·人工智能及识别技术· 文章编号：1000—3428(2013)05—0225—05 文献标识码：A 中图分类号：TP301.6

基于自适应梯度最速下降的分类字典训练算法

徐 健[1,2]，常志国[3]，赵小强[1]，马 祥[3]

(1. 西安邮电学院通信与信息工程学院，西安 710121；2. 西安交通大学电子与信息工程学院，西安 710049；
3. 长安大学信息工程学院，西安 710064)

摘 要：为提高稀疏表示字典用于图像分类时的正确率，提出一种基于自适应梯度最速下降的分类字典训练算法。该算法采用交替梯度下降法解决分类字典训练的双变量优化模型。为提高收敛速度，并保证稀疏表示残差和不同类别对应字典原子间的不相关性同时收敛，采用自适应步长，推导证明自适应步长的计算方法。通过固定字典，运用当前字典和训练样本计算出稀疏表示系数的下降方向和自适应步长，按照稀疏度约束将小系数置零，固定稀疏表示系数，利用稀疏表示系数和样本找到字典的下降方向和自适应步长。实验结果表明，该算法在手写字符识别中正确率能达到 96.51%。

关键词：图像分类；字典训练；稀疏表示；手写字符识别；双变量优化；自适应步长

Discriminate Dictionary Training Algorithm Based on Adaptive Steepest Descent

XU Jian[1,2], CHANG Zhi-guo[3], ZHAO Xiao-qiang[1], MA Xiang[3]

(1. School of Communication and Information Engineering, Xi'an University of Posts & Telecommunications, Xi'an 710121, China; 2. School of Electronic and Information Engineering, Xi'an Jiaotong University, Xi'an 710049, China; 3. School of Information Engineering, Chang'an University, Xi'an 710064, China)

【Abstract】To improve the correct rate of image classification using sparse representation dictionaries, a discriminate dictionary training algorithm based on adaptive steepest descent is proposed. This algorithm uses alternating gradient descent for bi-objective optimization model of discriminate dictionary training. To guarantee the convergence of sparse representation residual and the incoherence of atoms from dictionaries corresponding to different classes and increasing convergence rate, adaptive step size is adopted and the method of calculating the step size is proved. The steps of the algorithm are as following: Fix dictionaries and use current dictionaries and data samples to calculate the steepest descent directions and adaptive step sizes of the sparse representation coefficients. Small coefficients are set to zero according to sparseness constraints. Fix sparse representation coefficients, and use sparse representation coefficients and data samples to calculate the steepest descent direction and adaptive step size of dictionaries. Experimental results of hand-writing recognition show that the correct rate is 96.51%.

【Key words】image classification; dictionary training; sparse representation; hand-written character recognition; bi-variable optimization; adaptive step size

DOI: 10.3969/j.issn.1000-3428.2013.05.050

1 概述

基于样本学习的信号稀疏分解方法在图像处理方面有广泛的应用，例如图像去噪[1-2]、图像修补[3]、图像超分辨率放大[4]和图像分类[5]等。但是与传统的主成分分析法不同，近年来许多人提出的信号稀疏分解算法不再要求表示矩阵正交完备。非正交过完备的字典训练算法以其符合人眼视觉特性、鲁棒性和灵活等特点成为目前图像处理与模式识别领域的研究热点。

为了利用稀疏表示所产生的特征进行模式识别，人们对稀疏字典训练的传统模型进行了各种改进[6-7]，并提出了许多具体应用[8-9]。目前的图像分类算法主要针对提高正确

基金项目：国家自然科学基金资助项目(61101215)；中央高校基本科研业务费专项基金资助项目(CHD2012JC012)；陕西省社会发展科技攻关基金资助项目(2010K11-02-11)；陕西省教育厅科学研究计划基金资助项目(11JK0994)

作者简介：徐 健(1981－)，女，讲师、硕士，主研方向：模式识别，数字图像处理；常志国，讲师、博士；赵小强，副教授、硕士；马 祥，讲师、博士

收稿日期：2012-05-24 **修回日期：**2012-07-26 **E-mail：**xujian_paper@126.com

率和运算速度 2 个方向作努力。

为了提高分类的正确率，文献[10]将所有类别全部使用一个字典进行区分，提高了不同类别间的原子共用性。同时，将核函数引入字典训练模型，提高了算法鲁棒性。但是该模型在计算表示系数时也必须考虑约束条件，这对恢复算法提出了更高的要求。

文献[11]将对应不同类别字典间的相关性约束加入字典训练模型中。该模型通过训练加大对应不同类别字典原子间的夹角，降低它们互相的投影长度。这样可以在空间上分割不同类别的样本，达到利用稀疏表示残差对样本进行分类的目的。但是该文献没有描述步长的选取方法，并对迭代算法的收敛性缺乏分析。上述算法仅限于在固定拍摄条件的标准图像库上做实验。

文献[12]利用旋转不变特征进行稀疏字典训练，提高了对于拍摄条件不固定的自然图像分类的鲁棒性。文献[8]提出了一种利用 l_1 范数进行人脸识别的方法，该方法在正面人脸识别方面有出色表现，尤其对遮挡等退化具有较强的鲁棒性。但是没有考虑旋转、缩放等问题。

文献[13]提出了利用与类别相对应的字典识别人类行为的方法，该方法利用单帧中兴趣点的邻域空间信息和时间点上的连续分割动作作为特征，将行为的时间和空间特征相结合进行稀疏表示字典训练，在视频监控等应用领域有较好的表现。但是该算法在计算复杂度方面有待改进。

为了降低算法复杂度，文献[14]提出了基于聚类的稀疏字典训练算法，文献[15]提出了使用图形处理单元并行技术来解决较大图像分类的算法复杂度问题。

本文提出采用自适应步长的梯度下降分类字典训练算法。该算法模型将 3 个优化目标，即稀疏表示残差、稀疏度和字典间的相关系数同时进行优化。利用自适应步长保证了 3 个目标能够同时收敛，并从理论上证明了自适应步长的计算方法。

2 梯度交替最速下降分类字典训练算法

稀疏表示的重构字典训练模型通常被定义为：

$$\min_{\boldsymbol{D},\boldsymbol{A}} \|\boldsymbol{X}-\boldsymbol{D}\boldsymbol{A}\|_F^2, \|\boldsymbol{a}_i\|_0 < T \tag{1}$$

其中，$\boldsymbol{X}\in\boldsymbol{R}^{M\times N}$ 为训练样本；$\boldsymbol{D}\in\boldsymbol{R}^{M\times S}$ 为待训练的字典；$\boldsymbol{A}\in\boldsymbol{R}^{S\times N}$ 为表示系数；$\boldsymbol{a}_i\in\boldsymbol{R}^S$ 是组成 $\boldsymbol{A}$ 矩阵的列向量；$\|\cdot\|_F$ 表示矩阵的 F 范数；$\|\cdot\|_0$ 表示向量的 l_0 范数。该字典训练模型的意义是在满足 $\boldsymbol{a}_i$ 的 l_0 范数约束情况下能达到残差最小。

而用于分类的带约束字典训练的模型如式(2)[11]：

$$\min_{\boldsymbol{D}_i,\boldsymbol{A}_i} \|\boldsymbol{X}_i-\boldsymbol{D}_i\boldsymbol{A}_i\|_F^2+\eta\sum_{j\neq i}\left\|\boldsymbol{D}_j^{\mathrm{T}}\boldsymbol{D}_i\right\|, \text{s.t.}\|\boldsymbol{a}_{i,b}\|_0 < T \tag{2}$$

其中，$\boldsymbol{X}_i\in\boldsymbol{R}^{M\times N_i}$ 为第 i 类样本；$\boldsymbol{D}_i\in\boldsymbol{R}^{M\times S_i}$ 是对应第 i 类样本的字典；$\boldsymbol{A}_i\in\boldsymbol{R}^{S_i\times N_i}$ 是表示系数；$\boldsymbol{a}_{i,k}\in\boldsymbol{R}^{S_i}$ 是 $\boldsymbol{A}_i$ 的第 b 列；η 是拉格朗日参数。

该字典训练方法不但考虑了在固定稀疏度情况下的稀疏表示残差，还把字典间的相关性作为约束条件。这样可以加大对应不同类别字典原子间的夹角，降低它们互相的投影长度，使稀疏表示残差可以作为有效的分类特征。在相同的稀疏度约束情况下，样本在对应类的字典稀疏表示的残差应小于其他类的字典稀疏表示残差。这样就可以通过比较残差，判断测试样本的类别。

由于式(2)是一个双变量优化，有 2 个优化变量 $\boldsymbol{D}_i$ 和 $\boldsymbol{A}_i$，因此采用了基于梯度下降的优化方法。将式(2)表示成函数的形式。假设：

$$f(\boldsymbol{D}_i,\boldsymbol{A}_i)=\|\boldsymbol{X}_i-\boldsymbol{D}_i\boldsymbol{A}_i\|_F^2+\eta\sum_{j\neq i}\left\|\boldsymbol{D}_i^{\mathrm{T}}\boldsymbol{D}_j\right\|_F^2 \tag{3}$$

将 $f(\boldsymbol{D}_i,\boldsymbol{A}_i)$ 分别对 $\boldsymbol{D}_i$ 和 $\boldsymbol{A}_i$ 求梯度，其中，$\mathrm{tr}(\cdot)$ 表示矩阵的迹。

$$\begin{aligned}f(\boldsymbol{D}_i,\boldsymbol{A}_i)=&\mathrm{tr}(\boldsymbol{X}_i-\boldsymbol{D}_i\boldsymbol{A}_i)^{\mathrm{T}}(\boldsymbol{X}_i-\boldsymbol{D}_i\boldsymbol{A}_i)+\\&\eta\mathrm{tr}\sum_{j\neq i}(\boldsymbol{D}_i^{\mathrm{T}}\boldsymbol{D}_j)^{\mathrm{T}}(\boldsymbol{D}_i^{\mathrm{T}}\boldsymbol{D}_j)=\\&\mathrm{tr}(\boldsymbol{X}_i^{\mathrm{T}}\boldsymbol{X}_i)-2\mathrm{tr}(\boldsymbol{D}_i\boldsymbol{A}_i)^{\mathrm{T}}\boldsymbol{X}_i+\\&\mathrm{tr}(\boldsymbol{D}_i\boldsymbol{A}_i)^{\mathrm{T}}(\boldsymbol{D}_i\boldsymbol{A}_i)+\eta\mathrm{tr}\sum_{j\neq i}(\boldsymbol{D}_i^{\mathrm{T}}\boldsymbol{D}_j)^{\mathrm{T}}(\boldsymbol{D}_i^{\mathrm{T}}\boldsymbol{D}_j)\end{aligned} \tag{4}$$

$$\frac{\partial f(\boldsymbol{D}_i,\boldsymbol{A}_i)}{\partial\boldsymbol{D}_i}=-2\boldsymbol{X}_i\boldsymbol{A}_i^{\mathrm{T}}+2\boldsymbol{D}_i\boldsymbol{A}_i\boldsymbol{A}_i^{\mathrm{T}}+2\eta\sum_{i\neq j}(\boldsymbol{D}_j\boldsymbol{D}_j^{\mathrm{T}})\boldsymbol{D}_i \tag{5}$$

$$\frac{\partial f(\boldsymbol{D}_i,\boldsymbol{A}_i)}{\partial\boldsymbol{A}_i}=2\boldsymbol{D}_i^{\mathrm{T}}\boldsymbol{D}_i\boldsymbol{A}_i-2\boldsymbol{D}_i\boldsymbol{X}_i \tag{6}$$

算法步骤如下：

(1)初始化

1)L 类训练样本 $\{\boldsymbol{X}_i\}_{i=1}^L$，$\boldsymbol{X}_i$ 由第 i 类样本组成它的列。

2)初始化字典 $\{\boldsymbol{D}_i\}_{i=1}^L$，其中，$\boldsymbol{D}_i\in\boldsymbol{R}^{m_i\times n_i}$，随机选取 $\boldsymbol{X}_i$ 的某些列对 $\boldsymbol{D}_i$ 初始化。

(2)迭代直至收敛到满足停止条件(停止条件选择前一次和后一次的稀疏表示残差之差小于 ε)，$i=1,2,\cdots,L$：

1) $\boldsymbol{A}_i^{[k]}=\boldsymbol{A}_i^{[k-1]}-2\alpha_A\left(\boldsymbol{D}_i^{[k-1]\mathrm{T}}\boldsymbol{D}_i^{[k-1]}\boldsymbol{A}_i^{[k-1]}-\boldsymbol{D}_i^{[k-1]}\boldsymbol{X}_i\right)$ (7)

固定 $\boldsymbol{D}_i^{[k-1]}$ 以式(7)对 $\boldsymbol{A}_i^{[k]}$ 进行更新。其中，$\boldsymbol{A}_i^{[k]}$ 表示第 k 次迭代的第 i 类对应的稀疏表示系数；$\boldsymbol{D}_i^{[k]}$ 表示第 k 次迭代的第 i 类对应的字典。

2)按照稀疏度要求 $\|\boldsymbol{a}_{i,t}\|_0 < T$ 将 $\boldsymbol{A}_i^{[k]}$ 的每一列小系数置零。

3)根据如下公式：

$$\begin{aligned}\boldsymbol{D}_i^{[k]}=&\boldsymbol{D}_i^{[k-1]}-2\alpha_D(\boldsymbol{D}_i^{[k-1]}\boldsymbol{A}_i^{[k]}\boldsymbol{A}_i^{[k]\mathrm{T}}+\\&\eta\sum_{i\neq j}\left(\boldsymbol{D}_j^{[k-1]}\boldsymbol{D}_j^{[k-1]\mathrm{T}}\right)\boldsymbol{D}_i^{[k-1]}-\boldsymbol{X}_i\boldsymbol{A}_i^{[k]\mathrm{T}})\end{aligned} \tag{8}$$

固定 $\boldsymbol{A}_i^{[k]}$ 以对 $\boldsymbol{D}_i^{[k]}$ 进行更新。

4)将 $\boldsymbol{D}_i^{[k]}$ 的每一列归一化。

3　自适应步长计算

从实验情况来看，如果选取固定的步长 α_A 和 α_D 将会导致稀疏表示残差和不同类别对应字典间的相关性不能同时收敛。因此，本文设计了自适应步长。

命题 当固定 $\boldsymbol{D}_i^{[k-1]}$ 对 $\boldsymbol{A}_i^{[k]}$ 进行更新时，α_A 如式(9)所示可以使 $\left\|\boldsymbol{X}_i-\boldsymbol{D}_i^{[k-1]}\boldsymbol{A}_i^{[k]}\right\|_F^2$ 随 k 增加而减小。

$$\alpha_A=\frac{1}{\sigma_{\max}^2\left(\boldsymbol{D}_i^{[k-1]}\right)+\delta_A} \tag{9}$$

当固定 $\boldsymbol{A}_i^{[k]}$ 对 $\boldsymbol{D}_i^{[k]}$ 进行更新时，α_D 如式(10)所示可以使 $\sum_{j\neq i}\left\|\boldsymbol{D}_j^{[k]\mathrm{T}}\boldsymbol{D}_i^{[k]}\right\|$ 随 k 增加而减小。

$$\alpha_D=\frac{1}{1+2\sigma_{\max}^2\left(\boldsymbol{A}_i^{[k]}\right)} \tag{10}$$

证明：

(1)证明式(9)。

$$\boldsymbol{A}_i^{[k]}=\boldsymbol{A}_i^{[k-1]}-2\alpha_A\left(\boldsymbol{D}_i^{[k-1]\mathrm{T}}\boldsymbol{D}_i^{[k-1]}\boldsymbol{A}_i^{[k-1]}-\boldsymbol{D}_i^{[k-1]\mathrm{T}}\boldsymbol{X}_i\right)=\boldsymbol{A}_i^{[k-1]}-2\alpha_A\boldsymbol{D}_i^{[k-1]\mathrm{T}}\boldsymbol{D}_i^{[k-1]}\boldsymbol{A}_i^{[k-1]}+2\alpha_A\boldsymbol{D}_i^{[k-1]\mathrm{T}}\boldsymbol{X}_i \tag{11}$$

假设：

$$\boldsymbol{X}_i\approx\boldsymbol{D}_i^{[k-1]}\boldsymbol{A}_i^{*} \tag{12}$$

其中，$\boldsymbol{A}_i^{*}$ 是目标表示系数，即 $\boldsymbol{A}_i^{*}$ 满足 $\left\|\boldsymbol{X}_i-\boldsymbol{D}_i^{[k-1]}\boldsymbol{A}_i^{*}\right\|_F^2\approx 0$。因此，该算法的关键转化为怎样选取 α_A 能够得到 $\boldsymbol{A}_i^{*}$。

将式(12)代入式(11)得到：

$$\boldsymbol{A}_i^{k}\approx\boldsymbol{A}_i^{[k-1]}-2\alpha_A\boldsymbol{D}_i^{[k-1]\mathrm{T}}\boldsymbol{D}_i^{[k-1]}\boldsymbol{A}_i^{[k-1]}+2\alpha_A\boldsymbol{D}_i^{[k-1]\mathrm{T}}\boldsymbol{D}_i^{[k-1]}\boldsymbol{A}_i^{*} \tag{13}$$

给式(13)两边同时减去 $\boldsymbol{A}_i^{*}$ 可得：

$$\begin{aligned}\boldsymbol{A}_i^{k}-\boldsymbol{A}_i^{*}&\approx\boldsymbol{A}_i^{[k-1]}-2\alpha_A\boldsymbol{D}_i^{[k-1]\mathrm{T}}\boldsymbol{D}_i^{[k-1]}\boldsymbol{A}_i^{[k-1]}+2\alpha_A\boldsymbol{D}_i^{[k-1]\mathrm{T}}\boldsymbol{D}_i^{[k-1]}\boldsymbol{A}_i^{*}-\boldsymbol{A}_i^{*}=\\&\left(\boldsymbol{I}-2\alpha_A\boldsymbol{D}_i^{[k-1]\mathrm{T}}\boldsymbol{D}_i^{[k-1]}\right)\boldsymbol{A}_i^{[k-1]}-\left(\boldsymbol{I}-2\alpha_A\boldsymbol{D}_i^{[k-1]\mathrm{T}}\boldsymbol{D}_i^{[k-1]}\right)\boldsymbol{A}_i^{*}=\\&\left(\boldsymbol{I}-2\alpha_A\boldsymbol{D}_i^{[k-1]\mathrm{T}}\boldsymbol{D}_i^{[k-1]}\right)\left(\boldsymbol{A}_i^{[k-1]}-\boldsymbol{A}_i^{*}\right)\end{aligned}$$

假设 $\tilde{\boldsymbol{A}}_i^{k}=\boldsymbol{A}_i^{k}-\boldsymbol{A}_i^{*}$，所以：

$$\tilde{\boldsymbol{A}}_i^{k}=\left(I-2\alpha_A\boldsymbol{D}_i^{[k-1]\mathrm{T}}\boldsymbol{D}_i^{[k-1]}\right)\tilde{\boldsymbol{A}}_i^{k-1} \tag{14}$$

由于 $\boldsymbol{D}_i^{[k-1]\mathrm{T}}\boldsymbol{D}_i^{[k-1]}$ 是个方阵，对其作特征值分解可得：

$$\boldsymbol{D}_i^{[k-1]\mathrm{T}}\boldsymbol{D}_i^{[k-1]}=\boldsymbol{P}^{-1}\boldsymbol{\Lambda}\boldsymbol{P}=\boldsymbol{P}^{-1}\begin{bmatrix}\lambda_1&&&\\&\lambda_2&&\\&&\ddots&\\&&&\lambda_{n_i}\end{bmatrix}\boldsymbol{P} \tag{15}$$

因此：

$$\boldsymbol{I}-2\alpha_A\boldsymbol{P}^{-1}\begin{bmatrix}\lambda_1&&&\\&\lambda_2&&\\&&\ddots&\\&&&\lambda_{n_i}\end{bmatrix}\boldsymbol{P}=\boldsymbol{P}^{-1}\begin{bmatrix}1-2\alpha_A\lambda_1&&&\\&1-2\alpha_A\lambda_2&&\\&&\ddots&\\&&&1-2\alpha_A\lambda_{n_i}\end{bmatrix}\boldsymbol{P} \tag{16}$$

由于 $\lim\limits_{k\to\infty}\tilde{\boldsymbol{A}}_i^{k}=\lim\limits_{k\to\infty}\boldsymbol{A}_i^{k}-\boldsymbol{A}_i^{*}=0$，因此 $\left|1-2\alpha_A\lambda_i\right|<1$，得到 $0<\alpha_A<\dfrac{1}{\lambda_{\max}}$。又因为 $\lambda_{\max}=\sigma_{\max}^2\left(\boldsymbol{D}_i^{[k-1]}\right)$，为了防止分母为 0，在分母上加一个小的正数 δ：

$$\alpha_A=\frac{1}{\lambda_{\max}+\delta}=\frac{1}{\sigma_{\max}^2\left(\boldsymbol{D}_i^{[k-1]}\right)+\delta}$$

式(9)得证。

同理可证式(10)。

4　分类算法

对于每一个测试样本，可以执行以下算法判断其类别：

(1)将测试样本图像变成一列 $\boldsymbol{x}_{\text{test}}$，变化方式和训练样本相同。

(2)以相同的稀疏度计算测试样本在每个类别对应字典上的稀疏表示残差 $r_1,r_2,\cdots,r_{10}$。其中，r_i 表示在第 i 类字典 $\boldsymbol{D}_i$ 下的稀疏表示残差；a_i 为 $\boldsymbol{D}_i$ 对应的稀疏表示系数：

$$r_i=\left\|\boldsymbol{x}_{\text{test}}-\boldsymbol{D}_i\boldsymbol{a}_i\right\|_2^2$$

(3)所得稀疏表示残差最小的字典对应的类别 i^{*} 即为该测试样本对应的类别。

$$i^{*}=\min_i\left\|\boldsymbol{x}_{\text{test}}-\boldsymbol{D}_i\boldsymbol{a}_i\right\|_2^2,i=1,2,\cdots,10$$

5　实验与分析

实验选择了美国邮政管理)数据库，里面包括 7 291 个训练样本和 2 007 个测试样本。每个样本大小为 16×16。为了提高运算速度，把所有样本缩小为 10×10。实验采用字典列数为 500 列，稀疏度为 5。

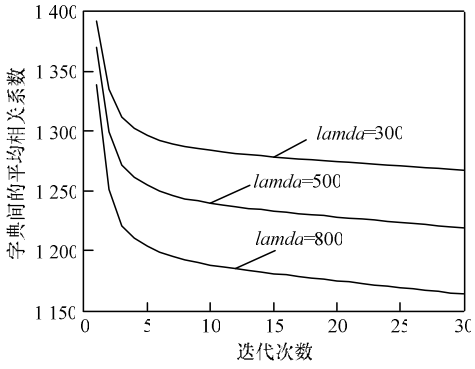
为了测试算法的收敛性，设计了参数 C 来衡量不同类别对应字典间的相关性，R 为稀疏表示残差。

$$C=\sum_{i=1}^{n}\sum_{j>i}\left\|\boldsymbol{D}_j^{\mathrm{T}}\boldsymbol{D}_i\right\|_F^2 \tag{17}$$

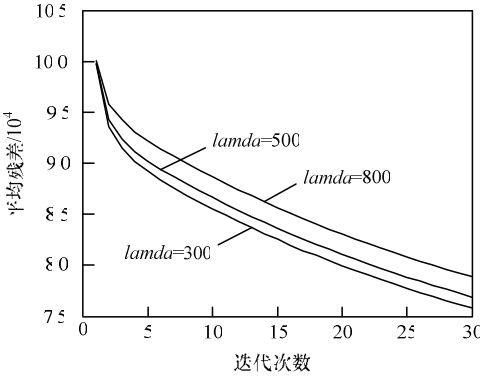
$$R=\sum_{i=1}^{n}\left\|\boldsymbol{X}_i-\boldsymbol{D}_i\boldsymbol{A}_i\right\|_F^2 \tag{18}$$

该实验需要设置的参数包括稀疏度、字典列数、拉格朗日乘子、迭代初值等。在每一个字典列数为 500 列，稀疏度为 5，不同拉格朗日乘子下字典间的相关性和残差的收

敛曲线如图 1 所示(图中 *lamda* 表示拉格朗日乘子)。由图 1 可以看出，该算法可以让稀疏表示的残差和不同类别对应字典间的相关性同时收敛。



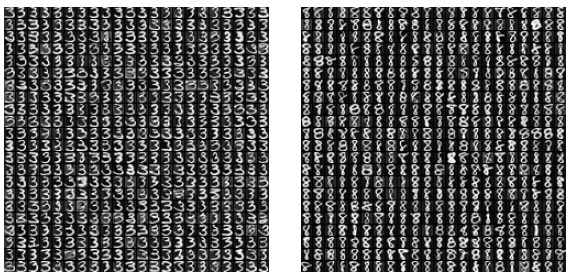
(a)相关性收敛曲线



(b)残差收敛曲线

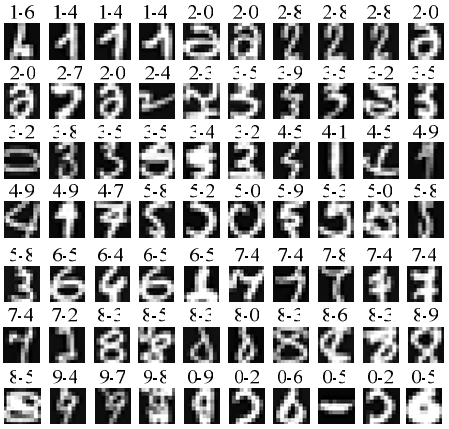
图 1 算法的收敛曲线

部分字典训练结果如图 2(a)所示。在 2007 个测试样本中，错误判决的字符总共 70 个，如图 3(a)所示，每个图片上方的标识指的是它们错判的情况，例如“1~6”表示本来是 1，错判为 6。一些错误字符写得过于不规整，在某些位置与别的字符非常相似，即使人眼也很难正确辨认。图 3(b)是固定稀疏度情况下部分错误样本在本类和错判类对应字典下稀疏表示恢复得到的图像。其中，最左边的一列是原始图像，中间列是在本类对应字典下稀疏表示恢复得到的图像，最右边的一列是错判类对应稀疏字典下稀疏表示恢复得到的图像。(例如最左边数字“1”表示字符本身为“1”，中间的数字“1”表示该字符在“1”对应字典下的稀疏表示恢复图像，右边的“6”表示该字符在“6”对应字典下的稀疏表示恢复图像。)

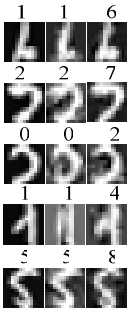


(a)“3”对应的字典 (b)“8”对应的字典

图 2 部分字典训练结果



(a)错误样本



(b)稀疏表示重构结果比较

图 3 错误判决的字符

本文方法所得到的实验结果与其他方法的对比如表 1 所示。test 是拉格朗日参数为 400、稀疏度为 5、字典列数为 500 时数据库测试样本得到的错误率；Data 是使用全体训练样本做字典时测试样本的错误率，A 是文献[10]中重构字典的错误率，B 是文献[10]中有监督字典错误率，C 是文献[11]中的错误率，支持向量机(Support Vector Machine, SVM)是使用高斯核 SVM 方法得到的错误率，K 近邻(K-Nearest Neighbor, K-NN)算法是使用 K 近邻算法得到的错误率。

表 1 手写字符识别错误率 (%)

方法	错误率
test	3.49
Data	4.14
A	4.38
B	3.56
C	3.98
SVM	4.20
K-NN	5.20

本文提出的算法之所以能提高正确率，是因为它在某种程度上强化了判决类别的特征。图 4 描述的是新算法和传统算法在其他条件完全相同的情况下，每个字符在其对应类字典下稀疏表示重构残差和在其不对应类字典下稀疏表示重构残差之间的差值(例如：字符“1”在“1”对应的字典下稀疏表示重构残差和在另外 9 个字典下稀疏表示重

构的平均残差的差的绝对值)。从图中曲线可以明显看到，新算法训练得到的字典所得残差在每个稀疏度情况下都大于传统的重构字典所得残差。根据模式识别的基本原则，所选特征在不同类别间的差别越大，分类错误率越低。因此，利用稀疏表示残差作为进行类别判别依据时，新算法具有一定优势。

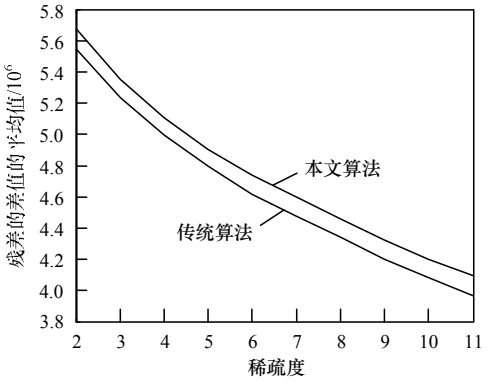


图 4　2 种算法训练字典残差的差值比较

6　结束语

本文在传统的基于重构的稀疏表示字典训练方法上加入约束，提高了以稀疏表示残差为特征的分类器的正确率。并且该分类器的训练过程运用了自适应最速下降的方法，不但克服了传统的最速下降法对于稀疏表示残差和字典的相关系数不能同时收敛的问题，而且能够适应训练样本较少的分类字典训练问题。在样本维数较高时，其算法复杂度远远低于 K 奇异值分解等传统算法。从语义分析角度来说，如何能够采用字典训练算法提取多级语义特征用于不同级别的分类，是今后需要努力的方向。

参考文献

[1] Cherian A, Sra S, Papanikolopoulos N. Denoising Sparse Noise via Online Dictionary Learning[C]//Proc. of IEEE International Conference on Acoustics, Speech, and Signal Processing. [S. l.]: IEEE Press, 2011.

[2] Elad M, Aharon M. Image Denoising via Sparse and Redundant Representations over Learned Dictionaries[J]. IEEE Transactions on Image Processing, 2006, 15(12): 3736-3745.

[3] Xu Zongben, Sun Jian. Image in Painting by Patch Propagation Using Patch Sparsity[J]. IEEE Transactions on Image Processing, 2010, 19(5): 1153-1165.

[4] Zhang Haichao, Zhang Yanning, Huang T S. Efficient Sparse Representation Based Image Super Resolution via Dual Dictionary Learning[C]//Proc. of IEEE International Conference on Multimedia and Expo. [S. l.]: IEEE Press, 2011.

[5] Mairal J, Bach F, Ponce J, et al. Discriminative Learned Dictionaries for Local Image Analysis[C]//Proc. of IEEE Conference on Computer Vision and Pattern Recognition. [S. l.]: IEEE Press, 2008.

[6] Wright J, Ma Yi, Mairal J, et al. Sparse Representation for Computer Vision and Pattern Recognition[J]. Proceedings of the IEEE, 2010, 98(6): 1031-1044.

[7] Mairal J, Bach F, Ponce J. Task-driven Dictionary Learning[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2012, 34(4): 791-804.

[8] 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.

[9] Gangeh M J, Ali G, Kamel MS. Dictionary Learning in Texture Classification[C]//Proc. of International Conference on Image Analysis and Recognition. Burnaby, Canada: [s. n.], 2011.

[10] Mairal J, Bach F, Ponce J, et al. Supervised Dictionary Learning[C]//Proc. of Annual Conference on Neural Information Processing Systems. Vancouver, Canada: [s. n.], 2008.

[11] Ramirez I, Sprechmann P, Sapiro G. Classification and Clustering via Dictionary Learning with Structured Incoherence and Shared Features[C]//Proc. of IEEE Conference on Computer Vision and Pattern Recognition. [S. l.]: IEEE Press, 2010.

[12] Yang Jianchao, Yu Kai, Gong Yihong, et al. Linear Spatial Pyramid Matching Using Sparse Coding for Image Classification[C]//Proc. of IEEE Conference on Computer Vision and Pattern Recognition. [S. l.]: IEEE Press, 2009.

[13] Wang Haoran, Yuan Chunfeng, Hu Weiming, et al. Supervised Class-specific Dictionary Learning for Sparse Modeling in Action Recognition[J]. Pattern Recognition, 2012, 45(11): 3902-3911.

[14] 徐　健, 常志国. 基于聚类的自适应图像稀疏表示算法及其应用[J]. 光子学报, 2011, 40(2): 316-320.

[15] Nagesh P, Gowda R, Li B. Fast GPU Implementation of Large Scale Dictionary and Sparse Representation Based Vision Problems[C]//Proc. of IEEE International Conference on Acoustics, Speech, and Signal Processing. [S. l.]: IEEE Press, 2010.

编辑　顾逸斐