

一种新的保类内核 Fisher 判别法及说话人辨别应用

郑建炜 王万良

(浙江工业大学信息学院 杭州 310023)

摘 要 在保留数据本质特征的前提下,降低数据维度是一种重要的分类预处理手段。深入分析了核 Fisher 判别(KFD)方法与核化全局局部保持 Fisher 投影(KLFDA)方法的相互关系与优缺点,提出了一种新的基于类内特性保持的核化 Fisher 判别分析方法(LW-KFD)。在保留 KFD 全局最优投影能力的同时,解决了 KLFDA 的过度局部保持问题,从而对重叠(离群)样本与多态分簇样本都能实现有效的分类投影。提出了快速训练算法,解决了大量训练样本时的内存溢出问题。仿真实验与说话人辨别应用表明,该方法具有很强的适应性,并提高了说话人识别率与识别速度。

关键词 Fisher 判别分析,局部保持投影,说话人辨别,核技巧,维度削减

中图法分类号 TP391.4 **文献标识码** A

Novel Local Within-class Features Preservation Kernel Fisher Discriminant Algorithm and Applied in Speaker Identification

ZHENG Jian-wei WANG Wan-liang

(School of Information, Zhejiang University of Technology, Hangzhou 310023, China)

Abstract Dimensionality reduction without losing intrinsic information on original data is an important technique for succeeding tasks such as classification. A novel local within-class features preservation kernel fisher discriminant algorithm was proposed after deeply analyzing the relationship between kernel fisher discriminant and kernel local fisher projection. The new method keeps the ability of KFD's global projection and solves the over-fitting of KLFDA's local preservation problem, which can work well on overlapped or multimodal labeled data. The training algorithm is improved for resolving out-of-memory problem when applied in large sample situation. The simulation and speaker identification application show that the proposed algorithm has more adaptability as well as advanced recognition rate and speed.

Keywords Fisher discriminant analysis, Local preservation projection, Speaker identification, Kernel trick, Dimensionality reduction

对高维数据进行有效特征提取,在降低维度的同时保留其本质信息,能很好地应用于分类任务^[1]。本文主要针对有监督降维问题,并应用于说话人辨别领域。说话人辨别是说话人识别的一个重要分支。利用包含在说话人语音波形中特有的个体信息,自动分析某一说话人属于语料库中哪一个身份的过程,为非线性判别问题。这在网络安全、银行证券系统、公安等领域有着广泛的应用前景。

核 Fisher 判别技术^[2]是基于 Fisher 线性判别提出的一种非线性降维方法,其基本思想是首先把样本映射到更高维的特征空间,然后在此特征空间进行 Fisher 线性判别。这样,在输入空间就隐含地产生了很强的非线性判别能力,使得样本在投影后达到最大的类间离散度和最小的类内离散度,亦即寻找一组最佳鉴别矢量,构成子空间,以原始样本在该子空间内的投影矢量作为鉴别特征用于分类识别。Belhumeur^[3], Yang^[4]等用其来解决人脸识别问题,邢玉娟^[5]、谭萍^[6]等将之应用于说话人识别领域。但是, KFD 方法应用于

具有类内分簇的数据时,效能急剧下降,而语音数据正好具有多态性,所以直接应用 KFD 于说话人识别,效果不是最优。

保持原信号的局部结构,对于多态性数据准确降维非常重要。局部保持投影(LPP)^[7]依据这个原理进行降维操作。LPP 在保持原空间数据对结构的前提下,寻求嵌入变换,降低维度而保留局部特性。然而, LPP 是一种无监督维度削减方法, Sugiyama^[8]将 LPP 思想应用到 KFD 中去,在最大化 Fisher 准则公式的同时,保持了类内、类间散度矩阵的局部特性,能对多态性数据进行很好的投影分类,但对于重叠样本却由于过多考虑局部特性而抑制了全局最优能力,性能有所下降。

本文分析了全局投影与局部保持的优势与弊端,提出了类内特性保持的核化 Fisher 判别分析方法(LWFD),该方法对于重叠样本与类内分簇样本都能进行准确的降维投影,引入核技巧^[9],可将之转变到非线性应用场合,如说话人识别。基于核函数的学习算法,其计算复杂度会随着样本规模的增

到稿日期:2009-08-10 返修日期:2009-11-01 本文受国家自然科学基金项目(60573123)资助。

郑建炜(1982—),男,博士生,主要研究方向为语音识别, E-mail: weibaby2007@163.com; 王万良(1957—),男,博士,教授,博士生导师,主要研究方向为计算机控制与智能化、人工智能及其应用。

大而急剧变大。在说话人辨别中,语音数据量大,直接使用 Fisher 准则公式进行判别,存在计算复杂度太大的问题,从而导致系统的实时性差,甚至内存溢出。基于此,提出了改进的训练算法来实现大数据量应用场合,使之可以成功应用于说话人辨别。

本文首先概述降维分类方式并介绍典型算法 KFD 与 LPP;然后提出类内保持核 Fisher 判别分析方法,并依据大样本训练数据,提出改进的训练算法;分别从仿真数据与说话人辨别两个方面验证所提算法的有效性;最后对本文工作进行总结,并展望后续工作重点。

1 降维分类

设定 $x_i \in R^d (i=1,2,\dots,N)$ 是 d 维样本数据, $y_i \in \{1,2,\dots,c\}$ 是相应的类别标签,其中 N 是样本总数, c 是类别总数。 c_l 是类 l 中的样本数,则

$$\sum_{l=1}^c c_l = N$$

X 是样本矩阵,即

$$X = (x_1 | x_2 | \dots | x_N)$$

降维分类的首要步骤是将原数据 x_i 投影为 $y_i \in R^r (1 \leq r \leq d)$,其中 r 是削减后的维度。采用变换矩阵 T 的投影公式

$$y_i = T^T x_i$$

式中,上标 T 为转置符号。原 c 类数据空间的分类投影维度为 $c-1$,之后求取每一类投影后的数据的中心值并规范化。将待分类数据投影到子空间并规范化后,计算其与子空间中的每一类数据中心点的欧氏距离,最近者判为识别结果。如原数据为二维两类信号,则先投影成一维数据,并计算两类数据各自的中心点。待分类数据投影成一维并规范化后,计算与两个中心点的欧氏距离,最近者判为识别类。

2 核 Fisher 判别法

核 Fisher 准则鉴别分析已被公认为特征抽取的最好方法之一^[10]。非线性映射的思想是将原始空间线性不可分的样本经过一个非线性映射函数投影到高维空间,转变成高维空间的线性可分情况^[11,12],即

$$\Phi: R^d \rightarrow H \quad x \mapsto \phi(x) \quad (1)$$

式(1)表示将一个低维空间 R^d 的原始样本通过非线性映射 Φ 映射到高维空间 H 。利用核技巧,算法实现的时候无需知道具体的映射函数,只要进行内积运算即可。

$$K(x, y) = (\Phi(x) \cdot \Phi(y))$$

高维空间 H 中再利用 Fisher 鉴别准则得到最优鉴别矢量集,此时就相当于得到了原始空间的非线性最优鉴别矢量集。高维空间 H 中的 Fisher 鉴别准则如下:

$$J^*(\varphi) = \frac{\varphi^T S_b^* \varphi}{\varphi^T S_w^* \varphi}, \varphi \neq 0 \quad (2)$$

式中,

$$S_b^* = \frac{1}{N} \sum_{l=1}^c c_l (\bar{\phi}_l - \bar{\phi})(\bar{\phi}_l - \bar{\phi})^T \quad (3)$$

$$S_w^* = \frac{1}{N} \sum_{l=1}^c \sum_{i,j=1}^{c_l} (\varphi(x_i^l) - \bar{\phi}_l)(\varphi(x_j^l) - \bar{\phi}_l)^T \quad (4)$$

分别代表了高维空间 H 中的类间散度矩阵和类内散度矩阵; $\bar{\phi}_l$ 是第 l 类样本在高维空间的平均值, $\bar{\phi}$ 代表高维空间所有样本的平均值。根据再生核空间理论^[12],式(2)中的 $\varphi \in H$

可表达成

$$\varphi = \sum_{i=1}^N \alpha_i \phi(x_i) \quad (5)$$

则式(2)优化准则等价于表达式

$$J(\alpha) = \frac{\alpha^T K_b \alpha}{\alpha^T K_w \alpha} \quad (6)$$

核类内散度矩阵 K_w 和核类间散度矩阵 K_b 定义为

$$K_b = \frac{1}{N} \sum_{l=1}^c c_l (\mu_l - \mu_0)(\mu_l - \mu_0)^T \quad (7)$$

$$K_w = \frac{1}{N} \sum_{l=1}^c \sum_{i,j=1}^{c_l} (\eta_{x_i^l} - \mu_l)(\eta_{x_j^l} - \mu_l)^T \quad (8)$$

式中,

$$\eta_x = (k(x_1, x), \dots, k(x_N, x))^T \quad (9)$$

$$\mu_l = (\frac{1}{c_l} \sum_{m=1}^{c_l} k(x_1, x_m), \dots, \frac{1}{c_l} \sum_{m=1}^{c_l} k(x_1, x_m))^T \quad (10)$$

$$\mu_0 = (\frac{1}{N} \sum_{m=1}^N k(x_1, x_m), \dots, \frac{1}{N} \sum_{m=1}^N k(x_1, x_m))^T \quad (11)$$

根据式(6)求得最佳核鉴别矢量 $\alpha_1, \alpha_2, \dots, \alpha_N$,则特征空间 H 中最佳鉴别矢量集构成的投影矩阵为

$$\varphi_{opt} = (\varphi_1, \varphi_2, \dots, \varphi_N) \quad (12)$$

至此,待识别向量 x 的核 Fisher 判别特征为 $v = (\varphi_1 \cdot \varphi(x)) = \sum_{i=1}^N \alpha_i k(x_i, x)$ 。

3 保局部投影法

LPP^[14]是一种无监督特征降维提取工具,利用训练样本之间的相关性进行投影。设 A 是一个亲和矩阵,其中元素 $A_{i,j}$ 代表样本 x_i 与 x_j 之间的亲和力,设定 $A_{i,j} \in [0, 1]$ 。当 x_i 与 x_j 越接近时, $A_{i,j}$ 就越大,反之相反。LPP 变换矩阵 T_{LPP} 定义为

$$T_{LPP} \equiv \argmin_{T \in R^{d \times r}} (\frac{1}{2} \sum_{i,j=1}^n A_{i,j} ||T^T x_i - T^T x_j||^2)$$

使得

$$T^T X D X^T T = I, \quad (13)$$

式中, D 是 n 维对角矩阵。第 i 个对角元素为

$$D_{i,i} = \sum_{j=1}^n A_{i,j} \quad (14)$$

从式(13)可以看出, LPP 旨在寻找变换矩阵 T 使得原空间中的近邻数据对在变换子空间中仍能保持相邻,其最优投影向量 $\{\Psi_k\}_{k=1}^d$ 是特征值 $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_d$ 相应的特征向量,由下式获得。

$$X(D - A)X^T \Psi = \lambda X D X^T \Psi \quad (15)$$

即

$$T_{LPP} = (\Psi_d | \Psi_{d-1} | \dots | \Psi_{d-r+1}) \quad (16)$$

Sugiyama^[8]将 LPP 的亲和矩阵思想应用到 Fisher 的类内、类间散度矩阵,提出全局特征保持的 Fisher 判别方法(LFDA),该方法对多态性数据能进行很好的特征提取,但对重叠数据却表现出较弱的性能,其算法计算复杂度与内存需求随训练样本的增长而剧增,不适用于现实中的多样本应用问题。

4 类内保持核 Fisher 判别

现实分类数据很多兼具类内簇与重叠样本的情况。如说话人识别应用,说话人音高、音色、音量在不同时期呈现出丰富多彩的表现形式,语音特征参数往往具有多态性,离群样

本也比皆是。直接应用 KFD 或者 LPP 方法都无法获得理想的识别结果,且语音信号特征样本众多。LFDA 方法虽然识别率比前两者较优,但也不甚理想。又因为计算量太大,现实应用明显不合理。本节将亲和矩阵与类内散度矩阵结合,提出保类内特性的 Fisher 判别方法(LFWF),利用核技巧非线性化,并提出快速训练算法,以适应大样本应用场合。

4.1 LWFD 定义

保持原 Fisher 判别分析法的类间散度矩阵 S_b 不变,类内散度矩阵调整如下:

$$S_w = \sum_{i=1}^n \sum_{j=1}^{n_i} (x_i - \frac{1}{n_i} \sum_{j=1}^{n_i} x_j) (x_i - \frac{1}{n_i} \sum_{j=1}^{n_i} x_j)^T$$

$$= \sum_{i=1}^n x_i x_i^T - \sum_{l=1}^c \frac{1}{n_l} \sum_{k=1}^{n_l} \sum_{m=1}^{n_l} x_k x_m^T \quad (17)$$

基于上述调整,将亲和因子融入类内散度矩阵,使之成为

$$\tilde{S}_w = \sum_{i=1}^n x_i x_i^T - \sum_{l=1}^c \frac{1}{n_l} \sum_{k=1}^{n_l} \sum_{m=1}^{n_l} A_{k,m} x_k x_m^T \quad (18)$$

也就是说,依据亲和矩阵 $A_{i,j}$,对同类中的样本对进行加权,减少同类样本中远距离数据对类内散度矩阵的作用,突出近邻数据的影响力,即保持类内特性。将 $\tilde{S}_w$ 应用于 Fisher 判别分析,则 Fisher 准则公式由式(2)变为

$$J(\varphi) = \frac{\varphi^T S_b \varphi}{\varphi^T \tilde{S}_w \varphi}, \varphi \neq 0 \quad (19)$$

式(19)就是本文所提的鉴别准则公式,其中亲和矩阵 A 的取法多种多样。如果将所有样本对的 $A_{i,j}$ 设为 1,则类内特性保持 Fisher 判别将退化为原 Fisher 判别方法。

4.2 亲和矩阵 A 求取

亲和矩阵的作用是凸显近邻样本点,抑制离群样本点。一般计算方法为

$$A_{i,j} = \exp(-\frac{\|x_i - x_j\|^2}{\sigma^2}) \quad (20)$$

式中, $\sigma > 0$ 是可调整参数^[15]。这样所求的亲和矩阵是非稀疏的,如果直接采用,会极大地增加算法计算量。引入欧式近邻法,将样本中的相邻点指定对应的亲和力和值,而非近邻点的亲和力和值取为零。如样本 x_i ,先在样本空间中以欧式距离搜寻其最近邻的 k 个样本,记第 k 近邻距离为 N_i^k ,假如另一样本 $x_j \in N_i^k$,认为 x_j 属于 x_i 的近邻,取亲和力和值如下:

$$A_{i,j} = \exp(-\frac{\|x_i - x_j\|^2}{N_i^k \times N_j^k}) \quad (21)$$

而其他非近邻点的亲和力和值设为零。一般取 k 值为 7 能获得较优的结果^[16]。这种方法得到的亲和矩阵具有稀疏性,能极大地提升计算性能。

4.3 快速训练算法

直接计算 LWFD 的最优值,类似 KFD 法,要求 K_w^{-1} K_b 矩阵最大特征值相应的特征向量。应用于说话人辨别时,样本特征向量动辄数千,计算量相当惊人,无法实时应用,必须改进训练算法。

LWFD 方法的最优解即依式(19)计算而得的最佳投影向量组,假设 $null B$ 与 $null \tilde{W}$ 分别代表 S_b 与 $\tilde{S}_w$ 的零空间,则式(19)的最佳鉴别子空间取自 $null B^\perp \cap null \tilde{W}$,其中 $null B^\perp$ 为 $null B$ 的正交补。首先将 S_b 投影到 $null B^\perp$,将式(3)改写为

$$S_b = \sum_{i=1}^c (\sqrt{\frac{c_i}{N}} (\bar{\phi}_i - \bar{\phi})) (\sqrt{\frac{c_i}{N}} (\bar{\phi}_i - \bar{\phi}))^T$$

$$= \sum_i \bar{\phi}'_i \bar{\phi}'_i^T = \Phi_b \Phi_b^T \quad (22)$$

式中, $\bar{\phi}'_i = \sqrt{\frac{c_i}{N}} (\bar{\phi}_i - \bar{\phi})$, $\Phi_b = [\bar{\phi}'_1 \cdots \bar{\phi}'_c]$ 。矩阵 S_b 的秩为 $c-1$, $\Phi_b \Phi_b^T$ 与 $\Phi_b^T \Phi_b$ 有相同的非零特征值,而零特征值所对应的特征子空间即为 S_b 的零空间,需要滤除。因此用 $\Phi_b^T \Phi_b$ 替代 $\Phi_b \Phi_b^T$ 并采取核技巧进行推导,可以简化运算量。

$$\Phi_b^T \Phi_b = [\bar{\phi}'_1 \cdots \bar{\phi}'_c]^T [\bar{\phi}'_1 \cdots \bar{\phi}'_c] = (\bar{\phi}'_i^T \bar{\phi}'_j)_{i,j=1,\dots,c} \quad (23)$$

式中,

$$\bar{\phi}'_i^T \bar{\phi}'_j = \frac{\sqrt{c_i c_j}}{N} (\phi_i^T \phi_j - \phi_i^T \bar{\phi} - \bar{\phi}^T \phi_j + \bar{\phi}^T \bar{\phi}) \quad (24)$$

将多项式(24)中每一项利用核函数转换为矩阵,得

$$\Phi_b^T \Phi_b = \frac{1}{N} B \times (A_{LC}^T \times K \times A_{LC} - \frac{1}{N} (A_{LC}^T \times K \times 1_{LC}) - \frac{1}{N} (1_{LC}^T \times K \times A_{LC}) + \frac{1}{N^2} (1_{LC}^T \times K \times 1_{LC})) \times B \quad (25)$$

式中, $B = \text{diag}[\sqrt{c_1} \cdots \sqrt{c_c}]$, 1_{LC} 是一个所有元素为 1 的 $L \times C$ 矩阵, $A_{LC} = \text{diag}[a_{c_1} \cdots a_{c_c}]$ 是一个 $L \times C$ 分块对角矩阵,块 a_{c_i} 是一个所有元素为 $\frac{1}{c_i}$ 的 $c_i \times 1$ 列向量。

设 λ_i 与 e_i ($i=1, \dots, c$) 是 $\Phi_b^T \Phi_b$ 第 i 个特征值与特征向量,并以特征值降序排列,则 $v_i = \Phi_b e_i$ 是原类内散度矩阵 S_b 的特征向量。为获得最优投影,必须去除 S_b 的零空间,即摒弃特征值为零的相应特征向量,保留 v_i 中前 $c-1$ 个特征向量 $V = [v_1 \cdots v_{c-1}] = \Phi_b E_m = \Phi_b [e_1 \cdots e_{c-1}]$,则 $V^T S_b V = \Lambda_b$, $\Lambda_b = \text{diag}[\lambda_1 \cdots \lambda_{c-1}]$ 是一个 $(c-1) \times (c-1)$ 对角矩阵。

得到 $null B^\perp$ 空间后,将类间散度矩阵与类内散度矩阵依 $U = V \Lambda_b^{-1/2}$ 投影进子空间,其中 $U^T S_b U = I$, $U^T \tilde{S}_w U = (E_m \Lambda_b^{-1/2})^T (\Phi_b^T \tilde{S}_w \Phi_b) (E_m \Lambda_b^{-1/2})$ 。利用核矩阵 K ,将 $\Phi_b^T \tilde{S}_w \Phi_b$ 进行核化转换:

$$\Phi_b^T \tilde{S}_w \Phi_b = [\bar{\phi}'_1 \cdots \bar{\phi}'_c]^T \tilde{S}_w [\bar{\phi}'_1 \cdots \bar{\phi}'_c]$$

$$= (\bar{\phi}'_i^T \tilde{S}_w \bar{\phi}'_j)_{i,j=1,\dots,c} \quad (26)$$

式中,

$$\bar{\phi}'_i^T \tilde{S}_w \bar{\phi}'_j = \sum_{l=1}^n \bar{\phi}'_i^T \phi(x_l) \phi(x_l)^T \bar{\phi}'_j - \sum_{l=1}^c \frac{1}{n_l} \sum_{k=1}^{n_l} \sum_{m=1}^{n_l} A_{k,m} \bar{\phi}'_i^T x_k x_m^T \bar{\phi}'_j \quad (27)$$

第一项

$$\sum_{l=1}^n \bar{\phi}'_i^T \phi(x_l) \phi(x_l)^T \bar{\phi}'_j = J_1 = \frac{1}{L} B (A_{LC}^T K K A_{LC} - \frac{1}{L} (A_{LC}^T K K 1_{LC}) - \frac{1}{L} (1_{LC}^T K K A_{LC}) + \frac{1}{L^2} (1_{LC}^T K K 1_{LC})) B \quad (28)$$

第二项

$$\sum_{l=1}^c \frac{1}{c_l} \sum_{k=1}^{n_l} \sum_{m=1}^{n_l} A_{k,m} \bar{\phi}'_i^T x_k x_m^T \bar{\phi}'_j = J_2 = \frac{1}{L} B (A_{LC}^T K W K A_{LC} - \frac{1}{L} (A_{LC}^T K W K 1_{LC}) - \frac{1}{L} (1_{LC}^T K W K A_{LC}) + \frac{1}{L^2} (1_{LC}^T K W K 1_{LC})) B \quad (29)$$

式中, $W = \text{diag}[w_1 \cdots w_c]$ 是一个 $N \times N$ 分块矩阵, w_i 是一个元素为 $\frac{1}{c_i} \times A_{k,m}$ 的 $c_i \times c_i$ 矩阵,因此 $\Phi_b^T \tilde{S}_w \Phi_b = J_1 - J_2$,也是一个 $c \times c$ 矩阵,则 $U^T \tilde{S}_w U$ 是一个维度为 $(c-1) \times (c-1)$ 的简单矩阵。计算其特征向量 p_i 与特征值 λ'_i 并以升序排列,

取前 m 个向量的特征提取变换矩阵 $Q=UP=U[p_1 \cdots p_{m-1}]$, 其中 $1 \leq m \leq c-1$, 可得 $Q^T \tilde{S}_w Q = \Lambda_w$ 。 $\Lambda_w = \text{diag}[\lambda'_1 \cdots \lambda'_m]$ 是一个 $m \times m$ 对角矩阵。

综上所述,类内保持 Fisher 判别的最优判别特征向量为 $\Gamma = Q\Lambda_w^{-1/2}$ 。变换后特征构成 H 空间中的一个低维子空间,并具有最大的可分离度。

对任意待分类的输入模式 z ,依 Γ 投影到特征子空间,计算如下

$$y = \Gamma^T \phi(z) = (E_m \Lambda_w^{-1/2} P \Lambda_w^{-1/2})^T (\Phi_b^T \phi(z)) \quad (30)$$

式中,

$$\Phi_b^T \phi(z) = [\phi_1' \cdots \phi_c']^T \phi(z) \quad (31)$$

由于

$$\begin{aligned} \bar{\phi}_i' \phi(z) &= (\sqrt{\frac{c_i}{N}} (\bar{\phi}_i - \phi)) \phi(z) \\ &= \sqrt{\frac{c_i}{N}} \left(\frac{1}{c_i} \sum_{m=1}^{c_i} \phi_{im}^T \phi(z) - \frac{1}{N} \sum_{p=1}^{c_i} \sum_{q=1}^{c_i} \phi_{pq}^T \phi(z) \right) \end{aligned} \quad (32)$$

可得

$$\Phi_b^T \phi(z) = \frac{1}{\sqrt{N}} B (A_{LC}^T \gamma(\phi(z)) - \frac{1}{N} 1_{LC}^T \gamma(\phi(z))) \quad (33)$$

式中, $\gamma(\phi(z)) = [\phi_{11}^T \phi(z) | \phi_{12}^T \phi(z) | \cdots | \phi_{cc}^T \phi(z)]^T$ 是一个 $N \times 1$ 核向量。联合式(30)与式(32),最终的特征向量取值为:

$$y = \frac{1}{\sqrt{N}} (E_m \Lambda_w^{-1/2} P \Lambda_w^{-1/2})^T (B (A_{LC}^T - \frac{1}{N} 1_{LC}^T)) \gamma(\phi(z)) \quad (34)$$

通过式(34),原输入特征向量 z 的低维表达 y 具有更强的判别特征,适宜于分类任务。

5 仿真测试与应用分析

为验证所提方法 LWFD 的性能,分别进行仿真数据实验及说话人辨别应用,并与经典降维分类方法进行对比,体现本文分类法的适应性与高效率。

5.1 数值仿真

构造典型的任意二维分类数据,为体现不同算法的特性,分别将算法应用于线性可分两类数据、重叠样本数据、类内分簇数据。分类投影结果如图 1 所示,图中圆点与星点分别代表两类不同数据,直线、虚线与点线相应为 LFDA, FDA, LWFD 的分类投影方向。当分类数据线性可分时,3 种方法都能成功投影到近似最优的方向,对数据进行有效分离,如图 1(a)所示。当两类数据中有部分交叉到对方区域,可分离度较低时,LFDA 方法同时考虑类内局部保持及类间局部保持,导致过度局部保持,投影方向偏离最优较多,而 FDA 与 LWFD 都给出了最优投影,这个方向上原混合数据达到最佳区分,如图 1(b)所示。当数据中的某一类或多类呈多态分簇时,传统 FDA 方法由于考虑数据的全局分类性能,因此应用于这些呈多态特性的数据时显得力不从心,捉襟见肘,而 LWFD 与 LFDA 算法在追求全局分类最优的同时还保留了同类数据之间的局部特性,所以对分簇同类数据分离时能保持优越的性能,如图 1(c)和图 1(d)所示。

综上所述,类内局部特征保持的 Fisher 判别法兼具了 Fisher 判别分析与全局特征保持 Fisher 分析的特点,扬长补短,能适应不同的数据特征形式。通过对比,可以发现它具有更强的适应性、更高的分类效率。

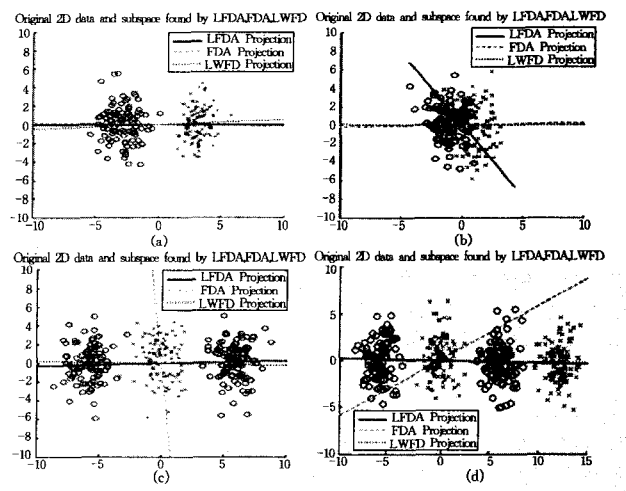


图 1 FDA, LFDA, LWFD 3 种降维分类方法举例对比

5.2 说话人辨别应用

说话人辨别属于非线性分类, FDA, LFDA, LWFD 应用时首先要借助核技巧非线性化,随着训练样本数量增多, KF-D A 与 KLFDA 两者的训练算法内存需求以 $O(n^3)$ 的量级增大,而说话人识别的训练样本动辄以数千计数,不适宜直接应用。实验采用文中所提训练算法应用 KLWFD 方法于说话人辨识,并与经典的说话人识别技术 GMM (Gaussian Mixture Model) 高斯混合模型法与 SVM (Support Vector Machine) 支持向量机法进行对比。其中高斯混合模型的高斯分量为 100 个,支持向量机多分类模型采用一对一多个二元分类器。

实验采用自己录制的语料库,录音总人数为 20 个,其中男 12 个、女 8 个。数据通过采样频率 8000Hz、量化位数 16bit、单声道 A/D 转化获得。每个人的语音信号通过不同时期录制合成。每人混合提取不同时期的语音片段总长度 15s 作为训练信号,不同时期的 20 个长度 1.5s 语音片段作为测试信号,即 20 个训练语音、400 个测试语音。语音信号先经高频提升、中心削减等预处理,再通过 VAD (Voice Activity Detection) 声音活性检测,提取其中有效的语音段,去除冗余的无声段,以 30ms 为长度分帧提取 12 维的 MFCC (Mel Frequency Cepstral Coefficient) 特征参数,即 mel 频率倒谱系数,作为分类参数。

首先,将 3 种典型的核函数应用于 LWFD 算法,对 10 名说话人进行分类测试。参数设置分别为:径向基核函数,参数 σ 取值 0.5;三阶多项式核函数,参数 σ 取值为 0.7;Cauchy 核函数,参数 σ 取值为 2。表 1 列出了基于不同核函数的说话人辨别识别率结果。

表 1 KLWFD 应用不同核函数的说话人辨别对比

Kernel Function	Recognition Rate	Function Expression
RBF	100%	$K(x, y) = \exp(- x - y ^2 / \sigma)$
Poly	86%	$K(x, y) = (x^T \times \sigma \times y + 1)^3$
Cauchy	92.5%	$K(x, y) = 1 / (1 + \sigma \times x - y ^2)$

如表 1 所列,几种经典核函数对比测试中,径向基函数具有明显的性能优势。因此在后续的实验中,均采用径向基内核进行非线性分类。

然后,应用 GMM, SVM 与 KLWFD 进行说话人辨别对比测试, KLWFD 与 SVM 采用相同的径向基核函数。表 2 列

出了对比结果。

表 2 KLWFD 与经典说话人识别方法 GMM,SVM 的对比结果

	5 person		10 person		20 person	
	ER	Speed(s)	ER	Speed(s)	ER	Speed(s)
GMM	1%	1.24	2%	2.15	3.5%	3.56
SVM	0%	0.21	0.5%	0.52	2.75%	0.86
KLWFD	0%	0.43	0%	0.85	2.5%	1.52

表 2 中速率为语音测试时每个语音片段平均识别需花费的时间。可以看出,基于核类内特性保持 Fisher 判别方法的说话人辨别不论从识别率还是从识别速度上均全面优于传统的经典说话人识别技术高斯混合模型法。在与支持向量机的对比测试中,本文方法的识别率占优,但识别速度略逊一筹,这是由支持向量机的稀疏性所致。但支持向量机是一种二元分类器,应用于多元分类时,必须构建多个二元分类器以间接实现多分类。如本例进行 20 人辨别实验时,采用一对多分类方法的支持向量机必须构建 $C_{20}^2=190$ 个不同的二元分类器,这大大地增加了训练模型的复杂度与繁琐性。而核类内特性保持 Fisher 判别方法可直接实现多元分类,训练一次就能实现分类模型,相比支持向量机,更加便捷与人性化。

结束语 Fisher 判别分析与局部保持 Fisher 投影都是降维分类的热门技术,但面对具有不同分布特性的分类数据时,都不能全面兼顾。本文折衷考虑全局投影与局部保持的平衡,将样本之间亲和力以权值形式融入类内散度矩阵,提出类内保持 Fisher 判别方法,以适应不同分布的分类数据。并利用核技巧非线性化,使之能应付线性不可分的分类场合。核化过程导致算法计算复杂度剧增,当训练样本增加时,内存量与计算量过大使实时应用性降低。本文对训练算法进行修改,使应用于说话人辨别等实际问题成为可能。但是核矩阵本身随样本量增多而耗费大量内存空间,说话人辨别应用时总人数受到限制,因此改进这类问题将是我们今后着力研究的方向。

参 考 文 献

[1] Hinton G E,Salakhutdinov R R. Reducing the dimensionality of data with neural networks[J]. Science, 2006, 313(5786): 504-507

[2] Baudat G, Anouar F. Generalized discriminate analysis using a kernel approach[J]. Neural Computation, 2000, 12(10): 2385-2404

[3] Belhumeur P N, Hespanha J P, Kriegam D J. Eigenfaces vs.

Fisherfaces; Recognition using class specific linear projection [J]. IEEE Trans. on Pattern Anal. Machine Intell,1997,19(7): 711-720

[4] Yang J, Yang J-Y. Why can LDA be performed in PCA transformed space? [J]. Pattern Recognition,2003,36(2):563-566

[5] 邢玉娟,李 明,张亚芬. 基于 PCA 和核 Fisher 判别的说话人确认[J]. 计算机工程与设计,2008,29(15):3983-3985

[6] 谭萍,邢玉娟,李明. 基于核 Fisher 判别的说话人辨认[J]. 科学技术与工程,2008,8(8):2230-2233

[7] He X F,Niyogi P. Locality Preserving Projections[A]// Thrun S, Saul L, Scholkopf B, eds. Advances in Neural Information Processing Systems 16[C]. Cambridge,MA: MIT Press,2004

[8] Sugiyama M. Dimensionality Reduction of Multimodal Labeled Data by Local Fisher Discriminant Analysis[J]. Machine Learning Research,2007,8:1027-1061

[9] 徐勇,杨静宇,金忠,等. 一种基于核的快速非线性鉴别分析方法[J]. 计算机研究与发展,2005,42(3):267-374

[10] 高秀梅,杨静宇,杨健. 一种最优的核 Fisher 鉴别分析与人脸识别[J]. 系统仿真学报,2004,16(12):2864-2868

[11] Liu Qing-Shan,Huang Rui. Face recognition using kernel based fisher discriminant analysis[A]//Proceedings of the 5th International Conference on Automatic Face and Gesture Recognition [C]. Washington,DC, USA,2002:187-191

[12] Liu Wei, Wang Yun-hong. Null space based kernel fisher discriminant analysis for face recognition[A]// Proceeding of the 6th International Conference on Automatic Face and Gesture Recognition[C]. Seoul,Korea,2004:369-374

[13] Mika S,Ratsch G, Weston J, et al. Fisher discriminant analysis with kernels[A]//Proceedings of IEEE International Workshop on Neural Networks for Singal Processing [C]. Madison, Wisconsin, August 1999:41-48

[14] He X F, Yan S, Hu Y, et al. Learning a Locality Preserving Subspace for Visual Recognition[A]// Proceedings of Ninth IEEE International Conference on Computer Vision[C]. Nice, France, 2003:385-392

[15] Belkin M,Niyogi P. Laplacian eigenmaps for dimensionality reduction and data representation[J]. Neural Computation, 2003, 15(6):1373-1396

[16] Zelnik-Manor L,Perona P. Self-tuning spectral clustering[A]// Saul L K, Weiss Y, Bottou Y, eds. Advances in Neural Information Processing Systems 17[C]. Cambridge, MA: MIT Press, 2005:1601-1608

(上接第 159 页)

[3] 金雪云,徐卫东,刘红卫. 基于 .NET 的代码生成工具的设计与实现[J]. 计算机工程与应用,2004,21:123-125

[4] 陈翔,王学斌,吴泉源. 代码生成技术在 MDA 中的实现[J]. 计算机应用研究,2006,1:147-150

[5] 轩兴涛. 基于 PowerDesigner 模型驱动机制下的全程建模研究 [J]. 西安石油大学学报:自然科学版,2008,23(6):104-106

[6] 何天,侯宗浩. 基于 Petshop 与 Duwamish 的多层架构设计与实

现[J]. 计算机应用,2006,26:257-259

[7] 李梦,李凡,李京. Web 应用自动构建框架——WACF[J]. 计算机工程,2007,33(9):97-99

[8] 张海英,万建成. Web 用户界面自动生成系统中的界面模板 [J]. 计算机工程与设计,2006,27(18):3491-3493

[9] Griffiths T, Barclay P. Teallach: A model-based user inter-face development environment for object database[Z]. Los Ala-mitos,CA: User Interfaces to Data Intensive Systems,1999:86-96