

# 融合稀疏保持的成对约束投影

齐鸣鸣<sup>1,2</sup> 向 阳<sup>1</sup>

(同济大学计算机科学与技术系 上海 201804)<sup>1</sup> (绍兴文理学院元培学院 绍兴 312000)<sup>2</sup>

**摘 要** 提出一种融合稀疏保持的成对约束投影(Pairwise Constraint Projections inoculating Sparsity Preserving, SPPCP)。该算法在成对约束指导的降维过程中,通过平衡参数引入稀疏保持投影(Sparsity Preserving Projections, SPP),在保持成对约束特征的同时,也继承了稀疏保持所蕴含的几何结构保持和近邻保持特性。在 UCI 数据集和 AR 人脸库上的实验表明,该算法有效地融合了稀疏保持投影的优点,与典型的成对约束的半监督降维算法相比,提高了基于最短欧氏距离的分类算法的精度和稳定性。

**关键词** 成对约束,稀疏保持,半监督降维

**中图法分类号** TP391 **文献标识码** A

## Pairwise Constraint Projections Inoculating Sparsity Preserving

QI Ming-ming<sup>1,2</sup> XIANG Yang<sup>1</sup>

(Department of Computer Science and Technology, Tongji University, Shanghai 201804, China)<sup>1</sup>

(College of Yuanpei, Shaoxing University, Shaoxing 312000, China)<sup>2</sup>

**Abstract** A kind of algorithm called Pairwise Constraint Projections inoculating Sparsity Preserving (SPPCP) was proposed, which introduces Sparsity Preserving Projections (SPP) with trade-off parameter in the process of pairwise constraint-guided dimensional reduction, preserving pairwise constraint feature and heriting the special character of geometrical structure preserving and neighbor preserving from sparsity preserving. Experiments operated on UCI databases and AR face database show that the algorithm inoculates merits of SPP effectively. Compared with the typical semi-supervised dimensional reduction algorithms based on pairwise constraint, the algorithm can improve the accuracy and stability of classified algorithms based on the shortest Euclidean distance.

**Keywords** Pairwise constraints, Sparsity preserving, Semi-supervised dimensional reduction

充分挖掘标注样本的监督分类信息和未标注样本所蕴涵的无监督分类信息是半监督学习的重要研究内容。通常监督标注方式可以分为样本类别标注和样本之间的成对约束标注。成对约束是指属于同一类的样本间的正约束和不属于同一类的样本间的负约束。与类别标注相比,成对约束标注可以不需要指定具体的样本类别,能通过比较判定两者是否属于同类来完成,甚至可以自动获取<sup>[1]</sup>。然而当成对约束标注信息过少和数据维数过高时,成对约束指导的特征投影(Pairwise Constraints-guided Feature Projections, PCFP)<sup>[2]</sup>的分类识别容易陷入过度拟合现象。因此,成对约束的半监督降维是当前机器学习的研究热点。

目前,主要相关研究成果有:

1) 文献[3]提出约束 Fisher 判别分析(CFDA)。该算法先通过正约束对数据进行聚类处理得到若干个簇,然后构建簇内散布矩阵和总体散布矩阵的比并将其最大化。但 CFDA 忽视负约束的作用。

2) Zhang 等人提出约束的半监督降维框架(SSDR)<sup>[4]</sup>。

该框架在降维过程中除了保持数据约束信息,还通过加权方式引入数据的方差来保持数据内部结构信息。但散布矩阵和数据的方差不能有效地发现数据的非线性结构。

3) Cevikalp 等人提出约束的局部保持投影(CLPP)<sup>[5]</sup>。该算法利用成对约束信息设置邻接权值矩阵并对约束信息进行传播。但该算法的性能比较依赖于数据邻接信息的建立。

4) 邻域保持半监督降维(NPSSDR)<sup>[6]</sup>在保持成对约束信息的同时,通过添加正则化项保持数据的局部结构信息。然而基于局部线性嵌入的正则化策略会导致 NPSSDR 陷入局部结构塌陷问题。文献[7]利用 NPSSDR 思想和二分搜索算法指导度量学习。但其计算复杂度较大且可能不收敛。

最近提出的稀疏保持投影(Sparsity Preserving Projections, SPP)<sup>[8]</sup>是一种新型无监督降维算法。SPP 的目的是将高维空间的稀疏重构关系保持到低维嵌入空间。稀疏重构具有良好的非线性结构保持和近邻保持特性,尤其对残缺、化妆以及遮挡等人脸图像表现出很好的鲁棒性<sup>[9]</sup>。研究表明<sup>[8]</sup>,与 PCA<sup>[10]</sup>、LPP<sup>[11]</sup>和 NPE<sup>[12]</sup>相比,SPP 在降维过程中能较

到稿日期:2011-09-06 返修日期:2012-01-21 本文受浙江省教育厅科研项目(Y201122544),省新世纪教改项目(YB2010092),国家自然科学基金项目(70771077)资助。

齐鸣鸣(1974-),男,博士生,讲师,主要研究方向为机器学习、生物特征鉴别,E-mail:webqmm1974@sohu.com;向阳(1962-),男,博士后,教授,博士生导师,主要研究方向为数据挖掘、Web 语义。

好地保持高维空间数据的非线性结构和局部近邻。然而缺乏监督信息的 SPP 的降维分类性能易受整体模式差异的影响<sup>[8]</sup>。

受到上述分析的启发,本文提出一种融合稀疏保持的成对约束投影(Pairwise Constraint Projections insculating Sparsity Preserving, SPPCP)。该算法在降维过程中保持高维空间的数据成对约束特征关系,同时通过平衡参数融合稀疏重构关系,保持非线性结构和局部近邻关系,有效地继承了 SPP 的特性。本文在若干 UCI 数据集和 AR 人脸库上进行实验分析,验证了该算法的有效性。

本文第 1 节介绍成对约束指导的特征投影算法;第 2 节介绍稀疏保持投影算法;第 3 节提出融合稀疏保持的成对约束投影算法;第 4 节给出实验和结果分析;最后为结论。

## 1 成对约束指导的特征投影(PCFP)

假定标注样本  $X = \{x_1, x_2, x_3, \dots, x_n\}$ , 存在相关的正约束集合  $ML = \{(x_i, x_j) | x_i \text{ 和 } x_j \text{ 属于同一类}\}$  和负约束集合  $CL = \{(x_i, x_j) | x_i \text{ 和 } x_j \text{ 不属于同一类}\}$ 。成对约束指导的特征投影的目标函数<sup>[2]</sup>:

$$\max_T \left( \sum_{(x_i, x_j) \in CL} \| (T^T x_i - T^T x_j) \|^2 - \sum_{(x_i, x_j) \in ML} \| (T^T x_i - T^T x_j) \|^2 \right) \quad (1)$$

式(1)可以认为:正约束集合的高维空间内两个样本在降维后的空间内应该保持近的距离,负约束集合的高维空间内两个样本在降维后的空间内应该保持远的距离。

## 2 稀疏保持投影(SPP)

### 2.1 稀疏表示和稀疏重构

给定训练样本集  $X = \{x_1, x_2, x_3, \dots, x_n\}$ , 每个样本  $x_i \in R^{d \times n}$ 。以  $X$  作为一个过完备字典,对于任何一个样本  $x_j \in R^{d \times 1}$ , 都可以找出分解最为稀疏的一组原子  $s_i$  并满足:

$$\min_{s_i} \| s_i \|_0 \quad (2)$$

s. t.  $x_j = X s_i$

式中,  $\| s_i \|_0$  表示  $s_i$  的  $\ell_0$  范数,代表  $s_i$  中非零的个数。

稀疏重构就是使用尽量少的训练样本来重构每一个测试样本。为了降低计算复杂度,以  $\| s_i \|_1$  取代  $\| s_i \|_0$ 。求得近似解。函数目标<sup>[9]</sup>为:

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

s. t.  $x_j = X s_i$   
 $1 = 1^T s_i$

式中,  $\| s_i \|_1$  表示  $s_i$  的  $\ell_1$  范数,  $s_i = [s_{i1}, \dots, s_{i\bar{a}-1}, 0, s_{i\bar{a}+1}, \dots, s_{im}]^T \in R^m$  和  $1 \in R^m$  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:

$$x_i = s_{i1} x_1 + \dots + s_{i\bar{a}-1} x_{\bar{a}-1} + s_{i\bar{a}+1} x_{\bar{a}+1} + \dots + s_{im} x_m \quad (4)$$

### 2.2 稀疏保持投影

稀疏重构权值反映了内在的几何性质且包含了判别信息<sup>[9]</sup>。稀疏保持投影就是将高维空间的稀疏重构特性保持到低位嵌入空间。

假设  $T$  是稀疏保持投影矩阵,  $T^T x_{s_i}$  是高维空间  $x_i$  的稀疏重构的投影点,则稀疏保持投影的目标函数为<sup>[8]</sup>:

$$\min_T \sum_{i=1}^m \| T^T x_i - T^T x_{s_i} \|^2 \quad (5)$$

将式(5)进行转换计算,得:

$$\begin{aligned} & \sum_{i=1}^m \| T^T x_i - T^T x_{s_i} \|^2 \\ &= T^T \left( \sum_{i=1}^m (X e_i - X s_i) (X e_i - X s_i)^T \right) T \\ &= T^T X \left( \sum_{i=1}^m (e_i - s_i) (e_i - s_i)^T \right) X^T T \\ &= T^T X \left( \sum_{i=1}^m e_i e_i^T - s_i s_i^T - e_i s_i^T + s_i s_i^T \right) X^T T \\ &= T^T X (I - S - S^T + S^T S) X^T T \end{aligned} \quad (6)$$

根据式(6),目标函数式(5)可以转换为:

$$\min_T T^T X (I - S - S^T + S^T S) X^T T \quad (7)$$

## 3 融合稀疏保持的成对约束投影(SPPCP)

### 3.1 基本思想

为了能够确切地描述 SPP 和 PCFP 的问题,生成一个二维二类数据集。图 1 给出了 SPP 和 PCFP 在该数据集以及改变的数据集上的一维投影。

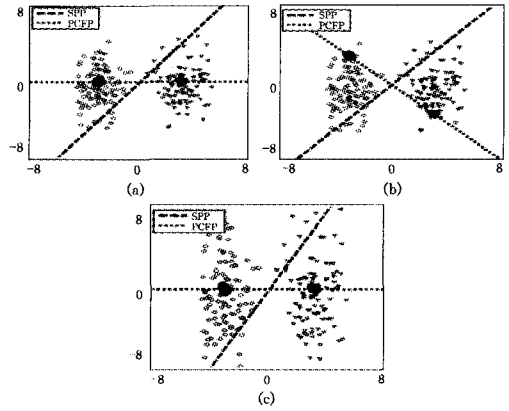


图 1 SPP 和 PCFP 在二维二类的数据集的一维投影

在图 1 中,圆点代表第一类,三角点代表第二类。每类数据样本都有相同数量的实心样本构成的成对约束数据集。从图(a)可以看到,SPP 和 PCFP 都具有较好的一维投影效果。在图(b)中,改变成对约束集的标注样本,SPP 的一维投影没有变化。约束监督信息的变化导致 PCFP 的一维投影发生变化,降维投影分类效果下降。通过放大图(a)的垂直标量改变训练样本的分布形态结构得到图(c),PCFP 的一维投影没有改变,仍然具有良好的分类效果。但训练样本结构的变化降低了 SPP 的投影分类效果。因此,监督降维 PCFP 容易因构成成对约束集的样本的变化而产生投影分类效果的不稳定,但不受训练样本的整体结构的影响。训练样本分布形态结构的改变容易导致无监督降维 SPP 的分类效果的不确定,因此,融合它们各自的优势从而克服相应的缺陷是一个有效的办法。

### 3.2 目标函数

1)根据式(1),得:

$$\begin{aligned} & \sum_{(x_i, x_j) \in CL} \| (T^T x_i - T^T x_j) \|^2 - \sum_{(x_i, x_j) \in ML} \| (T^T x_i - T^T x_j) \|^2 \\ &= \sum_{(x_i, x_j) \in CL} \sum_l T_l^T (x_i - x_j) (x_i - x_j)^T T_l - \sum_{(x_i, x_j) \in ML} \sum_l T_l^T (x_i - x_j) (x_i - x_j)^T T_l \\ &= \sum_l T_l^T \left( \sum_{(x_i, x_j) \in CL} (x_i - x_j) (x_i - x_j)^T - \sum_{(x_i, x_j) \in ML} (x_i - x_j) (x_i - x_j)^T \right) T_l \end{aligned}$$

$$= T^T P_\alpha T \quad (8)$$

其中

$$P_\alpha = \sum_{(x_i, x_j) \in C_L} (x_i - x_j)(x_i - x_j)^T - \sum_{(x_i, x_j) \in M_L} (x_i - x_j)(x_i - x_j)^T \quad (9)$$

引入  $T^T I T = I$ , 得:

$$\max_T \frac{T^T P_\alpha T}{T^T I T} \quad (10)$$

2) 引入约束条件  $T^T X X^T T = I$ , 转换 SPP 的目标函数式(7), 得:

$$\min_T \frac{T^T X(I - S - S^T + S^T S) X^T T}{T^T X X^T T} \quad (11)$$

将式(11)等价转换为:

$$\max_T \frac{T^T X S_\alpha X^T T}{T^T X X^T T} \quad (12)$$

其中

$$S_\alpha = S + S^T - S^T S \quad (13)$$

3) 根据式(10)和式(12), 引入平衡参数  $\beta (0 \leq \beta \leq 1)$  得到 SPPCP 的目标函数, 即:

$$\max_T \frac{T^T (\beta X S_\alpha X^T + (1 - \beta) P_\alpha) T}{T^T (\beta X X^T + (1 - \beta) I) T} \quad (14)$$

式中,  $\beta$  代表 SPP 在 SPPCP 的贡献度,  $1 - \beta$  代表 PCFP 在 SPPCP 的贡献度。如果  $\beta = 0$ , SPPCP 则变为 PCFP。如果  $\beta = 1$ , SPPCP 则变为 SPP。如果  $0 < \beta < 1$ , SPPCP 则部分继承 SPP 和 PCFP 的降维分类性能。在实际应用中, SPPCP 可以根据具体情况选择不同参数  $\beta$  来增加应用的灵活性和适应性。

### 3.3 算法步骤

1) 输入: 训练样本  $X = \{x_1, x_2, x_3, \dots, x_n\}$ , 每个样本  $x_i \in R^{d \times n}$ , 平衡参数  $\beta (0 \leq \beta \leq 1)$ 。

2) 根据式(9)构建  $P_\alpha$ 。

3) 利用式(3)获得稀疏重构系数矩阵  $S$ 。

4) 利用式(13)获得  $S_\alpha$ 。

5) 将式(14)转换为广义矩阵求解问题

$$(\beta X S_\alpha X^T + (1 - \beta) P_\alpha) \varphi = \lambda (\beta X X^T + (1 - \beta) I) \varphi$$

得到投影矩阵  $T^{d \times r} (r < d)$ 。

6) 输出: 投影矩阵  $T^{d \times r} (r < d)$ 。

## 4 实验

表 1 各种算法的具体参数设置

算法名称	参数设置
SPP	无
PCFP	无
SPPCP(0.9)	$\beta = 0.9$
SPPCP(0.1)	$\beta = 0.1$
CLPP	$\kappa = 5$
SSDR	$\alpha = 1, \beta = 20$
NPSSDR	$\alpha = 0.1, \kappa = 5$

本文选择若干个 UCI 数据集<sup>[13]</sup>以及 AR 人脸库<sup>[14]</sup>作为实验数据对象。除了选择 SPP 和 PCFP 作为对比算法, 本文还选择成对约束的典型半监督降维 SSSDR、CLPP 和 NPSSDR。本文采用最短欧氏距离的分类算法。为了验证算法的融合性和降维分类性能, 在实验数据集上进行 SPP、PCFP、SPPCP、CLPP、SSDR 和 NPSSDR 验证, 通过逐步增样本数目

或保留特征维数目, 对比分析 SPPCP 的融合性能和降维分类性能。表 1 给出各种算法具体参数的设置。

本实验是在 Windows7 和 MatlabR2010a 下完成的, 硬件环境是内存 2G, CPU 酷睿 E8400(3.0GHz)。

### 4.1 UCI 数据集上的实验

#### 4.1.1 UCI 数据集

本文从 UCI 数据集选出 Iris、Letter、Vehicle、Waveform\_V1、Soybean\_Large 和 Wine 6 个数据集。它们的描述如表 2 所列。

表 2 用于测试的 UCI 数据集的统计描述

数据集	样本维数	样本数量	类别数
Iris	4	150	3
Letter	16	20000	26
Vehicle	18	946	4
Waveform_V1	21	5000	3
Soybean_Large	35	307	19
Wine	13	178	3

#### 4.1.2 实验设置

从 6 个不同 UCI 数据集随机抽取一定数量的训练标注样本。测试样本为剩下未标注样本或特定的全部测试样本。在训练标注样本中, 构建成对约束集。每次实验使用数量相同但内容不同的训练样本、测试样本和成对约束集, 逐步增样本数目并重复 50 次取分类错误率的平均值。

#### 4.1.3 结果与分析

为了确切地对比不同半监督算法的性能, 表 3—表 8 列出不同训练样本数目下的不同降维算法的分类错误率。表中的粗体表示在训练样本数量下的最小分类错误率。

表 3 各种算法在 Iris 的分类错误率(%) vs. 样本数目

算法	样本数目					
	30	40	50	70	80	90
SPP	12.38	11.23	11.05	8.69	9.00	9.10
PCFP	5.00	5.55	4.85	4.19	3.58	4.10
SPPCP(0.9)	5.50	5.27	5.15	4.56	4.36	4.10
SPPCP(0.1)	<b>4.37</b>	<b>4.28</b>	<b>4.65</b>	4.13	<b>3.50</b>	<b>3.40</b>
CLPP	6.00	4.36	5.15	4.50	3.93	3.60
SSDR	5.88	4.43	4.80	3.94	3.64	4.10
NPSSDR	5.88	5.27	4.75	<b>3.88</b>	3.79	3.67

表 4 各种算法在 Letter 上的分类错误率(%) vs. 样本数目

算法	样本数目					
	20	40	80	200	400	800
SPP	81.60	69.38	63.93	48.53	35.20	25.80
PCFP	81.90	73.45	63.68	48.30	34.93	25.53
SPPCP(0.9)	83.35	72.05	62.20	47.30	33.65	22.93
SPPCP(0.1)	<b>80.30</b>	<b>68.35</b>	<b>60.93</b>	<b>45.93</b>	<b>31.40</b>	<b>21.78</b>
CLPP	83.73	72.60	64.38	48.20	34.28	25.30
SSDR	81.90	73.45	63.68	48.33	34.93	25.53
NPSSDR	84.43	71.87	62.11	47.39	33.23	24.59

表 5 各种算法在 Vehicle 上的分类错误率(%) vs. 样本数目

算法	样本数目					
	20	40	80	100	200	400
SPP	57.08	46.35	37.05	37.68	33.80	29.90
PCFP	52.63	49.40	40.80	39.58	38.13	34.00
SPPCP(0.9)	65.43	49.15	45.93	38.13	36.73	34.50
SPPCP(0.1)	50.53	<b>37.78</b>	<b>28.70</b>	<b>29.48</b>	<b>28.60</b>	<b>24.90</b>
CLPP	<b>48.98</b>	41.05	30.58	30.05	29.30	27.60
SSDR	52.63	49.40	40.80	39.58	38.13	34.00
NPSSDR	58.85	43.08	31.48	30.47	29.07	26.90

表6 各种算法在 Waveform\_V1 上的分类错误率(%)vs. 样本数目

算法 \ 样本数目	40	80	100	200	400	800
SPP	58.80	52.15	43.68	27.78	24.98	24.98
PCFP	30.40	26.58	26.80	25.85	25.55	24.90
SPPCP(0.9)	34.53	34.0	32.43	28.18	27.03	25.95
SPPCP(0.1)	<b>28.50</b>	<b>25.78</b>	<b>25.03</b>	<b>24.75</b>	<b>24.78</b>	<b>22.18</b>
CLPP	45.35	36.08	37.35	33.93	32.10	29.63
SSDR	30.40	26.58	26.80	25.85	25.55	24.90
NPSSDR	47.87	39.29	35.61	33.05	30.33	26.73

表7 各种算法在 Soybean\_Large 上的分类错误率(%)vs. 样本数目

算法 \ 样本数目	20	40	60	80	120	160
SPP	49.33	36.00	27.67	20.67	18.00	11.00
PCFP	42.00	26.33	25.67	15.33	17.33	13.33
SPPCP(0.9)	81.00	72.0	50.67	50.00	34.00	22.67
SPPCP(0.1)	<b>40.00</b>	<b>24.33</b>	<b>22.67</b>	<b>15.00</b>	<b>11.67</b>	<b>10.33</b>
CLPP	46.67	38.00	23.33	19.00	19.00	12.08
SSDR	42.67	25.33	26.00	18.33	19.67	17.33
NPSSDR	83.53	42.73	25.40	19.37	14.60	11.60

表8 各种算法在 Wine 上的分类错误率(%)vs. 样本数目

算法 \ 样本数目	20	40	60	70	80	90
SPP	33.20	19.02	19.83	19.31	20.26	15.50
PCFP	30.60	23.28	23.94	23.19	21.28	19.66
SPPCP(0.9)	24.46	23.9	25.21	21.06	19.74	18.69
SPPCP(0.1)	<b>12.03</b>	<b>7.33</b>	<b>6.57</b>	<b>6.30</b>	<b>6.12</b>	<b>5.63</b>
CLPP	22.44	16.39	15.72	15.05	13.42	12.78
SSDR	30.60	23.28	23.94	23.19	21.28	19.66
NPSSDR	19.68	9.13	7.37	6.62	6.17	6.02

可以得出结论:1) SPPCP(0.1)的分类错误率普遍低于SPP和PCFP,而SPPCP(0.9)不一定。2)与CLPP、SSDR和NPSSDR相比,SPPCP(0.1)具有更低的分类错误率,SPPCP(0.9)则不能确定。说明将较小的平衡参数 $\beta$ 引入SPP的分类信息能极大提升SPPCP的降维分类性能。

表9  $L=2$  下的各种算法在 AR 上的分类错误率(%)vs. 特征数目

算法 \ 特征数目	20	80	140	200	260	320	380	440	500	560	620	680	740	800
SPP	71.10	55.72	50.74	47.86	46.10	45.33	44.80	44.36	43.63	43.28	43.29	43.14	43.03	43.01
PCFP	57.91	50.96	49.23	48.69	48.44	48.44	48.44	48.44	48.44	48.44	48.44	48.44	48.44	48.44
SPPCP(0.9)	75.22	67.49	48.44	43.61	40.70	40.70	40.70	40.70	40.70	40.70	40.70	40.70	40.70	40.70
SPPCP(0.1)	66.65	43.33	<b>39.85</b>	<b>38.69</b>	<b>38.10</b>	<b>38.10</b>	<b>38.10</b>	<b>38.10</b>	<b>38.10</b>	<b>38.10</b>	<b>38.10</b>	<b>38.10</b>	<b>38.10</b>	<b>38.10</b>
CLPP	50.91	43.57	41.75	41.35	41.19	41.19	41.19	41.19	41.19	41.19	50.91	43.07	41.75	41.35
SSDR	64.36	54.11	50.65	48.84	48.44	48.44	48.44	48.44	48.44	48.44	48.44	48.44	48.44	48.44
NPSSDR	<b>48.67</b>	<b>41.67</b>	41.67	41.67	41.67	41.67	41.67	41.67	41.67	41.67	40.56	40.80	40.95	40.80

表10  $L=4$  下的各种算法在 AR 上的分类错误率(%)vs. 特征数目

算法 \ 特征数目	20	80	140	200	260	320	380	440	500	560	620	680	740	800
SPP	56.88	30.67	23.41	21.54	18.91	18.56	17.19	17.04	16.81	17.00	17.00	16.35	15.70	15.58
PCFP	<b>37.89</b>	27.85	25.32	24.45	23.99	23.95	23.80	23.80	23.83	23.83	23.83	23.83	23.83	23.83
SPPCP(0.9)	81.93	62.15	51.64	35.03	23.11	21.24	19.67	18.37	17.46	17.46	17.46	17.46	17.46	17.46
SPPCP(0.1)	47.56	<b>20.09</b>	<b>16.0</b>	<b>15.16</b>	<b>14.63</b>	<b>13.67</b>	<b>13.75</b>	<b>13.48</b>	<b>13.33</b>	<b>13.33</b>	<b>13.33</b>	<b>13.33</b>	<b>13.33</b>	<b>13.33</b>
CLPP	50.53	40.30	38.66	37.99	37.76	37.62	37.55	37.55	37.53	37.53	50.53	40.30	38.66	37.99
SSDR	48.20	30.98	27.92	26.20	25.29	24.68	24.22	23.87	23.83	23.83	23.83	23.83	23.83	23.83
NPSSDR	40.55	28.34	22.45	21.54	21.16	20.60	20.40	20.36	20.20	20.20	20.20	20.20	20.20	20.20

**结束语** 针对现有的成对约束的半监督降维的问题,本文在保持成对约束指导的特征投影(PCFP)基础上,提出一种融合稀疏保持的成对约束投影(SPPCP)。通过在若干个UCI数据集和AR人脸库上的实验验证表明,SPPCP较好地融合了SPP的非线性结构保持和近邻保持特性。同时与现有典

## 4.2 AR人脸库上的实验

### 4.2.1 AR人脸库

AR人脸库包含4000多幅人脸图像,共126人组,每个人组共有26幅人脸图像,分别在2个时间阶段进行采集,2个时间阶段的间隔时间为2周。AR人脸图像的最大特点是在不同光照、不同表情和不同遮挡条件下拍摄的。图2给出AR人脸库中一组人脸图样本。



图2 AR的一组人脸图样本

### 4.2.2 实验设置

将AR人脸库中的人脸图像的尺寸缩为 $30 \times 30$ ,从每个人脸库中随机抽取一定数量的图像作为训练样本集,其余的人脸图像作为测试集。在训练样本集中构造成对约束集。

为了考察SPPCP在AR人脸库的降维分类性能,从每个人脸组抽取 $L(L=2,4)$ 幅图像作为训练集,逐步递增保留特征数目并计算分类错误率。所有实验重复20次,取分类错误率的平均值。

### 4.2.3 实验结果与分析

表9、表10给出不同训练样本数目下的各种算法在AR人脸数据集上的不同保留特征数目对应的分类错误率。表中的粗体表示保留特征数目下的最小分类错误率。

分析可以得出结论:1)SPPCP(0.1)的分类错误率明显小于SPP和PCFP。2)SPPCP(0.1)和SPPCP(0.9)的分类错误率低于CLPP、SSDR和NPSSDR,但SPPCP(0.1)更为明显。说明在高维人脸图像数据集上,SPPCP继承了SPP在图像识别上的良好分类特性。

型的成对约束的半监督降维算法相比,SPPCP具有更好的基于最短欧氏距离的分类算法的准确率和稳定性能。然而,平衡参数和成对约束集对SPPCP的降维分类性能具有重要作用,如何设置平衡参数和选取成对约束集是下一步的研究工作。

(下转第242页)

$T$  独立的变量,构建一个关于感兴趣变量的局部模型;然后,基于非对称信息熵从局部模型中选择一个对其它结点影响最大的结点进行扰动;最后联合观测数据和扰动数据,在参数满足 Dirichlet 分布的条件下,计算边缘似然函数,进而计算边的后验概率,从而获得一个关于目标结点的局部因果网络。实验结果表明:CSI-LCSL 方法能够解决学习全局的贝叶斯网络 NP 难题问题,同时能够有效地学习到感兴趣变量的局部因果结构。

### 参考文献

- [1] Cooper G, Yoo C. Causal discovery from a mixture of data[C]// Proceeding of the 15th Annual Conference on UAI, San Francisco, CA; Morgan Kaufmann Publishers Inc. ,1999;116-125
- [2] Tsamardinos I, Aliferis C F, Statnikov A. Time and sample efficient discovery of Markov blankets and direct causal relations [C]//Proceedings of the 9th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. Washington; ACM Press,2003;673-678
- [3] Tsamardinos I, Aliferis C F, Statnikov A R. Algorithms for Large Scale Markov Blanket Discovery[C]//Proceedings of the Sixteenth International Florida Artificial Intelligence Research Society Conference. Florida; AAAI Press,2003;376-380
- [4] Aliferis C F, Tsamardinos I, Statnikov A R. HITON: A novel Markov blanket algorithm for optimal variable selection[C]// American Medical Informatics Association Annual Symposium. 2003;21-25
- [5] Eberhardt F, Glymour C, Scheines R. On the number of experiments sufficient and in the worst case necessary to identify all causal relations among  $N$  variables[C]//Proceedings of the 21th Annual Conference on UAI US; AUAI Press,2005;178-184
- [6] Korb K B, Hope L, Nicholson A E, et al. Varieties of causal in-

tervention[C]// Proceedings of the Pacific Rim International Conference on AI. Berlin; Springer,2004;322-331

- [7] Eaton D, Murphy K. Exact Bayesian structure learning from uncertain interventions[C]// Proceeding of the 12th on Artificial Intelligence and Statistics. JMLR Press,2007;107-114
- [8] Good I J. A causal calculus[J]. The British Journal for the Philosophy Science,1961,11;305-318
- [9] Cheng P W. From covariation to causation; A causal power theory[J]. Psychological Review,1997,104(2);367-405
- [10] Cover T M, Thomas J A. Elements of Information Theory[M]. Wiley,1991
- [11] Li Guo-liang, Leong T-Y. Active Learning for Causal Bayesian Network Structure with Non-symmetrical Entropy[C]// Proceedings of the 13th Pacific-Asia Conference on Advances in Knowledge Discovery and Data Mining. Berlin; Springer-Verlag, 2009;290-301
- [12] Pearl J. Probabilistic Reasoning in Intelligent Systems[M]. San Francisco, CA; Morgan Kaufmann Publishers Inc. ,1988;383-408
- [13] Aliferis C, Statnikov A, et al. Local Causal and Markov Blanket Induction for Causal Discovery and Feature Selection for Classification, Part I; Algorithms and Empirical Evaluation[J]. Journal of Machine Learning Research,2010,11;171-234
- [14] Tian J, Pearl J. Causal discovery from changes[C]// Proceeding of the 17th conference on UAI. San Francisco, CA; Morgan Kaufmann Publishers Inc. ,2001;512-521
- [15] Tong S, Koller D. Active Learning for Structure in Bayesian Networks[C]// IJCAI. Washington; Morgan Kaufmann, 2001; 863-869
- [16] www. ai. nit. edu/murphyk/Software/BNT/bnt. html
- [17] Eaton D, Murphy K. BDAGL: Bayesian DAG learning [EB/OL]. www. cs. ubc. ca/~murphyk /Software/BDAGL/,2007

(上接第 215 页)

### 参考文献

- [1] Yan R, Zhang J, Yang J, et al. A discriminative learning framework with pairwise constraints for video object classification[C]// Proceedings of the IEEE computer society conference on computer vision and pattern recognition. Washinton, USA, 2004, 2; 284-291
- [2] Tang W, Zhong S. Pairwise constraints-guided dimensionality reduction[C]// Proceedings of the Workshop on Feature Selection for DataMining (SDM 2006). Bethesda, USA, 2006; 295-311
- [3] Bar-Hillel A, Hertz T, Shental N, et al. Learning a mahalanobis metric from equivalence constraints [J]. Journal of Machine Learning Research,2006,6(6);937-965
- [4] Zhang D, Zhou Z, Chen S. Semi-Supervised dimensionality reduction[C]// Proceedings of the SDM 2007, Minneapolis, USA, 2007;629-634
- [5] Cevikalp H, Verbeek J, Jurie F, et al. Semi-Supervised dimensionality reduction using pairwise equivalence constraints[C]// Proceedings of the VISAPP 2008. Funchal,2008;489-496
- [6] Wei J, Peng H. Neighbourhood preserving based semi-supervised dimensionality reduction[J]. Electronics Letters,2008,44(20):

1190-1191

- [7] Baghshah M S, Shouraki S B. Semi-Supervised metric learning using pairwise constraints[C]// Proceedings of the IJCAI 2009. San Francisco, USA, 2009;1217-1222
- [8] Qiao L, Chen S, Tan X. Sparsity preserving projections with applications to face recognition[J]. Pattern Recognition,2010,43(1);331-341
- [9] 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
- [10] Turk M, Pentland A. Eigenfaces for recognition[J]. Journal of Cognitive Neuroscience,1991,3(1);71-86
- [11] Min W, Lu K, He X. Locality preserving projection[J]. Pattern Recognition,2004,37(4);781-788
- [12] He X, Cai D, Yan S, et al. Neighborhood preserving embedding [C]// Proceedings in International Conference on Computer Vision (ICCV). Beijing, China, 2005;1208-1213
- [13] Blake C, Keogh E, Merz C J. UCI repository of machine learning databases[Z]. Department of Information and Computer Science, University of California, Irvine,1998
- [14] Martinez A M, Benavente R. The AR Face Database[R]. CVC Technical Report 24, June 1998