

基于子编码和全编码联合惩罚的稀疏表示字典学习方法

董俊健 毛启容 胡素黎 詹永照

(江苏大学计算机科学与通信工程学院 镇江 212013)

摘要 针对目前稀疏表示字典学习的惩罚函数版本不一且各有优势的问题,提出基于子编码和全编码联合惩罚的稀疏表示字典学习方法,该方法在字典学习的目标函数中同时加入子编码惩罚函数和全编码惩罚函数。子编码惩罚函数使得学习后的字典在稀疏表示识别时可以用子字典的重构误差和子字典上编码系数的大小来识别,全编码惩罚函数则能直接利用整个字典上的编码系数来识别,通过联合这两个惩罚函数可以获得非常好的识别效果。为了验证所提方法的有效性,在语音情感库和人脸库上与最新的基于字典学习的稀疏表示识别方法 DKSVD 和 FDDL 进行对比,并与著名的识别方法 SVM 和 SRC 进行比较,实验结果显示所提方法具有更好的识别性能。

关键词 稀疏表示识别,结构化学字典学习,惩罚函数,稀疏编码,语音情感识别,人脸识别

中图分类号 TP391 **文献标识码** A **DOI** 10.11896/j.issn.1002-137X.2014.10.028

Sub-coding and Entire-coding Jointly Penalty Based Sparse Representation Dictionary Learning

DONG Jun-jian MAO Qi-rong HU Su-li ZHAN Yong-zhao

(School of Computer Science and Communication Engineering, Jiangsu University, Zhenjiang 212013, China)

Abstract Currently, the penalty function of dictionary learning (DL) used for sparse representation classification has many versions and each of them has its own advantages. This paper presented a new dictionary learning method called Sub-coding and Entire-coding jointly penalty based dictionary learning, which jointly adds sub-coding penalty functions and entire-coding penalty functions into the dictionary learning objective function. Sub-coding penalty function makes the dictionary after learning can use its reconstruction error and sub-coding for classification, and entire-coding penalty function makes the dictionary after learning can directly use its whole coding for classification at the same time. By combining these two penalty function, good recognition effect can be got. The proposed method is extensively evaluated on emotion speech database and face database in comparison with famous DL based sparse representation classification methods DKSVD and FDDL, and other famous recognition method SRC and SVM. The experimental results show that the proposed method has better recognition performance.

Keywords Sparse representation based classification, Structured dictionary learning, Penalty function, Sparse coding, Emotion speech recognition, Face recognition

1 引言

近年来,稀疏表示(编码)的理论和算法迅速发展,已经成功应用于图像重构^[1-3]和压缩感知^[4]等领域。最新的稀疏表示技术又在图像识别中取得很好的效果,例如人脸识别^[5-11,17]、数字和文本识别^[12,13]等。Wright等^[6]提出的基于稀疏表示分类(Sparse Representation based Classification,简称SRC)方法在人脸识别中获得了较高的识别率,该方法直接使用人脸训练样本组成一个结构化的字典来对待识别的人脸图片进行稀疏编码,然后根据得到的稀疏编码计算各个子字典(对应着各个不同类别)的重构误差,最后将待识别的人脸图片识别为重构误差最小的子字典类。SRC方法在人脸

识别中显示出较好的识别性能,但简单地将训练样本作为字典,当训练样本很多时,在一个很大的字典上进行稀疏编码是非常耗时和不现实的;而且识别结果的好坏极大地依赖于所选择的训练样本,只有好的训练样本(基函数,原子)才能获得较好的识别性能。因此基于稀疏表示识别的方法需要通过字典学习来自动地获得一个更好且更精简的字典。

字典学习(Dictionary learning, DL)最初旨在对训练样本进行学习从而使待处理的信号能更好地被字典重构。许多字典学习方法已经用于图像处理^[1,3,14]和识别^[7-9,11-13]。Aharon等^[14]提出的K-Means Singular Value Decomposition(简称K-SVD)算法是一个具有代表性的字典学习算法,该算法从训练样本集学习得到一个过完备的字典,然而用K-SVD学习得到

到稿日期:2013-06-30 返修日期:2013-08-09 本文受国家自然科学基金面上项目(61272211,61170126),江苏省自然科学基金面上项目(BK2011521),高级人才启动基金项目(10JDG065)资助。

董俊健(1989-),男,硕士生,主要研究方向为语音情感识别、稀疏表示识别,E-mail:974600685@qq.com;毛启容(1975-),女,博士,副教授,主要研究方向为多媒体信息处理、人机交互技术;胡素黎(1988-),女,硕士生,主要研究方向为语音情感识别、多分类器融合;詹永照(1962-),男,博士,教授,博士生导师,主要研究方向为人机交互、分布式计算。

的字典并不适合用来识别,因为该算法仅要求学习得到的字典能较好地稀疏表示训练样本。Mairal 等^[13]直接对稀疏编码系数进行学习得到一种分类器,提出了一种具有识别力的字典学习方法,并在数字识别和文本分类上验证了该方法的有效性。Zhang 等^[8]在 K-SVD 字典学习的目标函数中加入与类别标签信息相关的线性分类器惩罚函数,提出 Discriminative K-SVD(简称 D-KSVD)算法,该算法能得到一个具有较强识别力的字典,但 D-KSVD 学习得到的字典是一个被所有类别共享的字典,即学习后字典中的原子没有与之对应的类别标签信息,这样就不能用 SRC 的重构误差识别准则来识别,失去了重构误差识别准则的优势。Yang 等^[9]单独为每一个类别的训练样本学习得到一个子字典,然后将所有子字典组合成一个结构化的字典,在实验中获得了比 SRC 更好的识别性能。Ramirez 等^[12]在字典学习的目标函数中加入不同类别原子的相干性惩罚函数,使得不同类别的字典原子尽可能地不相干。后来 Yang 等^[11]提出了 Fisher Discrimination Dictionary Learning(简称 FDDL),该方法引入 Fisher 识别准则来学习得到一个结构化的字典,FDDL 同时加入了有识别力的保真度(Discriminative fidelity term)惩罚函数和有识别力的编码系数(Discriminative coefficient term)惩罚函数,并通过权衡两个识别准则(对应于两个惩罚函数)获得了较高的识别精度。受这一思想的启发,提出基于子编码和全编码联合惩罚的字典学习方法,该字典学习模型联合两个具有识别力的惩罚函数,一个是子编码差惩罚函数,另一个是全编码惩罚函数。子编码惩罚函数使得学习后的字典在稀疏表示识别时可以用子字典的重构误差和子字典上编码系数的大小来识别,全编码惩罚函数则能直接利用整个字典上的编码系数来识别,通过联合这两个惩罚函数可以获得非常好的识别效果,而普通的稀疏表示识别方法只能用子字典的重构误差来识别,FDDL 方法学习后的字典在稀疏表示识别时可以同时使用子字典的重构误差和每个类别的编码系数均值来识别。

2 相关工作

Wright 等^[5]提出基于稀疏表示分类(SRC)的方法。假设有 c 个类别, $A=[A_1, A_2, \dots, A_c]$ 表示所有的训练样本集合,其中 A_i 表示类别为 i 的训练样本集合, y 是一个测试样本,那么 SRC 方法过程如下:

a)首先在字典 A 上对测试样本 y 进行稀疏编码, α 为编码系数,即 $y \approx A\alpha$,具体为解式(1)所示的最优化问题:

$$\hat{\alpha} = \arg \min_{\alpha} \{ \|y - A\alpha\|_2 + \lambda \|\alpha\|_1 \} \quad (1)$$

其中, λ 是标量参数,用于权衡编码的稀疏度。

b)然后根据式(2)所示的识别准则进行识别:

$$\text{identity}(y) = \arg \min_i \{ \|y - A_i \alpha_i\|_2 \} \quad (2)$$

其中, $\alpha = [\alpha_1; \alpha_2; \dots; \alpha_c]$, α_i 是 y 在子字典 A_i 上的编码系数。SRC 的识别准则即测试样本 y 的类别为重构 y 误差最小的子字典对应的类别。

SRC 方法在人脸识别等领域获得了较好的效果,但 SRC 简单地将训练样本组成一个结构化的字典,这种方法得到的字典的识别性能不是最佳的,识别结果的好坏取决于所选择的训练样本(基函数,原子);同时当训练样本很多时,在一个

很大的字典上进行稀疏编码是非常耗时和不现实的,因此需要通过字典学习来弥补这两个缺陷。基于字典学习的稀疏表示识别方法在字典学习的目标函数中加入有识别力的惩罚函数来使得学习后的字典间接具有较强的识别力,目前惩罚函数版本不一且各有优势,究竟什么样的惩罚函数最佳还没有定论。

3 基于子编码和全编码联合惩罚的字典学习识别框架

基于子编码和全编码联合惩罚的字典学习的识别框架包括两个过程:字典学习过程和稀疏表示识别过程。如图 1 所示,图的上半部分是字典学习过程,下半部分是稀疏表示识别过程。其中训练样本特征矩阵 A 的每一列为某个训练样本的特征向量。首先对训练样本特征矩阵 A 用基于子编码和全编码联合惩罚的字典学习算法进行学习,得到字典 \hat{D} ,然后将测试样本 y 在学习得到的字典 \hat{D} 上进行稀疏编码,并根据相应的字典学习识别准则进行识别。在 SRC 方法的过程中则没有字典学习这一过程。

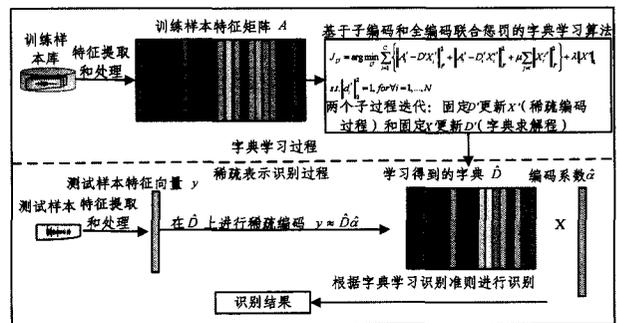


图1 基于子编码和全编码联合惩罚的字典学习识别框架

4 基于子编码和全编码联合惩罚的字典学习模型

假设有 c 个类别, $A=[A_1, A_2, \dots, A_c]$ 表示所有的训练样本集合,其中 A_i 表示类别为 i 的训练样本的集合。字典 $D=[D_1, D_2, \dots, D_c]=[d_1, d_2, \dots, d_N]$,其中 D_i 表示与类别 i 相关的子字典, d_i 表示字典的第 i 个原子,一共有 N 个原子。 X 表示训练样本 A 在字典 D 上的编码系数,也就是 $A \approx DX$,所以训练样本 A 的编码系数 X 可以表示为 $X=[X_1, X_2, \dots, X_c]$,其中 X_i 表示 A_i 在字典 D 上的编码系数。这里的字典学习不仅要求字典 D 能很好地重构训练样本 A ,还要求字典 D 对于 A 中的训练样本有较强的识别力。为了同时满足这两个要求,提出式(3)所示的字典学习模型:

$$J_{(D, W)} = \arg \min_{D, W} \{ \|A - DX\|_2 + r(A, H, D, W, X) \} \quad (3)$$

$$\text{s. t. } \|d_i\|_2 = 1, \text{ for } \forall i = 1, \dots, N$$

其中, $r(A, H, D, W, X) = r_1(A, D, X) + r_2(H, W, X)$, H 是训练样本 A 对应的类别标签矩阵, W 是线性分类器的参数矩阵。该字典学习模型包括两个部分:第一个部分 $\|A - DX\|_2$ 要求字典 D 能较好地重构训练样本,是重构精度要求,是字典学习的基本要求;第二个部分 $r(A, H, D, W, X)$ 则要求字典 D 和线性分类器参数 W 具有较好的识别性能,是识别精度要求。下面将详细讨论识别精度惩罚函数 $r(A, H, D, W, X)$ 。

4.1 子编码惩罚函数 $r_1(A, D, X)$

$$r_1(A, D, X) = \sum_{i=1}^C \{ \|A_i - D_i X_i\|_2^2 + \mu \sum_{j \neq i} \|X_j\|_2^2 \} + \lambda \|X\|_1 \quad (4)$$

其中, μ 和 λ 是权衡各个惩罚项的标量参数。引入该惩罚函数是由于 SRC 方法的重构误差识别准则取得了较好的识别效果。

子训练样本 A_i 在字典 D 上的编码系数 X_i 可以表示成 $X_i = [X_i^1; X_i^2; \dots; X_i^j]$, 其中 X_i^j 表示子训练样本 A_i 在子字典 D_j 上的编码系数, 因此子训练样本 A_i 可以表示成 $A_i \approx DX_i = D_1 X_i^1 + D_2 X_i^2 + \dots + D_j X_i^j$, 这样就可以考察各个子字典 D_i (对应着各个类别) 的重构误差来间接使字典 D 具有较高的识别精度。SRC 方法的重构误差识别准则希望子训练样本 A_i 仅能被子字典 D_i 很好地表示, 而不是 $D_j, j \neq i$ 。因此该惩罚函数的第一个惩罚项 $\|A_i - D_i X_i^i\|_2^2$ 就要求子训练样本 A_i 可以被子字典 D_i 很好地表示, 第二个惩罚项 $\mu \sum_{j \neq i} \|X_j\|_2^2$ 要求子训练样本 A_i 在其他子字典 $D_j (j \neq i)$ 上的系数非常小, 那么子字典 $D_j (j \neq i)$ 就不能很好地重构 A_i , 第三个惩罚项 $\lambda \|X\|_1$ 是编码系数稀疏度惩罚项。

子编码惩罚函数的第二个惩罚项 $\mu \sum_{j \neq i} \|X_j\|_2^2$ 和 FDDL 的有识别力的保真度惩罚函数中的惩罚项 $\mu \sum_{j \neq i} \|D_j X_j^i\|_2^2$ 相比是有优势的, 式(4)所示的惩罚函数不仅对子字典的重构误差有约束, 还对字典上的编码系数大小有约束, 因此字典学习后不仅可以用每个子字典的重构误差来识别, 还可以用每个子字典上编码系数的大小来识别, 而 FDDL 则只能用每个子字典的重构误差来识别。 $\mu \sum_{j \neq i} \|X_j\|_2^2$ 比 $\mu \sum_{j \neq i} \|D_j X_j^i\|_2^2$ 更强大且更符合基于稀疏表示识别的要求, 即大的编码系数几乎都集中在自己所属类别的子字典上。称 $r_1(A, D, X)$ 为子编码惩罚函数是因为该函数考察的都是子字典的编码系数。

对于子编码惩罚函数的直观解释如图 2 所示, 子字典 D_i 的重构精度记为 R_i , 则子训练样本 A_i 可以表示成 $A_i \approx DX_i = D_1 X_i^1 + D_2 X_i^2 + \dots + D_j X_i^j = R_1 + R_2 + \dots + R_j$ 。图 2(a) 是加入第一个惩罚项 $\|A_i - D_i X_i^i\|_2^2$ 之后 A_i 可能的重构情况, 发现 R_i 和 R_j 对 A_i 都有很好的重构精度, 这是不利于子字典的重构精度(重构误差)来识别的。因此加入第二个惩罚项 $\mu \sum_{j \neq i} \|X_j\|_2^2$, 如图 2(b) 所示, 加入第二个惩罚项后, 只可能 R_i 对 A_i 有很好的重构精度, 因为其它子字典的重构精度 $R_j (j \neq i)$ 都非常小, 而且只有子字典 D_i 上的编码系数较大, 这样就可以轻松将属于 A_i 的样本的类别判为 i 。



(a) 加入第一个惩罚项后 A_i 的重构情况 (b) 加入第二个惩罚项后 A_i 的重构情况

图 2 子编码惩罚函数的直观解释

4.2 全编码惩罚函数 $r_2(H, W, X)$

$$r_2(H, W, X) = \|H - WX\|_2^2 + \beta \|W\|_2^2 \quad (5)$$

其中, W 是线性分类器 $H = WX$ 的参数矩阵, $\|W\|_2^2$ 是正则化惩罚项, H 是训练样本 A 对应的类别标签矩阵, X 是 A 的

编码系数, 因此 X 也具有类别标签矩阵 H 。 H 的每一列都是一个这样的向量 $h_i = [0, 0, \dots, 1, \dots, 0, 0]^T$, 其中非零元素 1 的位置标志着类别信息。该惩罚函数直接将编码系数 X 和对应的类别标签矩阵 H 建立一个线性分类器, 通过学习后该分类器直接对编码系数具有较好的识别力, 这样就能用学习得到的线性分类器来进行识别了。引入该惩罚函数是因为 DKSV D 方法学习得到的线性分类器获得了很好的识别效果。称 $r_2(H, W, X)$ 是全编码惩罚函数是因为该函数考察的是整个字典上的编码系数。

4.3 基于子编码和全编码联合惩罚的字典学习模型

联合有识别力的惩罚函数式(4)和式(5), 得到式(6)所示的字典学习模型:

$$J_{(D, W)} = \arg \min_{D, W} \sum_{i=1}^C \{ \|A_i - DX_i\|_2^2 + \|A_i - D_i X_i^i\|_2^2 + \mu \sum_{j \neq i} \|X_j\|_2^2 + \eta \|H_i - WX_i\|_2^2 + \beta \|W\|_2^2 \} + \lambda \|X\|_1 \quad (6)$$

s. t. $\|d_i\|_2 = 1, \text{ for } \forall i = 1, \dots, N$

式(6) $\|A_i - DX_i\|_2^2$ 为字典学习的重构进度要求, $\|A_i - D_i X_i^i\|_2^2 + \mu \sum_{j \neq i} \|X_j\|_2^2$ 为子编码惩罚函数要求, $\eta \|H_i - WX_i\|_2^2 + \beta \|W\|_2^2$ 为全编码惩罚函数要求。其中 $\mu, \lambda, \eta, \beta$ 是权衡各个惩罚项的标量参数, 样本矩阵 A 和对应的类别标签矩阵 H 是给定的常量矩阵, 字典矩阵 D 和编码系数矩阵 X 在字典学习过程中是相互约束和不断更新的, 线性分类器参数矩阵 W 和编码系数矩阵 X 在字典学习中是相互约束和不断更新的。

5 基于子编码和全编码联合惩罚的字典学习算法

5.1 字典学习模型的优化

令 $A' = \begin{pmatrix} A \\ \sqrt{\eta} H \end{pmatrix}, D' = \begin{pmatrix} D \\ \sqrt{\eta} W \end{pmatrix}$, 原字典学习模型式(6)变为求解式(7)所示的最优化问题:

$$J_{D'} = \arg \min_{D'} \sum_{i=1}^C \{ \|A_i' - D' X_i^i\|_2^2 + \|A_i' - D_i' X_i^i\|_2^2 + \mu \sum_{j \neq i} \|X_j^i\|_2^2 \} + \lambda \|X'\|_1 \quad (7)$$

s. t. $\|d_i'\|_2 = 1, \text{ for } \forall i = 1, \dots, N$

其中, μ 和 λ 是权衡各个惩罚项的标量参数。由于对 D' 的每一列都做归一化操作, 因此惩罚项 $\|W\|_2^2$ 可以省略。通过上述换元方法, 训练样本矩阵 A 转化为带标签信息的样本矩阵 A' , 将字典 D 转化为带线性分类器参数的字典 D' , 将 3 个变量 D, W, X 转变为两个变量 D' 和 X' , 大大降低了字典的求解速度。

5.2 优化后的字典学习模型的迭代解法

同 K-SVD、FDDL 等著名字典学习算法一样, 式(7)的求解也可以看成是两个子过程的迭代: 固定 D' 更新 X' (稀疏编码过程) 和固定 X' 更新 D' (字典求解过程)。具体见表 1, 每次迭代求解式(8)和式(9)所示的问题, 直至目标函数式(7)收敛或达到指定迭代次数。从字典学习的目标函数分析可以看出, 该方法比 DKSV D 的时间复杂度要稍大一点, 因为 DKSV D 学习得到的字典要更小(不是结构化的字典); 但比 FDDL 的时间复杂度要稍小一些, 因为这里字典学习的目标函数比 FDDL 的目标函数更简洁。

表1 基于子编码和全编码联合惩罚的字典学习算法

基于子编码和全编码联合惩罚的字典学习算法	
1.	初始化 A' 和 D' , 将 A' 和 D' 的每一列都归一化;
2.	固定 D' 更新 X' , 具体为求解式(8)所示的问题;
3.	固定 X' 更新 D' , 具体为求解式(9)所示的问题;
4.	重复步骤 2,3 直到式(7)所示的目标函数收敛或达到指定迭代次数, 输出 D' .

首先, 固定字典 D' 更新编码系数 X' , 更新 $X' = [X_1', X_2', \dots, X_c']$ 时是逐个类别进行更新的, 即计算 X_i' 时, $X_j' (j \neq i)$ 是不变的。这时式(7)简化为:

$$J_{X_i'} = \arg \min_{X_i'} \{ \|A_i' - D'X_i'\|_2^2 + \|A_i' - D_i'X_i'\|_2^2 + \mu \sum_{j \neq i} \|X_j'\|_2^2 + \lambda \|X_i'\|_1 \} \quad (8)$$

式(8)是一个结构化的稀疏正则化问题, 可以用迭代的投影方法 (Iterative Projection Methods for Structured Sparsity Regularization, 简称 IPM)^[15] 来解决 (具体见 IPM^[15] 中的算法 2)。

然后, 固定刚更新的编码系数 X' 来更新字典 D' , 更新 $D' = [D_1', D_2', \dots, D_c']$ 时也是逐个类别 (逐个字典) 来更新的, 即计算 D_i' 时, $D_j' (j \neq i)$ 是不变的。这时式(7)简化为:

$$J_{D_i'} = \arg \min_{D_i'} \{ \|A' - \sum_{j=1, j \neq i}^c D_j' X_j' - D_i' X_i'\|_2^2 + \|A_i' - D_i' X_i'\|_2^2 \} \quad (9)$$

$$\text{s. t. } \|d_i'\|_2 = 1, \text{ for } \forall i=1, \dots, N$$

其中, X_i' 是样本矩阵 A' 在字典 D_i' 上的编码系数。更新 $D_i' = [d_1', d_2', \dots, d_N']$ 时是逐个原子进行更新的, 即计算 d_i' 时, $d_j' (j \neq i)$ 是不变的。因此式(9)变成了一个带约束的二次规划问题, 可以用拉格朗日乘子法求解 d_i' , 并最终更新字典 D' 。

6 字典学习后的识别准则

通过基于子编码和全编码联合惩罚的字典学习算法对训练样本进行学习得到 D' , D' 由学习得到的字典 D 和线性分类器参数 W 组成, 即

$$D' = \begin{pmatrix} D \\ \sqrt{\gamma}W \end{pmatrix} = \begin{pmatrix} d_1 & d_2 & \dots & d_N \\ w_1 & w_2 & \dots & w_N \end{pmatrix} \quad (10)$$

由于 D' 是列归一化的, 因此学习得到的字典 D 不是列归一化的。为了得到能够用于稀疏编码的字典 \hat{D} 和线性分类器参数 \hat{W} , 对 D' 做如下处理:

$$\hat{D} = \left(\frac{d_1}{\|d_1\|}, \frac{d_2}{\|d_2\|}, \dots, \frac{d_N}{\|d_N\|} \right) \quad (11)$$

$$\hat{W} = \left(\frac{w_1}{\|d_1\|}, \frac{w_2}{\|d_2\|}, \dots, \frac{w_N}{\|d_N\|} \right) \quad (12)$$

在字典 \hat{D} 上就可以对测试样本 y 进行稀疏编码, 然后通过得到的稀疏编码来识别。基于子编码和全编码联合惩罚的字典学习方法学习后能同时得到两个具有较高识别力的识别准则, 即子编码识别准则和全编码识别准则 (对应着字典学习中两个有识别力的惩罚函数)。其中的子编码识别准则不但可以用每个子字典的重构误差来识别, 还可以用每个子字典上的编码系数大小来识别, 全编码识别准则根据整个字典上的编码系数来识别。因此得到式(13)所示的识别准则:

$$\text{identity}(y) = \arg \min_i \{ \|y - \hat{D}_i \hat{\alpha}_i\|_2 - \theta \|\hat{\alpha}_i\|_2 + \omega \|h_i - \hat{W} \hat{\alpha}_i\|_2 \} \quad (13)$$

其中, θ 和 ω 是权衡各个项的标量参数, $\omega \in [0, +\infty)$ 越大表

示全编码识别准则越准确, $\omega=0$ 表示只采用子编码识别准则。当 $\omega=0, \theta=0$ 时就回到了只用字典的重构误差来识别, 这就成了 SRC 的识别准则。

7 基于子编码和全编码联合惩罚的字典学习方法用于语音情感识别和人脸识别

7.1 实验数据介绍

实验数据来源于两个情感语料库和一个人脸库: 课题组自制的情感语料库 (简称 UJS 情感语料库)、CASIA 汉语情感语料库^[18] 和 AR 人脸库。

UJS 情感语料库共包括 6 种典型的情感类别: 高兴、伤心、惊奇、生气、害怕、厌恶。采用 CoolEdit 作为音频录制编辑软件, 共有 4 个男生 5 个女生参与录制, 每种情感每人录制 13 个语料, 音频采样率为 11025Hz, 量化精度为 16bit, 样本都经过手工标注, 以 wav 格式存储, 共有 $6 \times (4+5) \times 13 = 702$ 个语料样本。每次实验从每个表演者每类情感中随机选择 8 句作为训练样本, 剩余的 2 句作为测试样本, 因此, 共得到 $6 \times (4+5) \times 8 = 432$ 个训练样本和 $6 \times (4+5) \times 2 = 108$ 个测试样本。实验中提取的特征为能量相关、基因频率相关和共振峰系数等 101 特征^[16], 并采用 SFFS 方法进行特征选择, 最终确定特征向量的维度为 101, 最终得到 101×432 的训练样本特征矩阵 A 和 101×108 的测试样本特征矩阵 Y 。

CASIA 汉语情感语料库考虑 6 种情感, 分别是高兴、悲哀、生气、惊吓、难过、中性。共包括 4 个专业发音人, 每种情感有 50 句语料, 即对相同的文本赋以不同的情感来阅读, 这些语料可以用来对比分析不同情感状态下的声学及韵律表现。在纯净录音环境进行数据采集, 信噪比约为 35db。以 16000 采样率, 16bit, wav 格式存储。实验中选择所有人的所有语料来测试, 每次实验都随机选取其中 40 个语料作为训练样本, 剩余的 10 个作为测试样本, 得到 $6 \times 4 \times 40 = 960$ 个训练样本和 $6 \times 4 \times 10 = 240$ 个测试样本。实验中提取的特征为能量相关、基因频率相关和共振峰系数等 101 维特征^[16]。最终得到 101×960 的训练样本特征矩阵 A 和 101×240 的测试样本特征矩阵 Y 。

AR 人脸库由 Martinez 等在美国计算机视觉中心创建, 共包括 126 人 (70 名男性和 56 名女性), 总计超过 4000 幅彩色图像, 其中包含了面部表情和光照的变化, 图像预处理分辨率为 50×40 像素。与 Yang 等^[11] 中所述一致, 实验中选择一个包含 50 男和 50 女的子库, 每个人选取 14 张图像, 其中的 7 张图像作为训练样本, 剩下的 7 张图像作为测试样本, 每幅图像的特征维度为 300 维。最终得到 300×700 的训练样本特征矩阵 A 和 300×700 的测试样本特征矩阵 Y 。

7.2 实验结果与分析

7.2.1 算法收敛性讨论与分析

为了考察基于子编码和全编码联合惩罚的字典学习算法的收敛性, 图 3 给出了字典学习过程中各个惩罚函数的收敛情况, 图 4 显示了每一次迭代后在当前字典上进行稀疏表示识别的识别率。从图 3 可以清楚地看到, 在每一次迭代后字典学习的目标函数和各惩罚函数都不断下降, 直至收敛; 同时如图 4 所示, 稀疏表示识别的识别率随着字典学习迭代次数的增加呈现上升趋势, 并在字典学习的目标函数收敛时获得较高的识别率。图 4(a) 显示仅使用子编码识别准则的最终识别率为 65.74%, 图 4(b) 显示仅使用全编码识别准则的最终识别率为 60.19%, 而图 4(c) 显示使用式(10)给出的识别

准则的最终识别率为67.59%，清晰地说明了综合字典学习的各个识别准则来识别的优势。为了表述清晰，给下面的惩罚函数加上公式标号，具体如式(14)~式(17)所示。

$$\sum_{i=1}^C \{ \| A_i' - D' X_i' \|_F^2 \} \quad (14)$$

$$\sum_{i=1}^C \{ \| A_i' - D_i' X_i' \|_F^2 \} \quad (15)$$

$$\sum_{i=1}^C \{ \mu \sum_{j \neq i} \| X_j' \|_F^2 \} \quad (16)$$

$$\lambda \| X' \|_1 \quad (17)$$

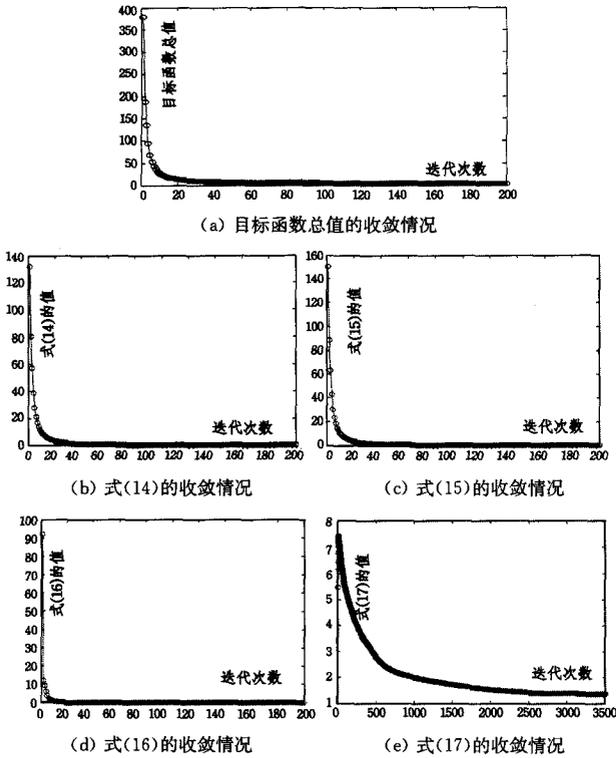


图3 字典学习过程中各个惩罚项的收敛情况

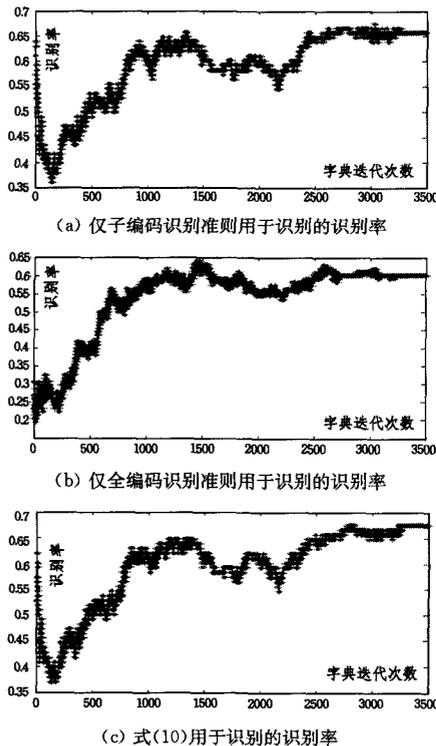


图4 各种识别准则在字典学习时的识别率

7.2.2 识别率讨论与分析

为了清楚地说明基于子编码和全编码联合惩罚的字典学习方法 (Sub-coding and Entire-coding Jointly Penalty based Dictionary Learning, 简称 ScEcDL) 在稀疏表示识别中突出的识别性能, 将该方法和两个最新的基于字典学习的稀疏表示识别方法 FDDL 和 DKSVD 进行比较, 并与两个著名的识别方法 SVM 和 SRC 进行比较。为了公平地与 SVM 和 SRC 进行比较, 将所有基于字典学习的稀疏表示识别方法所用字典的大小取为训练样本特征矩阵 A 的大小。

字典学习过程中需要用到 3 个标量参数 λ, η, μ , 识别过程中需要用到两个标量参数 θ, ω , 这些标量参数都是为了权衡各个项的大小, 均采用 5 倍交叉验证方法来获得。其中稀疏度标量参数 λ 用来权衡编码的稀疏度, 是稀疏表示识别最重要的参数; η 用来权衡两种惩罚函数的有效性, η 越大则表示全编码惩罚函数越有效; ω 用来权衡两种识别准则, ω 越大则表示全编码识别准则越准确。参数的具体选取方法为先根据参数的意义估计一些初值, 再用 5 倍交叉验证方法来选取最佳的参数值。

实验过程中, 在 UJS 情感语料库上选取的参数为 $\lambda=0.003, \eta=0.25, \mu=100, \theta=1, \omega=1$, 如表 2 所列, 实验结果显示 ScEcDL 比其他字典学习方法高 2%~6%, 比 SVM 高 4%, 比 SRC 高 1%。在 CASIA 汉语情感语料库上选取的参数为 $\lambda=0.01, \eta=0.25, \mu=100, \theta=1, \omega=1$, 从表 3 可以看出, 与其它方法相比, 在 CASIA 汉语情感语料库上 ScEcDL 的识别率依然最高, 比其它字典学习方法高 1%~8%, 比 SVM 高 9%, 比 SRC 高 2%。在 AR 人脸库上选取的参数为 $\lambda=0.003, \eta=0.25, \mu=100, \theta=1, \omega=1$, 从表 4 可以看出, ScEcDL 的识别率依然最高, 比其它字典学习方法高 1%~8%, 比 SVM 高 6%, 比 SRC 高 4%。

表2 UJS 语音情感库上各方法识别率比较

方法	SVM	SRC	DKSVD	FDDL	ScEcDL
识别率(%)	63.89	66.67	61.11	65.74	67.59

表3 CASIA 汉语情感语料库上各方法识别率比较

方法	SVM	SRC	DKSVD	FDDL	ScEcDL
识别率(%)	50.42	57.5	51.25	58.75	59.58

表4 AR 人脸库上各方法识别率比较

方法	SVM	SRC	DKSVD	FDDL	ScEcDL
识别率(%)	87.1	88.8	85.4	92.0	93.42

结束语 为了在稀疏表示识别中自动地获得一个简洁的并具有较强识别力的字典, 提出基于子编码和全编码联合惩罚的字典学习方法。通过在字典学习的目标函数中同时加入有识别力的子编码惩罚函数和全编码惩罚函数, 使得学习后的字典间接具有较强的识别力。子编码惩罚函数使得学习后的字典在稀疏表示识别时可以同时使用子字典的重构误差和子字典上编码系数的大小来识别; 全编码惩罚函数则能直接利用整个字典上的编码系数来识别, 通过权衡两个识别准则 (对应着字典学习中两个有识别力的惩罚函数) 可以获得很好的识别效果。接着给出了该字典学习模型的求解算法, 然后联合子编码识别准则和全编码识别准则来设计稀疏表示识别时的识别准则。该字典学习方法的优点是可同时使用子字典的重构误差、子字典的编码系数大小和整个字典的编码系数

来识别,三者都具有较强的识别性能,权衡好三者的优劣能得到一个很好的识别效果。最后在两个情感语料库和一个人脸库上的实验结果清楚地显示了 ScEcDL 方法优于一些最新的基于字典学习的稀疏表示识别方法。本文下一步工作是对比其他可能的有识别力的惩罚函数,从而求得最佳的字典来提高识别精度。

参 考 文 献

[1] Elad M, Aharon M. Image denoising via sparse and redundant representations over learned dictionaries[J]. IEEE TIP, 2006, 15(12):3736-3745

[2] Mairal J, Elad M, Sapiro G. Sparse representation for color image restoration[J]. IEEE TIP, 2008, 17(1):53-69

[3] Bryt O, Elad M. Compression of facial images using the K-SVD algorithm[J]. Journal of Visual Communication and Image Representation, 2008, 19(4):270-282

[4] Candes E. Compressive sampling[J]. Int. Congress of Mathematics, 2006, 3: 1433-1452

[5] Wright J, Yang A Y, Ganesh A, et al. Robust Face Recognition via Sparse Representation[J]. IEEE TPAMI, 2009, 31(2): 210-227

[6] Wagner A, Wright J, Ganesh A, et al. Towards a Practical Face Recognition System: Robust Registration and Illumination by Sparse Representation[C]//CVPR. 2009

[7] Yang M, Zhang L. Gabor Feature based Sparse Representation for Face Recognition with Gabor Occlusion Dictionary[C]//EC-CV. 2010

[8] Zhang Q, Li B X. Discriminative K-SVD for dictionary learning in face recognition[C]//CVPR. 2010

[9] Yang M, Zhang L, Yang J, et al. Metaface learning for sparse representation based face recognition[C]//ICIP. 2010

[10] Yang M, Zhang L, Yang J, et al. Robust sparse coding for face recognition[C]//CVPR. 2011

[11] Yang M, Zhang L, Feng X, et al. Fisher discrimination dictionary learning for sparse representation[C]//ICCV. 2011

[12] Ramirez I, Sprechmann P, Sapiro G. Classification and clustering via dictionary learning with structured incoherence and shared features[C]//CVPR. 2010

[13] Mairal J, Bach F, Ponce J, et al. Supervised dictionary learning [C]//NIPS. 2009

[14] Aharon M, Elad M, Bruckstein A. K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation[J]. IEEE TSP, 2006, 54(11):4311-4322

[15] Rosasco L, Verri A, Santoro M, et al. Iterative Projection Methods for Structured Sparsity Regularization[R]. MIT Technical Reports, MIT-CSAIL-TR-2009-050, CBCL-282. 2009

[16] Mao Qi-rong, Wang Xiao-Jia, Zhan Yong-zhao. Speech Emotion Recognition Method Based on Improved Decision Tree and layered feature selection [J]. International Journal of Humanoid Robotics, 2010, 7(2):245-261

[17] Yang M, et al. Gabor feature based robust representation and classification for face recognition with Gabor occlusion dictionary[J]. Pattern Recognition, 2012, 46(7):1865-1878

[18] 中科院自动化研究所人机语音交互课题组. CASIA 汉语情感语料库[OL]. Http://www.datatang.com/data/39277

(上接第 105 页)

围。本文方法的出发点是:对一维信号,从局部来看,主体分量是单调的,而细节分量不是单调的。为了分段进行单调逼近并把局部平滑的结果拼接成全局的平滑结果,本文提出了极值约束。大体上说,基于极值约束的平滑算法首先找出一组点 E ,然后在所有仅仅在这些点处取得极值的信号中选出和原信号最接近者作为平滑结果。为了得到极值点集 E ,可以先对原信号用其他方法平滑,得到一个初步的平滑结果 G ,再把 G 的极值点集作为 E 。 G 可以是最普通的高斯平滑或者是 WLS 滤波器平滑的结果。通过约束极值可以把 G 中平滑了的边缘复原,同时保持对细节的抑制。本文通过引入辅助变量和交替最小化来近似求解得到的优化问题。关于图像细节增强的实验表明本文方法可以更好地保持边缘,因而处理后图像中的虚假结构更少。

参 考 文 献

[1] Farbman Z, Fattal R, Lischinski D, et al. Edge-preserving decompositions for multi-scale tone and detail manipulation [J]. ACM Transactions on Graphics(TOG), ACM, 2008, 27(3): 67

[2] Rudin L I, Osher S, Fatemi E. Nonlinear total variation based noise removal algorithms [J]. Physica D: Nonlinear Phenomena, 1992, 60(1):259-268

[3] Xu Li, Lu Ce-wu, Xu Yi, et al. Image smoothing via L_0 gradient

minimization [C]// ACM SIGGRAPH Asia, Hong Kong, November 2011:174,1-174:11

[4] Xu Li, Yan Qiong, Xia Yang, et al. Structure extraction from texture via relative total variation [J]. ACM Transaction on Graphics(TOG), 2012, 31(6):139

[5] Tomasi C, Manduchi R. Bilateral filtering for gray and color images [C]// International Conference on Computer Vision, Bombay, January 1998: 839-846

[6] Kass M, Solomon J. Smoothed local histogram filters [C]// ACM SIGGRAPH. Los Angeles, CA, July 2010:100,1-100:10

[7] Gastal E S L, Oliveira M M. Domain transform for edge-aware Image and video processing [J]. ACM Transactions on Graphics (TOG), ACM, 2011, 30(4): 69

[8] Perona P, Malik J. Scale-space and edge detection using anisotropic diffusion [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1990, 12(7): 629-639

[9] Subr K, Soler C, Durand F. Edge-preserving multiscale image decomposition based on local extrema [J]. ACM Transactions on Graphics(TOG), 2009, 28(5):147

[10] Acton S T. Locally monotonic diffusion [J]. IEEE Transactions on Signal Processing, 2000, 48(5):1379-1389

[11] Wang Yi-lun, Yang Jun-feng, Yin Wo-tao, et al. A new alternating minimization algorithm for total variation image reconstruction[J]. SIAM Journal on Imaging Sciences, 2008, 1(3):248-272