文章编号: 1001-0920 (2014) 06-1115-05 DOI: 10.13195/j.kzyjc.2013.0274

基于监督学习的稀疏编码及在数据表示中的应用

舒振球, 赵春霞, 张浩峰

(南京理工大学 计算机科学与工程学院，南京 210094)

摘 要: 针对稀疏编码在数据表示时没有利用样本类别信息的问题, 提出一种基于监督学习的稀疏编码算法, 并应用于数据表示. 首先利用样本的类别信息构建图, 直接提取样本的鉴别结构信息; 然后利用基向量拟合鉴别结构特性向量, 进而在基向量中嵌入样本的鉴别信息; 最后对样本逐个进行稀疏表示. 在COIL20和PIE图像库的实验结果表明, 相比几种无监督矩阵分解算法, 所提出的算法更利于样本的表示和分类.

关键词: 矩阵分解；鉴别分析；稀疏编码；数据表示；拟合

中图分类号: TP391 **文献标志码:** A

Sparse coding based supervised leaning and its application to data representation

SHU Zhen-qiu, *ZHAO Chun-xia*, *ZHANG Hao-feng*

(School of Computer Science and Engineering，Nanjing University of Science and Technology，Nanjing 210094, China．Correspondent：SHU Zhen-qiu，E-mail：shuzhenqiu@163.com)

Abstract: For the problem of the sparse coding methods not making full use of the label information in data representation, an algorithm, the supervised learning sparse coding, is proposed which can be applied to data representation. Firstly, the proposed algorithm can build the graph via the label information. Thus it directly extracts the discriminate information of the data and then tries to learn the basis which can best fit the discriminate vector. Therefore, it can find a basis set embedding the discriminant information of the samples which are individually for sparse representation. The experiments on the COIL20 and PIE image data sets demonstrate that the proposed algorithm can provide a better representation and classification than the traditional unsupervised matrix factorization algorithms.

Key words: matrix factorization；discriminate analysis；sparse coding；data representation；fit

0 引 言

数据表示是图像处理和模式识别领域的重要课题, 也是目前研究的热点问题之一. 在过去几十年里, 数据表示技术得到了很大进步, 研究者相继提出了多种数据表示方法. 目前, 数据表示算法已广泛应用于机器学习、模式识别和计算机视觉等领域[1-4].

Lee 等[1]提出了非负矩阵分解 (NMF) 算法, 将一个高维矩阵分解成两个非负低秩矩阵的乘积, NMF 的非负约束导致了对数据的基于部分的表示. 由于 NMF 的目标函数非凸, 只能找到局部最优值, 不能找到全局最优解; 同时, 在实际应用中, NMF 的收敛性和收敛速度也是研究的一个难点. 稀疏编码 (SC)[2]是近期提出的一种矩阵分解算法, 将稀疏性作为先验知识, 用尽可能少的基向量表示样本. Zheng 等[3]提出了一种图正则化的稀疏表示 (GSC) 算法, 并应用于图像表示, GSC 算法通过对稀疏编码增加图正则化项, 使得在稀疏编码时考虑样本的几何流形结构信息, 从而获得较为光滑的稀疏表示. Jiang 等[4]提出类别一致的 KSVD 算法 (LC-KSVD), 并应用于人脸识别, LC-KSVD 算法在目标函数中增加样本类别的限制项, 使得在字典学习时保持样本类别信息的一致性. 上述改进的稀疏编码算法都通过增加正则项的方式引入样本的几何结构信息或者类别信息, 但是采用交替迭代算法求解目标函数中的 l_1 范数问题, 计算复杂度较高.

鉴于此, 为了利用样本的类别信息, 本文提出一

收稿日期: 2013-03-16；**修回日期:** 2013-05-16.

基金项目: 国家自然科学基金项目(61272220, 61101197)；江苏省自然基金青年项目(BK2012399)；江苏省普通高校研究生创新计划项目(CXLX13_19)；哈尔滨工程大学水下机器人技术国家科技重点实验室开放基金项目.

作者简介: 舒振球(1985−), 男, 博士生, 从事图像处理、模式识别的研究；赵春霞(1964−), 女, 教授, 博士生导师, 从事机器人、人工智能、图像处理与模式识别等研究.

种监督学习的稀疏编码(SSC)算法. 该算法考虑样本的类别信息, 将样本的鉴别结构信息嵌入到基向量中, 更有利于样本的表示与分类. 实验结果表明了所提出算法的有效性.

1 相关工作

1.1 稀疏编码算法

稀疏编码是一种基于生物视觉系统的数据表示算法, 它利用尽可能少的基向量的线性组合来表示数据. 稀疏编码的模型表示为

$$\min_x \|x\|_0;$$
$$\text{s.t. } y = Dx. \tag{1}$$

其中: y 为观察样本, x 为表示系数, D 为过完备字典, $\|\cdot\|_0$ 为 l_0 范数. 求解 l_0 范数是NP问题, 压缩感知理论[5-7]表明, 当满足一定条件时, 求解 l_0 范数等价求解 l_1 范数, 因此式(1)转换为

$$\min_x \|x\|_1;$$
$$\text{s.t. } y = Dx. \tag{2}$$

由于式(2)是凸的, 可以采用经典的优化理论进行求解. 目前已存在一些求解工具包, 如 l_1-magic、SPAMS、Sparselab 等.

稀疏编码的字典是过完备的, 已有研究表明, 在过完备字典下的信号稀疏表示更为有效, 这是稀疏编码的优点之一. 另外, 稀疏编码的系数是稀疏的, 其应用领域更广, 如数据压缩、图像去噪等. 同时, 与稀疏编码的不同在于, 稀疏子空间方法(如SparseLDA等)的稀疏项是 D, 其低维表示子空间仍然是稠密的, 而稀疏编码的低维表示子空间是稀疏的. 但是, 稀疏编码也存在缺点, 如求解 l_1 范数非常耗时等, 因此, 如何快速地求解 l_1 范数也是研究的热点之一.

1.2 稀疏概念编码算法

Cai 等[8]提出了一种SCC(sparse concept coding)算法, 并应用于视觉分析. 该算法结合流形学习算法和稀疏编码算法的思想, 只需求解两个回归问题, 计算方便简单. SCC算法首先对样本的最近邻图进行谱分析, 使得到的特征向量嵌入样本几何流形结构信息; 然后利用基向量进行拟合, 使基向量中嵌入样本流形的几何结构信息; 最后利用LARs(least angle regress)[9]算法对每个样本进行稀疏表示学习, 得到样本的表示系数矩阵.

SCC算法不但能使样本的几何结构信息嵌入到基向量中, 而且使得表示系数是稀疏的. 但SCC算法没有考虑样本的类别信息, 不利于图像的表示与分类.

2 基于监督的稀疏编码算法

2.1 SSC算法原理

针对SCC算法没有利用样本类别信息的缺陷, 本文提出的SSC算法根据样本的标签建立类别信息矩阵图, 方便地引入样本的类别信息, 并直接提取样本鉴别结构信息. 利用基向量进行拟合, 从而使学习到的基向量具有鉴别性, 并逐个对样本进行稀疏表示学习. 相比于SCC算法, SSC算法不仅引入了类别信息, 而且在处理大规模图像表示问题时, 不需要进行特征值分解, 计算方便简单, 具有明显的优势.

2.2 SSC算法介绍

SSC算法主要包括3部分: 提取鉴别特征信息、基学习和稀疏表示学习. 下面将对这3个部分进行详细说明.

2.2.1 提取鉴别特征信息

假设样本集 $\{x_i\}_{i=1}^m$ 属于 c 类不同的样本, m_t 表示第 t 类样本的数目, 其中 $\sum_{t=1}^{c} m_t = m$. 矩阵 W 表示样本的类别图, 根据样本的类别信息, 定义为

$$W_{ij} = \begin{cases} 1/m_t, \ x_i, x_j \text{ both belong to the } t\text{-th class;} \\ 0, \ \text{otherwise.} \end{cases} \tag{3}$$

其中矩阵 D 为对角矩阵且 $D = \sum_j W_{ij}$, $L = D - W$. 根据谱图理论, 对下式进行特征值分解:

$$Ly = \lambda Dy, \tag{4}$$

其中 y 为特征值 λ 对应的特征向量. 取 k 个最小特征值对应的特征向量组成低维表示向量, 即

$$Y = [y_1, y_2, \cdots, y_k].$$

定理1 假设 y 为特征值 λ 对应的特征向量, 如果 $X^{\mathrm{T}}a = y$, 其中 a 为投影向量, 则下式特征值 λ 对应的特征向量 a 与式(4)相同特征值 λ 对应的特征向量 y 等价:

$$XWX^{\mathrm{T}}a = \lambda XDX^{\mathrm{T}}a. \tag{5}$$

证明 因为 $Wy = \lambda Dy$, 所以由式(5)得

$$XWX^{\mathrm{T}}a = XWy = X\lambda Dy = \lambda XDX^{\mathrm{T}}a. \quad \square$$

式(5)是经典的图嵌入框架, 由定理1可见, 采用谱图分析提取的低维特征向量与基于监督的图嵌入特征提取算法等价. 但是, 在处理大规模图像表示问题时, 为了计算方便, 根据参考文献[10], 本文直接求取具有鉴别信息的特征向量, 而不需要特征值分解, 能够有效地提高计算效率. 假设 $\{x_i\}_{i=1}^m$ 按类别排列,

式 (3) 的类别图 W 可以等价表示为

$$W=\begin{bmatrix} W^{(1)} & 0 & \cdots & 0 \\ 0 & W^{(2)} & \cdots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & W^{(c)} \end{bmatrix},$$

其中 $W^{(j)}$ 为 $m_j \times m_j$ 维矩阵, 且元素全为 $1/m_t$. 式 (4) 的特征值 λ 和特征向量 y 可以由其块矩阵的特征值和特征向量组合, 即

$$W^{(t)}y^{(t)}=\lambda D^{(t)}y^{(t)}. \tag{6}$$

其中: $y^{(t)}$ 为第 t 类样本的特征向量, $D^{(t)}=\sum\limits_j W_{ij}^{(t)}=I$. 很明显, 式 (6) 的特征值为 1, 其对应的特征向量为 $e^{(t)}\in \boldsymbol{R}^{m_t}$, 且 $e^{(t)}=[1,1,\cdots,1]^{\mathrm{T}}$. Rank$(W^{(t)})=1$, 即 $W^{(t)}$ 中只有一个非零的特征值, 因此, 式 (4) 中的 c 个特征向量可以表示为

$$y_t=[\underbrace{0,\cdots,0}_{\sum_{i=1}^{t-1} m_i},\underbrace{1,\cdots,1}_{m_t},\underbrace{0,\cdots,0}_{\sum_{i=t+1}^{c} m_i}]^{\mathrm{T}},$$

其中 $t=1,2,\cdots,c$. 这些特征向量对应的特征值为 1, 选择任意 c 个正交向量张成 y_t, 但是向量投影函数的所有样本均嵌入到同一个点, 因此这些向量不适合用作投影子空间. 令 $y_0=[1,1,\cdots,1]^{\mathrm{T}}$, 由于 y_0 在 $\{y_t\}$ 张成的子空间中, 将 y_0 作为第 1 个向量, 并通过 Gram-Schmidt 求解 $c-1$ 个正交的特征向量, 即

$$\{\bar{y}_k\}_{k=1}^{c-1},\ \bar{y}_k y_0=0,\ \bar{y}_i y_j=0,\ i\neq j,$$

从而得到 $Y=[y_1,y_2,\cdots,y_k]$, 其中 y_i 每一行均嵌入了样本的鉴别信息.

2.2.2 基 学 习

在实际应用中, 当样本数小于特征向量时存在过拟合的情况, 因此, 满足 $Y=X^{\mathrm{T}}U$ 的基向量 U 可能是不存在的, 常用的方法是利用最小二乘法最大限度地逼近

$$\min_U \|Y-X^{\mathrm{T}}U\|_F^2+\alpha\|U\|^2, \tag{7}$$

其中 α 为正则项参数, $\alpha\|U\|^2$ 正则项可以避免过拟合. 式 (7) 中求解得到的 U 的最优解为

$$U^*=(XX^{\mathrm{T}}+\alpha I)^{-1}XY, \tag{8}$$

其中 I 为 m 阶单位矩阵. 通常采用迭代算法 (如 LSQR[11]算法) 求解类似式 (7) 的回归问题.

2.2.3 稀疏表示学习

根据稀疏编码的模型, 用 a_i 表示 A 的第 i 列, 通过式 (9) 得到稀疏表示系数 a_i, 有

$$\min_{a_i} \|x_i-Ua_i\|_2+\beta|a_i|, \tag{9}$$

其中 $|a_i|$ 为 a_i 的 l_1 范数, 表示系数 a_i 是稀疏的.

将式 (9) 转换为稀疏性约束的非线性逼近模型

$$\min_U \|x_i-Ua_i\|_2,\quad \text{s.t. } |a_i|\leqslant M, \tag{10}$$

其中 a_i 的非零个数小于等于 M. 通常称这类问题为 M 项稀疏逼近问题, 本文采用 LARs 算法进行求解.

2.3 SSC 具体步骤描述

SSC 算法的具体实现如下.

输入: 样本集 $X=(x_1,x_2,\cdots,x_m)\in \boldsymbol{R}^{m\times n}$, 样本的类别数为 c, m_k 为第 k 类样本的数目;

输出: 样本的稀疏表示系数矩阵 A.

Step 1: 令 $y_0=[1,1,\cdots,1]^{\mathrm{T}}$, 根据式 (6) 求取具有鉴别结构信息的特征向量 Y.

Step 2: 基学习. 利用式 (8) 求解具有鉴别性的基向量 U.

Step 3: 稀疏表示学习. 设置表示系数的非零个数, 利用 LARs 对式 (10) 进行求解, 得到每个样本的稀疏表示系数 a_i.

2.4 算法复杂度分析

SSC 算法的计算主要包括 3 部分: 1) 提取鉴别结构信息的复杂度 $O(mc^2-c^3/3)$; 2) 当样本矩阵稠密时, 基学习的复杂度为 $2kcmn+O(m)+O(n)$, 当样本矩阵稀疏时, 复杂度为 $2ksmn+5kcn+O(m)$, 其中 s 为每个特征向量中非零个数且 $s\ll n$, k 为 LSQR 的迭代次数; 3) 稀疏表示学习的复杂度为 $O(c^3+mc^2)$. 因为 $c\ll \min(m,n)$ 且 $k\ll n$, 所以当样本矩阵稠密时, SSC 算法的总体复杂度为 $O(2kcmn+c^3+mc^2)$, 当样本矩阵稀疏时, SSC 算法的总体复杂度为 $O(2kcms+5kcn+c^3+mc^2)$.

3 实验结果分析

由于矩阵分解算法已在聚类中得到了广泛应用[12], 本节在聚类实验中验证 SSC 算法的有效性.

3.1 数据集介绍

实验 1 在 COIL20 图像库进行, 包括共 20 个物体 1 440 幅灰度图像, 图 1 为部分 COIL20 库图像.

图 1 部分 COIL20 库图像

实验2在PIE库上进行, 包括共2 856幅人脸图像, 图2为部分PIE人脸图像.

图 2 部分PIE库图像

3.2 实验结果分析

实验用准确率(AC)和归一化互信息(NMI)评估聚类效果, 具体定义见文献[13]. 将本文提出的SSC算法与K-mean、NMF、PCA、SCC算法进行对比分析, 结果如表1～表4所示. 实验从样本集中随机选k类样本进行聚类, 并重复实验20次取平均值. 实验中SSC算法需要确定正则化参数α, 通常设置为0.1.

表 1 COIL20库上的AC

K	AC				
	K-mean	NMF	PCA	SCC	SSC
8	0.749	0.722	0.723	0.859	**0.996**
10	0.717	0.703	0.734	0.808	**0.975**
12	0.674	0.676	0.674	0.824	**0.874**
14	0.666	0.681	0.652	0.814	**0.851**
16	0.638	0.645	0.651	0.795	**0.833**
18	0.622	0.628	0.630	0.795	**0.844**
20	0.631	0.622	0.607	0.810	**0.878**
avg	0.671	0.668	0.667	0.815	**0.893**

表 2 COIL20库上的NMI

K	NMI				
	K-mean	NMF	PCA	SCC	SSC
8	0.734	0.708	0.714	0.877	**0.997**
10	0.744	0.728	0.738	0.868	**0.985**
12	0.734	0.735	0.732	0.874	**0.945**
14	0.740	0.737	0.732	0.880	**0.959**
16	0.732	0.724	0.739	0.868	**0.928**
18	0.728	0.723	0.735	0.878	**0.946**
20	0.746	0.727	0.733	0.885	**0.956**
avg	0.737	0.726	0.731	0.876	**0.959**

表 3 PIE库上的AC

K	AC				
	K-mean	NMF	PCA	SCC	SSC
10	0.296	0.473	0.306	0.843	**0.899**
20	0.284	0.442	0.288	0.809	**0.869**
30	0.260	0.418	0.263	0.778	**0.814**
40	0.251	0.413	0.260	0.769	**0.804**
50	0.253	0.408	0.2528	0.801	**0.818**
60	0.242	0.401	0.235	0.769	**0.786**
68	0.245	0.385	0.229	0.751	**0.782**
avg	0.262	0.420	0.262	0.789	**0.824**

表 4 PIE库上的NMI

K	NMI				
	K-mean	NMF	PCA	SCC	SSC
10	0.353	0.566	0.359	0.855	**0.952**
20	0.450	0.640	0.450	0.877	**0.947**
30	0.472	0.652	0.477	0.877	**0.938**
40	0.498	0.667	0.502	0.890	**0.924**
50	0.522	0.687	0.520	0.917	**0.948**
60	0.529	0.695	0.525	0.900	**0.927**
68	0.538	0.698	0.531	0.904	**0.937**
avg	0.480	0.658	0.481	0.889	**0.939**

由表1～表4可见:

1) SCC算法利用样本的几何流形结构信息, 实验结果明显优于K-mean、NMF和PCA算法, 表明在稀疏编码过程中嵌入样本的几何结构信息, 能有效地提高聚类的AC和NMI.

2) SSC算法优于其他几种无监督学习的矩阵分解算法, 表明在引入样本的类别信息后, 能更好地表示数据, 更好地利用样本的聚类.

3) 利用样本信息的SSC算法在数据表示时明显优于利用样本几何流形结构信息的SCC算法, 表明在稀疏编码时基向量嵌入的类别信息比几何结构信息更有利于样本的表示和分类.

3.3 评估稀疏表示系数中非零个数对实验结果的影响

利用LARs求解SCC算法和SSC算法时, 需要指定表示系数中的非零个数, 即基数. 因此, 通过实验分析基数对聚类实验的影响, 结果如图3和图4所示. 由图3和图4可见, 当表示系数中含有4个非零数时, 聚类的AC和NMI比较稳定, 表明算法具有较强的鲁棒性.

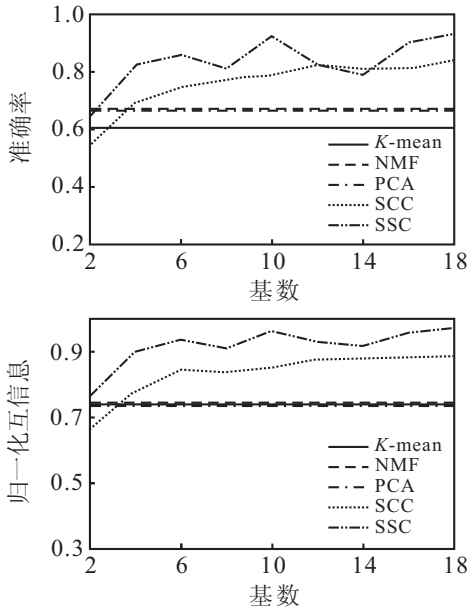

图 3 系数中基数的评估实验(COIL20库)

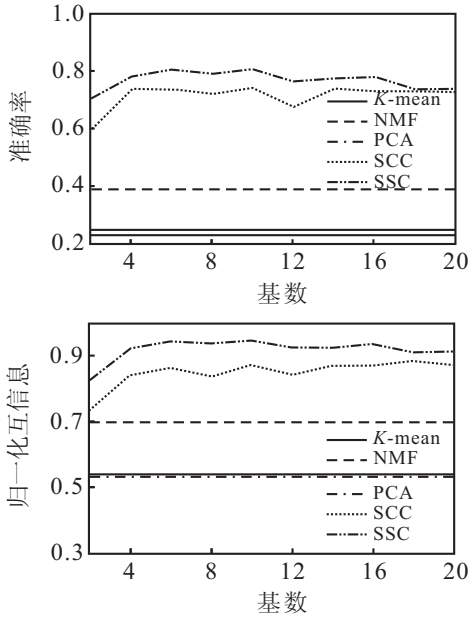

图 4　系数中基数的评估实验 (PIE 库)

4　结　论

本文针对稀疏编码在数据表示时没有利用样本类别信息的问题, 提出了一种基于监督学习的 SSC 算法. 首先提取样本的鉴别特征信息, 并将鉴别结构信息嵌入到基向量中, 使学习到的基向量更具有鉴别性, 有利于图像的表示, 通过实验也表明了 SSC 算法的有效性. 相对于传统的稀疏编码, SSC 只需求解两个回归问题, 计算代价小, 但是没有考虑样本流形的几何结构信息, 同时利用样本的类别信息和样本流形的几何结构信息是下一步的研究方向.

参考文献(References)

[1] Lee D D, Seung H S. Learning the parts of objects by non-negative matrix factorization[J]. Nature, 1999, 401(6755): 788-791.

[2] Lee H, Battle A, Rainna R, et al. Efficient sparse coding algorithms[C]. Advances in Neural Information Processing Systems. Cambridge: MIT Press, 2006: 801-808.

[3] Zheng M, Bu J J, Chen C, et al. Graph regularized sparse coding for image representation[J]. IEEE Trans on Image Processing, 2011, 20(5): 1327-1336.

[4] Jiang Z L, Lin Z, Davis L S. Learning a discriminative dictionary for sparse coding via label consistent K-SVD[C]. IEEE Conf on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2011: 1697-1704.

[5] Donoho D. For most large underdetermined systems of linear equations the minimal L1-norm solutions is also the sparsest solution[J]. Communications on Pure and Applied Mathemati, 2006, 59(6): 797-829.

[6] Candes E, Romberg J, Tao T. Stable signal recovery from incomplete and inaccurate measurements[J]. Communications on Pure and Applied Mathemati, 2006, 59(8): 1207-1223.

[7] Candes E, Tao T. Near-optimal signal recovery from random projections: Universal encoding strategies?[J]. IEEE Trans on Information Theory, 2006, 52(12): 5406-5425.

[8] Cai D, Bao H J, He X F. Sparse concept coding for visual analysis[C]. IEEE Conf on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2011: 2905-2910.

[9] Efron B, Hastie T, Johnstone I, et al. Least angle regression[J]. Annals of Statistics, 2004, 32(2): 407-499.

[10] Cai D, He X F, Han J W. SRDA: An afficient algorithm for large-scale discriminant analysis[J]. IEEE Trans on Knowledge and Data Engineering, 2008, 20(1): 1-12.

[11] Paige C C, Saunders M A. LSQR: An algorithm for sparse linear equations and sparse least squares[J]. ACM Transactions on Mathematical Software, 1982, 8(1): 43-71.

[12] Wang Y X, Zhang Y J. Nonnegative matrix factorization: A comprehensive review[J]. IEEE Trans on Knowledge and Data Engineering, 2013, 25(6): 1336-1353.

[13] Xu W, Liu X, Gong Y. Document clustering based on non-negative matrix factorization[C]. Proc of 2003 Int Conf Research and Development in Information Retrieval. New York: ACM Press, 2003: 267-273.

(责任编辑: 郑晓蕾)