Neurocomputing, 2013, 101: 104-115.
- [21] ZHANG L, YANG M, FENG X C. Sparse representation or collaborative representation: which helps face recognition?[C]// Proceedings of the IEEE 2011 International Conference on Computer Vision. Piscataway: IEEE Press, 2011: 471-478.
- [22] HE R, ZHENG W S, HU B G. Maximum correntropy criterion for robust face recognition[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2011, 33(8): 1561-1576.
- [23] ZHOU Z H, WAGNER A, MOBAHI H, et al. Face recognition with contiguous occlusion using Markov random fields[C]//Proceedings of the 2009 IEEE 12th International Conference on Computer Vision. Piscataway: IEEE Press, 2009: 1050-1057.
- [24] LIU W F, POKHAREL P P, PRINCIPE J C. Correntropy: properties and applications in non-Gaussian signal processing[J]. IEEE Transactions on Signal Processing, 2007, 55(11): 5286-5298.
- [25] SANTAMARIA I, POKHAREL P P, PRINCIPE J C. Generalized correlation function: definition, properties, and application to blind equalization[J]. IEEE Transactions on Signal Processing, 2006, 54(6): 2187-2197.
- [26] JEONG K H, LIU W F, HAN S, et al. The correntropy MACE filter[J]. Pattern Recognition, 2009, 42(5): 871-885.
- [27] YUAN X T, HU B G. Robust feature extraction via information theoretic learning[C]//Proceedings of the 26th Annual International Conference on Machine Learning. New York: ACM Press, 2009: 1193-1200.
- [28] SHI Q, ERIKSSON A, HENGEL A V D, et al. Is face recognition really a compressive sensing problem?[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2011: 553-560.
- [28] SHI Q F, ERIKSSON A, VAN DEN HENGEL A, et al. Is face recognition really a compressive sensing problem?[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2011: 553-560.
- [29] NIE F P, HUANG H, CAI X, et al. Efficient and robust feature selection via joint $L_{2,1}$ -norms minimization[C]//Proceedings of the 23rd International Conference on Neural Information Processing Systems. Piscataway: IEEE Press, 2010: 1813-1821.
- [30] TENENBAUM J B, DE SILVA V, LANGFORD J C. A global geometric framework for nonlinear dimensionality reduction[J]. Science, 2000, 290(5500): 2319-2323.
- [31] ROWEIS S T. Nonlinear dimensionality reduction by locally linear embedding[J]. Science, 2000, 290(5500): 2323-2326.
- [32] BELKIN M, NIYOGI P. Laplacian eigenmaps for dimensionality reduction and data representation[J]. Neural Computation, 2003, 15(6): 1373-1396.
- [33] YAN S C, XU D, ZHANG B, et al. Graph embedding and extensions: a general framework for dimensionality reduction[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2007, 29(1): 40-51.
- [34] ZHANG L Y, DENNEY B, LU J W. A collaborative approach for face verification and attributes refinement[J]. Information Sciences, 2014, 281: 620-634.
- [35] LIU F, TANG J H, SONG Y, et al. Local structure based multi-phase collaborative representation for face recognition with single sample per person[J]. Information Sciences, 2016, 346/347: 198-215.
- [36] ZHENG Z L, HUANG X Q, CHEN Z Y, et al. Regression analysis of locality preserving projections via sparse penalty[J]. Information Sciences, 2015, 303: 1-14.
- [37] DORNAIKA F, BOSAGHZADEH A. Adaptive graph construction using data self-representativeness for pattern classification[J]. Information Sciences, 2015, 325: 118-139.
- [38] YU K, ZHANG T, GONG Y H. Nonlinear learning using local coordinate coding[C]//Proceedings of the Advances in Neural Information Processing Systems. [S.l.:s.n.], 2009: 2223-2231.
- [39] WANG J J, YANG J C, YU K, et al. Locality-constrained linear coding for image classification[C]//Proceedings of the 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2010: 3360-3367.

[作者简介]



余沁茹(1997-),女,安徽工程大学计算机与信息学院硕士生,主要研究方向为图像处理。



卢桂馥(1976-),男,博士,安徽工程大学计算机与信息学院教授,主要研究方向为计算机图形学及图像处理。