

基于协同表示标签传播的半监督正交鉴别分析算法

杨俊川 蒋 同 张国庆

(南京信息工程大学 计算机与软件学院,江苏 南京 210044)

摘 要 近年来由于其表示的有效性,基于表示的分类方法,例如稀疏表示分类器(SRC)或者协同表示分类器(CRC)被广泛的应用于各种各样的识别任务.但是, SRC 或者 CRC 的性能严重依赖于每类带标签训练样本的个数,当每类带标签的训练样本不够充分, SRC 或者 CRC 的识别性能将会显著地下降.为了解决这个问题,文章[34]把协同表示技术引入到半监督学习方法中,提出了一种基于协同表示的标签传播算法,有效的利用了大量未标记样本的信息来进行标签传播.受此启发,把这种标签传播算法与正交鉴别分析算法相结合,提出了一种基于协同表示标签传播的半监督正交鉴别分析算法,目的是为了学习一个更好的鉴别子空间.不同于传统的半监督降维方法,所提算法首先利用这种标签传播算法将部分有标签数据的标签信息传递给不带标签的数据,之后利用传播后获得的全体软标签信息构造离散度矩阵实现鉴别分析,其次对鉴别投影施加正交约束,采用一种更加有效快速的迹比优化算法进行鉴别分析.大量的实验结果验证了所提算法的有效性.尤其在只存有少量标签样本的情况下,算法仍能保持良好的分类性能.

关键词 协同表示;标签传播;半监督;鉴别分析

中图分类号 TP391.4

文献标识码 A

0 引言

在机器学习、计算机视觉以及模式识别领域中,许多实际应用,比如基于外观的人脸识别任务,往往数据的维数非常高而样本非常少,这将不可避免地引发“维数灾难”问题^[1].会导致许多实际处理技术失效并且降低其分类性能.针对这一问题,降维(或者特征抽取)是一种行之有效的解决方法.降维或者特征提取是人脸识别问题中最基本和最终的问题之一,而提取模式中的有效特征是提高识别率的一种有效途径.目前特征提取方法已经广泛的用于降低计算方法的复杂度和提高目标算法的分类效果.特征提取的目的就是为了寻求一组线性或者非线性变换,在尽量保持原始数据蕴含的有效信息下,将处于高维空间中的原始数据压缩到一个低维子空间中,以便对数据做进一步地处理.在过去的几十年中,研究者们提出许多基于子空间的特征提取方法,其中最受欢迎且被广泛应用的特征提取方法包括主成分分析(PCA)^[2]和线性鉴别分析(LDA)^[3]方法.根据训练数据是否带有标签可将降维方法分为监督学习方法和非监督学习方法.监督学习方法通常用于数据分类^[4-6],无监督方法则多用于聚类,这类方法通过特定的准则来指导其降维过程.主成分分析(PCA)是一种典型无监督降维方法,其目的是寻求一个投影方法使得投影后数据的方差最大.与 PCA 不同, LDA 是一种监督降维方法,其目的通过最大化类间散度矩阵的同时最小化类内散度矩阵来获得最优鉴别子空间.

近年来,随着生物学、心理学揭示的生物视觉系统中神经元响应的稀疏特性,稀疏表示技术在计算机视觉与模式识别领域得到了广泛的应用^[7-12].其基本思想是将样本数据表示为超完备集(或者字典)中基元素的线性组合,在这些组合中找到对样本数据最稀疏的表示. Wright 等人将稀疏表示应用于人脸识别提出了一种稀疏表示分类器(Sparse Representation based Classification, SRC)^[7].稀疏表示主要解决了人脸识别中

收稿日期:2018-11-26

基金项目:国家自然科学基金青年基金项目(61806099)资助

通讯作者:张国庆,男,汉族,博士,讲师,研究方向:模式识别,机器学习, E-mail:xiayang14551@163.com.

的特征抽取问题,另外使用稀疏表示进行人脸识别可以一定程度解决遮挡干扰对识别的影响.正是由于稀疏表示对噪声数据展现了其独特的鲁棒性,研究者们提出了一些基于稀疏表示的特征提取方法^[13-16]. Qiao 等人提出了一种稀疏保持投影方法(SPP)^[13],其目的是为了寻找一个子空间,使得原始数据的稀疏重构关系得以保持. Zang 和 Zhang 等人标签信息和 ℓ_1 图相结合^[14],提出了一种新的特征提取方法,使得 SRC 可以实现更好的效果. 基于 SRC, Yang 等人提出了一种 SRC 引导的鉴别投影方法(SRC-DP)^[16].

上面提到的所有方法都需要有充足的标签训练数据来实现特征提取. 而在实际应用中采集大量的带标签的训练数据需要耗费大量的人力劳动,代价是昂贵和耗时的. 特别是在人脸识别领域,通常只能获得每个人少量的带有标签的样本. 在这种情况下,上述方法都不能得到令人满意的性能. 另一方面收集大量未标签的数据是容易且廉价的,因此如何有效地利用少量带标签数据和大量无标签数据已经变成一个非常重要的问题. 实际应用中较为常见的情形是数据集中同时存在少量带标签数据和大量不带标签的数据. 如果仅使用带标签数据进行训练,则学习出的模型可能会出现过拟合现象而缺乏泛化能力;如果仅使用不带标签的数据进行训练,则不利于分类. 为了解决这一问题,学者们提出了一些半监督学习(Semi-supervised learning, SSL)算法,即同时利用带标签和不带标签的数据来促进人脸识别性能^[17-22]. 基于图的半监督学习方法因其优雅的数学形式和独特的效能,已发展成为模式识别领域热门研究主题之一^[17,21]. 在数据分析中,图是一种比较有用的工具,基于图的算法也被广泛应用于各种领域,例如流型嵌入,半监督学习,和图像匹配^[23,24]. 现有的大部分降维方法都可以纳入图嵌入的框架中^[21],在这个框架中,半监督降维方法利用带标签和不带标签的数据构建图上的权重矩阵. 但是这些算法仅仅以一种简单的方式利用所有的训练数据(带标签和不带标签)构建图上的权重矩阵,而没有充分利用大量未标签数据的分布信息. 常用的构图方法主要有 k 近邻图, ϵ -近邻图^[23],这两种构图方法方都需要手动设置特定的参数,这些参数对构建的图结构具有很大影响. 另一种半监督学习方法为标签传播(Label propagation, LP)算法,根据带标签样本和未标签样本的分布把标签信息从带标签数据传播到未标签数据^[22,25-27]. 标签传播算法中一个重要的步骤就是在图中定义用于度量不同定点相似性的边权矩阵.

标签传播算法中最流行的权重分配方法就是高斯核相似性^[30],但是估计一个合适的核参数是比较困难的^[25]. 因此,为了在标签传播算法中构建权重矩阵时实现自适应近邻,基于稀疏编码的方法引起了许多学者的关注^[22,25-29]. Chen 等人提出另一种 ℓ_1 图构建方法^[27]. 通过 ℓ_1 范数最小化问题, Zang 等人使用测试样本预先计算的稀疏近邻作为字典计算稀疏编码^[22]. 文献^[28]中提出了一种基于 ℓ_1 图的标签传播算法并用于人脸识别. 尽管这些方法都取得了不错的效果,但是这些方法有两个缺点:首先,每一个数据样本的稀疏编码是单独计算的,在求解过程中没有全局约束,因此数据中的全局结构实际上没有充分的捕获^[31]. 其次,如果没有充分的标签数据,仅使用带标签的数据作为字典会降低最终的识别性能.

除此之外,基于稀疏表示的方法都要求解 ℓ_1 范数优化问题,这是极端耗时的. 近来的研究表明,协同表示(Collaborative Representation)技术在稀疏表示分类器中比稀疏约束起了更重要的作用^[32-34]. 因此,通过使用 ℓ_2 范数正则项取代 ℓ_1 范数正则项, Zhang 等人提出了一种协同表示分类器(CRC)并用于图像识别^[32]. CRC 利用所有训练样本的线性组合对测试样本进行编码,之后使用最小重构残差对测试样本进行分类. 与 SRC 相比, CRC 使用更小的计算成本获得了类似的分类性能. 因此基于协同表示的特征提取方法也收到了研究者的关注^[35,36]. 当带标签训练样本不够充分时,协同表示方法的效果也同样会下降. 因此,在半监督学习中,带标签数据通常比较少,而未标签数据通常是冗余的并且很容易获得,并且未标签数据包含很多没有被充分利用的有用信息. 特别是在人脸识别应用中,由于同一类的样本位于同一个子空间中,因此每一个带标签样本可以被同一类中未标签样本很好的表示. 基于这一假设, Zhang 等人提出了一种基于协同表示的标签传播(LPCR)^[34]算法. LPCR 利用大量未标签样本作为字典重构带标签样本,之后利用协同表示系数传播标签信息,从而为每一个未标签样本得到一个软标签. 如何有效的利用传播后地数据学习一个具有鉴别性的子空间是文章^[34]没有很好的解决的一个问题.

基于以上考虑,本文将 LCPR 与正交鉴别分析相结合,提出了一种基于协同表示标签传播的半监督正交鉴别分析方法. 通过标签传播,将标签信息根据总体样本的分布情况传播给未标签样本,获得未标签样本的软标签. 之后利用所有的样本构造新的类间散度矩阵 S_b 和类内散度矩阵 S_w ,学习一个更加有效和可靠的

鉴别子空间,其本质就是学习一组投影矩阵,使得在投影后的子空间中, \mathbf{S}_w 最小而 \mathbf{S}_b 最大,使识别性能更加准确.除此之外,本文对投影矩阵施加正交约束,使学习到的投影具有正交属性,因此可以采用更加有效的迹比优化算法进行求解,避免了高维小样本情形下,类内散度矩阵 \mathbf{S}_w 奇异的问题.

文章组织结构如下:第一节介绍了基于迹比优化的正交鉴别分析与协同表示分类器,第二节介绍了基于协同表示的半监督正交鉴别分析算法,第三节展示了实验结果,第四节给出结论.

1 基于迹比优化的正交鉴别分析算法 (TR-LDA)

1.1 线性鉴别分析(LDA)

线性鉴别分析(LDA)方法是为了学习一个线性转换矩阵 $\mathbf{W} \in \mathbf{R}^{d \times m}$ ($d < m$),利用线性变换 $\mathbf{y} = \mathbf{W}^T \mathbf{x}$,输入空间中的每一个高维样本 $\mathbf{x} \in \mathbf{R}^d$ 都可以转换为一个低维向量 $\mathbf{y} \in \mathbf{R}^m$.给定一组输入空间的训练样本 $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n] \in \mathbf{R}^{d \times n}$,其中 $\mathbf{X}_i = [\mathbf{x}_{i,1}, \mathbf{x}_{i,2}, \dots, \mathbf{x}_{i,n_i}] \in \mathbf{R}^{d \times n_i}$ 为第 i 类训练样本, n_i 表示第 i 类训练样本的个数, $n = \sum_{i=1}^c n_i$, c 为样本类别数.令 $x_{i,j}$ ($i = 1, \dots, c, j = 1, \dots, n_i$) 表示第 i 类训练数据中的第 j 个样本. LDA 定义类内散度矩阵 \mathbf{S}_w 和类间散度矩阵 \mathbf{S}_b

$$\mathbf{S}_w = \sum_{i=1}^c \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i) (x_{ij} - \bar{x}_i)^T, \quad (1)$$

$$\mathbf{S}_b = \sum_{i=1}^c \frac{n_i}{n} (\bar{x}_i - \bar{x}) (\bar{x}_i - \bar{x})^T, \quad (2)$$

其中 $\bar{x}_i = \sum_{j=1}^{n_i} x_{i,j}$ 为第 i 类样本均值, $\bar{x} = \sum_{i=1}^c x_i$ 为全体样本均值. LDA 的目的是同时最大化类间散度与最小化类内散度,因此 LDA 的目标函数表示

$$J(\mathbf{W}_{LDA}) = \max_{\mathbf{W}} \frac{\text{tr}(\mathbf{W}^T \mathbf{S}_b \mathbf{W})}{\text{tr}(\mathbf{W}^T \mathbf{S}_w \mathbf{W})}. \quad (3)$$

在 LDA 中,通常假设 \mathbf{S}_w 是非奇异的,因此公式(3)可以通过求解广义特征分解问题 $\mathbf{S}_b \mathbf{w} = \lambda \mathbf{S}_w \mathbf{w}$ 获得投影矩阵, \mathbf{W}_{LDA} 由 \mathbf{S}_b 和 \mathbf{S}_w 前 m 个特征值所对应的广义特征向量组成.

1.2 基于迹比优化的正交鉴别分析(TR-LDA)

许多降维算法都可以视为一个迹比最大化或者迹比最小化问题.迹比问题的形式描述如

$$J(\mathbf{W}) = \max_{\mathbf{W}^T \mathbf{C} \mathbf{W} = \mathbf{I}} \frac{\text{tr}(\mathbf{W}^T \mathbf{A} \mathbf{W})}{\text{tr}(\mathbf{W}^T \mathbf{B} \mathbf{W})}, \quad (4)$$

其中 $\mathbf{A}, \mathbf{B} \in \mathbf{R}^{d \times d}$ 为对称矩阵, $\mathbf{C} \in \mathbf{R}^{d \times d}$ 为一个正定矩阵, $\mathbf{W} \in \mathbf{R}^{d \times m}$ 为投影矩阵.通常情况下 \mathbf{C} 设定为单位矩阵,即 $\mathbf{W}^T \mathbf{W} = \mathbf{I}$, \mathbf{I} 为单位矩阵则 \mathbf{W} 满足正交性.

利用正交约束,公式(3)的迹比问题可以写为

$$J(\mathbf{W}_{TR-LDA}) = \max_{\mathbf{W}^T \mathbf{W} = \mathbf{I}} \frac{\text{tr}(\mathbf{W}^T \mathbf{S}_b \mathbf{W})}{\text{tr}(\mathbf{W}^T \mathbf{S}_w \mathbf{W})}, \quad (5)$$

实际上,为了避免过拟合,对 \mathbf{S}_w 添加正则项,优化问题(5)变为

$$J(\mathbf{W}_{TR-LDA}) = \max_{\mathbf{W}^T \mathbf{W} = \mathbf{I}} \frac{\text{tr}(\mathbf{W}^T \mathbf{S}_b \mathbf{W})}{\text{tr}(\mathbf{W}^T (\mathbf{S}_w + \mu \mathbf{I}_d) \mathbf{W})}, \quad (6)$$

其中 \mathbf{I}_d 为一个 $d \times d$ 的单位矩阵, $\mu > 0$ 为正则化参数.目标函数(6)可以通过迹比优化算法求解,这种算法称为基于迹比优化的正交鉴别分析算法^[37].

1.3 协同表示分类器(CRC)

稀疏表示分类器(SRC)在人脸识别中取得了令人印象深刻的效果.但是 SRC 需要解决一个 ℓ_1 范数最小化问题,因此计算成本昂贵.与 SRC 相比 CRC 可以使用较小的计算成本实现相似的结果^[49].

给定一个测试样本 $\mathbf{y} \in \mathbf{R}^{d \times 1}$,和结构字典 $\mathbf{D} = [\mathbf{D}_1, \mathbf{D}_2, \dots, \mathbf{D}_c] \in \mathbf{R}^{d \times k}$,其中 $\mathbf{D}_i \in \mathbf{R}^{d \times k_i}$ 为与第 i 类训练样本相关的子字典, $k = \sum_{i=1}^c k_i$, c 为样本类别数, k_i 为第 i 类字典中原子的个数.测试样本 \mathbf{y} 可以利用字典 \mathbf{D} 对其线性表示,即 $\mathbf{y} = \mathbf{D} \boldsymbol{\alpha}$,其中 $\boldsymbol{\alpha} = [\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \dots, \boldsymbol{\alpha}_c]$, $\boldsymbol{\alpha}_i$ 为与第 i 类相关的表示向量.假设测试样本 \mathbf{y} 属于类 c ,则 $\mathbf{y} = \mathbf{D}_c \boldsymbol{\alpha}_c$ 成立.这意味着向量 $\boldsymbol{\alpha}$ 中大部分元素都接近为 0,并且仅仅 $\boldsymbol{\alpha}_c$ 中有显著的非 0 元素.

编码系数 α 可以通过求解如下优化问题获得

$$\hat{\alpha} = \underset{\alpha}{\operatorname{argmin}} \|\alpha\|_0, \text{ s. t. } y = D\alpha, \quad (7)$$

其中 $\|\cdot\|_0$ 为 ℓ_0 范数, 统计向量中非 0 元素的个数.

求解公式(7)中的 ℓ_0 优化问题是一个 NP 难问题并且极端耗时的. 研究表明, 对于特定的字典如果结果足够稀疏, 求解 ℓ_0 优化问题与求解如下 ℓ_1 优化问题是等价的^[38]

$$\hat{\alpha} = \underset{\alpha}{\operatorname{argmin}} \|\alpha\|_1, \text{ s. t. } y = D\alpha. \quad (8)$$

实际上公式(8)通常转换为如下正则化问题

$$\hat{\alpha} = \underset{\alpha}{\operatorname{argmin}} \|\mathbf{y} - D\alpha\|_2^2 + \lambda \|\alpha\|_1, \quad (9)$$

其中 λ 为正则化参数. 求解公式(8)中的 ℓ_1 范数最小化问题仍然是比较耗时的. 近年来, Zhang 等人指出使用协同表示比 α 的 ℓ_1 稀疏更加关键^[32], 因此提出协同表示分类方法 (*Collaborative representation based classification, CRC*), 表示形式如下:

$$\hat{\alpha} = \underset{\alpha}{\operatorname{argmin}} \|\mathbf{y} - D\alpha\|_2^2 + \lambda \|\alpha\|_2^2, \quad (10)$$

其中 $\|\cdot\|_2$ 记为 ℓ_2 范数, λ 为正则化参数. $\hat{\alpha} \in R^{k \times 1}$ 为测试样本 y 在字典 D 下的协同表示系数. 正则项的作用主要体现在两个方面: 1) 正则项使最小二乘解稳定; 2) 对于解向量 $\hat{\alpha}$, 正则项引入了一定数量的稀疏性, 但是这种稀疏性比 ℓ_1 范数要小的多.

利用正则化最小二乘, 公式(10)中协同表示解很容易求得

$$\hat{\alpha} = (D^T D + \lambda I)^{-1} D^T y, \quad (11)$$

记 $P = (D^T D + \lambda I)^{-1} D^T$, 可以发现 P 与 y 是相互独立的, 因此 P 可以作为一个投影矩阵预先计算. 对于测试样本 y , 可以简单的在 P 上对 y 投影, 即 Py . 这将使协同表示节省很多时间.

分类策略类似于 SRC^[7], 计算每一类的重构误差 $r_i = \|\mathbf{y} - D_i \hat{\alpha}_i\|_2^2 / \|\hat{\alpha}_i\|_2$, $\hat{\alpha}_i$ 为与第 i 类相关的编码系数. 测试样本的类别标签为 $\operatorname{identify}(y) = \arg \min_i \{r_i\}$.

2 基于协同表示标签传播的半监督正交鉴别分析算法

2.1 基于协同表示的标签传播算法

标签传播算法^[39]是基于这样一种假设: 如果两个样本的相似度高, 则它们属于同一类别的概率就越大. 标签传播算法是通过任一节点的各加权边, 将带标签样本的标签信息循环地传递至临近的各个节点. 不同的边有不同的权值, 权值越大越容易传递信息.

在人脸识别场景中, 每一类的人脸图像通常位于同一个子空间中^[32, 34]. 假设有一系列训练样本, 仅有少量训练样本具有标签, 基于子空间假设, 带有标签训练样本可以被同一类的未标签训练样本很好的表示. 因此同一类未标签样本对带标签样本的表示系数应该大于其它类未标签样本对带标签样本进行表示所产生的系数. 因此, 同一类中更多的信息可以通过较大的表示系数从带标签数据传播给未标签数据. 除此之外, 在人脸识别中, 属于不同类的人脸图像具有一些相似性, 即来自于类 A 的一些样本有可能对表示类 B 中的样本有帮助作用^[32]. 因此在人脸识别中利用协同表示技术可能是一个更加合适的选择.

给定一组数据样本 $X = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n, \mathbf{x}_{n+1}, \dots, \mathbf{x}_{n+u}]$, $\{\mathbf{x}_i\}_{i=1}^n$ 与 $\{\mathbf{x}_j\}_{j=n+1}^{n+u}$ 分别为带标签和不带标签数据样本. $\{\mathbf{x}_i\}_{i=1}^n$ 类别标签是已知的, 假定为 $l_i \in \{1, 2, \dots, c\}$. 对于每一个带标签的样本 \mathbf{x}_i , 通过求解如下优化问题对其重构

$$\hat{\alpha}_i = \underset{\alpha_i}{\operatorname{argmin}} \|\mathbf{x}_i - \mathbf{X}_u \alpha_i\|_2^2 + \lambda \|\alpha_i\|_2^2, \quad (12)$$

其中 $\alpha_i \in R^{u \times 1}$ 为样本 \mathbf{x}_i 的表示系数向量, $\mathbf{X}_u = [\mathbf{x}_{n+1}, \dots, \mathbf{x}_{n+u}]$ 为未标签样本矩阵. 图 1 展示了一些带标签图像样本的系数.

向量 α_i 可以表示为 $\alpha_i = [\alpha_{i,1}, \alpha_{i,1}, \dots, \alpha_{i,u}]$, 其中 $\alpha_{i,j}$ 为未标签样本 \mathbf{x}_j 对 \mathbf{x}_i 进行重构所产生的表示系数. 定义 $\mu_{i,j} = |\alpha_{i,j}| / \sum_j |\alpha_{i,j}|$, $1 \leq i \leq n, 1 \leq j \leq u$, 明显的, $\mu_{i,j}$ 越大, 样本 \mathbf{x}_j 对重构 \mathbf{x}_i 所做的贡献越大. 当 $\mu_{i,j} = 1$ 时, 意味着 \mathbf{x}_i 可以完全由 \mathbf{x}_j 重构, 这表明 \mathbf{x}_i 与 \mathbf{x}_j 相同^[58].

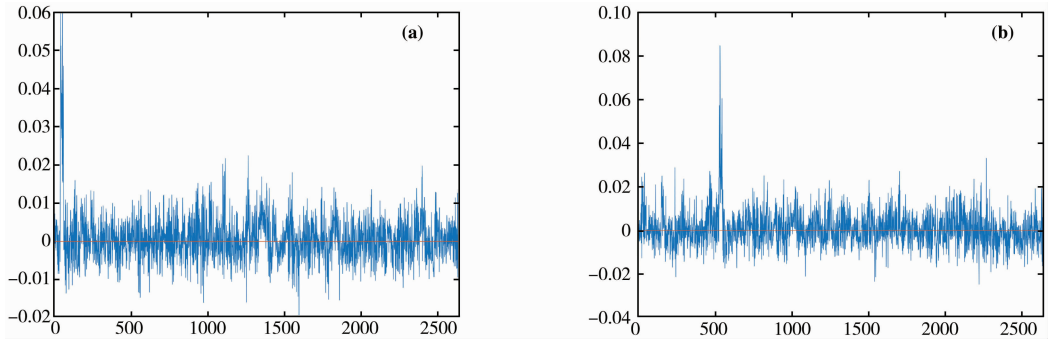


图 1 AR 数据集中不同带标签样本的协同表示系数 (a) 来自于第 1 类, (b) 来自于第 25 类

因为 \mathbf{x}_i 可以通过 \mathbf{x}_j 进行线性重构, 权重为 μ_{ij} , 因此可以通过权重系数对标签进行传播. 定义软标签向量 $\mathbf{f}_{x_j} = [f_{j,1}, f_{j,2}, \dots, f_{j,c}]$ 用于描述 \mathbf{x}_j 的类别信息, 其中 $f_{j,q}, q = 1, \dots, c$ 度量 \mathbf{x}_j 属于第 q 类的概率

$$f_{j,q} = \sum_{i=1}^n \mu_{ij} \sigma(l_{x_i} - q), \quad (13)$$

其中 l_{x_i} 表示样本 \mathbf{x}_i 的标签信息, $\sigma(\tau)$ 为一个离散单元函数, 定义

$$\sigma(\tau) = \begin{cases} 1, & n = 0, \\ 0, & n \neq 1. \end{cases} \quad (14)$$

标签传播之后, 获得每一个未标签样本 $\mathbf{x}_j (j = 1, 2, \dots, u)$ 的软标签, 记为 $\mathbf{f}_j = [f_{j,1}, f_{j,2}, \dots, f_{j,c}]$. 此时获得的标签概率矩阵为 $\mathbf{F}^u = [\mathbf{f}_1; \mathbf{f}_2; \dots; \mathbf{f}_u] \in \mathbf{R}^{u \times c}$, 其中矩阵 \mathbf{F}^u 中的第 j 行表示为 $\mathbf{x}_j (1 \leq j \leq u)$ 的标签向量 \mathbf{f}_j , 且其元素 F_{jq}^u 可视为 \mathbf{x}_j 属于类别 q 的概率.

2.2 基于软标签的离散度矩阵

标签传播算法结束之后, 得到每一个未标签数据样本 $\{\mathbf{x}_j\}_{j=n+1}^{n+u}$ 的软标签, 即 \mathbf{x}_j 属于类别 $q (q = 1, 2, \dots, c)$ 的概率为 F_{jq}^u . 在线性鉴别分析(LDA)中, 散度矩阵 \mathbf{S}_w 和 \mathbf{S}_b 是根据每个样本 $\{\mathbf{x}_i\}_{i=1}^n$ 的硬标签定义的, 即 \mathbf{x}_i 属于类别 q 的概率非 0 即 1. 这里, 可以在 LDA 中定义的离散度矩阵推广至基于软标签的离散度矩阵. 定义带标签数据 $\{\mathbf{x}_i\}_{i=1}^{n+u}$ 的概率矩阵为 $\mathbf{F}^l = [\mathbf{f}_1; \mathbf{f}_2; \dots; \mathbf{f}_n] \in \mathbf{R}^{n \times c}$, 其中 F_{iq}^l 表示 \mathbf{x}_i 属于类别 q 的概率, 非 0 即 1. 总的概率矩阵为 $\mathbf{F} = [\mathbf{F}^l; \mathbf{F}^u] \in \mathbf{R}^{(n+u) \times c}$, 基于软标签的类内离散度和类间离散度定义如下

$$\tilde{\mathbf{S}}_w = \frac{1}{n} \sum_{i=1}^c \sum_{j=1}^{n+u} F_{ji} (\mathbf{x}_j - \tilde{\mathbf{x}}_i) (\mathbf{x}_j - \tilde{\mathbf{x}}_i)^T = \frac{1}{n} \mathbf{X} (\mathbf{B} - \mathbf{F} \mathbf{D} \mathbf{F}^T) \mathbf{X}^T, \quad (15)$$

$$\tilde{\mathbf{S}}_b = \sum_{i=1}^c \frac{\tilde{n}_i}{n} (\tilde{\mathbf{x}}_i - \tilde{\mathbf{x}}) (\tilde{\mathbf{x}}_i - \tilde{\mathbf{x}})^T = \frac{1}{n} \mathbf{X} (\mathbf{F} \mathbf{D} \mathbf{F}^T - \frac{1}{n} \mathbf{B} \mathbf{I} \mathbf{I}^T \mathbf{B}) \mathbf{X}^T, \quad (16)$$

其中 $\tilde{n}_i = \sum_{j=1}^{n+u} F_{ji}$, $\tilde{n} = \sum_{i=1}^c \tilde{n}_i$, $\tilde{\mathbf{x}}_i = \sum_{j=1}^{n+u} F_{ji} \mathbf{x}_j / \tilde{n}_i$, $\tilde{\mathbf{x}} = \sum_{i=1}^c \sum_{j=1}^{n+u} F_{ji} \mathbf{x}_j / \tilde{n}$, $\mathbf{B} \in \mathbf{R}^{(n+u) \times (n+u)}$ 为一个对角矩阵, 对角元素 $B_{ii} = \sum_{j=1}^c F_{ji}$, $\mathbf{D} \in \mathbf{R}^{c \times c}$ 为对角矩阵, 对角元素 $D_{ii} = 1 / \sum_{j=1}^{n+u} F_{ji}$, $\mathbf{1} = [1, 1, \dots, 1]^T \in \mathbf{R}^{(n+u) \times 1}$, $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n, \mathbf{x}_{n+1}, \dots, \mathbf{x}_{n+u}] \in \mathbf{R}^{d \times (n+u)}$.

2.3 目标函数

构造完基于软标签概率矩阵的散度矩阵之后, 应用于监督线性鉴别分析(LDA)得到半监督线性鉴别分析算法. 为了避免过拟合, 在 $\tilde{\mathbf{S}}_w$ 中添加正则项. 本文关注基于迹比优化的正交鉴别分析, 因此对投影矩阵施加正交约束. CR-SODA 的目标函数就是求解如下优化问题

$$J(\mathbf{W}_{\text{CR-SODA}}) = \max_{\mathbf{W}^T \mathbf{W} = \mathbf{I}} \frac{\text{tr}(\mathbf{W}^T \tilde{\mathbf{S}}_b \mathbf{W})}{\text{tr}(\mathbf{W}^T (\tilde{\mathbf{S}}_w + \mu \mathbf{I}_d) \mathbf{W})}, \quad (17)$$

其中 \mathbf{I} 为 $d \times d$ 的单位矩阵, $\mu > 0$ 为正则化参数. 近来, 一种有效的算法^[37]被提出用于求解正交约束的迹比优化问题, CR-SODA 的优化过程如下所示

输入 训练样本 $\{\mathbf{x}_i\}_{i=1}^{n+u}$, 投影维数 m , 正则化参数 μ .

输出 投影矩阵 $\mathbf{W} \in \mathbf{R}^{d \times m}$.

初始化 初始化 \mathbf{W} 为一任意正交矩阵.

步骤 1 计算 $\tilde{\mathbf{S}}_w$, $\tilde{\mathbf{S}}_b$, 设置 $\rho := \text{tr}(\mathbf{W}^T \tilde{\mathbf{S}}_b \mathbf{W}) / \text{tr}(\mathbf{W}^T (\tilde{\mathbf{S}}_w + \mu \mathbf{I}_d) \mathbf{W})$

步骤 2 获取 $\tilde{\mathbf{S}}_b - \rho(\tilde{\mathbf{S}}_w + \mu \mathbf{I}_d)$ 前 m 个最大特征值所对应的特征向量 $\mathbf{W}^* = [\omega_1, \omega_2, \dots, \omega_m]$.

步骤 3 迭代执行步骤 1-2,直到算法收敛.

3 实验结果与分析

本章节,我们在 5 个标准人脸库上做了一系列的实验来验证所提算法(CR-SODA)的有效性,并与一些具有代表性的特征提取方法进行比较:包括主成分分析(PCA)^[2],线性鉴别法分析(LDA)^[3],稀疏保持投影(SPP)^[13],半监督鉴别分析(SDA)^[18],半监督稀疏保持鉴别分析(SPDA)^[41]. PCA 用于降维训练,对所有训练样本行单位 ℓ_2 范数归一化.

3.1 数据集描述

ORL^[42]人脸数据库包含 40 个人,每个人 10 幅图像,对于一些人的图像,有些拍摄于不同的时期,不同的光照条件;人脸脸部表情与脸部细节有变化,例如笑或不笑,睁眼或闭眼,戴眼镜与不戴眼镜.每一幅人脸图像手动裁剪为大小 32×32 像素的图像,如图 1(a)所示.

Yale 人脸数据库,该数据库有 15 个人,每个人 11 幅图像.这 11 幅图像取自不同的拍摄条件,如在正常光照条件下,不同光源,是否戴眼镜,不同表情下的图像.每一幅人脸图像手动裁剪为大小 64×64 像素的图像,其中一个人的人脸图像如图 1(b)所示.

CMU PIE^[43]人脸数据库由 68 个人的 41368 幅人脸图像组成,这些人脸图像在不同的拍摄环境下得到,比如不同的动作,光照和表情.每个人有 13 幅不同的动作图像,43 幅不同的光照图像,4 幅不同的表情图像.在本实验中选择正面姿态(C27)在不同光照下的 43 幅图像类进行测试.每一幅人脸图像手动裁剪为大小为 64×64 像素的图像,图 1(c)给出了从选择数据中裁剪后的人脸图像样本.

Extended Yale B^[44]数据库包含 38 个人的 2414 幅正面人脸图像.每一个人的大约 64 幅图像都是在实验可控制的光照条件下采集的.在我们的实验中,每一幅图像被手动裁剪为 32×32 像素大小的图像.如图 1(d)所示.



(a) ORL 人脸库的样本图像



(b) Yale 人脸库的样本图像



(c) CMU PIE 人脸库的样本图像



(d) Extended Yale B 人脸库的样本图像



(e) AR 人脸库的样本图像

图 2 不同人脸库上的样本图像

AR^[45]数据库包括 126 个人的 4000 幅图像,对于每个人的 26 幅图像拍摄于两个不同的场景,每个场景包含 13 幅图像,这些图像包含在不同表情,不同光照和有无遮挡下的正面人脸图像.在我们的实验中,仅使用没有遮挡的人脸图像.这个子集中包含 120 个人的 1680 幅图像,每个人有来自两个不同场景的 14 幅图像,前七幅和后七幅分别在不同的时间拍摄.这里,为了计算方便,每一幅图像裁剪成大小为 50×45 像素的图像,如图 2(e)所示

3.2 单一训练样本的人脸识别问题

人脸识别中最成功并且被充分研究的技术之一就是基于外貌的方法^[2].两种最重要的基于外貌的方法包括特征脸(PCA)^[2]和 Fisher 脸(LDA)^[3].通常情况下,人脸外貌并不能唯一确定人的身份.它也受光照和主观评价的影响.姿态和光照的变化都会引起人脸外貌的变化,因此,为了对应姿态和光照的变化,基于外貌的方法需要每类中都有一些训练图像.

在人脸识别中经典的挑战之一来自于单一人脸图像的识别.在这种情况下,原始的基于外貌的方法(例如特征脸和 Fisher 脸)都不可用.实际上,如果每一类中只有一个训练样本,则容易验证类间散度矩阵等同于总的类内散度矩阵.因此,LDA 方法不适用.为了表明本文提出算法的有效性,我们在每类只有一个训练样本的情况下做了一些实验,在我们的实验中使用一些比较流行的人脸数据库,包括 CMU PIE^[43],Extended Yale B^[44]和 AR^[45]人脸数据库.

对于每一个人脸数据库,在不同的训练样本集中随机选择一幅图像作为带标签的样本,其余的为不带标签的样本.对于所有的实验,实验随机产生 20 组不同的训练样本集,并统一采用最近邻分类器进行识别,实验结果取这 20 次实验的平均值.基于每一次划分,我们把本文算法 CR-SODA 与非监督的 PCA^[2]、LPP^[46]、SPP^[13],监督 LDA^[3]和半监督的 SDA^[18]、SPDA^[41]等算法进行比较.对于 LDA,由于每类只有一个训练样本,所以 LDA 方法不适用此次实验.实验结果如表 1 所示.

表 1 单一训练样本的性能比较

Database	PCA	LDA	SPP	SDA	SPDA	CR-SODA
PIE/%	29.44	-	43.98	41.15	45.56	66.57
Yale B/%	23.88	-	31.41	27.98	31.71	50.45
AR/%	25.26	-	38.33	34.25	40.04	65.24

从上述实验结果可以看出,在每类只有一个训练样本的情况下,所有方法都实现了比较低的识别率,但是我们的方法显著地优于其他所有的方法.尽管 SPP 是非监督的方法,但是识别性能优于 SDA,这也是因为稀疏表示具有自然的鉴别能力,在所有的实验中半监督的 SPDA 和我们的方法优于 SPP.这表明未标签的信息起了比较大的作用.

3.3 参数设置

在我们的算法中包含两个参数 λ 和 μ ,本文中我们设置 $\lambda=0.01$, $\mu=0.001$.PCA、LDA 和 SPP 参数自由,无参数需要设定.而对于半监督算法 SDA,近邻图中的权重通过高斯核 $K(x,y) = \exp\{-\|x-y\|^2/\delta^2\}$ 来计算,核参数 σ 取为所有样本均值的 $1/2$.参数 $\lambda_1=0.01$, $\lambda_2=0.1$,和文献[18]中的参数一样,所有含有局部信息的方法,设置近邻个数 $k=8$,SPDA 的参数设置与文献[41]中相同.

3.4 多标签人脸识别问题

3.4.1 ORL 人脸库.在每类人脸中随机选择 6 幅图像作为训练样本,其余的 4 幅作为测试样本.在选定的训练样本数目的情况下,实验随机产生 20 组不同的训练样本集.实验结果取这 20 次实验的平均值.另外,在实验过程中,每类训练样本随机选择 $p=(2,3,4,5)$ 个训练样本作为带标签样本,其余的为无标签样本.我们比较 PCA、LDA、SDA、SPP、SPDA 和本文方法的分类识别效果,20 次实验的平均结果如表 2 所示.图 3 显示了不同算法在不同带标签样本下的平均识别率.在本次实验中可以看出 SDA 结果最差,LDA 优于 PCA,SPP.我们的算法实现了最高识别率分别为 93.62%、96.26%、98.58%、99.09%.SPDA 获得了第二高的识别率.

表 2 不同方法在不同标签样本数下的实验结果

Dataset	Method	2label	3label	4label	5label
ORL	PCA/%	82.05	87.84	88.96	90.00
	LDA/%	79.17	88.23	92.44	93.98
	SPP/%	68.84	79.43	89.56	91.58
	SDA/%	74.00	78.88	83.56	87.16
	SPDA/%	88.56	91.84	92.56	94.37
	CR-SODA/%	93.62	96.26	98.58	99.09
Yale	Method	2label	3label	4label	5label
	PCA/%	82.44	89.13	89.86	89.28
	LDA/%	85.70	89.29	90.62	92.89
	SPP/%	89.04	92.13	89.43	91.61
	SDA/%	63.25	68.25	74.25	77.50
	SPDA/%	89.25	93.33	95.67	95.92
CR-SODA/%	93.72	96.38	97.75	98.08	
CMU PIE	Method	4label	7label	10label	13label
	PCA/%	51.56	65.84	72.87	77.55
	LDA/%	80.78	88.43	95.44	98.68
	SPP/%	89.82	92.39	94.98	95.94
	SDA/%	60.15	70.88	81.95	87.65
	SPDA/%	92.52	94.66	95.96	98.14
CR-SODA/%	96.69	98.03	98.68	99.16	

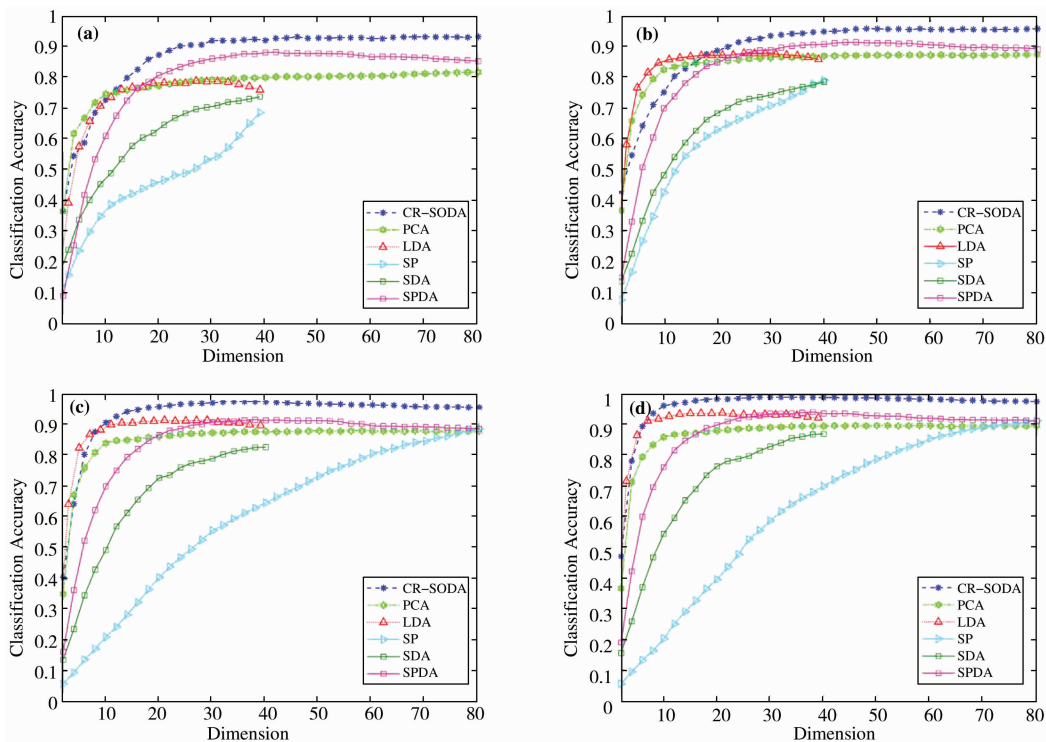


图 3 ORL 数据库中不同算法在不同维数下的识别精度,带标签样本的个数分别为(a)2、(b)3、(c)4、(d)5

3.4.2 Yale 人脸数据库. 实验中,在每类人脸中随机选择 7 幅图像作为训练样本,其余的 4 幅作为测试样本. 在选定的训练样本数目的情况下,实验随机产生 20 组不同的训练样本集. 实验结果取这 20 次实验的平均值. 另外,在实验过程中,每类训练样本随机选择 $p = (2, 3, 4, 5)$ 个训练样本作为带标签样本,其余的为无标签样本. 我们比较 PCA、LDA、SDA、SPP、SPDA 和本文方法的分类识别效果,20 次实验的平均结果如表 2 所示. 图 4 显示了不同算法在不同带标签样本下的平均识别率. 通过本次实验可以很明显的看出,我们的算法,尤其是只使用少量带标签人脸图像时 $p = (2, 3)$, 显著地优于其他方法. 最高识别率分别为 93.72%,

96.38%但是当带标签人脸图像比较多时 $p = (4,5)$,所提方法也获得了最高的识别性能.

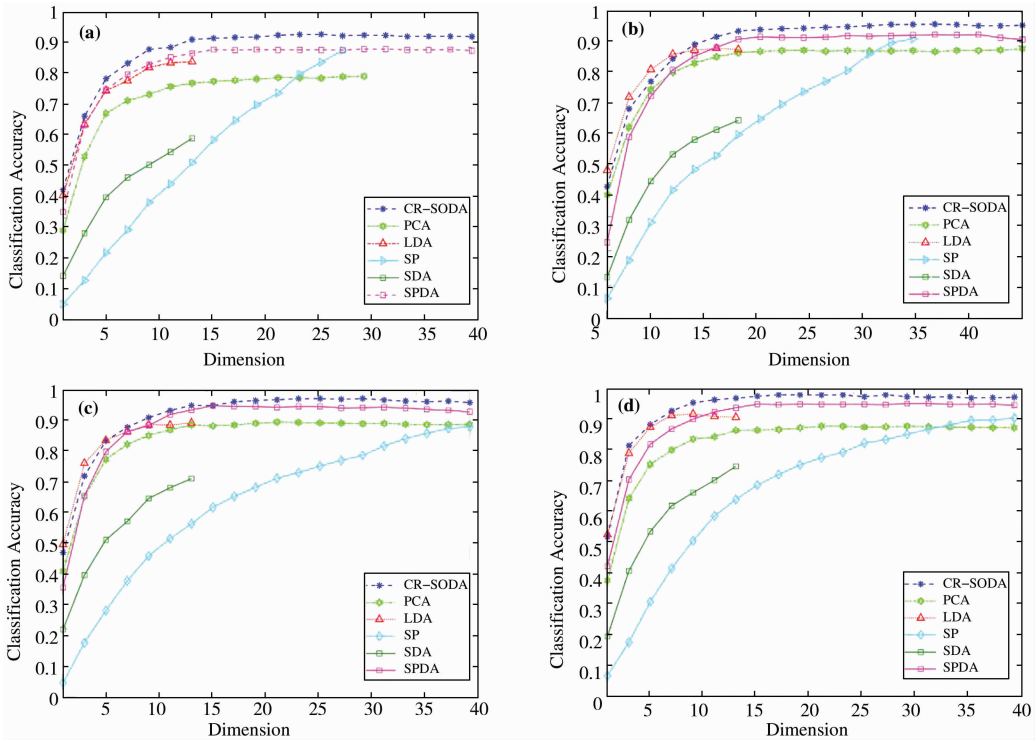


图 4 Yale 数据库中不同算法在不同维数下的识别精度,带标签样本的个数分别为(a)2、(b)3、(c)4、(d)5

3.4.3 CMU PIE 人脸数据库. 在本实验中我们选择在不同光照变化下的正面姿态图像(C27)进行测试,每类随机选取 24 幅图像作为训练样本,令 $p = (4,7,10,13)$ 幅图像为带标签样本,其余为不带标签样本. 独立重复 20 实验,计算平均识别率如表 2 所示. 图 5 显示了 PCA、LDA、SDA、SPP、SPDA 和本文方法的分类识别效果. 最高识别率为 96.69%、98.03%、98.68% 和 99.16%. 从实验结果可以看出与 ORL 和 Yale 数据库一样当标签样本比较少的时候,我们的方法 CR-SODA 明显优于其他方法,当带标签样本逐渐增多的时候,所有方法的识别性能都有所提升,但是 CR-SODA 仍然获得了最高的识别性能. 当 $p = 13$ 时,我们的方法获得了 99.16%的识别效果,优于第二识别性能 1% (SPDA).

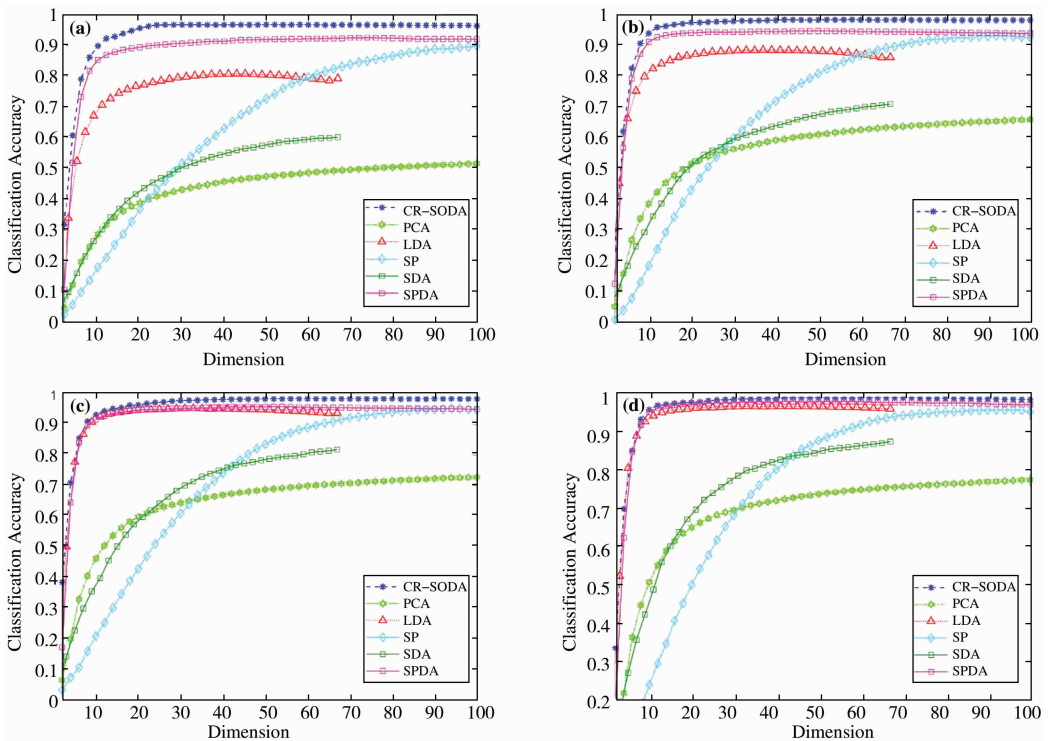


图 5 CMU PIE 数据库中不同算法在不同维数下的识别精度,带标签样本的个数分别为(a)4、(b)7、(c)10、(d)13

3.4.4 关于训练样本个数的分析. 通过之前的分析我们知道, 未标签样本对提高分类性能是有帮助作用的, 当可获得的未标签的样本数增加时, 一个比较有效的半监督学习方法应当促进分类性能. 在本节中, 我们做了一些实验来验证我们的算法在未标签的训练样本增加时的分类性能. 对于 Extended Yale B 和 Yale 数据库, 每类带标签的样本分别为 2、4. 每类的训练样本(包括带标签的和不带标签的样本)分别为 3、10、17、24、30 和 5、6、7、8、9. 实验运行 20 次, 平均识别率如图 6 所示. 在 Extended Yale B 人脸库中, 当所获得未标签样本的数量增加时, 我们算法的性能有了显著的提升. 这个实验结果显示了我们的算法能更有效的利用未标签样本的信息. 在 Yale 人脸库中, 当所获得未标签样本的数量增加时, 算法性能也有所提升, 但是比较小. 这是由于每类训练样本比较少(只有 11 幅). 这也说明了, 在 Yale 库中未标签样本对促进分类性能的作用比较小.

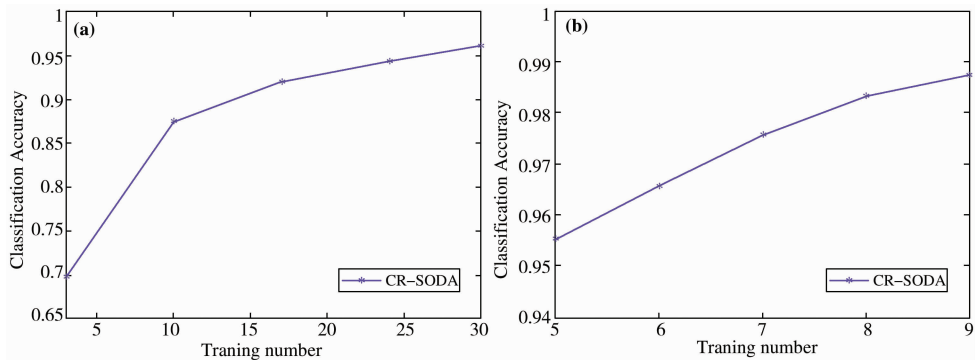


图 6 不同训练样本个数下的识别精度 (a)Extended Yale B 数据库, (b)Yale 数据库

4 结论

本文中我们提出了一种新的基于协同表示标签传播的正交半监督鉴别分析算法(CR-SODA). 我们利用大量未标签样本表示带标签样本, 之后利用表示系数将部分有标签数据的标签信息传递给无标签数据, 因此未标签样本的分布信息可以得到充分的利用. 之后利用传播后的全体软标签数据构造散度矩阵, 提出半监督正交鉴别分析算法, 目的是利用大量未标签样本学习出一个更加有效的鉴别子空间, 促进分类性能. 所提方法可采用更加有效的迹比优化算法求解. 在人脸库上的大量实验也验证了我们算法的有效性, 尤其是只有少量带标签样本时, 我们的方法明显的优于其他方法.

参 考 文 献

- [1] Jain A K, Duin W, Mao J. Statistical pattern Recognition[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2000, 22(1): 4-37.
- [2] Turk M, Pentland A. Eigenfaces for recognition[J]. J Cognitive Neuroscience, 1991, 3(1): 71-86.
- [3] Belhumeur P, Hespanha J, Kriegman D. Eigenfaces vs. fisherfaces: recognition using class specific linear projection[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1997, 19(7): 711-720.
- [4] 侯明, 张新新, 范丽亚. 四种基于支持向量机的多类分类器的性能比较[J]. 聊城大学学报(自然科学版), 2014, 27(3): 54-60.
- [5] 夏文静, 陈耿, 范丽亚. 八种最小二乘 SVM 型学习算法的优势比较[J]. 聊城大学学报(自然科学版), 2016, 29(2): 33-41.
- [6] 许颖春, 范丽亚. 基于 L1-范数的非线性 TSVR[J]. 聊城大学学报(自然科学版), 2017, 30(3): 6-11.
- [7] Wright J, Yang A, Ganesh A, et al. Robust face recognition via sparse representation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2009, 31(2): 210-227.
- [8] Bibi A, Ghanem B. High order tensor formulation for convolutional sparse coding[C]. //Proceedings of the Computer Vision and Pattern Recognition (CVPR), 2017.
- [9] Zhang G, Sun H, Porikli F, et al. Optimal couple projections for domain adaptive sparse representation-based classification[J]. IEEE Transactions on Image Processing, 2017, 26(12): 5922-5935.
- [10] Lu J, Wang G, Zhou J. Simultaneous feature and dictionary learning for image set based face recognition[J]. IEEE Transactions on Image Processing, 2017, 26(8): 4042-4054.

- [11] Li S, Shao M, Fu Y. Person Re-identification by cross-view multi-level dictionary learning[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2018, 40(12):2963-2977.
- [12] Zhu X, Li X, Zhang S, et al. Robust joint graph sparse coding for unsupervised spectral feature selection[J]. *IEEE Transactions on Neural Network and Learning Systems*, 2017, 28(6):1263-1275.
- [13] Qiao L, Chen S, Tan X. Sparsity preserving projections with applications to face recognition[J]. *Pattern Recognition*, 2010, 43(1):331-341.
- [14] Zang F, Zhang J. Discriminant learning by sparse representation for classification[J]. *Neurocomputing*, 2011, 74(12):2176-2183.
- [15] Zhang G, Sun H, Xia G, et al. Kernel dictionary learning based discriminant analysis[J]. *Journal of Vision Communication and Image Representation*, 2016, 40:470-484.
- [16] Yang J, Chu D, Zhang L, et al. Sparse representation classifier steered discriminative projection with applications to face recognition[J]. *IEEE Transactions on Neural Network and Learning Systems*, 2013, 24(7):1023-1035.
- [17] Zhu X. Semi-supervised learning literature survey[J]. *Computer Science Univ Wisconsin-Madison*, 2006, 2(3):4-9.
- [18] Cai D, He X, Han J. Semi-supervised discriminant analysis[C]. // *Processing of IEEE International Conference on Computer Vision (ICCV)*, 2007.
- [19] Lu J, Zhou X, Tan Y, et al. Cost-sensitive semi-supervised discriminant analysis for face recognition[J]. *IEEE Trans. Information Forensics and Security*, 2012, 7(3):944-953.
- [20] Zhang G, Jiang Z, Davis L S. Online semi-supervised discriminative dictionary learning for sparse representation[C]. // *Processing of Asian Conference on Computer Vision (ACCV)*, 2013.
- [21] Yan S, Xu D, Zhang B, et al. Graph embedding and extensions: a general framework for dimensionality reduction[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2007, 29(1):40-51.
- [22] Zang F, Zhang J S. Label propagation through sparse neighborhood and its application[J]. *Neurocomputing*, 2012, 97:267-277.
- [23] Belkin M, Niyogi P. Laplacian eigenmaps for dimensionality reduction and data representation[J]. *Neural Computation*, 2003, 15(6):1373-1396.
- [24] Zhu X, Ghahramani Z, Lafferty J. Semi-supervised learning using Gaussian fields and harmonic functions[C]. // *Proceedings of the 20th International conference on Machine learning*, 2003.
- [25] Wang F, Zhang C S. Label propagation through linear neighborhoods[J]. *IEEE Transactions on Knowledge and Data Engineering*, 2008, 20(1):55-67.
- [26] Liu Y, Nie F P, Wu J G, et al. Semi-supervised feature selection based on label propagation and subset selection[C]. // *Processing of International Conference on Computer and Information Application*, 2010.
- [27] Cheng B, Yang J, Yan S, et al. Learning with ℓ_1 -graph for image analysis [J]. *IEEE Transactions on Image Processing*, 2010, 19(4):858-866.
- [28] Yan S, Wang H. Semi-supervised learning by sparse representation[C]. // *Proceeding of SIAM International Conference on Data Mining (SDM)*, 2009.
- [29] Gao Y, Ma J, Yuille A L. Semi-supervised sparse representation-based classification for face recognition with insufficient labeled samples [J]. *IEEE Transactions on Image Processing*, 2017, 26(5):2545-2560.
- [30] Zhou D, Bousquet O, Lal T N, et al. Learning with local and global consistency[C]. // *Proceedings of the Conference of Advances in Neural Information Processing Systems (NIPS)*, 2004.
- [31] Liu G C, Lin Z C, Yan S C, et al. Robust recovery of subspace structures by low-rank representation[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2012, 34(3):663-670.
- [32] Zhang L, Yang M, Feng X. Sparse representation or collaborative representation: Which helps face recognition[C]. // *Proceedings of the International Conference on Computer Vision (ICCV)*, 2011.
- [33] Cai S, Zhang L, Zuo W, et al. A probabilistic collaborative representation-based approach for pattern classification[C]. // *Proceedings of the Computer Vision and Pattern Recognition (CVPR)*, 2016.
- [34] Zhang G, Sun H, Ji Z, et al. Label propagation based on collaborative representation for face recognition[J]. *Neurocomputing*, 2016, 171:1193-1204.
- [35] Yang W, Wang Z, Sun C. A collaborative representation-based projections method for feature extraction[J]. *Pattern Recognition*, 2015, 48:20-27.
- [36] Wei L, Xu F, Yin J, et al. Kernel locality-constrained collaborative representation-based discriminant analysis[J]. *Knowledge-Based Systems*, 2014, 70:212-220.
- [37] Ngo T T, Bellalij M, Saad Y. The trace ratio optimization problem[J]. *SIAM Review*, 2012, 54(3):545-569.

- [38] Donoho D. For most large underdetermined systems of linear equations the minimal ℓ_1 -norm solution is also the sparse solution[J]. *Communications on Pure and Applied Mathematics: A Journal Issued by the Courant Institute of Mathematical Sciences*, 2006, 59(6):797-829.
- [39] Nie F P, Xiang S M, Jia Y Q, et al. Semi-supervised orthogonal discriminant analysis via label propagation. *Pattern Recognition*, 2009, 42(11):2615-2627.
- [40] Lu J, Zhou X, Tan Y, et al. Cost-Sensitive Semi-supervised Discriminant Analysis for Face Recognition[J]. *IEEE Trans Information Forensics and Security*, 2012, 7(3):944-953.
- [41] Qiao L, Chen S, Tan X. Sparsity preserving discriminant analysis for single training image face recognition[J]. *Pattern Recognition Letters*, 2010, 31:422-429.
- [42] Samaria F, Harter A. Parameterisation of a stochastic model for human face identification[C]. // *Proceedings of the IEEE Workshop on Applications of Computer Vision*, December 1994.
- [43] Sim T, Baker S, Bsat M. The CMU pose, illumination, expression database[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2005, 25(12):1615-1618.
- [44] Georghiades A, Belhumeur P, Kriegman D. From few to many: Illumination cone models for face recognition under variable lighting and pose[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2001, 23(6):643-660.
- [45] Martinez A, benavente R. The AR face database[J]. *CVC Tech Report*, 1998, 24:30-35.
- [46] He X F, Yan S C, Hu Y X, et al. Face recognition using laplacianfaces[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2005, 27(3):328-340.

Collaborative Representation Label Propagation based Semi-supervised orthogonal Discriminative Analysis

YANG Jun-chuan JIANG Tong ZHANG Guo-qing

(School of Computer and Software, Nanjing University of Information Science & Technology, Nanjing 210044, China)

Abstract Recently, representation-based classification methods, such as sparse representation-based classification (SRC) and collaborative representation-based classification (CRC) have been developed and shown potential due to its effectiveness in various recognition tasks. However, the performance of SRC or CRC depend on the number of labeled training samples for each class. When the labeled training samples per class are insufficient, the recognition performance of SRC or CRC will decrease significantly. In order to address this problem, literature[34] introduced the collaborative representation into semi-supervised learning and proposed a semi-supervised label propagation method based on collaborative representation, which can explore the information of unlabeled data effectively. Inspired by this, we propose a collaborative representation label propagation based semi-supervised orthogonal discriminant analysis algorithm (CR-SODA), which combines the collaborative representation label propagation with the orthogonal discriminant analysis. Different from the existing semi-supervised dimensionality reduction algorithms, our method propagates the label information from the labeled data to the unlabeled data through this label propagation algorithm, and then construct the scatter matrices by using the soft label to perform discriminant analysis. In addition, we impose the projection matrix is orthogonal such that the solution of proposed method can be effectively obtained by using a trace ratio optimization algorithm. Extensive experiments on face dataset verify the effectiveness of our algorithm, especially when limited number of images were labeled, our method can still maintain its good classification performance.

Key words collaborative representation; label propagation; semi-supervised; discriminant analysis