

核鉴别稀疏保持嵌入

李贝贝

摘要

定义 0.1 If there are two or more ways to do something, and one of those ways can result in a catastrophe, then someone will do it.

1 前言

2 相关工作

2.1 稀疏表示

信号稀疏表示的最初目的是为了用比香农定理更低的采样率来表示和压缩信号 [2], 就像离散余弦变换和小波变换等。在过去几年里, 稀疏表示已经被用于很多信号处理, 模式识别的实际应用中。例如, 在信号的图片处理领域, 稀疏表示被用于信号压缩和编码 [5], 图片去噪 [3], 图片超像素图片超像素 [9]。在模式识别领域, 稀疏表示被用于目标识别和分类任务 [4, 1]。一些研究表明基于稀疏表示的分类器非常有效, 在一些人脸数据库上的识别率甚至是目前的最高水平。

正如在离散余弦变换和小波变换中, 原始信号可以用基函数表示, 稀疏表示则是用一个称之为字典的超完备冗余的函数系统来代替原始基函数, 并利用字典 \mathbf{X} 中最少的原子来表示一个原始信号 b 。

稀疏表示的思想如下 [10]: 假设样本集 X 有可以分为 k 类的共 m 个训练样本, $X_k = [x_{k1}, x_{k2}, \dots, x_{kl_k}] \in R^{n \times l_k}$ 表示属于第 k 类的 l_k 个样本集, 如果训练样本足够多, 则属于第 k 类的任意一个样本 $y \in R^n$ 可以近似地表示成此类的所有训练样本的线性组合

$$\mathbf{y} = h_{k1}\mathbf{X}_{k1} + h_{k2}\mathbf{X}_{k2} + \dots + h_{kl_k}\mathbf{X}_{kl_k}$$

目标函数如下:

$$\begin{aligned} \min_s \quad & \|s\|_0 \\ \text{s.t.} \quad & b = Xs \end{aligned}$$

假如有 \mathbf{N}

2.2 鉴别稀疏保持投影

稀疏保持投影 [6] 将所有样本作为字典原子进行稀疏重构, 鉴别稀疏保持投影则是增强了同类数据的稀疏表示权重 [11], 在保持局部结构的同时保持样本之间的总体信息, 通过局部合并, 以整体的方式发现和重建数据集的内在规律。并通过求解最小二乘问题来更新 SPP 中的稀疏权重矩阵, 得到了一个更能真实反映鉴别信息的鉴别稀疏权重。将无监督的 SPP 转换成有监督的 DSPE, 充分利用训练样本中的标签信息。样本训练集 $X = [X_1, X_2, \dots, X_c]$, 其中, $X_i = [x_{i1}, \dots, x_{ik}] \in R^{m \times k}$ 是类别标记为 $label(i)$ 的样本集合, 即第 i 类样本。通过最小二乘法获得最能反映类别鉴别信息的权重, 目标函数如下:

$$\begin{aligned} \min_t \quad & \|x_{ij} - \mathbf{X}_i \mathbf{t}\|_2 \\ \text{s.t.} \quad & \mathbf{I} \mathbf{t} = 1 \end{aligned}$$

其中 $\mathbf{t} = [t_1, t_2, \dots, t_{j-1}, 0, t_{j+1}, \dots, t_k]^T$, 第 j 个分量为 0, 即将任一向量利用同类中除自身以外的向量进行表示。经证明 [11], 该方法具有良好的鲁棒性, 具有旋转、平移、尺度不变的特性。

为了考虑样本集的全局几何结构, DSPE 进一步考虑异类样本对原样本重构的影响。将 $\hat{t} \in \mathbf{R}^k$ 扩展到 n 个分量 $\hat{h} = [\vec{0}, \dots, \vec{0}, \hat{t}, \vec{0}, \dots, \vec{0}] \in \mathbf{R}^n$, 即将除鉴别信息权重以外的系数设置为 0. 令 $er = x_{ij} - X\hat{t} = x_{ij} - X\hat{h}$, 对 er 进行在超完备字典 $\hat{X} = [X_1, X_2, \dots, X_c]$ 上的稀疏表示, 其中 $X_i = \vec{0}$, 将人脸样本由鉴别信息权重重构后得到的残差有其它类的所有样本进行稀疏重构:

$$\begin{aligned} \min_s \quad & \|s\|_1 \\ \text{s.t.} \quad & \|er - \hat{X}s\|_1 < \epsilon \\ & Is = 0 \end{aligned}$$

作如下变换:

$$\begin{aligned} s^+ &= \begin{cases} s, & s > 0 \\ 0, & s \leq 0 \end{cases} \\ s^- &= \begin{cases} -s, & s < 0 \\ 0, & s \geq 0 \end{cases} \end{aligned}$$

由于 s^+, s^- 非负, $s = s^+ + s^-$, $\|s\|_1 = \sum_i (s_i^+ + s_i^-)$, 目标函数转化为

$$\begin{aligned} \min_{s^+ + s^-} \quad & \sum_i (s_i^+ + s_i^-) \\ \text{s.t.} \quad & \|er - [\hat{X}, -\hat{X}][s^+, s^-]^T\| < \epsilon \\ & [I, -I][s^+, s^-]^T = 0 \end{aligned}$$

将优化结果 $\hat{s} = [\hat{s}_1, \dots, \hat{s}_{i-1}, \vec{0}, \hat{s}_{i+1}, \dots, \hat{s}_c]$ 与 \hat{t} 串联, 得鉴别稀疏权重 $\hat{d} = \hat{s} + \hat{h} = [\hat{s}_1, \dots, \hat{s}_{i-1}, \hat{t}, \hat{s}_{i+1}, \dots, \hat{s}_c] \in \mathbf{R}^n$. 由于 $I\hat{t} = 1, I\hat{s} = 1$, 故 $I\hat{d} = 1$, 该鉴别稀疏权重能够保持旋转, 平移, 尺度不变的特性 [11] 与 SPP 相比较, DSPE 中最小化同类样本重构数更加趋近于 0, 既增强了同类费劲林样本在重构中的作用, 又削弱了异类伪近邻样本对原样本重构的影响, 有利于保持同类样本相互靠近, 异类样本相互远离的本质。当 $er = 0$ 时, 鉴别稀疏权重便退化成了类似于 NPE 保持低维流形上同类数据近邻关系的重构权重, 表明 DSPE 能很好地反映数据在低维流形上的分布, 与此同时, 对 er 的稀疏表示有利于加强算法的鲁棒性 DSPE 的目标在于将数据间的鉴别稀疏重构特征进行保持并嵌入到低维流形上, 因此, 数学模型如下:

$$\begin{aligned} \min_W \quad & \sum_{i=1}^n \|W^T x_i - W^T X \hat{d}_i\|^2 \\ \text{s.t.} \quad & W^T X X^T W = 1 \end{aligned}$$

对目标函数进行变换, 目标函数转换成

$$\begin{aligned} \max_W \quad & W^T X M X^T W \\ \text{s.t.} \quad & W^T X X^T W = 1 \end{aligned}$$

其中, $M = D + D^T - D^T D$, 目标函数可以通过求解广义特征值得解

$$X M X^T W = \lambda X X^T W$$

选取最大的 d 个特征值所对应的特征向量 \mathbf{a}_i 构成特征子空间, 即可得到 DSPE 的线性降维映射 $W_{DSPE} = [a_1, a_2, \dots, a_d]$. DSPE 既反映了样本间的鉴别关系, 又能很好地排除伪类样本在重构人脸数据时带来的负面影响, 经实验, 在多个人脸库上效果超过 PCA, LDA, NPE, LPP, SPP 等算法。

2.3 核函数

在实际分类中, 样本本身可能存在线性不可分的情况, 而将特征映射到高维空间后, 往往就可分了。利用核函数可以将样本从低维空间变换到高维, 基于核的子空间特征提取算法是模式识别中的一个重要的研究方向, 核主成分分析 [8] 和核 Fisher 鉴别分析 [7] 是核函数的典型用例。

3 核鉴别稀疏保持嵌入

3.1 核鉴别信息权重

将样本信息 $X_i = [x_{i1}, \dots, x_{ik}]$ 映射到高维空间为 $B_i = [\phi(x_{i1}), \dots, \phi(x_{ik})]$, 只要证明 $\|x_{ij} - X_i t\|$ 越小, 则 $\|\phi(x_{ij}) - B_i t\|$ 越小最小二乘问题变为

$$\begin{aligned} \min_t & \|\phi(x_{ij}) - B_i t\|_2 \\ \text{s.t.} & \quad I t = 1 \end{aligned}$$

因为 B 和 ϕx 是未知的, 所以目标函数不能直接求解, 所以将问题转化成以下约束:

$$\begin{aligned} \min_t & \|B^T \phi(x_{ij}) - B^T B_i t\|_2 \\ \text{s.t.} & \quad I t = 1 \end{aligned}$$

求证 $\|x_{ij} - X_i t\|$,

$$\min_t \|\phi(x_{ij}) - B_i t\|_2$$

,

$$\min_t \|B^T \phi(x_{ij}) - B^T B_i t\|_2$$

同时达到最小值

$$B_i^T \phi(x_{ij}) = \begin{bmatrix} \phi(x_{i1}) \\ \phi(x_{i2}) \\ \vdots \\ \phi(x_{ik}) \end{bmatrix} \phi(x_{ij}) = \begin{bmatrix} k(x_{i1}, x_{ij}) \\ k(x_{i2}, x_{ij}) \\ \vdots \\ k(x_{ik}, x_{ij}) \end{bmatrix}$$

$$B_i^T B_i = \begin{bmatrix} \phi(x_{i1}) \\ \phi(x_{i2}) \\ \vdots \\ \phi(x_{ik}) \end{bmatrix} \begin{bmatrix} \phi(x_{i1}) & \phi(x_{i2}) & \dots & \phi(x_{ik}) \end{bmatrix} = \begin{bmatrix} k(x_{i1}, x_{i1}) & k(x_{i1}, x_{i2}) & \dots & k(x_{i1}, x_{ik}) \\ k(x_{i2}, x_{i1}) & k(x_{i2}, x_{i2}) & \dots & k(x_{i2}, x_{ik}) \\ \vdots & \vdots & \ddots & \vdots \\ k(x_{ik}, x_{i1}) & k(x_{ik}, x_{i2}) & \dots & k(x_{ik}, x_{ik}) \end{bmatrix}$$

目标函数转换成线性规划问题, 可以求解。

3.2 核稀疏重构权重

令 $C = [\phi(X_1), \phi(X_2), \dots, \phi(X_c)]$, 令 $C_i = [\phi(X_1), \dots, \phi(X_{i-1}), \phi(X_{i+1}), \dots, \phi(X_c)]$ 即从超完备字典中剔除样本 X_i 令 $er = \phi(x_{ij}) - B_i \hat{t} = \phi(x_{ij}) - C_i \hat{h}$, 即样本由鉴别信息权重重构后得到的残差, 对其对样本中的其他类进行重构:

$$\begin{aligned} \min_s & \|s\|_1 \\ \text{s.t.} & \quad \|er - C_i s\| < \epsilon \\ & \quad I s = 0 \end{aligned}$$

同理由于函数 ϕ 不能直接求解, 将 $er - C_i s$ 左乘 C^T 转换成 $C_i^T(er - C_i s)$, 即 $C^T \phi(x_{ij}) - C^T C_i \hat{h} - C^T C_i s$

求证

公理 3.1 对任意 $\epsilon \geq 0$ 都存在 $\delta \geq 0$, 只要 $C^T \phi(x_{ij}) - C^T C_i \hat{h} - C^T C_i s < \delta$ 就有 $er - C_i s < \epsilon$.

$$\begin{aligned}
C^T \phi(x_{ij}) &= \begin{bmatrix} \phi(x_{11}) & \phi(x_{12}) & \dots & \phi(x_{1k}) \\ \phi(x_{21}) & \phi(x_{22}) & \dots & \phi(x_{2k}) \\ \vdots & \vdots & \ddots & \vdots \\ \phi(x_{c1}) & \phi(x_{c2}) & \dots & \phi(x_{ck}) \end{bmatrix} \phi(x_{ij}) = \begin{bmatrix} k(x_{11}, x_{ij}) & k(x_{12}, x_{ij}) & \dots & (x_{1k}, x_{ij}) \\ k(x_{21}, x_{ij}) & k(x_{22}, x_{ij}) & \dots & (x_{2k}, x_{ij}) \\ \vdots & \vdots & \ddots & \vdots \\ k(x_{c1}, x_{ij}) & k(x_{c2}, x_{ij}) & \dots & (x_{ck}, x_{ij}) \end{bmatrix} \\
C^T C &= \begin{bmatrix} \sum_{j=1}^k k(x_{1j}, x_{1j}) & \sum_{j=1}^k k(x_{1j}, x_{2j}) & \dots & \sum_{j=1}^k k(x_{1j}, x_{cj}) \\ \sum_{j=1}^k k(x_{2j}, x_{1j}) & \sum_{j=1}^k k(x_{2j}, x_{2j}) & \dots & \sum_{j=1}^k k(x_{2j}, x_{cj}) \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{j=1}^k k(x_{cj}, x_{1j}) & \sum_{j=1}^k k(x_{cj}, x_{2j}) & \dots & \sum_{j=1}^k k(x_{cj}, x_{cj}) \end{bmatrix} \\
C^T C_i &= \begin{bmatrix} \sum_{j=1}^k k(x_{1j}, x_{1j}) & \dots & \sum_{j=1}^k k(x_{1j}, x_{i-1,j}) & \sum_{j=1}^k k(x_{1j}, x_{i+1,j}) & \dots & \sum_{j=1}^k k(x_{1j}, x_{cj}) \\ \sum_{j=1}^k k(x_{2j}, x_{1j}) & \dots & \sum_{j=1}^k k(x_{2j}, x_{i-1,j}) & \sum_{j=1}^k k(x_{2j}, x_{i+1,j}) & \dots & \sum_{j=1}^k k(x_{2j}, x_{cj}) \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \sum_{j=1}^k k(x_{cj}, x_{1j}) & \dots & \sum_{j=1}^k k(x_{cj}, x_{i-1,j}) & \sum_{j=1}^k k(x_{cj}, x_{i+1,j}) & \dots & \sum_{j=1}^k k(x_{cj}, x_{cj}) \end{bmatrix}
\end{aligned}$$

将 \hat{s} 按照不同类别分块,得 $\hat{s} = [\hat{s}_1, \dots, \hat{s}_{i-1}, \hat{s}_{i+1}, \dots, \hat{s}_c]$, 将 \hat{s} 稀疏到完备字典上得 $\hat{l} = [\hat{s}_1, \dots, \hat{s}_{i-1}, \vec{0}, \hat{s}_{i+1}, \dots, \hat{s}_c]$ 结合得到的鉴别权重 \hat{t} , 将系数串联得到 $\hat{d} = \hat{l} + \hat{h}$, 这样我们就得到了核鉴别稀疏权重。由于 $I\hat{l} = I\hat{s} = 0$, $I\hat{t} = 1$, 故 $I\hat{d} = 1$, 能够保持旋转, 平移, 尺度不变的特性, 证明过程见 [11]

3.3 KDSPE 的目标函数

与 DSPE 同理, KDSPE 的目标函数数学模型如下:

$$\begin{aligned}
\min_W \quad & \sum_{i=1}^n \|W^T x_i - W^T X \hat{d}_i\|^2 \\
s.t. \quad & W^T X X^T W = 1
\end{aligned}$$

与 DSPE 的推导相似, 可得到

$$X M X^T W = \lambda X X^T W$$

选取得到的最大的 d 个特征值所对应的特征向量 \mathbf{a}_i 构成的特征子空间, 即可得到 KDPE 的线性降维映射 $W_{DSPE} = [a_1, a_2, \dots, a_d]$.

4 实验

利用多个人脸库或者人脸库 + 行人库 + 玩具库 (算法需要样本数量比较多的限制)

4.1 Yale 人脸库

4.2 某库

5 结论

参考文献

- [1] Mark A Davenport, Marco F Duarte, Michael B Wakin, Jason N Laska, Dharmpal Takhar, Kevin F Kelly, and Richard G Baraniuk. The smashed filter for compressive classification and target recognition. In Electronic Imaging 2007, pages 64980H - 64980H. International Society for Optics and Photonics, 2007.

- [2] Yan De-Qin, Liu Sheng-Lan, and Li Yan-Yan. An embedding dimension reduction algorithm based on sparse analysis. *Acta Automatica Sinica*, 37(11):1306 - 1312, 2011.
- [3] Michael Elad and Michal Aharon. Image denoising via sparse and redundant representations over learned dictionaries. *Image Processing, IEEE Transactions on*, 15(12):3736 - 3745, 2006.
- [4] Ke Huang and Selin Aviyente. Sparse representation for signal classification. In *Advances in neural information processing systems*, pages 609 - 616, 2006.
- [5] Michael W Marcellin, Michael J Gormish, Ali Bilgin, and Martin P Boliek. An overview of jpeg-2000. In *Data Compression Conference, 2000. Proceedings. DCC 2000*, pages 523 - 541. IEEE, 2000.
- [6] Lishan Qiao, Songcan Chen, and Xiaoyang Tan. Sparsity preserving projections with applications to face recognition. *Pattern Recognition*, 43(1):331 - 341, 2010.
- [7] Volker Roth and Volker Steinhage. Nonlinear discriminant analysis using kernel functions. In *Advances in neural information processing systems*. Citeseer, 1999.
- [8] Bernhard Schölkopf, Alexander Smola, and Klaus-Robert Müller. Nonlinear component analysis as a kernel eigenvalue problem. *Neural computation*, 10(5):1299 - 1319, 1998.
- [9] Jianchao Yang, John Wright, Thomas Huang, and Yi Ma. Image super-resolution as sparse representation of raw image patches. In *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*, pages 1 - 8. IEEE, 2008.
- [10] 殷俊 and 杨万扣. 核稀疏保持投影及生物特征识别应用. *电子学报*, 41(4):639 - 645, 2013.
- [11] 马小虎 and 谭延琪. 基于鉴别稀疏保持嵌入的人脸识别算法. *自动化学报*, 40(1):73 - 82, 2014.