

成对约束指导的稀疏保持投影

齐鸣鸣^{1,2*}

(1. 绍兴文理学院 元培学院, 浙江 绍兴 312000; 2. 同济大学 计算机科学与技术系, 上海 201804)

(* 通信作者电子邮箱 qimingming2012@163.com)

摘要:针对稀疏保持投影的稀疏重构过程中监督信息不足的问题,提出一种成对约束指导的稀疏保持投影算法。该算法在训练样本数据的稀疏重构的过程中,通过引入正约束和负约束监督信息指导稀疏重构,使得稀疏保持投影有效地融合了约束监督信息。在 UMIST、YALE 和 AR 人脸库人脸数据集上的实验结果表明,与无监督的稀疏保持投影相比,该方法提高了基于最近邻分类算法的 5%~15% 识别准确率,有效地提高了降维分类性能。

关键词:降维;稀疏重构;成对约束;稀疏保持投影

中图分类号: TP391.41 **文献标志码:** A

Pairwise constraint-guided sparsity preserving projections

QI Ming-ming^{1,2*}

(1. School of Yuanpei, Shaoxing University, Shaoxing Zhejiang 312000, China;

2. Department of Computer Science and Technology, Tongji University, Shanghai 201804, China)

Abstract: Concerning the deficiency of supervision information in the process of sparse reconstruction in Sparsity Preserving Projection (SPP), Pairwise Constraint-guided Sparsity Preserving Projection (PCSPP) was proposed, which introduced supervision information of must-link constraints and cannot-link constraints to guide sparse reconstruction in the process of sparsity reconstruction of training samples, making SPP fuse constraint supervise information efficiently. The experimental results in UMIST, YALE and AR face datasets show, in contrast to unsupervised sparsity preserving projections, our algorithm achieves approximately 5%~15% increase in recognition accuracy based on the nearest neighbor classifier and promotes efficiently the performance of dimensionality reduction classification.

Key words: dimensionality reduction; sparse reconstruction; pairwise constraint; Sparsity Preserving Projection (SPP)

0 引言

由于良好的分类性能和较强的鲁棒性能,稀疏学习是最近几年机器学习领域的研究热点。当前稀疏学习的研究主要体现在两方面:一是基于稀疏基向量的子空间发现,典型的算法有稀疏主成分分析^[1]、稀疏非负矩阵分解^[2]和非负稀疏主成分分析^[3];二是基于稀疏重构权值保持的降维研究。文献[4-5]是此类典型的两个稀疏重构权值保持的降维算法。文献[4]提出稀疏保持投影(Sparsity Preserving Projection, SPP)旨在降维过程中保持样本之间的稀疏重构关系,是全局性的稀疏重构保持降维。文献[5]在近邻保持嵌入的算法的基础上,提出稀疏近邻保持嵌入(Sparse Neighborhood Preserving Embedding, SNPE)。SNPE利用稀疏重构替代近邻的线性重构,达到保持局部的近邻的稀疏重构关系。然而,监督信息的不足导致稀疏保持重构容易受到样本数据的整体结构的影响,为此,人们结合各种监督算法提出了半监督的稀疏保持投影降维算法。文献[6]在SNPE的基础上,引入类内的判别监督信息,提出判别稀疏近邻保持嵌入(Discriminant Sparse Neighborhood Preserving Embedding, DSNPE)。文献[7]在典型相关分析的基础上引入稀疏保持,提出一种稀疏保持典型相关分析方法,实现了两组特征集鉴别信息的有效融合,同时对提取特征间的稀疏重构性加以约束,增强了特征的表达和鉴别能力。文献[8]提出基于局部稀疏重构的特征描述与全

局判别分析的半监督降维算法。文献[9]结合稀疏表示和半监督判决分析对人脸表情进行识别。文献[10]提出通过平衡参数融合局部 Fisher 监督信息和稀疏重构信息的稀疏局部 Fisher 判别分析。然而这些算法普遍忽视稀疏重构自身过程中的监督信息的指导作用。

近年来,基于成对约束的监督学习得到了广泛的研究并作为监督信息进行特征选择^[11]。成对约束可以分为同类样本之间的正约束和不同类样本之间的负约束。与其他标注信息相比,成对约束信息具有更好的分类性能和简易获取的特性。

受到成对约束的启发,本文提出一种成对约束指导的稀疏保持投影(Pairwise Constraints-guided Sparsity Preserving Projections, PCSPP)算法。该算法在训练样本数据的稀疏重构的过程中,结合成对约束的监督信息,通过正约束和负约束设计惩罚项对稀疏重构权值进行调整。实验证明,与基本的稀疏重构算法相比,经过约束监督信息指导的稀疏保持投影后的数据具有更好的判别分类性能。

1 相关理论

1.1 稀疏表示

稀疏表示最初应用于信号领域。在过去几年里,稀疏表示广泛应用于信号处理、数理统计和模式识别。研究^[12]表明,稀疏表示对于残缺和强干扰的数据具有良好的鲁棒性能。

收稿日期:2012-08-06;修回日期:2012-09-06。

基金项目:国家自然科学基金资助项目(71171148);浙江省教育厅科研项目(Y201122544);绍兴文理学院元培学院科研项目(090610)。

作者简介:齐鸣鸣(1974-),男,江西景德镇人,讲师,博士研究生,主要研究方向:机器学习、图像处理。

假定样本 $X = \{x_1, x_2, x_3, \dots, x_n\} \in \mathbf{R}^{d \times n}$, 稀疏表示的目的尽可能用少量的 X 的样本来表示 $x_i \in X$, 具体的数学定义描述如下:

$$\min_{s_i} \|s_i\|_0 \quad (1)$$

$$\text{s.t. } x_i = Xs_i$$

其中: $\|s_i\|_0$ 表示 s_i 中非零的个数, s_{ij} 表示样本 x_j 对重构 x_i 的重构系数。为了降低式(1)计算复杂度,文献[11]提出了用 l_1 范式取代 l_0 近似的解决方法。即:

$$\min_{s_i} \|s_i\|_1 \quad (2)$$

$$\text{s.t. } x_i = Xs_i$$

但在很多实际数据挖掘的应用问题中,受到噪声等外部环境的影响,式(2)并不能总是成立。文献[11]提出了两个具体的解决方案: $\|x_i - Xs_i\| < \varepsilon$, ε 表示误差范围;将 X 替换为 $[X, I]$ 。其中 I 表示 n 阶单位矩阵。

1.2 稀疏重构

假设训练样本在式(2)的基础上,设置约束条件,得

$$\min_{s_i} \|s_i\|_1 \quad (3)$$

$$\text{s.t. } x_i = Xs_i; \mathbf{1} = \mathbf{1}^T s_i$$

其中: $\|s_i\|_1$ 表示 s_i 的 l_1 范数, $S_i = [s_{i1}, \dots, s_{i(i-1)}, 0, s_{i(i+1)}, \dots, s_{in}]^T \in \mathbf{R}^n$ 和 $\mathbf{1} \in \mathbf{R}^n$ 表示全 1 向量。 s_{ij} 表示样本 x_j 对重构 x_i 的重构系数,即:

$$x_i = s_{i1}x_1 + \dots + s_{i(i-1)}x_{i-1} + s_{i(i+1)}x_{i+1} + \dots + s_{in}x_n \quad (4)$$

对每一个 $x_i \in X$, 计算相应的 s_i , 就可以得到训练样本的稀疏重构矩阵 $S = [S_1, S_2, \dots, S_n]^T$ 。

1.3 稀疏保持投影

稀疏保持投影(SPP)的目的是在降维嵌入数据空间中保持输入数据的重构关系。假设 T 是稀疏保持投影矩阵, $T^T Xs_i$ 是高维数据空间的 X_i 的稀疏重构的投影点,可以得到:

$$\min_T \sum_{i=1}^n \|T^T X_i - T^T Xs_i\| = \min_T \|T^T X(I - S - S^T + S^T S)X^T T\| \quad (5)$$

为了避免产生退化解,设置约束条件 $YY^T = T^T XX^T T = I$, 得到 SPP 的目标函数^[4]:

$$\min_T \frac{T^T X(I - S - S^T + S^T S)X^T T}{T^T XX^T T} \quad (6)$$

将式(6)可等价转换为:

$$\max_T \frac{T^T X(S + S^T - S^T S)X^T T}{T^T XX^T T} \quad (7)$$

2 成对约束指导的稀疏保持投影

2.1 基本思想

输入样本数据的稀疏重构矩阵的构建是 SPP 的关键步骤。基本的稀疏重构算法是基于逼近理论的近似计算方法,是一种无监督优化的计算模式,监督信息的缺乏导致 SPP 容易受到样本数据整体结构影响。为了验证 SPP 的局限性,本文构造了一个二维样本数据集。图 1 给出了一个 SPP 在该数据集以及发生变化的数据集上的实验效果。

图 1 中有两类不同形状表示的二维数据。图中的粗体虚线表示经过 SPP 投影后的一维数据空间描述。与图 1(a) 的样本数据相比,图 1(b) 的样本数据发生二倍的纵向拉长,导致样本的整体结构发生变化,但图 1(b) 中的 SPP 产生的投影后的一维投影数据空间描述没有图 1(a) 中的投影的分类效果好。

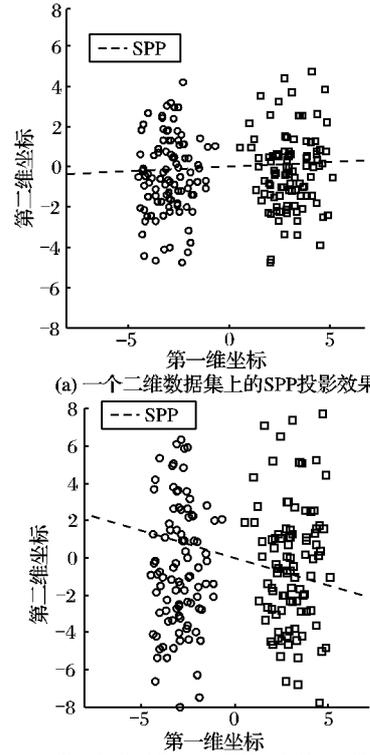


图 1 SPP 在一个二维数据集以及变化的数据集上的实验效果

2.2 目标函数

在稀疏重构的过程中,引入监督信息指导稀疏重构具有重要的意义。假设训练样本 $X = \{x_1, x_2, x_3, \dots, x_n\} \in \mathbf{R}^{d \times n}$, 存在正约束集合 $ML = \{x_i, x_j \mid x_i \text{ 和 } x_j \text{ 属于同一类}\}$ 和负约束集合 $CL = \{x_i, x_j \mid x_i \text{ 和 } x_j \text{ 不属于同一类}\}$ 。根据约束监督信息,结合式(3)和式(4),本文重新定义稀疏重构矩阵 S 。

文献[13]提出利用加权调节参数对成对约束监督信息进行建模,并取得不错的应用效果。受到文献[13]的启发,本文重新定义每一个 $x_i \in X$ 的 S_i , 对于样本 x_j 对重构 x_i 的重构系数 S_{ij} , 利用加权调节参数引入约束监督信息,具体描述如下:

$$S'_{ij} = \begin{cases} S_{ij} - \frac{\beta \times n_c}{n_M + n_c}, & (x_i, x_j) \in CL \\ S_{ij} + \frac{\alpha \times n_M}{n_M + n_c}, & (x_i, x_j) \in ML \\ S_{ij}, & \text{其他} \end{cases} \quad (8)$$

其中: n_M 表示正约束集合 ML 的个数; n_C 表示负约束集合 CL 的个数; S'_{ij} 表示调整后样本 x_j 对重构 x_i 的重构系数; α 和 β 表示调节参数。 α 和 β 的值不是越大越好或者越小越好,通常根据实际应用经验进行确定。

式(8)可以理解为:如果两个样本属于同类,则它们之间的重构系数越高越能体现它们的亲密关系。如果两个样本不属于同类,则它们之间的重构系数越低越能体现它们的互斥关系。

根据式(8),用经过调整后的稀疏重构矩阵 S' 替代 S , 得到目标函数:

$$\max_T \frac{T^T X(S' + S'^T - S'^T S')X^T T}{T^T XX^T T} \quad (9)$$

2.3 算法步骤

输入 人脸训练样本数据 $X = \{x_i \mid x_i \in \mathbf{R}^d\}_{i=1}^n$ 。

输出 投影矩阵 $T \in \mathbf{R}^{d \times l} (l \leq d)$ 。

算法步骤如下:

- 1) 利用式(3)构造稀疏重构矩阵 S 。
- 2) 根据式(8),利用正约束和负约束的监督信息调整构造稀疏重构矩阵 S' 。
- 3) 根据式(9),转化为广义矩阵的求解 $X(S' + S'^T - S'^T S')X^T t_i = \lambda_i X X^T t_i (1 \leq i \leq l)$,求得投影矩阵 $T = [t_1, t_2, \dots, t_l]$ 。

3 实验和分析

3.1 实验数据

本文选择 UMIST 人脸库、YALE 人脸库和 AR 人脸库作为实验数据。它们的具体描述如下:

1) UMIST 人脸库由 20 组人脸图像,共 564 幅人脸图像组成。每幅人脸图像具有不同角度和不同姿态。图 2 给出了一组的 UMIST 人脸图像。



图 2 一组 UMIST 人脸图像

2) YALE 人脸库由 15 个人,共 165 幅图像组成。在每个人的 11 幅图像中,1 幅图像戴有眼镜,3 幅图像光照分别是正面光、右侧光和左侧光,7 幅图像面部表情是不同的。图 3 给出了一组 YALE 人脸图像。



图 3 一组 YALE 人脸图像

3) AR 人脸库由 126 组人脸图像组成。每个人包含间隔为 2 周拍摄的 26 幅图像。这些人脸包含了不同的表情、光照和遮掩。图 4 给出了一组 AR 人脸图像。



图 4 一组 AR 人脸图像

3.2 实验设置

本文首先从每组人脸图像中随机抽取 L 幅图像作为训练样本,剩下的图像作为测试样本。此外,在训练样本中以随机方式构建大小为 PC 的成对约束集。

为了消除样本引发的奇异问题,训练样本将通过投影矩阵投影到 PCA^[14]子空间中,以最近邻分类算法作为识别算法并求得相应的降维后的识别准确率。选择无降维的 Baseline 和 SPP 作为对比算法。各算法的参数具体设置如下所示:Baseline 和 SPP 无需特殊设置;PCSPP 需设为 $\alpha = 10, \beta = 30$ 。

3.3 实验结果与分析

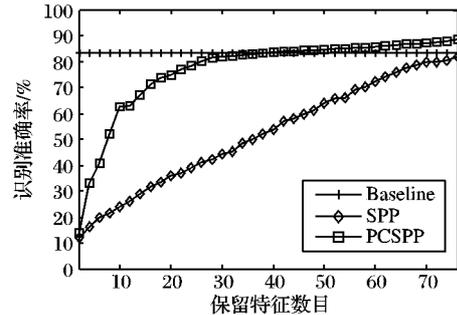
本文分别在不同数目的保留特征和不同大小的成对约束集的条件下对 PCSPP 进行实验和分析。

3.3.1 保留特征维数对性能的影响

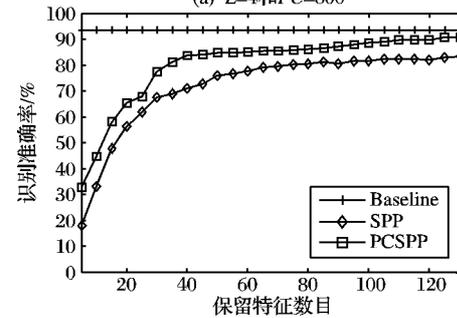
为了准确评价算法的降维分类性能,本文按一定步长取保留特征数目并分别计算相应的识别准确率。所有实验重复 30 遍并取平均值。图 5~7 给出了不同 L 和 PC 的保留特征数对应的平均识别准确率。

由图 5~7,可以得出以下结论:

1) 随着保留特征数目的增加,SPP 和 PCSPP 的平均识别正确率不断增加。当单组人脸训练样本数目和保留特征维数达到一定值时,SPP 和 PCSPP 的分类性能好于 Baseline。

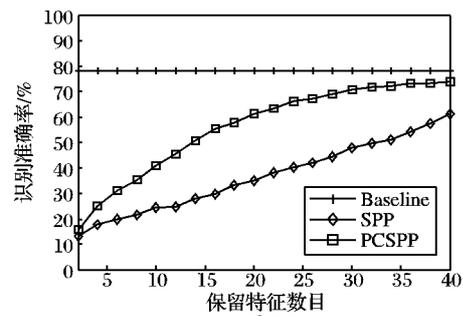


(a) $L=4$ 和 $PC=800$

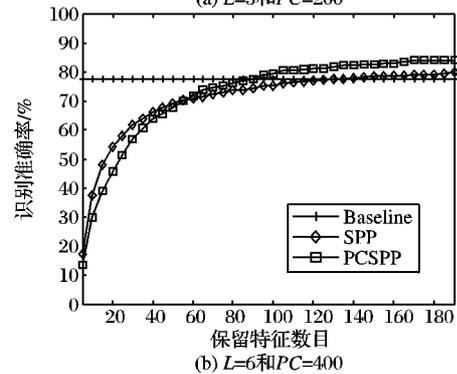


(a) $L=8$ 和 $PC=1600$

图 5 3 种算法在 UMIST 上的实验结果



(a) $L=3$ 和 $PC=200$



(b) $L=6$ 和 $PC=400$

图 6 3 种算法在 YALE 的实验结果

2) 与 SPP 算法相比,不同特点的人脸库数据集的 PCSPP 的识别准确率得到了提高,说明 PCSPP 的成对约束监督信息对稀疏重构的指导是有效的。

3) 与 UMIST 和 YALE 相比,尽管 AR 上的约束集大,但约束集大小与样本量的比值最小,PCSPP 较 SPP 的优势不是很明显,说明 PCSPP 中的监督信息量取决于约束集规模与样本规模的比值。

3.3.2 成对约束集大小对性能的影响

为了考察成对约束集对 PCSPP 的影响,实验根据不同人脸数据集和不同 L ,随机构建不同大小的约束集并计算相应的最高识别准确率。所有实验重复 30 遍并取平均结果值。表 1~3 给出了不同人脸数据集上的不同 L 和不同大小的 PC 及对应的识别准确率。

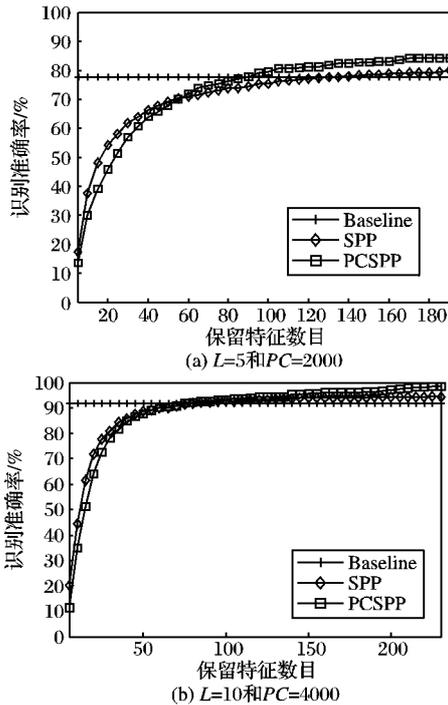


图 7 3 种算法在 AR 的实验结果

表 1 不同 L 时 UMIST 上 PC 对算法性能的影响

L	PC	识别准确率/%
4	400	81.50
	800	85.43
	1200	86.25
	1600	86.50
8	500	85.07
	1000	87.35
	1500	89.23
	2000	90.73

表 2 不同 L 时 YALE 上 PC 对算法性能的影响

L	PC	识别准确率/%
3	100	73.87
	200	74.87
	300	75.87
	400	76.16
6	200	81.06
	400	82.13
	600	83.73
	800	84.80

表 3 不同 L 时 AR 上 PC 对算法性能的影响

L	PC	识别准确率/%
5	1000	82.87
	2000	83.19
	3000	84.15
	4000	85.65
10	2000	97.08
	4000	97.19
	8000	98.27
	10000	98.85

从表 1~3 可以看出:

1) 随着约束集合的增大, PCSPP 的识别准确率呈递增变化, 说明约束信息越多, 式(8)中的稀疏重构权值的调整越有效。

2) 在 UMIST 和 YALE 上, 全部训练样本少, 尽管约束集增量小, 但 PCSPP 的性能对约束集的大小比较敏感, 性能变化大。但在 AR 上, 虽然约束集增量小, 但 PCSPP 的性能变化小, 说明 PCSPP 对约束集的绝对值大小变化不敏感, 而是对约束集大小与训练样本的大小的相对比值变化比较敏感。

4 结语

由于稀疏学习良好的可分类特性和较强的鲁棒性, 稀疏保持投影是一个有效的无监督降维方法。但和其他无监督降维方法类似, 缺乏监督信息指导的稀疏保持投影容易受到训练样本的整体结构影响。本文提出的成对约束指导的稀疏保持投影对稀疏重构矩阵进行了调整, 利用约束监督信息完成稀疏重构过程中的重构系数的调整, 解决了稀疏重构过程中监督信息不足的问题。在 UMIST、YALE 和 AR 人脸数据上的实验表明, 该算法具有更好的降维分类性能。然而, 该算法需要一定大小的成对约束集才能有效地指导稀疏重构。尽管成对约束的获取相对简洁, 但仍然耗费相当一定的代价。因此如何引入其他新型监督方式指导稀疏重构是下一步的研究工作。

参考文献:

- [1] ZOU H, HASTIE T, TIBSHIRANI R. Sparse principal component analysis[J]. *Journal of Computational and Graphical Statistics*, 2006, 15(6): 265-286.
- [2] HOYER P O. Non-negative matrix factorization with sparseness constraints[J]. *Journal of Machine Learning Research*, 2004, 5(11): 1457-1469.
- [3] ZASS R, SHASHUA A. Nonnegative sparse PCA[C]// *Proceedings of Neural Information Processing Systems*. Cambridge, MA: The MIT Press, 2006: 1561-1568.
- [4] QIAO LISHAN, CHEN SONGCAN, TAN XIAOYANG. Sparsity preserving projections with applications to face recognition[J]. *Pattern Recognition*, 2010, 43(1): 331-341.
- [5] CHENG BIN, YANG JIANCHAO, YAN SHUICHENG, *et al.* Learning with $l(1)$ -graph for image analysis[J]. *IEEE Transactions on Image Processing*, 2010, 19(4): 858-866.
- [6] GUI JIE, SUN ZHENAN, JIA WEI, *et al.* Discriminant sparse neighborhood preserving embedding for face recognition[J]. *Pattern Recognition*, 2012, 45(2): 2884-2893.
- [7] 侯书东, 孙权森. 稀疏保持典型相关分析及在特征融合中的应用[J]. *自动化学报*, 2012, 38(4): 599-665.
- [8] 黄勇. 稀疏判决分析在表情识别中的应用[J]. *计算机工程与应用*, 2012, 48(7): 172-174.
- [9] 许淑华, 齐鸣鸣. 稀疏局部 Fisher 判别分析[J]. *计算机工程与应用*, 2012, 48(4): 173-175.
- [10] ZHANG DAOQIANG, CHEN SONGCAN, ZHOU ZHIHUA. Constraint Score: A new filter method for feature selection with pairwise constraints[J]. *Pattern Recognition*, 2008, 41(5): 1440-1451.
- [11] WRIGHT J, YANG A, SASTRY S, *et al.* Robust face recognition via sparse representation[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2009, 31(2): 210-227.
- [12] DONOHO D. For most large underdetermined systems of linear equations the minimal l^1 -norm solution is also the sparsest solution[J]. *Communications on Pure and Applied Mathematics*, 2006, 59(6): 797-829.
- [13] 朱凤梅, 张道强. 张量图像上的半监督降维算法[J]. *模式识别与人工智能*, 2009, 22(4): 574-580.
- [14] TURK M, PENTLAND A. Eigenfaces for recognition[J]. *Journal of Cognitive Neuroscience*, 1991, 3(1): 71-86.