

基于低密度分割几何距离的半监督 KFDA 算法*

陶新民, 常瑞, 沈微, 王若彤, 李晨曦

(东北林业大学 工程技术学院, 黑龙江 哈尔滨 150040)

通讯作者: 陶新民, E-mail: taixinmin@nefu.edu.cn



摘要: 提出了一种基于低密度分割几何距离的半监督 KFDA(kernel Fisher discriminant analysis)算法(semi-supervised KFDA based on low density separation geometry distance, 简称 SemiGKFDA). 该算法以低密度分割几何距离作为相似性度量, 通过大量无标签样本, 提高 KFDA 算法的泛化能力. 首先, 利用核函数将原始空间样本数据映射到高维特征空间中; 然后, 通过有标签样本和无标签样本构建低密度分割几何距离测度上的内蕴结构一致性假设, 使其作为正则化项整合到费舍尔判别分析的目标函数中; 最后, 通过求解最小化目标函数获得最优投影矩阵. 人工数据集和 UCI 数据集上的实验表明, 该算法与 KFDA 及其改进算法相比, 在分类性能上有显著提高. 此外, 将该算法与其他算法应用到人脸识别问题中进行对比, 实验结果表明, 该算法具有更高的识别精度.

关键词: 费舍尔判别分析; 低密度分割几何距离; 半监督; 分类性能

中图法分类号: TP391

中文引用格式: 陶新民, 常瑞, 沈微, 王若彤, 李晨曦. 基于低密度分割几何距离的半监督 KFDA 算法. 软件学报, 2020, 31(2): 493-510. <http://www.jos.org.cn/1000-9825/5639.htm>

英文引用格式: Tao XM, Chang R, Shen W, Wang RT, Li CX. Semi-supervised KFDA algorithm based on low density separation geometry distance. Ruan Jian Xue Bao/Journal of Software, 2020, 31(2): 493-510 (in Chinese). <http://www.jos.org.cn/1000-9825/5639.htm>

Semi-supervised KFDA Algorithm Based on Low Density Separation Geometry Distance

TAO Xin-Min, CHANG Rui, SHEN Wei, WANG Ruo-Tong, LI Chen-Xi

(School of Engineering and Technology, Northeast Forestry University, Harbin 150040, China)

Abstract: In this study, a novel semi-supervised kernel Fisher discriminant analysis (KFDA) based on low density separation geometric distance is proposed. The method employs the low density separation geometric distance as the measure of similarity and thus improves the generalization ability of the KFDA through a large number of unlabeled samples. First, the original spatial data are implicitly mapped onto the high-dimensional feature space by kernel function. Then, both the labeled data and the unlabeled data are used to capture the consistence assumption of geometrical structure based on low density separation geometric distance, which are incorporated into the objection function of Fisher discriminant analysis as a regularization term. Finally, the optimal projection matrix is obtained by minimizing the objective function. Experiments on artificial datasets and UCI datasets show that the proposed algorithm has a significantly improvement in classification performance compared with the KFDA and its modified approaches. In addition, comparison results with other methods on face recognition problems demonstrate that the proposed algorithm has higher identification accuracy.

Key words: Fisher discriminant analysis; low density separation geometry distance; semi-supervised; classification performance

* 基金项目: 中央高校基本科研业务费专项资金(2572017EB02, 2572017CB07); 东北林业大学双一流科研启动基金(411112438); 哈尔滨市科技局创新人才基金(2017RAXXJ018); 国家自然科学基金(31570547)

Foundation item: Fundamental Research Funds for the Central Universities (2572017EB02, 2572017CB07); "Double-First Class" Research Start-Up Fund of Northeast Forestry University (411112438); Innovative Talents Fund of Harbin Municipal Bureau of Science and Technology (2017RAXXJ018); National Natural Science Foundation of China (31570547)

收稿时间: 2018-01-28; 修改时间: 2018-07-25; 采用时间: 2018-08-30

费舍尔判别分析(Fisher discriminant analysis,简称 FDA)作为一种基于统计学的维数约减和特征提取方法,在图像检索、人脸识别、图像分割等领域得到了广泛的应用^[1-4].该算法通过线性变化将高维样本数据投影到最佳鉴别向量空间,利用最佳判别向量进行特征提取,找到样本类间距离与样本类内距离比值达到最大的投影方向,即在该投影空间上有最佳分类效果^[5-7].但是,FDA 在处理非线性数据时存在一定的局限性.

为了打破这种局限,学者们提出了基于核的 KFDA(kernel Fisher discriminant analysis)非线性算法.该算法是一种非线性数据维数约简方法.它通过核函数将原空间数据映射到一个高维核空间,然后在映射后的核空间中进行 Fisher 判别分析.KFDA 有效地解决了原始数据间的非线性相关问题^[8-10],在故障分类、模式识别、人脸识别等领域取得了良好成果^[11-14].但是在处理分类问题过程中,面对实际应用中的大量无标注数据和少量有标注数据,KFDA 算法作为一种有监督学习方法,只能使用少量有标注数据的特征向量进行学习,泛化能力较弱,导致分类结果并不理想^[15-19].

为了解决上述问题,利用无标注样本数据指导有标注样本数据进行判别分析受到学者们的广泛关注^[20-22].文献[20]通过半监督的方法构造算法的目标函数,以正交方式求解特征值找出最优投影向量.文献[21]通过核函数从原始高维数据中提取低维特征,利用标注和未标注样本寻找最优低维特征,并将其输入到分类器中进行分类.文献[22]通过 KFDA 寻找局部流形结构的最优投影,同时利用支持向量机来反映数据集的低维几何结构.以上算法虽然充分利用了半监督学习的优势,但是不能准确反映样本的内在流形结构^[23].为了解决此问题,学者们相继提出了一些基于距离的半监督 KFDA 改进算法^[24,25].文献[24]通过格拉斯曼核度量空间的相似度将原始数据样本映射到格拉斯曼流形上,来进行基于格拉斯曼流形的非线性判别分析.文献[25]依据流形距离定义数据点与邻近点之间的相似度,并利用这种相似度构造算法的目标函数,从而保持了数据集的内在流形结构.但以上基于距离的改进算法得到的判别分析结果并不能同时满足样本空间的全局一致性与局部一致性要求^[26,27].

鉴于此,本文提出一种基于低密度分割几何距离的半监督 KFDA 算法.算法采用半监督学习方式,合理利用无标注样本的信息指导有监督 KFDA 算法的学习.考虑到样本集合的空间全局一致性和局部一致性假设,算法通过定义低密度分割几何距离作为相似性度量来获得正则化项,使高维空间中的邻域样本在降维空间中仍然保持流形结构,最后将正则化项整合到 Fisher 判别分析的目标函数中,通过求解最小化目标函数获得最优投影矩阵.实验部分,将本文算法同其他算法应用到人工数据和 UCI 数据分类问题中进行性能比较,结果表明,本文算法的分类性能均优于其他算法.此外,我们将本文算法应用到人脸识别实验中,通过 Yale 人脸数据库上的实验表明,相对于其他算法,本文算法有更高的识别精度.

1 KFDA 算法

1.1 传统FDA算法

FDA 作为一类有监督学习算法,其目标是通过样本的类别信息寻找一个投影矩阵,从而使投影后的类内散度尽量小,类间散度尽量大,即找到最易分类的投影向量^[2].算法的计算过程简单描述如下.

令 $X = \{x_1, x_2, \dots, x_i, \dots, x_n\}$ 为训练样本集合, $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$ 为总体样本均值, n 为训练样本个数, $\bar{x}^{(k)} = \frac{1}{n_k} \sum_{i=1}^{n_k} x_i^{(k)}$ 代表第 k 类别样本集合的均值, n_k 为第 k 类别样本集合的个数, C 为类别数.定义类间散度矩阵 S_B 为

$$S_B = \frac{1}{n} \sum_{k=1}^C n_k (\bar{x}^{(k)} - \bar{x})(\bar{x}^{(k)} - \bar{x})^T \quad (1)$$

样本集合的整体散度矩阵 S_T 为

$$S_T = \frac{1}{n} \sum_{i=1}^n n_k (x_i - \bar{x})(x_i - \bar{x})^T \quad (2)$$

令投影后样本 $y_i = v^T x_i, i=1, 2, \dots, n$, 则 $\bar{y} = \frac{1}{n} \sum_{i=1}^n v^T x_i = v^T \frac{1}{n} \sum_{i=1}^n x_i = v^T \bar{x}$, 同理可得 $\bar{y}^{(k)} = v^T \bar{x}^{(k)}$, v 是投影向量.则投影后的类间散度和类内散度可以定义为

$$\tilde{S}_B = \mathbf{v}^T \mathbf{S}_B \mathbf{v}, \tilde{S}_T = \mathbf{v}^T \mathbf{S}_T \mathbf{v} \quad (3)$$

$$\mathbf{v}_{opt} = \arg \max_{\mathbf{v}} \frac{\mathbf{v}^T \mathbf{S}_B \mathbf{v}}{\mathbf{v}^T \mathbf{S}_T \mathbf{v}} \quad (4)$$

广义特征值特征向量,即投影向量,可以通过 $(\mathbf{S}_T)^{-1} \mathbf{S}_B \mathbf{v} = \lambda \mathbf{v}$ 求得.

1.2 核空间FDA分析算法

虽然传统 FDA 在线性领域应用广泛,但是在现实生活中,不同样本数据之间的相互关系并不都是线性相关的.当数据分布呈现非线性分布时,FDA 算法无法判别分析数据集中的非线性内在结构分类信息.为此,可以通过核函数将原空间数据映射到一个高维核空间,然后在映射后的核空间中对投影后的数据进行判决分析.算法计算步骤及推导过程如下.

令 \mathbf{F} 为特征空间.映射函数为: $\phi: \mathbf{R}^r \rightarrow \mathbf{F}$, r 为原始空间维度,则特征空间的内积表示为

$$\langle \phi(\mathbf{x}), \phi(\mathbf{z}) \rangle = k(\mathbf{x}, \mathbf{z}) \quad (5)$$

这里, $k(\mathbf{x}, \mathbf{z})$ 表示为半正定核函数.著名的核函数如高斯核: $k(\mathbf{x}, \mathbf{z}) = \exp(-\|\mathbf{x} - \mathbf{z}\|^2 / \sigma^2)$, σ 为高斯核宽度.

在特征空间 \mathbf{F} 中,类间散度矩阵 \mathbf{S}_B 为

$$\mathbf{S}_B = \frac{1}{n} \sum_{k=1}^C n_k^2 \left(\frac{1}{n_k} \sum_{i=1}^{n_k} \phi(\mathbf{x}_i^{(k)}) - \frac{1}{n} \sum_{j=1}^n \phi(\mathbf{x}_j) \right) \left(\frac{1}{n_k} \sum_{i=1}^{n_k} \phi(\mathbf{x}_i^{(k)}) - \frac{1}{n} \sum_{j=1}^n \phi(\mathbf{x}_j) \right)^T \quad (6)$$

由于 $\frac{1}{n} \sum_{j=1}^n \phi(\mathbf{x}_j)$ 是个固定值,因此上式可进一步等价于

$$\mathbf{S}_B = \frac{1}{n} \sum_{k=1}^C n_k^2 \left(\frac{1}{n_k} \sum_{i=1}^{n_k} \phi(\mathbf{x}_i^{(k)}) - \frac{1}{n} \sum_{j=1}^n \phi(\mathbf{x}_j) \right) \left(\frac{1}{n_k} \sum_{i=1}^{n_k} \phi(\mathbf{x}_i^{(k)}) - \frac{1}{n} \sum_{j=1}^n \phi(\mathbf{x}_j) \right)^T \quad (7)$$

为了方便求解,对上式进行变换,将原式中的 n_k 变成 n_k^2 ,即考虑类别分布的平方项,同时为便于讨论,令

$$\Psi(\mathbf{x}_i) = \phi(\mathbf{x}_i) - \frac{1}{n} \sum_{i=1}^n \phi(\mathbf{x}_i) \quad (8)$$

则类间散度矩阵 \mathbf{S}_B 可表示为

$$\begin{aligned} \mathbf{S}_B &= \frac{1}{n} \sum_{k=1}^C n_k^2 \left(\frac{1}{n_k} \sum_{i=1}^{n_k} \phi(\mathbf{x}_i^{(k)}) - \frac{1}{n} \sum_{j=1}^n \phi(\mathbf{x}_j) \right) \left(\frac{1}{n_k} \sum_{i=1}^{n_k} \phi(\mathbf{x}_i^{(k)}) - \frac{1}{n} \sum_{j=1}^n \phi(\mathbf{x}_j) \right)^T \\ &= \frac{1}{n} \sum_{k=1}^C \left(\sum_{i=1}^{n_k} \Psi(\mathbf{x}_i^{(k)}) \sum_{i=1}^{n_k} (\Psi(\mathbf{x}_i^{(k)}))^T \right) \\ &= \frac{1}{n} \sum_{k=1}^C (\Psi(\mathbf{X})^{(k)} \mathbf{W}^{(k)} \Psi(\mathbf{X})^{(k)})^T \end{aligned} \quad (9)$$

这里, \mathbf{W}^k 是 $n_k \times n_k$ 的全部元素均为 1 的方阵, $\Psi(\mathbf{X})^{(k)} = [\Psi(\mathbf{x}_1^{(k)}), \dots, \Psi(\mathbf{x}_{n_k}^{(k)})]$ 代表第 k 类的特征空间均值化后的数据集.令 $\Psi(\mathbf{X}) = [\Psi(\mathbf{X})^{(1)}, \Psi(\mathbf{X})^{(2)}, \dots, \Psi(\mathbf{X})^{(C)}]$, 定义 $n \times n$ 矩阵 $\mathbf{W}_{n \times n}$ 为

$$\mathbf{W}_{n \times n} = \begin{bmatrix} \mathbf{W}^{(1)} & 0 & 0 & 0 \\ 0 & \mathbf{W}^{(2)} & 0 & 0 \\ \vdots & 0 & \ddots & 0 \\ 0 & 0 & 0 & \mathbf{W}^{(C)} \end{bmatrix} \quad (10)$$

得到类内散度和类间散度分别为

$$\mathbf{S}_B = \frac{1}{n} \sum_{k=1}^C (\Psi(\mathbf{X})^{(k)} \mathbf{W}^{(k)} \Psi(\mathbf{X})^{(k)})^T = \frac{1}{n} \Psi(\mathbf{X}) \mathbf{W}_{n \times n} \Psi(\mathbf{X})^T \quad (11)$$

$$\begin{aligned}
 S_T &= \frac{1}{n} \sum_{k=1}^C \sum_{i=1}^{n_k} \left(\phi(x_i^{(k)}) - \frac{1}{n} \sum_{j=1}^n \phi(x_j) \right) \left(\phi(x_i^{(k)}) - \frac{1}{n} \sum_{j=1}^n \phi(x_j) \right)^T \\
 &= \frac{1}{n} \sum_{k=1}^C \sum_{i=1}^{n_k} (\Psi(X)^{(k)})(\Psi(X)^{(k)})^T \\
 &= \frac{1}{n} \sum_{i=1}^n \Psi(x_i)\Psi(x_i)^T \\
 &= \frac{1}{n} \Psi(X)\Psi(X)^T
 \end{aligned} \tag{12}$$

给定一组 $\{v_i \in F | i=1, 2, \dots, d\}$ 投影向量, d 代表降维后的维度. 去均值化后的样本 $\Psi(x_i) \in F$ 投影到 v_1, v_2, \dots, v_d 的 R^d 空间中对应向量 y_i 为

$$y_i = (\langle v_1, \Psi(x_i) \rangle \langle v_2, \Psi(x_i) \rangle \dots \langle v_d, \Psi(x_i) \rangle)^T \tag{13}$$

则去均值化后的样本在 v 上的投影为

$$\Psi(Y) = v^T \Psi(X) \tag{14}$$

将公式(14)代入到 S_B, S_T 中, 得到核 Fisher 成分分析的目标函数为

$$S_B(\Psi(Y)) = \frac{1}{n} v^T \Psi(X) W_{n \times n} \Psi(X)^T v \tag{15}$$

$$S_T(\Psi(Y)) = \frac{1}{n} v^T \Psi(X) \Psi(X)^T v \tag{16}$$

$$v_{opt} = \arg \max_v \frac{v^T \Psi(X) W_{n \times n} \Psi(X)^T v}{v^T \Psi(X) \Psi(X)^T v} \tag{17}$$

为了确保解的唯一性, 令 $v^T \Psi(X) \Psi(X)^T v = 1$, 则上述问题的解可通过拉格朗日函数求得:

$$L(v, \lambda) = v^T \Psi(X) W_{n \times n} \Psi(X)^T v - \lambda (v^T \Psi(X) \Psi(X)^T v - 1) \tag{18}$$

对 v 求导并设置为 0, 可以得到:

$$\begin{cases}
 2\Psi(X) W_{n \times n} \Psi(X)^T v - \lambda 2\Psi(X) \Psi(X)^T v = 0 \\
 \Psi(X) W_{n \times n} \Psi(X)^T v = \lambda \Psi(X) \Psi(X)^T v
 \end{cases} \tag{19}$$

左右两边都同左乘以 $\Psi(X)^T$, 得:

$$\Psi(X)^T \Psi(X) W_{n \times n} \Psi(X)^T v = \lambda \Psi(X)^T \Psi(X) \Psi(X)^T v \tag{20}$$

其中, v 可由训练样本即 $\{\Psi(x_1), \Psi(x_2), \dots, \Psi(x_n)\}$ 张成:

$$v = \sum_{i=1}^n a_i \Psi(x_i) = \Psi(X) a \tag{21}$$

其中, a_i 为 $\Psi(x_i)$ 对应的系数, $a = (a_1, a_2, \dots, a_n)^T$.

将公式(21)代入到公式(20)中, 可以得到:

$$\Psi(X)^T \Psi(X) W_{n \times n} \Psi(X)^T \Psi(X) a = \lambda \Psi(X)^T \Psi(X) \Psi(X)^T \Psi(X) a \tag{22}$$

令 $\tilde{K} = \Psi(X)^T \Psi(X)$, 则 $\tilde{K} W_{n \times n} \tilde{K} a = \lambda \tilde{K} \tilde{K} a$ 广义特征值和特征向量可以通过 $(\tilde{K} \tilde{K})^{-1} \tilde{K} W_{n \times n} \tilde{K} a = \lambda a$ 求得. 为了保证 $\tilde{K} \tilde{K}$ 可逆, 通常增加 Tikhonov 正则化项对其进行处理:

$$(\tilde{K} \tilde{K} + \gamma I)^{-1} \tilde{K} W_{n \times n} \tilde{K} a = \lambda a \tag{23}$$

其中, $\gamma > 0$, 为 Tikhonov 正则化项系数; I 为 $n \times n$ 单位阵.

令 a_1, a_2, \dots, a_d 为上式的 d 个广义特征向量, 则 $\theta = [a_1, a_2, \dots, a_d] \in R^{n \times d}$, 结合公式(13)、公式(20)得原有训练样本投影到 d 维子空间的公式为

$$x \rightarrow y = \theta^T \tilde{K}(:, x) \tag{24}$$

这里, $\tilde{K}(:, x) = [k(x_1, x), k(x_2, x), \dots, k(x_n, x)]^T$. 其中,

$$\tilde{\mathbf{K}} = \begin{bmatrix} \Psi(\mathbf{x}_1)^T \Psi(\mathbf{x}_1) & \Psi(\mathbf{x}_2)^T \Psi(\mathbf{x}_1) & \cdots & \Psi(\mathbf{x}_n)^T \Psi(\mathbf{x}_1) \\ \Psi(\mathbf{x}_1)^T \Psi(\mathbf{x}_2) & \Psi(\mathbf{x}_2)^T \Psi(\mathbf{x}_2) & \cdots & \Psi(\mathbf{x}_n)^T \Psi(\mathbf{x}_2) \\ \vdots & \vdots & \cdots & \vdots \\ \Psi(\mathbf{x}_1)^T \Psi(\mathbf{x}_n) & \Psi(\mathbf{x}_2)^T \Psi(\mathbf{x}_n) & \cdots & \Psi(\mathbf{x}_n)^T \Psi(\mathbf{x}_n) \end{bmatrix}, \tilde{\mathbf{K}}_{ij} = \Psi(\mathbf{x}_i)^T \Psi(\mathbf{x}_j)^T \quad (25)$$

其中, $\tilde{\mathbf{K}}_{ij} = \Psi(\mathbf{x}_i)^T \Psi(\mathbf{x}_j)$ 的求解详见附录 A.

2 基于低密度分割几何距离的半监督 KFDA 算法

KFDA 虽然能够解决 FDA 在非线性领域存在的局限,但是在实际应用中,面对大量未标注数据和少量已标注数据,KFDA 算法作为一种有监督学习只能使用少量的有标注数据的特征向量进行学习,泛化能力较弱,分类结果并不理想.为了解决此问题,学者们提出了半监督 KFDA 算法.然而,现有的半监督 KFDA 算法所采用的距离测度均不能很好地体现数据集的空间全局一致性和局部一致性假设.为此,本文提出一种基于低密度分割几何距离的半监督 KFDA 算法.

2.1 基于低密度分割几何距离的相似性度量

众所周知,对象间的空间相似性需同时满足以下两个一致性关系.

- (1) 局部一致性,即空间位置上相邻的数据点具有较高的相似性.
- (2) 全局一致性,即位于同一流形上的数据点具有较高的相似性.

文献[24]采用的距离测度能够较好地满足数据的全局一致性特征,但对数据的局部一致性考虑较少.文献[25]采用的距离测度虽充分考虑了数据的局部一致性特征,但却无法满足数据的全局一致性要求.为了同时满足数据空间的全局一致性和局部一致性,本文采用低密度分割几何距离作为相似性度量.低密度分割几何距离的具体描述如下:将样本点定义为图 $G=(V,E)$ 的顶点 V ,令 $p \in V^l$ 表示为图上连接点 p_1 与 $p_{|p|}$ 的一条长度为 $l=|p|-1$ 的路径,其中,边 $(p_k, p_{k+1}) \in E$, E 是边集.令 p_{ij} 表示连接数据点 p_i, p_j 的所有路径的集合,其中, $p_i, p_j \in V$, 则 p_i 与 p_j 之间的几何距离按下式计算.

$$d_{ij}^\rho = \frac{1}{\rho} \ln \left(1 + \min_{p \in p_{ij}} \sum_{k=1}^{|p|} e^{\rho d(p_k, p_{k+1})} - 1 \right) \quad (26)$$

其中, $|p|$ 表示两点 p_i, p_j 间所含有的所有路径的个数, p_k, p_{k+1} 表示数据点 p_i, p_j 间路径上的任意两个相邻点, $\min_{p \in p_{ij}} \sum_{k=1}^{|p|} e^{\rho d(p_k, p_{k+1})} - 1$ 表示图 G 上顶点 p_i, p_j 之间最短路径距离, $d(p_k, p_{k+1})$ 是图 G 上顶点 p_i 到 p_j 最短路径上任意相邻两点的欧式距离.这里采用 $e^{\rho d(p_k, p_{k+1})} - 1$ 而非传统的欧式距离来计算最短路径距离,是考虑到在给定的图 G 中,同一流形体上任意相邻两点间的欧式距离较小,不同流形体上样本间的欧式距离较大.在计算最短路径距离时,位于同一流形上的两点之间会有许多较短边相连,而位于不同流形上的两点需要较长边相连,因此通过指数函数和伸缩因子 ρ 的值来实现放大位于不同流形体上的数据点间距离和缩短位于同一流形上的数据点间距离的目的,从而使算法能够在反映数据间局部一致性的同时,兼顾全局一致性.为了进一步说明该距离测度能够同时满足数据集的全局一致性和局部一致性特征,这里对距离测度进行了取极限处理,结果如下.

当 $\rho \rightarrow \infty$ 时,本文采用的低密度分割几何距离等同于连接距离.连接距离为连接任意两点 (p_i, p_j) 所有路径中的最大距离中的最小值,其表达式如下:

$$d_{ij}^\rho = \min_{p \in p_{ij}} \max_{k < |p|} d(p_k, p_{k+1}) \quad (27)$$

由于同一流体上的样本间通常会有较短边连接,而不同流体上的样本间会存在较长的边,因此通过该距离测度的计算,能使样本数据在同一个流体中的相似度高,而在不同流体中的相似度低,即满足了全局一致要求.

当 $\rho \rightarrow 0$ 时,本文采用的低密度分割几何距离等同于流形距离.流形距离通过度量流形上的最短路径来反映数据集的内在流形结构.流形距离的表达式如下:

$$d_{ij}^p = \min_{p \in P_{ij}} \sum_{k=1}^{|p|} d(p_k, p_{k+1}) \quad (28)$$

由于该距离测度的计算采用路径求和的方式,即考虑到路径长度的影响,因此满足空间局部一致性要求。

2.2 空间一致性假设

当样本集合存在未被标注的样本信息时,为了能够合理地利用未标注的样本信息指导有监督算法的学习,需要考虑样本集合的空间一致性假设.针对 FDA 降维方法而言,空间一致性即是要保证在原有高维空间中全局和局部邻域的样本在降维空间中仍然保持这个领域关系。

设 $\mathbf{y}_i, \mathbf{y}_j$ 是新基向量投影下的坐标,从总体样本空间一致性角度考虑,原来相近的样本 $\mathbf{x}_i, \mathbf{x}_j$ 在新的基坐标空间中也同样相近.本文利用高斯函数表示原有空间的样本间相近程度。

$$S_{ij} = \begin{cases} \exp\left(-\frac{d_{ij}^p(\mathbf{x}_i, \mathbf{x}_j)}{2\delta^2}\right), & \mathbf{x}_i \in knn(\mathbf{x}_j) \text{ or } \mathbf{x}_j \in knn(\mathbf{x}_i) \\ 0, & \text{others} \end{cases} \quad (29)$$

其中, δ 为表示样本间相似性的尺度参数,本文设置为 $\delta=3$; k 为体现相似度矩阵稀疏性而选择的近邻个数。

因此,满足 $J(\mathbf{v}) = \|\mathbf{y}_i - \mathbf{y}_j\|^2 S_{ij}$ 最小化即可实现保持空间一致性的目的.令投影向量为 \mathbf{v} ,定义含 n 个标识样本和 $m-n$ 个未标识样本总的样本集合为

$$\mathbf{X}_{m+n} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n, \mathbf{x}_{n+1}, \dots, \mathbf{x}_m\} \quad (30)$$

其中, $m > n$. 定义空间一致性假设正则化项 $J(\mathbf{v})$:

$$\left. \begin{aligned} J(\mathbf{v}) &= \sum_{i,j}^m (\mathbf{y}_i - \mathbf{y}_j)^2 S_{ij} \\ &= \sum_{i,j}^m (\mathbf{v}^T \Psi(\mathbf{x}_i) - \mathbf{v}^T \Psi(\mathbf{x}_j))^2 S_{ij} \\ &= \sum_{i,j}^m (\mathbf{v}^T \Psi(\mathbf{x}_i) \mathbf{v}^T \Psi(\mathbf{x}_i) - \mathbf{v}^T \Psi(\mathbf{x}_i) \mathbf{v}^T \Psi(\mathbf{x}_j) - \mathbf{v}^T \Psi(\mathbf{x}_j) \mathbf{v}^T \Psi(\mathbf{x}_i) + \mathbf{v}^T \Psi(\mathbf{x}_j) \mathbf{v}^T \Psi(\mathbf{x}_j)) S_{ij} \end{aligned} \right\} \quad (31)$$

因为 $\mathbf{v}^T \Psi(\mathbf{x}_i), \mathbf{v}^T \Psi(\mathbf{x}_j)$ 都是数值标,考虑到求和项,上式可以改为

$$\left. \begin{aligned} J(\mathbf{v}) &= \sum_{i,j}^m [(2\mathbf{v}^T \Psi(\mathbf{x}_i) \mathbf{v}^T \Psi(\mathbf{x}_i) - 2\mathbf{v}^T \Psi(\mathbf{x}_i) \mathbf{v}^T \Psi(\mathbf{x}_j))] S_{ij} \\ &= 2 \sum_{i,j}^m [(\mathbf{v}^T \Psi(\mathbf{x}_i) \mathbf{v}^T \Psi(\mathbf{x}_i) - \mathbf{v}^T \Psi(\mathbf{x}_i) \mathbf{v}^T \Psi(\mathbf{x}_j))] S_{ij} \end{aligned} \right\} \quad (32)$$

$d_{ii} = \sum_j S_{ij}$, 上面的公式固定 i 后, $\sum_{i,j}^m \mathbf{v}^T \Psi(\mathbf{x}_i) S_{ij} \mathbf{v}^T \Psi(\mathbf{x}_i) = \mathbf{v}^T \Psi(\mathbf{x}_i) d_{ii} \Psi(\mathbf{x}_i)$, $j=1, 2, \dots, m$, 因此 $q_{ii} = \sum_j S_{ij}$.

令 $\Psi(\mathbf{X}) = [\Psi(\mathbf{x}_1), \dots, \Psi(\mathbf{x}_m)]$, $\mathbf{S} \in \mathbf{R}^{m \times m}$, $\mathbf{Q} \in \mathbf{R}^{m \times m}$ 对角矩阵, m 为样本总数. 上述公式可进一步变为

$$\mathbf{v}^T \Psi(\mathbf{X}) \mathbf{Q} \Psi(\mathbf{X})^T \mathbf{v} - \mathbf{v}^T \Psi(\mathbf{X}) \mathbf{S} \Psi(\mathbf{X})^T \mathbf{v} = \mathbf{v}^T \Psi(\mathbf{X}) (\mathbf{Q} - \mathbf{S}) (\Psi(\mathbf{X}))^T = \mathbf{v}^T \Psi(\mathbf{X}) \mathbf{L} (\Psi(\mathbf{X}))^T \mathbf{v} \quad (33)$$

其中, $\mathbf{L} = \mathbf{Q} - \mathbf{S}$.

2.3 半监督规范化核Fisher成分分析

将上述空间一致性假设作为正则化项引入到 KFDA 的目标函数中,则半监督 KFDA 的目标函数定义为

$$\left. \begin{aligned} \mathbf{v}_{opt} &= \arg \max_{\mathbf{v}} \frac{\mathbf{v}^T \Psi(\mathbf{X}) \tilde{\mathbf{W}}_{m \times m} \Psi(\mathbf{X})^T \mathbf{v}}{\mathbf{v}^T \Psi(\mathbf{X}) \tilde{\mathbf{I}}_{m \times m} \Psi(\mathbf{X})^T \mathbf{v} + \alpha J(\mathbf{v})}, \\ \mathbf{v}_{opt} &= \arg \max_{\mathbf{v}} (\mathbf{v}^T \Psi(\mathbf{X}) \tilde{\mathbf{W}}_{n \times n} \Psi(\mathbf{X})^T \mathbf{v}) (\mathbf{v}^T \Psi(\mathbf{X}) \tilde{\mathbf{I}}_{m \times m} \Psi(\mathbf{X})^T \mathbf{v} + \alpha \mathbf{v}^T \Psi(\mathbf{X}) \mathbf{L} (\Psi(\mathbf{X}))^T \mathbf{v})^{-1} \end{aligned} \right\} \quad (34)$$

构造拉格朗日函数:

$$L(\mathbf{v}) = \mathbf{v}^T \Psi(\mathbf{X}) \tilde{\mathbf{W}}_{m \times m} \Psi(\mathbf{X})^T \mathbf{v} - \lambda (\mathbf{v}^T \Psi(\mathbf{X}) \tilde{\mathbf{I}}_{m \times m} \Psi(\mathbf{X})^T \mathbf{v} + \alpha \mathbf{v}^T \Psi(\mathbf{X}) \mathbf{L} (\Psi(\mathbf{X}))^T \mathbf{v} - 1) = 0 \quad (35)$$

对 \mathbf{v} 的拉格朗日函数求偏导:

$$\frac{\partial L(\mathbf{v})}{\mathbf{v}} = 2\Psi(\mathbf{X})\tilde{\mathbf{W}}_{n \times n}\Psi(\mathbf{X})^T\mathbf{v} - 2\lambda\Psi(\mathbf{X})\tilde{\mathbf{I}}_{m \times m}\Psi(\mathbf{X})^T\mathbf{v} - 2\boldsymbol{\alpha}\Psi(\mathbf{X})\mathbf{L}\Psi(\mathbf{X})^T\mathbf{v} = 0 \quad (36)$$

$$\Psi(\mathbf{X})\tilde{\mathbf{W}}_{n \times n}\Psi(\mathbf{X})^T\mathbf{v} = \lambda\Psi(\mathbf{X})\tilde{\mathbf{I}}_{m \times m}(\Psi(\mathbf{X}))^T\mathbf{v} + \boldsymbol{\alpha}\Psi(\mathbf{X})\mathbf{L}(\Psi(\mathbf{X}))^T\mathbf{v} \quad (37)$$

等式两边同时左乘 $(\Psi(\mathbf{X}))^T$, 将 $\mathbf{v} = \Psi(\mathbf{X})\mathbf{a}$ 代入公式(37)后得:

$$(\Psi(\mathbf{X}))^T\Psi(\mathbf{X})\tilde{\mathbf{W}}_{n \times n}\Psi(\mathbf{X})^T\Psi(\mathbf{X})\mathbf{a} = (\Psi(\mathbf{X}))^T\lambda\Psi(\mathbf{X})\tilde{\mathbf{I}}_{m \times m}\Psi(\mathbf{X})^T\Psi(\mathbf{X})\mathbf{a} + \boldsymbol{\alpha}(\Psi(\mathbf{X}))^T\Psi(\mathbf{X})\mathbf{L}\Psi(\mathbf{X})^T\Psi(\mathbf{X})\mathbf{a} \quad (38)$$

其中, \mathbf{a} 为空间一致性假设正则化项控制权重, 本文设置为 1. 定义 $\tilde{\mathbf{W}}_{m \times m} \in \mathbf{R}^{m \times m}$ 为

$$\tilde{\mathbf{W}}_{m \times m} = \begin{bmatrix} \mathbf{W}_{n \times n} & 0 \\ 0 & 0 \end{bmatrix} \quad (39)$$

其中, $\mathbf{W}_{n \times n}$ 如公式(10)中定义所示.

$$\tilde{\mathbf{I}}_{m \times m} \in \mathbf{R}^{m \times m}, \tilde{\mathbf{I}}_{m \times m} = \begin{bmatrix} \mathbf{I}_{n \times n} & 0 \\ 0 & 0 \end{bmatrix} \quad (40)$$

这里, $\mathbf{I}_{n \times n}$ 是 $n \times n$ 单位阵, $\tilde{\mathbf{I}}_{m \times m}$ 是 $m \times m$ 单位阵.

其中, \mathbf{v} 可由含标注和未含标注的全部训练样本 $\{\Psi(\mathbf{x}_1), \Psi(\mathbf{x}_2), \dots, \Psi(\mathbf{x}_n)\}$ 张成, 即

$$\mathbf{v} = \sum_{i=1}^m \mathbf{a}_i \Psi(\mathbf{x}_i) = \Psi(\mathbf{X})\mathbf{a} \quad (41)$$

其中, \mathbf{a}_i 为 $\Psi(\mathbf{x}_i)$ 对应的系数, $\mathbf{a} = (\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_m)^T$.

广义特征值特征向量等式为

$$(\tilde{\mathbf{I}}_{m \times m} \tilde{\mathbf{K}} + \boldsymbol{\alpha} \tilde{\mathbf{K}} \mathbf{L} \tilde{\mathbf{K}})^{-1} \tilde{\mathbf{K}} \tilde{\mathbf{W}}_{m \times m} \tilde{\mathbf{K}} \mathbf{a} = \lambda \mathbf{a} \quad (42)$$

其中, $\tilde{\mathbf{K}} \in \mathbf{R}^{m \times m}$, $\tilde{\mathbf{K}}_{ij} = \Psi(\mathbf{x}_i)^T \Psi(\mathbf{x}_j)$, $i, j = 1, \dots, m$.

为使求得的投影向量 $\mathbf{v}^T \mathbf{v} = 1$, 我们对其进行标准化处理:

$$(\Psi(\mathbf{X})\mathbf{a})^T \Psi(\mathbf{X})\mathbf{a} = 1, \mathbf{a}^T (\Psi(\mathbf{X}))^T \Psi(\mathbf{X})\mathbf{a} = 1, \mathbf{a}^T \tilde{\mathbf{K}} \mathbf{a} = 1 \quad (43)$$

将所求的 \mathbf{a} 除以 $\sqrt{\mathbf{a}^T \tilde{\mathbf{K}} \mathbf{a}}$, 则得到标准化后的 \mathbf{v} .

因 $\mathbf{v} = \sum_{i=1}^m \mathbf{a}_i \Psi(\mathbf{x}_i)$, $\Psi(\mathbf{Y}) = \mathbf{v}^T \Psi(\mathbf{X})$, 原有训练样本 \mathbf{x}_j 在 \mathbf{v} 上的投影坐标为

$$(\mathbf{v}, \Psi(\mathbf{x}_j)) = \sum_{i=1}^m \mathbf{a}_i (\Psi(\mathbf{x}_i), \Psi(\mathbf{x}_j)) = \sum_{i=1}^m \mathbf{a}_i \tilde{\mathbf{K}}_{ij} \quad (44)$$

新样本 \mathbf{t}_j 在 \mathbf{v} 上的投影坐标为

$$(\mathbf{v}, \Psi(\mathbf{t}_j)) = \sum_{i=1}^m \mathbf{a}_i (\Psi(\mathbf{x}_i), \Psi(\mathbf{t}_j)) = \sum_{i=1}^m \mathbf{a}_i \mathbf{K}_{ij}^{test} \quad (45)$$

其中, \mathbf{K}_{ij}^{test} 的求解详见附录 A.

2.4 投影维度分析

从上面公式(43)的推导可以看出, 降维的维度 d 除了要小于原始数据空间维度 r 以外, 还不能超过 $(\tilde{\mathbf{I}}_{m \times m} \tilde{\mathbf{K}} + \boldsymbol{\alpha} \tilde{\mathbf{K}} \mathbf{L} \tilde{\mathbf{K}})^{-1} \tilde{\mathbf{K}} \tilde{\mathbf{W}}_{m \times m} \tilde{\mathbf{K}}$ 非零特征值的个数, 因为矩阵非零特征值的个数与矩阵的秩相等, 故只需要知道 $(\tilde{\mathbf{I}}_{m \times m} \tilde{\mathbf{K}} + \boldsymbol{\alpha} \tilde{\mathbf{K}} \mathbf{L} \tilde{\mathbf{K}})^{-1} \tilde{\mathbf{K}} \tilde{\mathbf{W}}_{m \times m} \tilde{\mathbf{K}}$ 的秩, 这里假设 $\tilde{\mathbf{I}}_{m \times m} \tilde{\mathbf{K}} + \boldsymbol{\alpha} \tilde{\mathbf{K}} \mathbf{L} \tilde{\mathbf{K}}$ 矩阵可逆. 又因为 $\text{rank}((\tilde{\mathbf{I}}_{m \times m} \tilde{\mathbf{K}} + \boldsymbol{\alpha} \tilde{\mathbf{K}} \mathbf{L} \tilde{\mathbf{K}})^{-1} \tilde{\mathbf{K}} \tilde{\mathbf{W}}_{m \times m} \tilde{\mathbf{K}}) \leq \min\{\text{rank}(\tilde{\mathbf{K}} \tilde{\mathbf{W}}_{m \times m} \tilde{\mathbf{K}}), \text{rank}(\tilde{\mathbf{K}} \tilde{\mathbf{W}}_{m \times m} \tilde{\mathbf{K}})\}$, 则只需考虑 $\tilde{\mathbf{K}} \tilde{\mathbf{W}}_{m \times m} \tilde{\mathbf{K}}$ 的秩. $\text{rank}(\cdot)$ 表示矩阵的秩, $\text{rank}(\tilde{\mathbf{K}} \tilde{\mathbf{W}}_{m \times m} \tilde{\mathbf{K}}) \leq \text{rank}(\tilde{\mathbf{W}}_{m \times m})$. 由 $\tilde{\mathbf{W}}_{m \times m}$ 的表达式可知 $\text{rank}(\tilde{\mathbf{W}}_{m \times m}) = C$, 因此 $d \leq C$.

3 实验与分析

为了测试本文提出的基于低密度分割几何距离的半监督 KFDA 算法的性能, 实验部分利用仿真数据集和 UCI 数据集, 对本文提出的 SemiGKFDA 算法和 FDA 算法、KFDA 算法及其改进算法的性能进行对比分析, 并

采用 Yale 人脸数据集进行人脸识别的测试.

实验环境:Windows 7 操作系统,CPU:Intel i7,3.4G 处理器,仿真软件为 Matlab2010b.

3.1 判决边界的确定

实验过程中,为了便于显示比较,采用判决边界作为分类标准.要确定一个判决边界,只需在 \mathbf{v} 的投影坐标上确定一个阈值点即可,如对于两类样本点经投影后得到的 $\bar{\mathbf{y}}_1, \bar{\mathbf{y}}_2$, 则阈值点为

$$\tilde{\mathbf{y}} = \frac{1}{2}(\bar{\mathbf{y}}_1 + \bar{\mathbf{y}}_2) = \frac{1}{2} \left(\mathbf{v}^T \frac{1}{n_1} \sum_{i=1}^{n_1} \Psi(\mathbf{x}_i^{(1)}) + \mathbf{v}^T \frac{1}{n_2} \sum_{i=1}^{n_2} \Psi(\mathbf{x}_i^{(2)}) \right) = \frac{1}{2} \left(\frac{1}{n_1} \sum_{i=1}^{n_1} \mathbf{v}^T \Psi(\mathbf{x}_i^{(1)}) + \frac{1}{n_2} \sum_{i=1}^{n_2} \mathbf{v}^T \Psi(\mathbf{x}_i^{(2)}) \right) \quad (46)$$

将公式(43)代入上式后,得:

$$\tilde{\mathbf{y}} = \left(\frac{1}{n_1} \sum_{i=1}^{n_1} \left(\sum_{j=1}^m \mathbf{a}_j \tilde{\mathbf{K}}_{ij} \right) + \frac{1}{n_2} \sum_{i=1}^{n_2} \left(\sum_{j=1}^m \mathbf{a}_j \tilde{\mathbf{K}}_{ij} \right) \right) \quad (47)$$

3.2 仿真数据集下的算法性能分析

3.2.1 “两类样本”问题

为了验证本文算法的有效性,将传统 LFDA 算法、KFDA 算法和本文提出的 SemiGKFDA 算法分别通过 Twomoons、Twocircles、Spirals 这 3 个人工数据集进行如下对比实验:

Twomoons 人工样本集合的实验设置如下:实验中共选取 200 个人工样本,其中有标签样本 83 个,无标签样本 117 个.算法参数设置如下:核函数为高斯核,核参数为 $\sigma=1$,近邻点选取为 $k=12$.Twocircles 人工样本集合的实验设置如下:实验中共选取 500 个样本,其中有标签样本 305 个,无标签样本 195 个.算法参数设置如下:核函数为高斯核,核参数为 $\sigma=0.1$,近邻点选取为 $k=12$.Spirals 人工数据集的实验设置为:实验中共选取 1 000 个样本,其中有标签样本 900 个,无标签样本 100 个.算法参数设置如下:核函数为高斯核,核参数为 $\sigma=0.3$,近邻点选取为 $k=10$.判决边界图和在第 1 个投影向量的系数分布图分别如图 1、图 2 所示,其中,菱形表示为无标签样本.

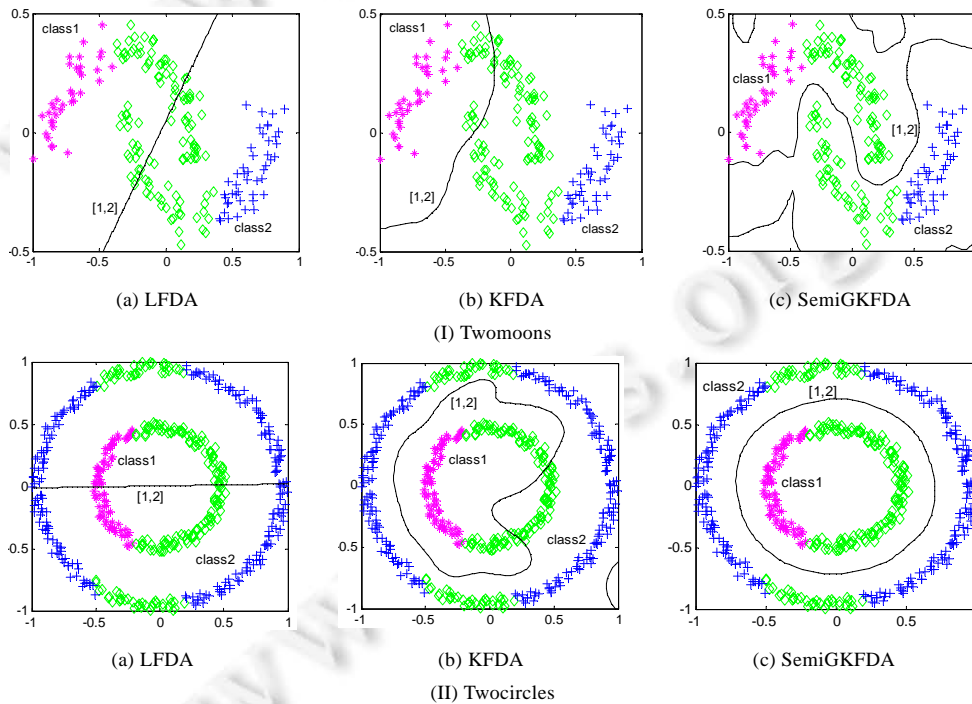


Fig.1 Discriminant interface of three two-class artificial datasets obtained by different FDA algorithms

图 1 不同 FDA 算法对 3 种两类人工数据集的判决边界

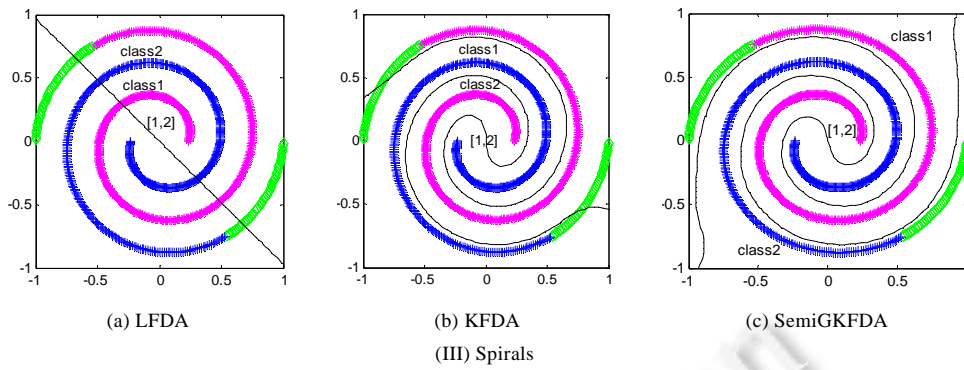


Fig.1 Discriminant interface of three two-class artificial datasets obtained by different FDA algorithms (Continued)

图1 不同 FDA 算法对 3 种两类人工数据集的判决边界(续)

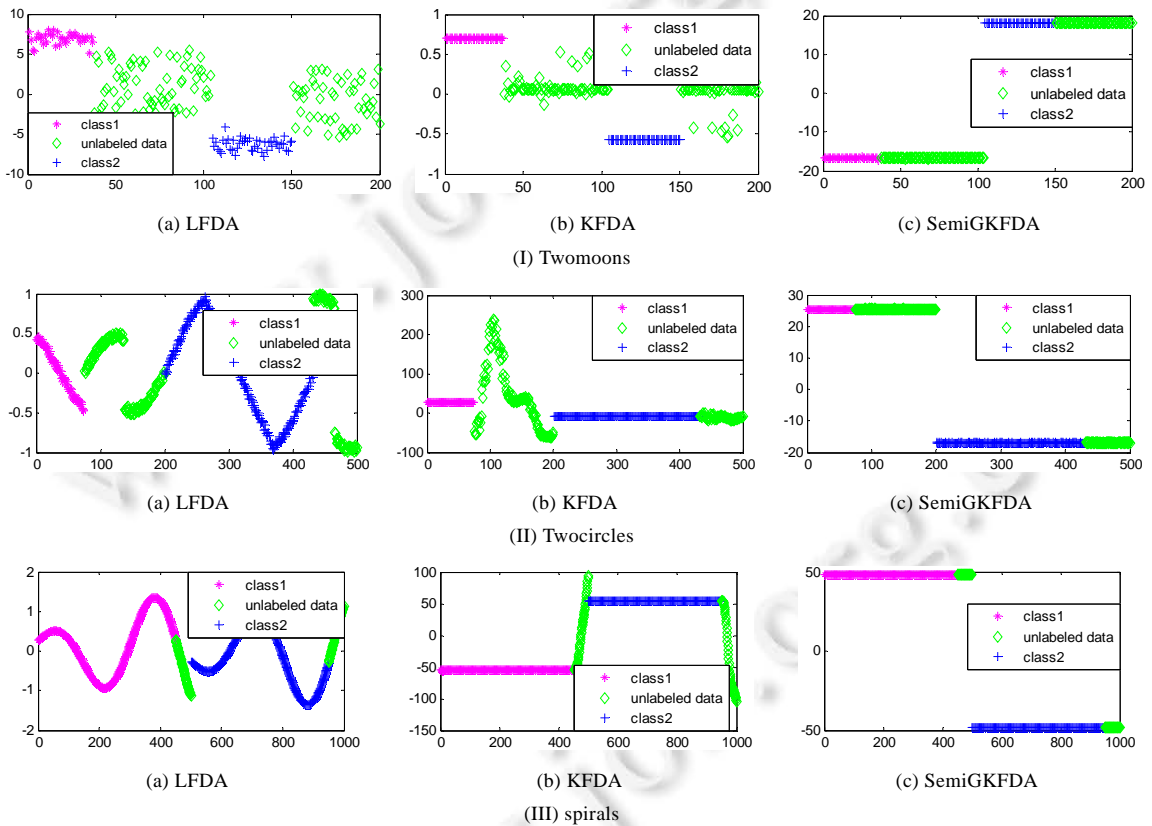


Fig.2 Coefficient distribution projected on the first projection vector of three two-class artificial datasets obtained by different FDA algorithms

图2 不同 FDA 算法对 3 种两类人工数据集的第 1 个投影向量系数分布

从图 1 的判决结果中可以看出,当面临线性可分问题时,LFDA 算法能够将有标签训练样本有效地区分开,如图 1(I)(a)所示;但当数据呈现非线性分布时,如图 1(II)(a)和图 1(III)(a)所示,LFDA 算法由于无法判别分析数据集中的非线性内在信息,因此无法实现 class1 和 class2 有标签训练样本的区分.KFDA 由于在映射后的核空间中

进行 Fisher 判别分析,能够有效地解决原始数据间的非线性相关问题,可实现非线性分布有标签样本的有效区分,如图 1(I)(b)、图 1(II)(b)和图 1(III)(b)所示.然而,进一步观察可知,KFDA 虽然能够解决 LFDA 在非线性的局限性,但是其作为一种有监督学习,只能使用少量有标签样本的特征向量进行学习,导致得到的判别边界只考虑了有标签样本的分布情况,无法实现对无标签样本的有效区分,即得到的判别边界不符合真实样本的分布情况.而本文算法由于采用半监督学习方式,合理利用无标签样本的信息指导有监督 KFDA 算法的学习,同时充分考虑了全局和局部空间一致性假设,使得到的判决边界更加符合数据的真实分布情况,如图 1(I)(c)、图 1(II)(c)和图 1(III)(c)所示.

图 2 显示了 3 种人工数据集经传统 LFDA 算法、KFDA 算法和本文提出的 SemiGKFDA 算法降维后在第一个投影向量系数的分布情况.观察图 2 可以进一步看出,除第 1 个数据集外,其他两个数据集的有标签样本经 LFDA 算法投影后的系数均存在严重重叠.该实验结果进一步验证了 LFDA 算法无法实现非线性数据有效区分的事实.观察 3 种人工数据集经 KFDA 算法投影后的系数分布情况可知,该算法有效地实现了 3 种数据集中两类有标签训练样本的区分,但由于 KFDA 算法只依赖于训练样本中存在的有标签样本,导致得到的投影向量无法反映真实的数据分布情况,使得大量的无标签样本经 KFDA 算法投影后的系数出现重叠,无法实现有效区分.相对而言,经本文提出的 SemiGKFDA 算法降维后,无论是有标签还是无标签样本均具有较大的类间距离和较小的类内距离,不存在类间重叠.该实验结果进一步说明,本文算法通过采用半监督学习方式,利用无标签样本指导降维,极大地提高了算法的判决性能.

为了显示地说明本文算法采用的距离测度能够有效地体现数据空间全局一致性与局部一致性的特点,实验中同时计算了 3 种人工数据在 SemiGKFDA 算法下的相似矩阵(如图 3 所示).其中,白色代表数据间有较高相似性,黑色代表数据间无相似性.图中方格顺序依次表示为 class1 有标签样本、class2 有标签样本、class1 无标签样本和 class2 无标签样本.从相似矩阵中可以明显看出,本文算法能够较好地表达出样本间的相似关系.这是由于本文算法采用低密度分割几何距离作为距离测度,在很好地反映全局结构一致性的同时,兼顾长度因素影响,使得空间位置相邻的样本点间相似度高,即满足局部空间一致性的要求.该实验结果可进一步说明,本文算法采用的距离测度能够有效地体现数据空间的全局一致性和局部一致性假设,从而使本文算法得到的相似矩阵更符合实际数据的分布情况.

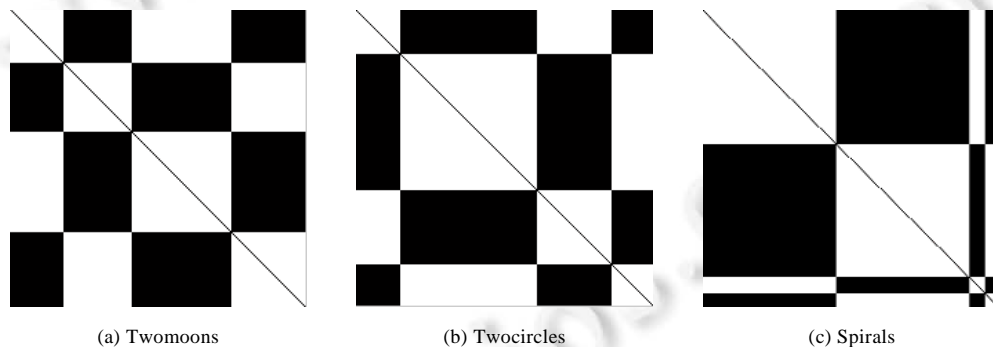


Fig.3 Similarity matrix of different artificial datasets obtained by SemiGKFDA algorithm

图 3 SemiGKFDA 算法对不同人工数据的相似矩阵

3.2.2 “多类样本”问题

为了考察本文算法解决多类样本判别问题的能力,采用 LFDA 算法、KFDA 算法和本文算法对 threecircles 人工数据样本进行对比实验.实验参数设置如下:共选取 300 个样本,其中标签样本 206 个,无标签样本 93 个,核函数选取为高斯核,核参数为 $\sigma=3$,近邻点选取为 $k=11$.判决边界图和在第 1 个投影向量的系数分布图分别如图 4、图 5 所示,其中,菱形表示为无标识样本.

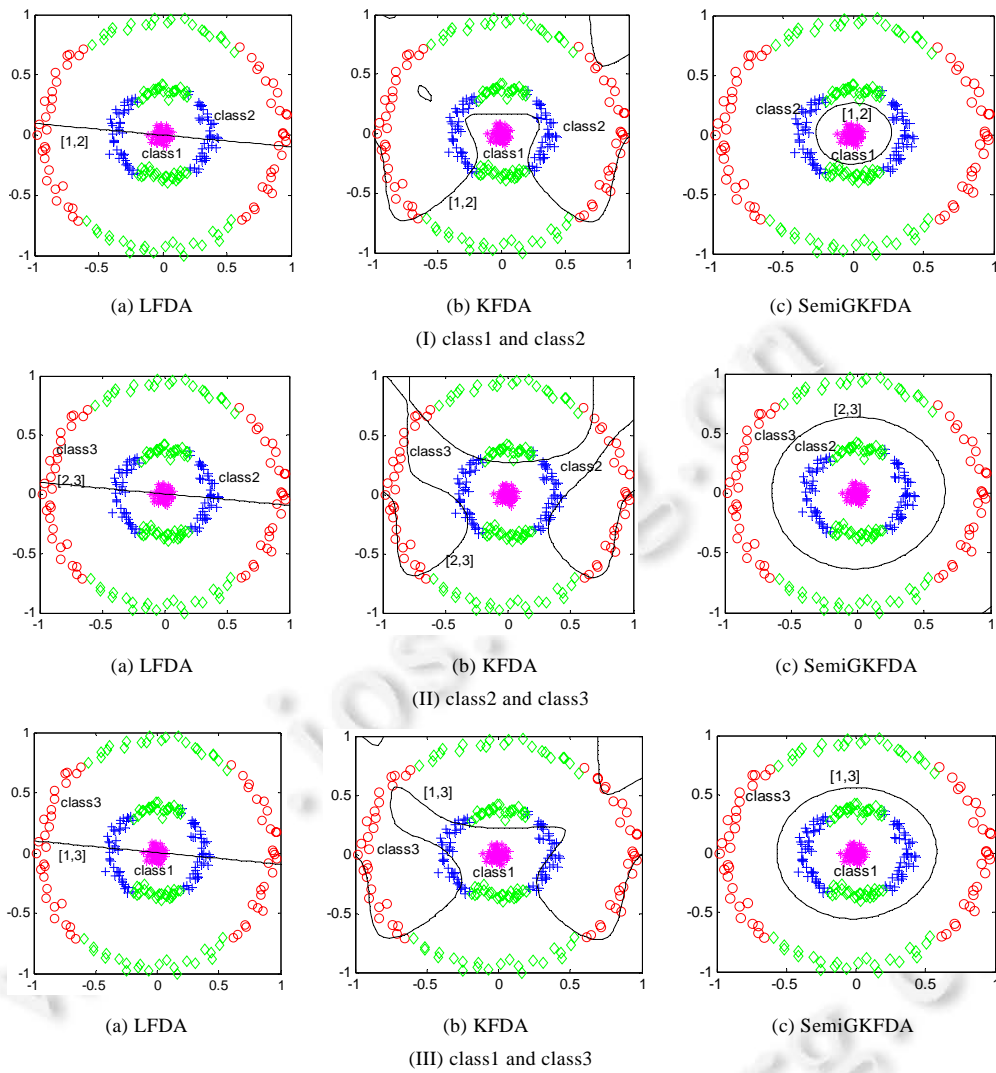


Fig.4 Discriminant interface of three multi-class artificial datasets obtained by different FDA algorithms

图 4 不同 FDA 算法对 3 个多类人工数据集的判决边界

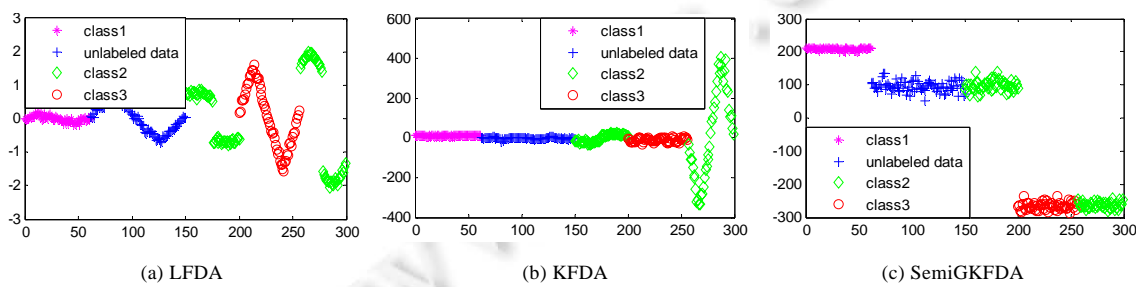


Fig.5 Coefficient distribution projected on the first projection vector of multi-class artificial datasets obtained by different FDA algorithms

图 5 不同 FDA 算法对多类人工数据集在第 1 个投影向量的系数分布

从不同类别样本间的判决边界图可以看出,当数据呈现非线性分布时,LFDA 算法无法判别分析数据集中的非线性结构信息,导致得到的判决边界无法对不同类别有标签样本进行有效区分.KFDA 算法是一种非线性特征提取方法,能够有效地解决 LFDA 在非线性空间中存在的局限性,进而使得到的判决边界部分实现了对不同类别有标签样本的区分.但是,由于 KFDA 仍属于有监督学习,只使用少量有标签样本训练,因此得到的判决边界很难具有好的泛化能力,即无法实现对无标签样本的区分,不符合实际的数据分布特征.从图 4(I)(c)、图 4(II)(c)和图 4(III)(c)可以看出,本文算法对不同类别的样本均实现了有效区分,其中包括有标签样本和无标签样本.这是由于本文算法在考虑空间一致性假设的前提下,利用无标签样本指导有标签样本学习,进而提高了算法的判决性能.

从图 5 投影向量系数分布情况可以明显看出,经过 LFDA 降维后,class1、class2 和 class3 之间的系数存在类间重叠,无法实现不同类别样本间的有效区分.KFDA 算法得到的判决界面可实现部分不同类别样本间的区分,然而通过观察图 5(I)(b)、图 5(II)(b)和图 5(III)(b)可知,不同类别有标签样本间的类间距离较小,泛化能力不强;同时,对无标签的样本无法实现区分.相对而言,经过本文提出的 SemiGKFDA 算法得到的不同类别样本投影向量系数具有较大的类间距离和较小的类内距离,其中包括有标签样本和无标签样本.这是由于本文算法采用半监督学习方式,兼顾了空间一致性假设,合理利用了无标签样本和有标签样本,使得到的投影向量更加具有鉴别能力.

为了显示地验证本文算法所采用的距离测度在多类别样本判别问题中体现数据空间全局一致性与局部一致性方面的优势,我们计算 threecircles 人工数据集在 SemiGKFDA 算法下的相似矩阵(如图 6 所示).其中,白色区域表示有相似性,黑色区域表示无相似性.

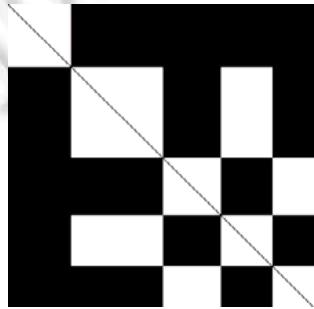


Fig.6 Similarity matrix of threecircles artificial datasets obtained by SemiGKFDA algorithm

图 6 SemiGKFDA 算法对 threecircles 人工数据的相似矩阵

从 threecircles 人工数据集的相似矩阵可以清晰地看出,样本集被分成 5 类,依次表示为 class1 有标签样本、class2 有标签样本、class3 有标签样本、class2 无标签样本以及 class3 无标签样本.由图 6 可知,由本文算法计算得到的相似矩阵与数据的实际分布情况一一对应.可见,本文提出的 SemiGKFDA 算法采用的距离测度能充分地体现出样本集全局空间一致性和局部空间一致性的特征.

3.3 Iris数据集实验性能对比

为了便于显示地观察本文算法对实际复杂结构数据的判决性能,采用 UCI 数据集中的 Iris 经典数据集进行实验,并与传统 LFDA 算法和 KFDA 算法进行对比.实验中,各选取 50 个 Iris 数据集中的 Iris-Setosa、Iris-Versicolor、Iris-Virginica 数据作为标签样本,共 150 有标签样本.无标签样本的个数设置为标签样本的 10%,即分别为 5 个.实验参数设置如下:核函数选取为高斯核,核参数为 $\sigma=0.7$,近邻点选取为 $k=6$.投影后的 2 维数据分布如图 7 所示.

由图中的实验结果可知,LFDA 基本没有实现不同类别样本间的区分.KFDA 虽然能够实现 Setosa 同其他两类样本的区分,但对于 Versicolor 和 Virginica 两类样本的区分则显得无能为力.相对而言,本文算法在不同类别

样本间区分能力上明显优于上述两种算法.甚至对原本重叠的 *Versicolor* 和 *Virginica* 两类样本也实现了有效区分.这是由于本文算法采用半监督学习方式,充分利用无标签样本的信息来指导有标签样本学习,同时利用低密度分割几何距离作为相似性度量,充分考虑空间的全局一致性和局部一致性假设,使得到的投影向量的鉴别能力大幅度提升.

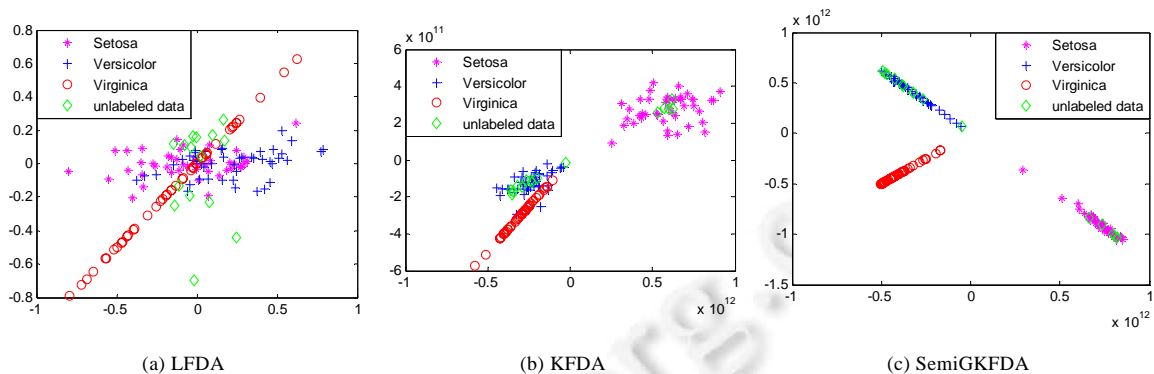


Fig.7 Two-dimensional projections data distribution of Iris dataset obtained by different methods

图7 不同算法对 Iris 数据集的 2 维投影数据分布图

为了量化地验证本文算法得到的投影向量判别分析的性能,将本文算法与 LFDA 算法和 KFDA 算法进行实验对比.为了保证实验具有统计意义,结果通过 10 次交叉验证策略获取,算法的分类性能采用分类错误率作为性能评价指标.其中,本文算法将测试样本作为无标签样本指导半监督学习.实验参数设置如下:本文算法采用的最优核参数与 KFDA 算法相同,核函数为 $\sigma=0.7$,伸缩因子为 $\rho=100$,近邻点选取为 $k=6$,相似度转换参数 $\delta=3$,降维维度为两维,分类器采用的是 1-NN 分类器.从图 8 中可以看出,本文算法的平均分类错误率要明显低于 LFDA 算法和 KFDA 算法.这是由于本文算法能够充分考虑到无标签样本具有的空间一致性假设信息,使得到的投影向量下的系数具有明显的区分性,从而极大地提高了以此为特征的分类器的分类精度.

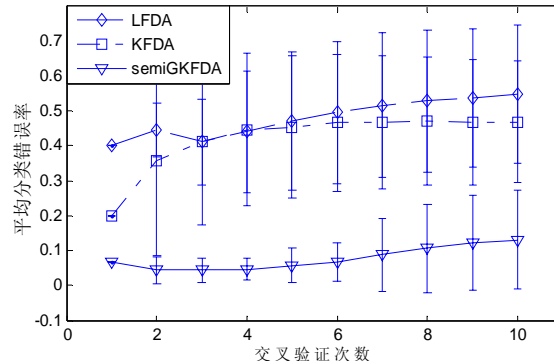


Fig.8 Classification error rate of different algorithms

图8 不同算法的分类错误率

3.4 UCI数据集实验性能对比

为了验证本文提出的基于低密度分割几何距离的半监督 KFDA 算法对不同结构数据集的判别分析能力,我们进行了下列实验.其中,实验数据均来源于国际机器学习标准数据库 UCI 中的 8 组不同的数据集,分别为 WINE、SEEDS、IONOSPHERE、GLASS、SPAMBASE、SONAR、HEART、VEHICLE 和 WPBC.

数据的特征信息见表 1.

Table 1 Description of experimental datasets**表 1** 实验数据集描述

数据集	属性	样本个数	分类数
WINE	13	178	3
SEEDS	7	210	3
IONOSPHERE	34	351	2
GLASS	10	214	6
SPAMBASE	57	4 601	2
SONAR	60	208	2
VEHICLE	18	846	4
WPBC	33	194	2

我们将本文算法与目前流行的判别分析算法进行对比,其中包括 LFDA 算法、KFDA 算法、基于格拉曼斯流形判别分析算法(semi-supervised discriminant analysis on Grassmannian manifold,简称 SemiGMDA)^[23]算法以及基于流形距离的半监督判别分析(semi-supervised discriminant analysis based on manifold distance,简称 SemiMDDA)^[24]算法.高斯核参数为 $\sigma=1$,伸缩因子为 $\rho=100$,其他参数同上.分类器为 1-NN 分类器,实验采用 10 次交叉验证,其中,每次产生的测试样本作为半监督算法的无标签样本进行学习.实验部分对比分析了不同算法的统计分布情况,结果见表 2.

Table 2 Comparison of classification performance among the different algorithms on different UCI datasets**表 2** 不同算法对不同 UCI 数据集的分类性能比较

	LFDA	KFDA	SemiGMDA	SemiMDDA	SemiGKFDA
Wine	0.7097±0.0829	0.6499±0.0587	0.5951±0.0691	0.6439±0.0666	0.5277±0.0207
Glass	0.7801±0.1693	0.6068±0.1721	0.5609±0.1553	0.5563±0.1769	0.4946±0.1340
Ionosphere	0.3314±0.0987	0.4240±0.1325	0.3681±0.1392	0.3762±0.1268	0.1978±0.1212
Seeds	0.5047±0.0784	0.7000±0.1209	0.4684±0.0857	0.4482±0.0807	0.2809±0.0759
Sonar	0.4473±0.1056	0.4939±0.1119	0.3831±0.1396	0.4524±0.1166	0.3831±0.1396
Spambase	0.4400±0.0614	0.4300±0.0856	0.3807±0.0752	0.3996±0.0570	0.3270±0.0553
Vehicle	0.7199±0.0588	0.6940±0.0917	0.6240±0.0572	0.6652±0.0659	0.6654±0.0561
Wpbc	0.3550±0.1141	0.3850±0.0579	0.2322±0.1225	0.2926±0.1025	0.2889±0.1239

通过实验结果可以发现,LFDA 和 KFDA 的分类错误率较高.这是由于 LFDA 和 KFDA 属于有监督学习,只依赖有标签的样本进行学习,泛化能力较弱.相对而言,采用基于距离测度的半监督学习方式的 SemiGMDA 算法和 SemiMDDA 算法的分类错误率有所降低.这是由于两种算法采用半监督学习方式,利用无标签样本的信息指导有监督 KFDA 算法的学习提升了投影向量的鉴别能力.然而,从统计结果对比来看,本文提出算法的平均分类错误率在 3 种半监督学习算法中最低.这是由于 SemiGMDA 采用的距离测度虽能够较好地满足全局一致性的要求,却没有考虑局部一致性特征;SemiGMDA 采用的距离测度虽充分考虑了数据的局部一致性特征,但对全局一致性特征兼顾较少.而本文算法通过采用低密度分割距离测度能同时考虑到样本集合空间的全局一致性和局部一致性假设的要求,使得相似度矩阵更符合实际的数据分布特征,算法提取的投影向量具有更高的鉴别能力,从而极大地提升了以此为特征的分类器的分类性能.

3.5 人脸识别实验

为了验证本文提出的基于低密度分割几何距离半监督 KFDA 算法对高维特征数据降维后的判别性能,我们采用 Yale 人脸数据集进行测试,该数据库由耶鲁大学计算视觉与控制中心创建,包含 15 位志愿者的 165 幅图片,每个志愿者包含 11 幅在不同光照、不同表情和不同姿态下拍摄得到的图片.图片的灰度值尺度化到 $[0,1]$,图像大小压缩到 32×32 像素.实验中随机取每个志愿者的 6 幅图片组成训练数据,剩余为测试样本.分别采用 PCA、LPP、LLE、KFDA、SemiGMDA、SemiMDDA 和本文提出的 SemiGKFDA 通过最近邻分类器对训练样本进行分类识别,比较它们在不同维度下的识别错误率,结果如图 9 所示.为了方便比较,LPP 和 LLE 算法的 k 近邻参数和核函数与本文算法设置相同.

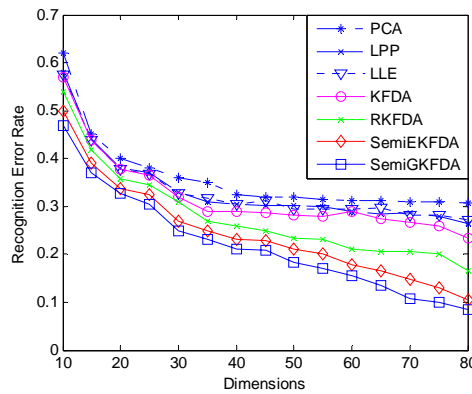


Fig.9 Recognition error rate on the Yale database

图9 Yale 数据库上的识别错误率

为了验证本文算法特征提取的优势,我们采用 k -NN 分类器、SVM 分类器和 SRC 分类器对 Yale 数据库的人脸样本集合进行识别.由于每个个体的训练样本的个数为 6,根据个体的训练样本个数,我们将 k -NN 分类器的 k 值设置为 6.SVM 算法参数为高斯核,针对不同的算法组合,根据 grid search 网格搜索法确定惩罚因子 C 和高斯核参数.从上述实验结果图 9 中可以发现,算法的识别错误率在 80 维时最低,降维效果最好.因此,为了保持降维效果,同时减少复杂度,在本实验中,我们统一将降维的维度设置为 80.实验结果与 PCA、LPP、LLE、KFDA、SemiGMDA 以及 SemiMDDA 进行比较,结果见表 3.

Table 3 Face recognition error rate of different feature extraction methods combined with different classifiers on Yale database

表 3 不同特征提取方法和分类器组合的 Yale 数据库的人脸识别错误率

	PCA	LPP	LLE	KFDA	SemiGMDA	SemiMDDA	SemiGKFDA
k -NN	0.291 3	0.262 2	0.246 3	0.215 7	0.186 5	0.199 97	0.125 5
SVM	0.170 2	0.099 9	0.090 5	0.087 3	0.087 9	0.092 7	0.029 1
SRC	0.050 1	0.044 7	0.043 7	0.042 3	0.038 3	0.056 7	0.026 5

从表中可以清晰地看到,无论与哪一种分类器组合,本文算法特征提取后的识别错误率相对于 PCA、LPP、LLE、KFDA、SemiGMDA 以及 SemiMDDA 方法均有明显的降低.由此可以看出,对处理光照、表情、姿态等变化比较大的人脸库时,本文提出的算法是有效的.此外可以发现,不同特征提取方法与 SRC 分类组合得到的识别效果更好.就本文算法而言,与 SRC 分类器组合时的错误率比 SVM 分类器降低 0.3%,比 K -NN 分类器降低 1%.这是由于 SRC 分类器能够在低维特征空间中保持样本的稀疏结构.

4 结 论

本文提出一种基于低密度分割几何距离半监督 KFDA 算法,结合实验得到以下结论.

- (1) 本文采用半监督学习方式,在考虑一致性假设的前提下,合理利用无标签样本的信息指导有监督 KFDA 算法的学习,使得到的投影向量更加具有鉴别能力.仿真数据和 IRIS 数据集实验结果表明,本文算法得到的判决结果更加符合数据的真实分布情况.
- (2) 为了进一步提升算法的分类性能,通过引入低密度分割几何距离作为相似性度量,考虑到样本集合的空间全局一致性和局部一致性假设,使高维空间中的邻域样本在降维空间中仍然保持流形体结构.实验部分通过与 FDA、KFDA、SemiGMDA 以及 SemiMDDA 算法进行对比的结果表明,经本文算法提取的特征能够有效地提高分类器的分类性能.
- (3) 通过在 Yale 人脸数据库上的实验表明,相对于其他 FDA 算法,本文算法对高维特征数据降维后能够

取得更高的识别率.同时,与不同分类器组合的实验结果表明,本文算法采用的低密度分割几何距离更能降低基于距离算法的分类错误率.实验中发现, k 近邻的个数选择和核参数的设置对算法性能影响较大.因此,如何选择最优 k 近邻图以及核参数来提升算法的鉴别性能,将是本课题下一阶段研究的重点.

References:

- [1] Puthenputhussery A, Liu QF, Liu CJ. A sparse representation model using the complete marginal Fisher analysis framework and its applications to visual recognition. *IEEE Trans. on Multimedia*, 2017,19(8):1757–1770. [doi: 10.1109/TMM.2017.2685179]
- [2] Zheng JW, Wang WL, Yao XM, Shi HY. Face recognition using tensor local fisher discriminant analysis. *Acta Automatica Sinica*, 2012,38(9):1485–1495 (in Chinese with English abstract).
- [3] Wu L, Shen CH, Van HA. Deep linear discriminant analysis on fisher networks: A hybrid architecture for person re-identification. *Pattern Recognition*, 2017,65:238–250. [doi: 10.1016/j.patcog.2016.12.022]
- [4] He JR, Ding LX, Li ZK, Hu QH. Margin discriminant projection for dimensionality reduction. *Ruan Jian Xue Bao/Journal of Software*, 2014,25(4):826–838 (in Chinese with English abstract). <http://www.jos.org.cn/1000-9825/4571.htm> [doi: 10.13328/j.cnki.jos.004571]
- [5] Ye HH, Li YJ, Chen C, Zhang ZH. Fast Fisher discriminant analysis with randomized algorithms. *Pattern Recognition*, 2017,72: 82–92. [doi: 10.1016/j.patcog.2017.06.029]
- [6] Dieu T, Bui NDH. A Bayesian framework based on a Gaussian mixture model and radial-basis-function Fisher discriminant analysis (BayGmmKdaV1.1) for spatial prediction of floods. *Geoscientific Model Development*, 2017,10(9):3391–3409. [doi: 10.5194/gmd-10-3391-2017]
- [7] Hoai ALT, Duy NP. DC programming and DCA for sparse Fisher linear discriminant analysis. *Neural Computing & Applications*, 2017,28(9):2809–2822. [doi: 10.1007/s00521-016-2216-9]
- [8] Hu WP, Hu HF. Heterogeneous face recognition based on modality-independent kernel Fisher discriminant analysis joint sparse auto-encoder. *Electronics Letters*, 2016,52(21):1753–1755. [doi: 10.1049/el.2016.2661]
- [9] Fan ZZ, Xu Y, Ni M. Individualized learning for improving kernel Fisher discriminant analysis. *Pattern Recognition*, 2016,58: 100–109. [doi: 10.1016/j.patcog.2016.03.029]
- [10] Dong SQ, Wang ZZ, Zeng LB. Lithology identification using kernel Fisher discriminant analysis with well logs. *Journal of Petroleum Science and Engineering*, 2016,143:95–102. [doi: 10.1016/j.petrol.2016.02.017]
- [11] Boussaad L, Benmohammed M, Benzid R. Age invariant face recognition based on DCT feature extraction and kernel Fisher analysis. *Journal of Information Processing Systems*, 2016,12(3):392–409.
- [12] Yin Y, Hao YF, Bai Y, Yu HC. A Gaussian-based kernel Fisher discriminant analysis for electronic nose data and applications in spirit and vinegar classification. *Journal of Food Measurement and Characterization*, 2017,11(1):24–32. [doi: 10.1007/s11694-016-9367-3]
- [13] Ge ZQ, Zhong SY, Zhang YW. Semisupervised kernel learning for FDA model and its application for fault classification in industrial processes. *IEEE Trans. on Industrial Informatics*, 2016,12(4):1403–1411. [doi: 10.1109/TII.2016.2571680]
- [14] Wang GQ, Shi NF, Shu YX. Embedded manifold-based kernel Fisher discriminant analysis for face recognition. *Neural Processing Letters*, 2016,43(1):1–16. [doi: 10.1007/s11063-014-9398-x]
- [15] Shi HT, Liu JC, Wu YH, Zhang K, Zhang LX, Xue P. Fault diagnosis of nonlinear and large-scale processes using novel modified kernel Fisher discriminant analysis approach. *Int'l Journal of Systems Science*, 2016,47(5):1095–1109. [doi: 10.1080/00207721.2014.912780]
- [16] Liu YH, Wang SH, Hu MR. A self-paced P300 healthcare brain-computer interface system with SSVEP-based switching control and kernel FDA + SVM-based detector. *Applied Sciences*, 2016,6(5):Article No.142. [doi: 10.3390/app6050142]
- [17] Van M, Kang HJ. Bearing defect classification based on individual wavelet local Fisher discriminant analysis with particle swarm optimization. *IEEE Trans. on Industrial Informatics*, 2016,12(1):124–135. [doi: 10.1109/TII.2015.2500098]
- [18] Jiang L, Xuan JP, Shi TL. Feature extraction based on semi-supervised kernel marginal Fisher analysis and its application in bearing fault diagnosis. *Mechanical Systems and Signal Processing*, 2013,41(1-2):113–126. [doi: 10.1016/j.ymssp.2013.05.017]

- [19] Blaschko MB, Shelton JA, Bartels A. Semi-supervised kernel canonical correlation analysis with application to human fMRI. *Pattern Recognition*, 2011,32(11):1572–1583. [doi: 10.1016/j.patrec.2011.02.011]
- [20] Wang Y, Liu HL, Su WJ. Face recognition algorithm based on kernel orthogonal semi-supervised discriminant analysis. *Computer Engineering and Applications*, 2014,50(12):120–124 (in Chinese with English abstract).
- [21] Jiang L, Xuan JP, Shi TL. Feature extraction based on semi-supervised kernel marginal Fisher analysis and its application in bearing fault diagnosis. *Mechanical Systems and Signal Processing*, 2013,41(1-2):113–126. [doi: 10.1016/j.ymsp.2013.05.017]
- [22] Huang SC, Tang, YC, Lee CW, Chang MJ. Kernel local Fisher discriminant analysis based manifold-regularized SVM model for financial distress predictions. *Expert Systems with Applications*, 2011,39(3):3855–3861. [doi: 10.1016/j.eswa.2011.09.095]
- [23] Dhamecha TI, Singh R, Vatsa M. On incremental semi-supervised discriminant analysis. *Pattern Recognition*, 2016,52: 135–147. [doi: 10.1016/j.patcog.2015.09.030]
- [24] Jiang W, Lu Y, Yang BR. Semi-supervised discriminant analysis on Grassmannian manifold. *Journal of Image and Graphics*, 2013, 18(8):944–952 (in Chinese with English abstract).
- [25] Wei L, Wang SJ. Semi-supervised discriminant analysis based on manifold distance. *Ruan Jian Xue Bao/Journal of Software*, 2010, 21(10):2445–2453 (in Chinese with English abstract). <http://www.jos.org.cn/1000-9825/3629.htm> [doi: 10.3724/SP.J.1001.2010.03629]
- [26] Lu CH, Xiao SQ, Gu XF. Hyperplane distance neighbor clustering based on local discriminant analysis for complex chemical processes monitoring. *Korean Journal of Chemical Engineering*, 2014,31(11):1943–1953. [doi: 10.1007/s11814-014-0153-9]
- [27] Jing XY, Wu F, Dong XW. An improved SDA based defect prediction framework for both within-project and cross-project class-imbalance problems. *IEEE Trans. on Software Engineering*, 2017,43(4):321–339. [doi: 10.1109/TSE.2016.2597849]

附中文参考文献:

- [2] 郑建伟,王万良,姚晓敏,石海燕,张量局部 Fisher 判别分析的人脸识别. *自动化学报*, 2012,38(9):1485–1495.
- [4] 何进荣,丁立新,李照奎,胡庆辉.基于边界判别投影的数据降维. *软件学报*, 2014,25(4):826–838. <http://www.jos.org.cn/1000-9825/4571.htm> [doi: 10.13328/j.cnki.jos.004571]
- [20] 王燕,刘花丽,苏文君.基于核正交半监督鉴别分析的人脸识别算法. *计算机工程与应用*, 2014,50(12):120–124.
- [24] 姜伟,陆瑶,杨炳儒.格拉斯曼流形上的半监督判别分析. *中国图像图形学报*, 2013,18(8):944–952.
- [25] 魏莱,王守觉.基于流形距离的半监督判别分析. *软件学报*, 2010,21(10):2445–2453. <http://www.jos.org.cn/1000-9825/3629.htm> [doi: 10.3724/SP.J.1001.2010.03629]

附录 A

(1) $\tilde{\mathbf{K}}_{ij} = \Psi(\mathbf{x}_i)^T \Psi(\mathbf{x}_j)$ 的求法如下:

设 $\mathbf{I}_m \in \mathbf{R}^{n \times n}$, $\mathbf{I}_{ij} = 1, i=1, 2, \dots, n, j=1, 2, \dots, n$:

$$\left. \begin{aligned} \tilde{\mathbf{K}}_{ij} &= \Psi(\mathbf{x}_i)^T \Psi(\mathbf{x}_j) = \left(\phi(\mathbf{x}_i) - \frac{1}{n} \sum_{l=1}^n \phi(\mathbf{x}_l) \right)^T \left(\phi(\mathbf{x}_j) - \frac{1}{n} \sum_{k=1}^n \phi(\mathbf{x}_k) \right) \\ &= \phi(\mathbf{x}_i)^T \phi(\mathbf{x}_j) - \frac{1}{n} \sum_{l=1}^n \phi(\mathbf{x}_i)^T \phi(\mathbf{x}_l) - \frac{1}{n} \sum_{k=1}^n \phi(\mathbf{x}_k)^T \phi(\mathbf{x}_j) + \frac{1}{n^2} \sum_{l,k=1}^n \phi(\mathbf{x}_l)^T \phi(\mathbf{x}_k) \\ &= \mathbf{K}_{ij} - \frac{1}{n} \sum_{l=1}^n \mathbf{I}_{ij} \mathbf{K}_{il} - \frac{1}{n} \sum_{k=1}^n \mathbf{K}_{ik} \mathbf{I}_{kj} + \frac{1}{n^2} \sum_{l,k=1}^n \mathbf{I}_{il} \mathbf{K}_{lk} \mathbf{I}_{kj} \end{aligned} \right\} \quad (\text{A1})$$

其中, $\mathbf{K}_{ij} = \phi(\mathbf{x}_i)^T \phi(\mathbf{x}_j)$, \mathbf{K} 为核矩阵, 所以,

$$\tilde{\mathbf{K}} = \mathbf{K} - \mathbf{I}_n \mathbf{K} - \mathbf{K} \mathbf{I}_n + \mathbf{I}_n \mathbf{K} \mathbf{I}_n, \mathbf{I}_n = \frac{1}{n} \mathbf{I} \quad (\text{A2})$$

(2)

$$\begin{aligned} \mathbf{K}_{ij}^{test} &= \Psi(\mathbf{x}_i)^T \Psi(\mathbf{t}_j) = \left(\phi(\mathbf{x}_i) - \frac{1}{m} \sum_{l=1}^m \phi(\mathbf{x}_l) \right)^T \left(\phi(\mathbf{t}_j) - \frac{1}{m} \sum_{k=1}^m \phi(\mathbf{x}_k) \right) \\ &= \phi(\mathbf{x}_i)^T \phi(\mathbf{t}_j) - \frac{1}{m} \sum_{l=1}^m \phi(\mathbf{x}_l)^T \phi(\mathbf{t}_j) - \frac{1}{m} \sum_{k=1}^m \phi(\mathbf{x}_i)^T \phi(\mathbf{x}_k) + \frac{1}{m^2} \sum_{l,k=1}^m \phi(\mathbf{x}_l)^T \phi(\mathbf{x}_k) \end{aligned} \quad (A3)$$

训练样本集 $\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_m\}$, out-of-sample 新样本集合 $\{\mathbf{t}_1, \mathbf{t}_2, \dots, \mathbf{t}_L\}$, L 为新样本集合的个数:

$$\tilde{\mathbf{K}}^{test} = \mathbf{K}_{ij}^{test} - \frac{1}{m} \sum_{l=1}^m \mathbf{I}_{il} \mathbf{K}_{ij}^{test} - \frac{1}{m} \sum_{k=1}^m \mathbf{K}_{ik} \mathbf{I}_{ik} + \frac{1}{m^2} \sum_{l,k=1}^m \mathbf{I}_{il} \mathbf{K}_{ik} \mathbf{I}_{kj} \quad (A4)$$

其中, $\mathbf{K}_{ij}^{test} = \phi(\mathbf{x}_i)^T \phi(\mathbf{t}_j)$, \mathbf{K} 为核矩阵.

所以得到 $\tilde{\mathbf{K}}^{test}$ 的表达式为

$$\tilde{\mathbf{K}}^{test} = \mathbf{K}^{test} - \frac{1}{m} \mathbf{I}_m \mathbf{K}^{test} - \frac{1}{m} \mathbf{K} \mathbf{I}_{m \times L} + \frac{1}{m^2} \mathbf{I}_m \mathbf{K} \mathbf{I}_{m \times L} \quad (A5)$$

$\mathbf{I}_m \in \mathbf{R}^{m \times m}$ 和 $\mathbf{I}_{m \times L} \in \mathbf{R}^{m \times L}$ 每一个元素都是 1.



陶新民(1973—),男,安徽蚌埠人,博士,教授,博士生导师,主要研究领域为人工智能,大数据挖掘,智能优化.



王若彤(1994—),女,博士生,主要研究领域为人工智能,聚类分析.



常瑞(1995—),女,博士生,CCF 专业会员,主要研究领域为人工智能,大数据挖掘,降维算法.



李晨曦(1993—),女,博士生,主要研究领域为故障诊断.



沈微(1977—),男,博士,讲师,主要研究领域为数据分析,物流系统规划与管理咨询,系统建模与优化.