

融合极端学习机的判别性分析字典学习算法



王军浩 闫德勤 刘德山 邢钰佳

辽宁师范大学计算机与信息技术学院 辽宁 大连 116033

(lnnu_junhaowang@163.com)

摘要 研究表明,端学习机和判别性字典学习算法在图像分类领域极具有高效和准确的优势。然而,这两种方法也具有各自的缺点,极端学习机对噪声的鲁棒性较差,判别性字典学习算法在分类过程中耗时较长。为统一这种互补性以提高分类性能,文中提出了一种融合极端学习机的判别性分析字典学习模型。该模型利用迭代优化算法学习最优的判别性分析字典和极端学习机分类器。为验证所提算法的有效性,利用人脸数据集进行分类。实验结果表明,与目前较为流行的字典学习算法和极端学习机相比,所提算法在分类过程中具有更好的效果。

关键词: 极端学习机;判别性字典学习;分析字典学习

中图分类号 TP391

Algorithm with Discriminative Analysis Dictionary Learning by Fusing Extreme Learning Machine

WANG Jun-hao, YAN De-qin, LIU De-shan and XING Yu-jia

School of Computer and Information Technology, Liaoning Normal University, Dalian, Liaoning 116033, China

Abstract Recent researches have shown that the speed advantage of extreme learning machine (ELM) and the accuracy advantage of discriminative dictionary learning (DDL) in the area of image classification. However these two methods have their respective drawbacks, in general, ELM is known to be less robust to noise while DDL is known to be time-consuming. In order to unify such mutual complementarity and further enhance the classification performance, we propose a discriminative analysis dictionary learning fusing extreme learning machine model in this paper. More precisely, the iterative optimization algorithm is used to learn the most optimal discriminative analysis dictionary and extreme learning machine classifier. In order to verify the effect of the proposed algorithm, the face data is used for classification. Experiments demonstrate that our method achieves a better performance than the state-of-the-art dictionary learning algorithms and extreme learning machine in a variety of image classification tasks.

Keywords Extreme learning machine, Discriminative dictionary learning, Analysis dictionary learning

1 引言

在计算机视觉、压缩感知和模式识别领域,稀疏表示已成为主流技术之一,主要用于图像去噪^[1]、图像恢复^[2-3]和图像分类^[4-13]等。传统的稀疏表示通过一个过完备字典线性表示输入信号,稀疏编码的性能主要依赖于过完备字典的重构性和判别性,因此构造一个判别性字典对稀疏表示至关重要。Elad等^[7]提出了K-SVD字典学习算法,该算法有效地从一组训练样本中学习一个过完备的学习字典,该字典具有比预先指定的字典更优的性能;Jiang等^[6]在文献[7]的基础上增加标签一致性约束,提出了LC-KSVD算法,使学习字典更具判别性;传统的字典学习算法在训练样本数目较少或人脸图像随着面部姿势、表情以及照明条件变化时,很难获得用于人

脸识别的判别性字典。Xu等^[14]为了解决以上问题,提出了样本多样性稀疏表示字典学习算法(Sample Diversity Dictionary Learning, SD-DL)。Li等^[15]提出了一种基于局部信息和标签信息的新字典学习算法(Locality Constrained and Label Embedding Dictionary Learning, LCLE-DL),该算法将训练样本的图拉普拉斯矩阵和标签一致性引入其中。Guo等^[16]将拓扑结构和表示相似性约束与分析字典相结合,提出了一种判别性的分析字典学习算法(Discriminative Analysis Dictionary Learning, DADL)。Wang等^[17]在变换稀疏编码模型和LC-KSVD算法的基础上,提出了基于分析字典的合成线性分类模型(Synthesis Linear Classifier Analysis Dictionary Learning, SLC-ADL),该算法对标准数据集的依赖性较强,在随机人脸数据集上的鲁棒性较差。Wang等^[18]在合成变换稀

到稿日期: 2019-06-18 返修日期: 2019-10-05 本文已加入开放科学计划(OSID),请扫描下方二维码获取补充信息。

基金项目: 辽宁省自然科学基金(20170540574); 辽宁省教育厅科学研究项目(LJ2019014)

This work was supported by the Natural Science Foundation of Liaoning Province, China(20170540574) and Scientific Research Project of Liaoning Education Department(LJ2019014).

通信作者: 闫德勤(yandeqin@163.com)

疏编码模型和 K-SVD 算法的基础上,提出了基于分析字典的合成 K-SVD 算法。Tang 等^[19]为提高学习过程的计算效率和字典的判别性能,在分析字典的基础上结合类的结构信息提出结构化的分析字典学习算法(Structured analysis dictionary Learning, SADL)。Wei 等^[20]为降低 0 范数和 1 范数作为约束在判别性字典学习过程中产生较长时耗的影响,提出了一种基于非线性分析融合稀疏表示模型的判别性字典学习算法,并将其应用于 SAR 图像的分类任务中。为进一步提高学习字典的判别能力,大多数字典学习算法、分析字典学习算法都在训练阶段对目标函数进行额外的限制,这些限制将在判别性字典的学习过程中产生额外的计算复杂度。

Huang 等^[21-22]提出的极端学习机(Extreme Learning Machine, ELM)是一种训练单隐层前馈神经网络(Single-hidden Layer Feedforward Neural Network, SLFN)的有效方法,为多类分类和回归提供了统一框架。基本的极端学习机模型可以简单地看作基于最小二乘的线性回归随机特征映射。极端学习机的主要优势在于可以随机生成输入层和隐含层之间的输入权重和偏置值,进而有效地学习隐含层和输出层之间的输出权重。这种改进极大地减轻了广泛使用反向传播算法^[23](Back Propagation, BP)引起的权重调整负担,从而保证了极端学习机的学习速度和泛化能力。此外,最近的一些研究^[24-25]验证了极端学习机随机生成的输入权重的合理性,这将进一步扩大极端学习机在多类分类任务^[26-29]中的应用。

最近,Cao 等^[30]将极端学习机与稀疏表示结合起来进行人脸图像分类,在分类过程中稀疏表示算法与极端学习机被单独进行优化。极端学习机和判别性字典学习(Discriminative Dictionary Learning, DDL)尽管在图像分类过程中呈现出了较好的分类特性,但仍存在着实际应用受限的缺点。研究表明,ELM 运行速度极快,但对噪声的鲁棒性较差;而 DDL 表现出了对噪声的极强鲁棒性,但在图像分类过程中产生了很高的计算成本。值得注意的是,在图像分类模型的设计上,不仅要考虑模型的预测精度,还要考虑模型的效率。本文为统一极端学习机和判别性字典学习的互补性,提出了一种融合极端学习机的判别性分析字典学习模型(Analysis Dictionary Learning Fusing Analysis Extreme Learning Machine, ADL-AELM),以在人脸图像分类过程中增强分析字典的重建性和判别性,进而提高人脸图像的分类性能和分类效率,为接下来极端学习机在稀疏表示分类领域的应用和研究提供借鉴。为证明本文模型在图像分类过程中具有的性能,本文在 4 个人脸数据库上将其与 5 组对比算法进行了综合比较,结果表明本文算法在图像分类中表现出了较强的鲁棒性和较高的分类性能。

2 ADL-AELM

文献[6,8,31-34]为提高学习字典在分类过程中的精确度并规避过拟合现象,将分类误差作为字典学习目标函数中的一个正则项,以确保学习出一个最具判别性的学习字典。文献[35-37]结合分析字典的特点,从不同角度进行研究并对其进行改进。基于这样的研究启发,本文结合极端学习机提出了新的分类方法。

极端学习机是一种训练单隐层前馈神经网络的有效方

法,为多类分类提供了统一框架。本节提出了一个融合极端学习机分类器 $\sum_{i=1}^L \beta_i g(\omega_i \cdot x_j + b_i) = t_j$ 的判别性分析字典学习模型,旨在利用输入样本的相关信息来重构和学习判别性分析字典。所提算法的框架由训练和测试两个阶段组成。在训练阶段,迭代优化极端学习机分类器 β 和判别性分析字典 D ;在测试阶段,利用分析字典 D 对测试样本进行稀疏表示,同时利用极端学习机分类器 β 对稀疏表示矩阵进行分类。对于各阶段的 ELM 而言,我们将激活函数 g 设置为 1,这意味着 $g(\omega_i \cdot x_j + b_i) = \omega_i \cdot x_j + b_i$ 。作为线性极端学习机分类器 β 在 $T^T - (X^T W^T + B)\beta$ 中的对偶观点,分析极端学习机分类器 β 采用以下形式 $(X^T W^T + B) - T^T \beta$,它建立了稀疏编码系数和数据标签矩阵(期望输出矩阵)之间的对应关系。

给定训练样本集 $Y = [y_1, \dots, y_N] \in R^{n \times N}$,假设训练样本集中的所有样本均来自 C 个不同的类别。令 $D = [d_1, \dots, d_n] \in R^{H_c \times n}$ 表示分析字典矩阵, d_i 表示字典中的一个原子; $X = [x_1, \dots, x_N] \in R^{H_c \times N}$ 为获得的稀疏表示矩阵。ADL-AELM 模型可表示为:

$$\langle D, X, \beta \rangle = \min_{D, X, \beta} \|X - DY\|^2 + \alpha \|D\|^2 + \gamma \|(X^T W^T + B) - T^T \beta\|^2 + \tau \|\beta\|^2 \quad (1)$$

其中, $\|X - DY\|^2 + \alpha \|D\|^2$ 表示字典学习重构误差项, $\gamma \|(X^T W^T + B) - T^T \beta\|^2 + \tau \|\beta\|^2$ 表示分析字典的分类项,参数 γ 是控制相关项在目标函数中的相对贡献量, α 和 τ 是惩罚参数。 $W = (\omega_1, \dots, \omega_{H_c})^T \in R^{L \times H_c}$ 表示输入权重矩阵, $B = (b_1, \dots, b_L)^T \in R^{N \times L}$ 表示隐含层节点的偏移值, $\beta = (\beta_1, \dots, \beta_{H_c}) \in R^{L \times H_c}$ 表示输出权重向量, L 表示 ELM 隐含层节点的个数, H_c 表示稀疏表示矩阵的长度。 $T = (t_1, \dots, t_N) \in R^{H_c \times N}$ 表示训练样本的期望输出矩阵, $t_i = [0, \dots, 0, 1, 0, \dots, 0]$ 表示训练样本的标签向量,其中非零位置表示训练样本的类别。本文将初始化稀疏表示矩阵,用期望输出矩阵来赋值。

例如,假设训练样本 $Y = [y_1, \dots, y_5] \in R^{n \times 5}$ 和样本标签矩阵 $I = [t_1, \dots, t_5] \in R^{1 \times 5}$ 已知,其中 y_1 属于第一类样本, y_2 和 y_3 属于第二类样本, y_4 属于第三类样本。根据以上信息可以定义期望输出矩阵 $T = [t_1, \dots, t_5] \in R^{3 \times 5}$ 为:

$$T = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 \end{bmatrix}_{3 \times 5} \quad (2)$$

文献[17]提出,对于分析字典而言,稀疏表示矩阵的长度往往大于输入样本的维度,因此需要对期望输出矩阵进行扩展。如果不扩展期望输出矩阵,则在分类过程中只由一个元素在期望输出矩阵中的位置来确定测试样本的标签,这就更易导致错误的判别。两个任意矩阵 A 和 B 的 Kronecker 积通常用 $A \otimes B$ 表示。给定 $A \in R^{a \times b}$ 和 $B \in R^{c \times d}$, $A \otimes B$ 是一个 $ac \times bd$ 的块矩阵:

$$A \otimes B = \begin{bmatrix} A_{11}B & \cdots & A_{1b}B \\ \vdots & & \vdots \\ A_{a1}B & \cdots & A_{ab}B \end{bmatrix}_{a \times b} \quad (3)$$

简单起见,我们用期望输出矩阵与一个全一向量的 Kronecker 积来代替期望输出矩阵,其中全一向量的长度就是期望输出矩阵维数的扩展倍数。例如,给定一个全一向量 $Z =$

$[1,1]^T$,初始化期望输出矩阵可改写为:

$$\mathbf{T} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 \end{bmatrix}_{H_c \times N} \quad (4)$$

最后,基于所获得的判别性分析字典 $\mathbf{D} \in \mathbb{R}^{H_c \times n}$,利用 $\mathbf{X}_{\text{test}} = \mathbf{D} * \mathbf{Y}_{\text{test}}$ 算法将测试样本 $\mathbf{Y}_{\text{test}} = [y_1, \dots, y_M] \in \mathbb{R}^{n \times M}$ 转换为稀疏编码矩阵 $\mathbf{X}_{\text{test}} = [x_1, \dots, x_M] \in \mathbb{R}^{H_c \times M}$,然后利用极端学习机分类器 $\beta \in \mathbb{R}^{L \times H_c}$ 进行分类。

3 算法优化

近年来,学者提出了许多新的字典学习优化策略,如交替方向法^[38]、基于半二次方的迭代最小化法^[39]和交替迭代法^[7,40]。本文为有效地学习一个判别性分析字典,采用交替迭代法来优化所提出的模型。

3.1 ADL-AELM 算法优化

对于 ADL-AELM 算法的优化而言,首先利用 Kronecker 积扩展期望输出矩阵 \mathbf{T} 并初始化稀疏表示矩阵 $\mathbf{X}_0, \beta_0, \mathbf{D}_0$; 然后对式(1)进行最优化,以求解判别性分析字典 $\mathbf{D}^{(*)}$ 、极端学习机分类系数 $\beta^{(*)}$ 和稀疏矩阵 $\mathbf{X}^{(*)}$ 。最后利用以下 3 个步骤多次迭代学习 $\beta, \mathbf{D}_{\text{train}}$ 和 $\mathbf{X}_{\text{train}}$ (此处以第一次迭代操作为例)。

步骤 1 固定稀疏矩阵 \mathbf{X}_0 ,更新学习字典 \mathbf{D}_1 。

为获得 ADL-AELM 算法中关于参数 \mathbf{D} 的最优解,式(1)可以转换为:

$$\mathbf{D}^{(*)} = \arg \min_{\mathbf{D}} \|\mathbf{X} - \mathbf{D}\mathbf{Y}\|^2 + \alpha \|\mathbf{D}\|^2 \quad (5)$$

其中, $\|\mathbf{X} - \mathbf{D}\mathbf{Y}\|^2$ 表示分析字典学习的重构误差, α 是惩罚参数, \mathbf{Y} 表示训练样本, \mathbf{D} 表示分析字典, \mathbf{X} 表示稀疏表示矩阵。

令目标函数(5)的一阶导数为零,分析字典 \mathbf{D} 的最优解,如式(6)所示:

$$\frac{\partial \mathcal{D}^{(*)}(\mathbf{X}_0, \mathbf{D}_0, \mathbf{Y})}{\partial \mathbf{D}} = 0$$

$$-\mathbf{X}_0 \mathbf{Y}^T + \mathbf{D}_1 \mathbf{Y} \mathbf{Y}^T + \alpha \mathbf{D}_1 = 0 \quad (6)$$

$$\mathbf{D}_1 = \mathbf{X}_0 \mathbf{Y}^T (\mathbf{Y} \mathbf{Y}^T + \alpha \mathbf{I})^{-1}$$

步骤 2 固定稀疏表示 \mathbf{X}_0 ,更新极端学习机分类系数 β_1 。

为获得 ADL-AELM 算法中关于参数 β 的最优解,式(1)可以转换为:

$$\beta^{(*)} = \arg \min_{\beta} \gamma \|(X^T \mathbf{W}^T + \mathbf{B}) - \mathbf{T}^T \beta\|^2 + \tau \|\beta\|^2 \quad (7)$$

其中, $\|(X^T \mathbf{W}^T + \mathbf{B}) - \mathbf{T}^T \beta\|^2$ 表示分类误差项,参数 γ 用于控制该项在目标函数中的相对贡献, τ 是惩罚参数, \mathbf{T} 表示训练样本的期望输出矩阵, \mathbf{W} 表示输入权重矩阵, \mathbf{B} 表示隐含层节点的偏移值。

令目标函数(7)的一阶导数为零,极端学习机分类器 β 的最优解如式(8)所示:

$$\frac{\partial \beta^{(*)}(\mathbf{X}_0, \beta_0, \mathbf{T}, \mathbf{W}, \mathbf{B})}{\partial \beta} = 0$$

$$\gamma(-\mathbf{T} \mathbf{X}_0^T \mathbf{W}^T - \mathbf{T} \mathbf{B} + \mathbf{T} \mathbf{T}^T \beta_1) + \tau \beta_1 = 0 \quad (8)$$

$$(\gamma \mathbf{T} \mathbf{T}^T + \tau \mathbf{I}) \beta_1 = \gamma(\mathbf{T} \mathbf{X}_0^T \mathbf{W}^T + \mathbf{T} \mathbf{B})$$

$$\beta_1 = (\gamma \mathbf{T} \mathbf{T}^T + \tau \mathbf{I})^{-1} \times \gamma(\mathbf{T} \mathbf{X}_0^T \mathbf{W}^T + \mathbf{T} \mathbf{B})$$

步骤 3 固定极端学习机分类系数 β_1 、学习字典 \mathbf{D}_1 ,更新稀疏矩阵 \mathbf{X}_1 。

为获得 ADL-AELM 算法中关于参数 \mathbf{X} 的最优解,式(1)可以转换为:

$$\mathbf{X}^{(*)} = \arg \min_{\mathbf{X}} \|\mathbf{X} - \mathbf{D}\mathbf{Y}\|^2 + \alpha \|\mathbf{X}\|^2 + \gamma \|(X^T \mathbf{W}^T + \mathbf{B}) - \mathbf{T}^T \beta\|^2 \quad (9)$$

同理,令目标函数(9)的一阶导数为零,稀疏矩阵 \mathbf{X} 的最优解如式(10)所示:

$$\frac{\partial \mathbf{X}^{(*)}(\mathbf{X}_0, \beta_1, \mathbf{D}_1, \mathbf{T}, \mathbf{W}, \mathbf{B})}{\partial \mathbf{X}} = 0$$

$$(\mathbf{X}_1 - \mathbf{D}_1 \mathbf{Y}) + \gamma \mathbf{W}^T (\mathbf{W} \mathbf{X}_1 + \mathbf{B}^T - \beta_1^T \mathbf{T}) = 0 \quad (10)$$

$$\mathbf{X}_1 + \gamma \mathbf{W}^T \mathbf{W} \mathbf{X}_1 = \mathbf{D}_1 \mathbf{Y} - \gamma \mathbf{W}^T (\mathbf{B}^T + \beta_1^T \mathbf{T})$$

$$\mathbf{X}_1 = (\gamma \mathbf{W}^T \mathbf{W} + \alpha \mathbf{I})^{-1} \times 2 \times (\mathbf{D}_1 \mathbf{Y} - \gamma \mathbf{W}^T (\mathbf{B}^T + \beta_1^T \mathbf{T}))$$

上述迭代步骤如算法 1 所示。

算法 1 ADL-AELM 算法

输入:训练样本 \mathbf{Y} ,训练样本标签 \mathbf{H} ,期望输出矩阵 \mathbf{T} ,极端学习机隐含层节点数 L ,迭代次数 t ,稀疏度 $sparsity$ 以及惩罚参数 α, γ 和 τ
输出:判别性分析字典 $\mathbf{D}_{\text{train}}$ 和极端学习机分类器 β

1. 初始化学习字典 \mathbf{D}_0 、稀疏矩阵 \mathbf{X}_0 、极端学习机输入权重矩阵 \mathbf{W} 、隐含层节点的偏移值 \mathbf{B} 和极端学习机分类系数 β_0 。设置当前迭代次数 $t_0 = 0$ 。
2. While($t_0 \leq t$)
3. do
4. $t_0 = t_0 + 1$;
5. 根据式(6)固定稀疏表示 \mathbf{X}_{t_0-1} ,更新学习字典 \mathbf{D}_{t_0} ;
6. 根据式(8)固定稀疏矩阵 \mathbf{X}_{t_0-1} ,更新极端学习机分类器 β_{t_0} ;
7. 根据式(10)固定极端学习机分类系数 β_{t_0} 、学习字典 \mathbf{D}_{t_0} ,更新稀疏矩阵 \mathbf{X}_{t_0} ;
8. end While.

3.2 分类方案的设计

在测试样本的分类过程中,判别性分析字典 $\mathbf{D}_{\text{train}}$ 和极端学习机分类器 β 将会起到至关重要的作用。给定训练样本 \mathbf{Y} 、训练样本标签 \mathbf{H} 、期望输出矩阵 \mathbf{T} 、极端学习机隐含层节点数 L 、迭代次数 t 、稀疏度 $sparsity$ 以及惩罚参数 α, γ 和 τ ,通过算法 1 学习判别性分析字典 $\mathbf{D}_{\text{train}}$ 和极端学习机分类器 β 。对于测试样本 \mathbf{Y}_{test} 而言,首先利用判别性分析字典 $\mathbf{D}_{\text{train}}$ 对式(11)进行稀疏表示求解,计算测试样本最优的稀疏表示矩阵 \mathbf{X}_{test} 。

$$\langle \mathbf{X}_{\text{test}} \rangle = \min_{\mathbf{X}_{\text{test}}} \|\mathbf{X}_{\text{test}} - \mathbf{D}_{\text{train}} \mathbf{Y}_{\text{test}}\|^2 \quad (11) \text{kWTI}$$

$$\text{s. t. } \forall i, \|\mathbf{x}_i\|_0 \leq T$$

然后,利用式(12)预测测试样本 \mathbf{Y}_{test} 的标签矩阵 \mathbf{A}_{Pre} :

$\mathbf{Pre}_{\text{test}}$:

$$\mathbf{A}_{\text{Pre}_{\text{test}}} = \arg \max_{\mathbf{Pre}_{\text{test}}} (\mathbf{H}_{\text{test}} = (\beta \beta^T)^{-1} (\beta \mathbf{W} \mathbf{X}_{\text{test}} + \beta \mathbf{B}^T)) \quad (12)$$

其中, \mathbf{H}_{test} 表示测试样本的标签矩阵, $\mathbf{A}_{\text{Pre}_{\text{test}}}$ 表示测试样本的预测标签矩阵。

4 实验对比与分析

本节将采用 4 个经过标准化处理的人脸数据集(即 Extended YaleB^[41], AR^[42], ORL 和 LFW Crop^[43])进行实验,以评估 ADL-AELM 稀疏模型的性能。同时,为了进一步验证

ADL-AELM 的优势,选择传统的合成字典学习算法 K-SVD, LC-KSVD, LCLE-DL, 分析字典学习算法 SLC-ADL, 以及最近提出的 SADL 算法进行综合比较,并选取固定训练样本个数与 ELM 进行不同隐含层节点个数分类性能的比较。另外,除考虑分类准确率这个因素以外,本文还考虑这些对比算法在实验中的训练时间和测试时间。为确保实验结果具有说服力,本文参考文献[6,14]对实验参数进行设置,其中迭代次数设置为 10,稀疏度设置为 30。所有实验的运行平台均为 4 GB 内存,3.20 GHz Intel 处理器的惠普计算机,实验环境为 Matlab2016。

4.1 Extended YaleB 人脸数据库上的实验结果分析

Extended YaleB 人脸数据库(以下简称 YaleB 人脸数据库)包含 38 个人在不同光照和表情下获得的 2 414 张正面人脸图像,其中每个人大约拥有分辨率为 32×32 像素的 59~64 张标准化人脸图像。我们随机筛选每个人的 30 张人脸图像作为训练样本,将剩余图像作为测试样本。YaleB 人脸数据库的部分人脸图像如图 1 所示。在 ADL-AELM 实验中,惩罚参数 α, γ 和 τ 分别设置为 50, 0.01, 0.01;极端学习机隐含层的节点个数设置为 1 000。重复进行 5 次对比实验,平均准确率、训练时间、测试时间如表 1 所列。此外,训练样本数目分别为 20, 25, ..., 45, 50 时,ADL-AELM 算法的分类准确率变化如图 2 所示;极端学习机隐含层节点个数分别为 200, 400, 600, ..., 1 800, 2 000 时,ADL-AELM 算法的分类准确率变化如图 3 所示。



图 1 YaleB 人脸数据库
Fig. 1 YaleB face database

表 1 YaleB 人脸数据库上相关算法的分类准确率与计算时间
Table 1 Classification accuracy and computing time on YaleB face database

算法	准确率/%	训练时间/s	测试时间/s
SRC	95.21	0.00	186.80
LC-KSVD	96.60	43.29	0.32
LCLE-DL	97.31	4.81	0.36
SLC-ADL	97.80	1.15	0.07
SADL	95.05	325.61	0.06
ADL-AELM	98.00	1.29	0.05

从表 1 可以看出,ADL-AELM 具有最优的分类效果,与 SDL 相比准确率大约提高了 0.7%~1.4%,与 ADL 相比准确率大约提高了 0.1%~3.5%。同时,本文算法在训练时间和测试时间上相比合成字典学习也有巨大的缩减。

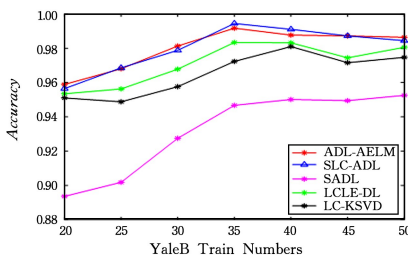


图 2 在 YaleB 人脸数据库上分类准确率与训练样本的关系
Fig. 2 Classification accuracy with different numbers of training on YaleB face database

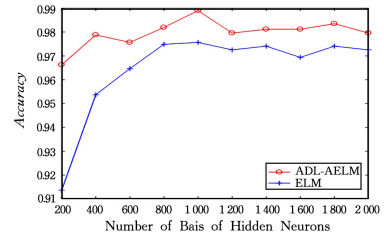


图 3 YaleB 人脸数据库上分类准确率与隐含层节点的关系
Fig. 3 Classification accuracy with different numbers of hidden neurons on YaleB face database

从图 2 可以看出,随着训练样本数量的增加,本文算法的分类准确率逐渐增高,直到达到较高分类准确率时逐渐平稳,当训练样本个数为 35 时具有最优的分类准确率。从图 3 可以看出,本文算法具有较优的分类准确率且较稳定,当隐含层节点个数为 1 000 时具有最优的分类准确率。

4.2 AR 实验结果与分析

AR 人脸数据库包含 126 人的 4 000 多张人脸图像。实验中使用的 AR 人脸数据库是经过后期标准化处理的子数据库,其中包含 50 名男性和 50 名女性的 2 600 张人脸图像。随机筛选每个人的 20 张人脸图像作为训练样本,将剩余图像作为测试样本。AR 人脸数据库的部分人脸图像展示如图 4 所示。在 ADL-AELM 实验中,惩罚参数 α, γ 和 τ 分别设置为 50, 0.01, 50;极端学习机隐含层节点个数设置为 1 000。重复进行 5 次对比实验,平均准确率、训练时间、测试时间如表 2 所列。此外,训练样本数目分别为 12, 14, ..., 22, 24 时,ADL-AELM 算法的平均准确率变化如图 5 所示;极端学习机隐含层节点个数分别为 200, 400, 600, ..., 1 800, 2 000 时,ADL-AELM 算法的平均准确率变化如图 6 所示。



图 4 AR 人脸数据库
Fig. 4 AR face database

表 2 AR 人脸数据库上相关算法的分类准确率与计算时间
Table 2 Classification accuracy and computing time on AR Face database

算法	准确率/%	训练时间/s	测试时间/s
SRC	92.33	0	110.97
LC-KSVD	89.50	10.39	0.15
LCLE-DL	91.38	3.87	0.14
SLC-ADL	97.20	1.41	0.14
SADL	95.83	753.80	0.05
ADL-AELM	97.20	5.09	0.16

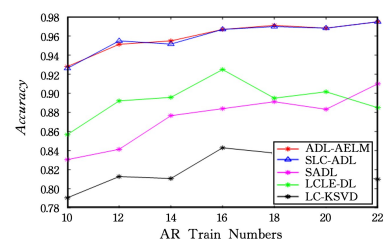


图 5 AR 数据库上分类准确率与训练样本的关系
Fig. 5 Classification accuracy with different numbers of training on AR face database

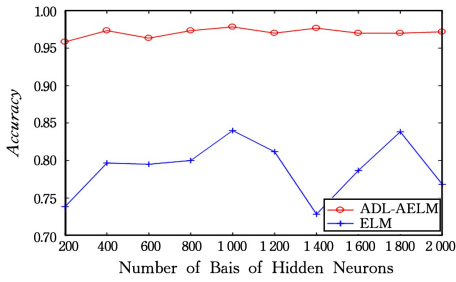


图 6 AR 数据库上分类准确率与隐含层节点的关系

Fig. 6 Classification accuracy with different numbers of hidden neurons on AR face database

从表 2 可以看出,本文算法和 SLC-ADL 算法具有最优的分类效果,分类准确率达到 97.2%,与 SDL 相比准确率大约提高了 6%,与 ADL 相比准确率大约提高了 1.5%。从图 5 可以看出,随着训练样本数量的增加,本文算法的准确率逐渐提高。随着训练样本的变化,ADL-ELM 算法与 SLC-ADL 算法的准确率总体持平,但是当训练样本个数分别为 14 和 18 时,本文算法具有更优的分类准确率。从图 6 可以看出,ADL-AELM 算法具有更优的分类准确率且较稳定,当隐含层节点个数设置为 1000 时具有最优的分类准确率。

4.3 在 ORL 人脸数据库上的实验结果与分析

ORL 人脸数据库包含 40 个人在不同表情和姿态变化下获得的 400 张正面人脸图像,其中每个人拥有图像标准化大小为 32×32 像素的 10 张人脸图像,随机选择每个类别的 5 张图像作为训练样本,将剩余图像作为测试样本。ORL 人脸数据库的部分人脸图像如图 7 所示。在 ADL-AELM 实验中,惩罚参数 α, γ 和 τ 分别设置为 0.1, 0.001, 0.1; 极端学习机隐含层节点个数设置为 1000。重复进行 5 次对比实验,平均准确率、训练时间、测试时间如表 3 所列。此外,训练样本数目分别为 2, 3, ..., 8, 9 时,ADL-AELM 算法的平均识别率变化如图 8 所示; 极端学习机隐含层节点个数分别为 200, 400, ..., 1800, 2000 时,ADL-AELM 算法的平均识别率变化如图 9 所示。



图 7 ORL 人脸数据库

Fig. 7 ORL face database

表 3 ORL 人脸数据库上相关算法的分类准确率与计算时间
Table 3 Classification accuracy and computing time on ORL Face database

算法	准确率/%	训练时间/s	测试时间/s
SRC	93.50	0	9.31
LC-KSVD	96.00	0.65	0.03
LCLE-DL	89.00	0.16	0.02
SLC-ADL	96.50	1.50	0.11
SADL	87.50	22.88	0.01
ADL-AELM	98.00	7.37	0.21

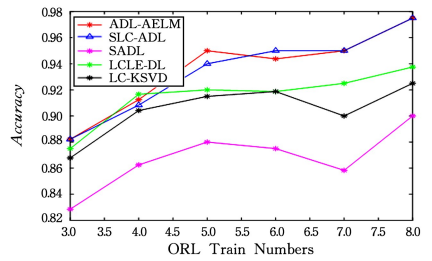


图 8 ORL 数据库上分类准确率与训练样本的关系

Fig. 8 Classification accuracy with different numbers of training on ORL face database

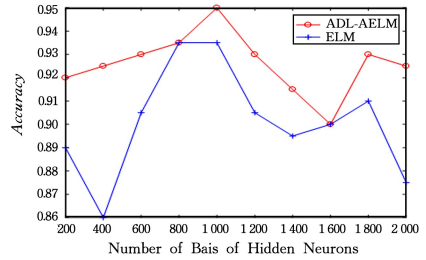


图 9 在 ORL 数据库上分类准确率与隐含层节点的关系

Fig. 9 Classification accuracy with different numbers of hidden neurons on ORL face database

从表 3 可以看出,本文算法 ADL-AELM 具有最优的分类效果,与 SDL 相比准确率大约提高了 2%,与 ADL 相比准确率大约提高了 1.5%~10.5%。从图 8 可以看出,随着训练样本数量的增加,本文算法的分类准确率逐渐增高,当训练样本个数为 8 时其具有最优的分类准确率。从图 9 可以看出,本文算法具有较优的分类准确率,当隐含层节点个数设置为 1000 时具有最优的分类准确率。

4.4 在 LFW Crop 人脸数据库上的实验结果与分析

LFW Crop 人脸数据库是 LFW 数据集中具有标签人脸的裁剪版本,仅保留每个图像的中心部分,在绝大多数的图像中几乎所有的背景都被省略。由于 LFW 中人脸的位置和大小是通过自动人脸定位器来确定的,因此 LFW Crop 人脸数据库可能呈现不对齐、尺度变化、平面内或平面外旋转等真实情况。LFW Crop 包含 86 个人在不同状态下获得的 1251 张人脸图像,其中每个人大约拥有图像标准化大小为 64×64 像素的 11~20 张人脸图像,随机选择每个类别的 8 张图像作为训练样本,将剩余图像作为测试样本。LFW 人脸数据库的部分人脸图像如图 10 所示。在 ADL-AELM 实验中,惩罚参数 α, γ 和 τ 分别设置为 1, 1, 1; 极端学习机隐含层节点个数均设置为 2000。重复进行 5 次对比实验,平均准确率、训练时间、测试时间如表 4 所列。此外,训练样本数目分别为 3, 4, ..., 9, 10 时,ADL-AELM 算法的平均识别率变化如图 11 所示; 极端学习机隐含层节点个数分别为 200, 400, ..., 1800, 2000 时,ADL-AELM 算法的平均识别率变化如图 12 所示。



图 10 LFW Crop 人脸数据库

Fig. 10 LFW Crop face database

表 4 LFW Crop 人脸数据库上相关算法的分类准确率与计算时间
Table 4 Classification accuracy and computing time on LFW Crop

Face database			
算法	准确率/%	训练时间/s	测试时间/s
SRC	39.96	0	62.16
LC-KSVD	29.48	10.78	0.20
LACLE-DL	31.79	6.49	0.24
SLC-ADL	37.83	1.00	0.02
SADL	19.79	60.75	0.02
ADL-AELM	37.52	1.45	0.04

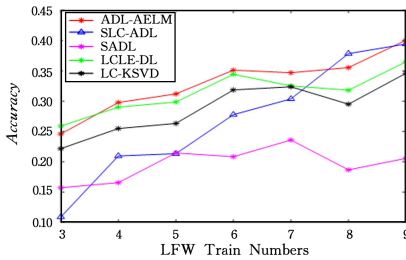


图 11 LFW Crop 数据库上分类准确率与训练样本的关系

Fig. 11 Classification accuracy with different numbers of training on LFW Crop face database

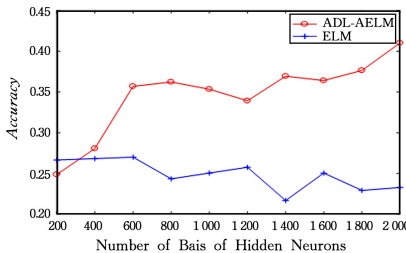


图 12 LFW Crop 数据库上分类准确率与隐含层节点的关系

Fig. 12 Classification accuracy with different numbers of hidden neurons on LFW Crop face database

从图 11 可以看出,随着训练样本数量的增加,本文算法的准确率逐渐提高,当训练样本为 9 时其具有最优的分类效果,分类准确率达 40.9%,与 SDL 相比准确率大约提高了 5%,与 ADL 相比准确率大约提高了 1%。从图 12 可以看出,ADL-AELM 算法具有更优的分类准确率,当隐含层节点个数设置为 2000 时具有最优的分类准确率。

结束语 本文提出了一种新的融合极端学习机的判别性分析字典学习算法,该算法将极端学习机和分析字典相融合,以学习最优的判别性学习字典。与传统的稀疏表示结合极端学习机分类算法相比,本文算法在融合时充分考虑了稀疏表示与极端学习机的自适应性和契合度,为接下来极端学习机在高光谱图像、场景图像稀疏表示分类等领域的应用提供了借鉴。本文所提算法的优点如下:1)将极端学习机正则项加入分析字典的目标函数中,以增强稀疏表示与标签信息的关联性;2)在字典重构的同时,通过极端学习机约束增强学习字典的判别性;3)在人脸图像分类过程中,与传统的字典学习算法相比,所提出的判别性分析字典学习算法对噪声更具鲁棒性。本文在 4 个人脸数据库上的实验也进一步证明了以上优点。

参考文献

[1] ELAD M, AHARON M. Image Denoising Via Sparse and Re-

dundant Representations Over Learned Dictionaries[J]. IEEE Transactions on Image Processing, 2006, 15(12): 3736-3745.

- [2] MAIRAL J, ELAD M, SAPIRO G. Sparse Representation for Color Image Restoration[J]. IEEE Transactions on Image Processing, 2007, 17(1): 53-69.
- [3] RANZATO M, POULTNEY C, CHOPRA S, et al. Efficient Learning of Sparse Representations with an Energy-Based Model [C] // Advances in Neural Information Processing Systems, 2006: 1137-1144.
- [4] WRIGHT J, YANG A Y, GANESH A, et al. Robust face recognition via sparse representation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2008, 31(2): 210-227.
- [5] YANG J, YU K, GONG Y, et al. Linear spatial pyramid matching using sparse coding for image classification [C] // Cvpr, 2009: 1794-1801.
- [6] JIANG Z, LIN Z, DAVIS L S. Label Consistent K-SVD: Learning a Discriminative Dictionary for Recognition[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2013, 35(11): 2651-2664.
- [7] AHARON M, ELAD M, BRUCKSTEIN A. K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation[J]. IEEE Transactions on Signal Processing, 2006, 54(11): 4311-4322.
- [8] ZHANG Q, LI B. Discriminative K-SVD for dictionary learning in face recognition [C] // 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. IEEE, 2010: 2691-2698.
- [9] CAI S, ZUO W, ZHANG L, et al. Support vector guided dictionary learning [C] // European Conference on Computer Vision. Cham: Springer, 2014: 624-639.
- [10] RUBINSTEIN R, PELEG T, ELAD M. Analysis K-SVD: A Dictionary-Learning Algorithm for the Analysis Sparse Model[J]. IEEE Transactions on Signal Processing, 2013, 61(3): 661-677.
- [11] SHEKHAR S, PATEL V M, CHELLAPPA R. Analysis sparse coding models for image-based classification [C] // 2014 IEEE International Conference on Image Processing (ICIP). IEEE, 2014: 5207-5211.
- [12] TANG W, PANNAHI A, KRIM H, et al. Analysis Dictionary Learning: An Efficient and Discriminative Solution [J]. arXiv: 1903.03058, 2019.
- [13] PENG J, JIANG X, CHEN N, et al. Local adaptive joint sparse representation for hyperspectral image classification [J]. Neurocomputing, 2019, 334: 239-248.
- [14] XU Y, LI Z, ZHANG B, et al. Sample diversity, representation effectiveness and robust dictionary learning for face recognition [J]. Information Sciences, 2017, 375: 171-182.
- [15] LI Z, LAI Z, XU Y, et al. A Locality-Constrained and Label Embedding Dictionary Learning Algorithm for Image Classification [J]. IEEE Transactions on Neural Networks and Learning Systems, 2015: 1-16.
- [16] GUO J, GUO Y, KONG X, et al. Discriminative analysis dictionary learning [C] // Thirtieth AAAI Conference on Artificial Intelligence, 2016.
- [17] WANG J, GUO Y, GUO J, et al. Synthesis linear classifier based

- analysis dictionary learning for pattern classification[J]. *Neurocomputing*, 2017, 238:103-113.
- [18] WANG Q, GUO Y, GUO J, et al. Synthesis K-SVD based analysis dictionary learning for pattern classification[J]. *Multimedia Tools and Applications*, 2018, 77(13):17023-17041.
- [19] TANG W, PANAHI A, KRIM H, et al. Structured analysis dictionary learning for image classification[C]// 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2018:2181-2185.
- [20] WEI Y, JIAO L, LIU F, et al. Fast DDL Classification for SAR Images with l_1, ∞ Constraint[J]. *IEEE Access*, 2019, 7:68991-69006.
- [21] HUANG G B, ZHU Q Y, SIEW C K. Extreme learning machine: a new learning scheme of feedforward neural networks [C]// 2004 IEEE International Joint Conference on Neural Networks. 2004:985-990.
- [22] HUANG G B, ZHU Q Y, SIEW C K. Extreme learning machine: theory and applications[J]. *Neurocomputing*, 2006, 70(1/2/3):489-501.
- [23] RUMELHART D E, HINTON G E, WILLIAMS R J. Learning representations by back-propagating errors[J]. *Cognitive Modeling*, 1988, 5(3):1.
- [24] LIU X, LIN S, FANG J, et al. Is extreme learning machine feasible? A theoretical assessment (Part I)[J]. *IEEE Transactions on Neural Networks and Learning Systems*, 2014, 26(1):7-20.
- [25] LIN S, LIU X, FANG J, et al. Is extreme learning machine feasible? A theoretical assessment (Part II)[J]. *IEEE Transactions on Neural Networks and Learning Systems*, 2014, 26(1):21-34.
- [26] YAN D, CHU Y, ZHANG H, et al. Information discriminative extreme learning machine[J]. *Soft Computing*, 2018, 22(2):677-689.
- [27] LI Q, LIU Y, WANG S, et al. Image Classification Using Low-Rank Regularized Extreme Learning Machine[J]. *IEEE Access*, 2019, 7:877-883.
- [28] PARK Y, YANG H S. Convolutional neural network based on an extreme learning machine for image classification[J]. *Neurocomputing*, 2019, 339:66-76.
- [29] SUN Y, LI B, YUAN Y, et al. Big graph classification frameworks based on Extreme Learning Machine[J]. *Neurocomputing*, 2019, 330:317-327.
- [30] CAO J, ZHANG K, LUO M, et al. Extreme learning machine and adaptive sparse representation for image classification[J]. *Neural Networks*, 2016, 81:91-102.
- [31] YANG J, YU K, HUANG T. Supervised translation-invariant sparse coding[C]// 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. IEEE, 2010:3517-3524.
- [32] PHAM D S, VENKATESH S. Joint learning and dictionary construction for pattern recognition[C]// 2008 IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2008:1-8.
- [33] MAIRAL J, PONCE J, SAPIRO G, et al. Supervised dictionary learning[C]// *Advances in Neural Information Processing Systems*. 2009:1033-1040.
- [34] YANG C, LIU H, WANG S, et al. Remote sensing image classification using extreme learning machine-guided collaborative coding[J]. *Multidimensional Systems and Signal Processing*, 2017, 28(3):835-850.
- [35] SONG X, CHEN Y, FENG Z H, et al. Collaborative representation based face classification exploiting block weighted LBP and analysis dictionary learning[J]. *Pattern Recognition*, 2019, 88:127-138.
- [36] WANG H H, TU C W, CHIANG C K. Sparse representation for image classification via paired dictionary learning[J]. *Multimedia Tools and Applications*, 2019, 78(12):16945-16963.
- [37] SONG J, XIE X, SHI G, et al. Multi-layer discriminative dictionary learning with locality constraint for image classification [J]. *Pattern Recognition*, 2019, 91:135-146.
- [38] RAKOTOMAMONJY A. Applying alternating direction method of multipliers for constrained dictionary learning[J]. *Neurocomputing*, 2013, 106:126-136.
- [39] HE R, ZHENG W S, TAN T, et al. Half-quadratic-based iterative minimization for robust sparse representation [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2013, 36(2):261-275.
- [40] YAGHOUBI M, BLUMENSATH T, DAVIES M E. Dictionary learning for sparse approximations with the majorization method [J]. *IEEE Transactions on Signal Processing*, 2009, 57(6):2178-2191.
- [41] GEORGHIADES A S, BELHUMEUR P N, KRIEGMAN D J. From few to many: Illumination cone models for face recognition under variable lighting and pose[J]. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 2001(6):643-660.
- [42] MARTINEZ A, BENAVENTE R. The AR face database; 24 CVC Technical Report[R]. 1998.
- [43] HUANG G B, MATTAR M, BERG T, et al. Labeled faces in the wild: A database for studying face recognition in unconstrained environments[C]// *Workshop on Faces in 'Real-Life' Images: Detection, Alignment, and Recognition*. 2008.



WANG Jun-hao, born in 1994, postgraduate. His main research interests include machine learning, dictionary learning and remote sensing image classification.



YAN De-qin, born in 1962, Ph.D, professor. His main research interests include machine learning, dictionary learning, deep learning and remote sensing image classification.