软件学报 ISSN 1000-9825, CODEN RUXUEW
Journal of Software,2013,24(11):2597−2609 [doi: 10.3724/SP.J.1001.2013.04473]
©中国科学院软件研究所版权所有.
E-mail: jos@iscas.ac.cn
http://www.jos.org.cn
Tel/Fax: +86-10-62562563

一种引入成对代价的子类判别分析*

万建武[1,2], 杨 明[1]

[1](南京师范大学 计算机科学与技术学院,江苏 南京 210023)

[2](常州大学 信息科学与工程学院,江苏 常州 213164)

通讯作者: 杨明, E-mail: m.yang@njnu.edu.cn

摘 要: 传统的降维方法追求较低的识别错误率,假设不同错分的代价相同,这个假设在一些实际应用中往往不成立.例如,在基于人脸识别的门禁系统中,存在入侵者类和合法者类,将入侵者错分成合法者的损失往往高于将合法者错分成入侵者的损失,而将合法者错分成入侵者的损失又大于将合法者错分成其他合法者的损失.为此,首先通过对人脸识别门禁系统进行分析,将其归为一个代价敏感的子类学习问题,然后将错分代价以及子类信息同时注入判别分析的框架中,提出一种近似于成对贝叶斯风险准则的降维算法.在人脸数据集Extended Yale B以及ORL上的实验结果表明了该算法的有效性.

关键词: 代价敏感降维;人脸识别;子类学习

中图法分类号: TP181 **文献标识码**: A

中文引用格式: 万建武,杨明.一种引入成对代价的子类判别分析.软件学报,2013,24(11):2597−2609. http://www.jos.org.cn/1000-9825/4473.htm

英文引用格式: Wan JW, Yang M. Pairwise costs in subclass discriminant analysis. Ruan Jian Xue Bao/Journal of Software, 2013, 24(11):2597−2609 (in Chinese). http://www.jos.org.cn/1000-9825/4473.htm

Pairwise Costs in Subclass Discriminant Analysis

WAN Jian-Wu[1,2], YANG Ming[1]

[1](School of Computer Science and Technology, Nanjing Normal University, Nanjing 210023, China)

[2](School of Information Science and Engineering, Changzhou University, Changzhou 213164, China)

Corresponding author: YANG Ming, E-mail: m.yang@njnu.edu.cn

Abstract: Conventional dimensionality reduction algorithms aim to attain low recognition errors, assuming the same misclassification loss from different misclassifications. In some real-world applications, however, this assumption may not hold. For example, in the door-locker syetem based on face recognition, there are impostor and gallery person. The loss of misclassifying an impostor as a gallery person is larger than misclassifying a gallery person as an impostor, while the loss of misclassifying a gallery person as an impostor can be larger than misclassifying a gallery person as other gallery persons. This paper recognizes the door-locker system based on face recognition as a cost-sensitive subclass learning problem, incorporates the subclass information and misclassification costs into the framework of discriminant analysis at the same time, and proposes a dimensionality reduction algorithm approximate to the pairwise Bayes risk. The experimental results on face datasets Extended Yale B and ORL demonstrate the superiority of the proposed algorithm.

Key words: cost-sensitive dimensionality reduction; face recognition; subclass learning

人脸识别[1−17]是机器学习和模式识别领域的一个难点问题.目前,人脸识别被认为是一个代价敏感的学习问题.例如,在基于人脸识别的门禁系统中[4−8,16],有入侵者类和合法者类.将入侵者错分成合法者,并让其进入房

* 基金项目: 国家自然科学基金(60873176, 61272222, 61003116); 江苏省自然科学基金(BK2011782, BK2011005); 江苏省创新基金(CXZZ12_0386)

收稿时间: 2013-04-18; 修改时间: 2013-07-12, 2013-08-02; 定稿时间: 2013-08-27

© 中国科学院软件研究所 http://www.jos.org.cn

间所带来的损失要远大于将合法者拒之门外的损失;而将合法者拒之门外的损失又大于将合法者识别为其他类别的合法者并让其进入房间的损失.因此,不同错分的损失不同.

为了解决不同的错分损失问题,有学者对不同的错分赋予不同的错分代价,提出了代价敏感的分类算法[4,8,18−23].其中,应用于人脸识别门禁系统的有:Zhang 等人[4]提出的代价敏感的 logistic 回归(multi-class cost-sensitive kernel logistic regression,简称 McKLR)和 k 近邻分类器(multi-class cost-sensitive k-nearest neighbor,简称 McKNN).Man 等人提出的代价敏感的稀疏表示分类器[8].人脸识别是一个典型的小样本问题[24],文献[4,8]仅将错分代价融入人脸识别的分类部分,这将在降维部分丢失大量的代价信息,导致局部最优的分类结果[5−7].为此,Lu 等人提出了代价敏感的线性判别分析(cost sensitive linear discriminant analysis,简称 CSLDA)等降维算法[5−7].由于 CSLDA 等代价敏感降维算法未能从理论上保证其模型满足最小贝叶斯风险准则,Wan 等人提出了代价敏感的加权局部保持降维(weighted cost-sensitive local preserving projection,简称 WCSLPP)[16],其模型满足最小错分损失准则.

上述代价敏感的学习方法都将代价信息注入了类结构粒度(CSLDA,McKLR,McKNN)或者局部结构粒度中(WCSLPP).根据 Yang 等人的观点[25],数据的结构粒度有 3 种:类结构粒度、局部结构粒度和聚类结构粒度.数据的聚类结构粒度介于类结构粒度和局部结构粒度之间.使用聚类结构粒度可以有效避免使用类结构粒度带来的欠拟合数据问题,也可以避免使用局部结构粒度带来的过拟合数据问题.为此,本文基于数据的聚类粒度结构提出了一种引入成对代价的子类判别分析(pairwise costs in subclass discriminant analysis,简称 PCSCDA),并将其应用于基于人脸识别的门禁系统.PCSCDA 算法的创新之处在于:

(1) 将基于人脸识别的门禁系统归为子类学习问题.本文对门禁系统分析后发现,门禁系统中的入侵者类是由多个入侵者组成的混合类,因此,入侵者类的数据分布满足混合高斯分布,是一个典型的子类学习问题.

(2) 将聚类信息同时注入判别分析的类内和类间协方差矩阵中,充分利用了聚类结构信息.

(3) 在各个成对的子类上设计了近似于成对贝叶斯风险准则的模型,使得投影方向更具判别能力.

在人脸数据集 Extended Yale B[10]以及 ORL[17]上的实验结果表明了本文 PCSCDA 算法的有效性.

1 基于聚类结构粒度的降维算法

目前,基于聚类结构粒度的降维算法大都基于判别分析框架,都试图将聚类结构信息注入判别分析的类内和类间协方差矩阵.最具代表性的工作有:子类判别分析(subclass discriminant analysis,简称 SDA)[26]、将聚类信息同时引入类内和类间协方差矩阵的子类判别分析 SWBDA[25].下面本文分别介绍 SDA 以及 SWBDA 算法.

SDA 算法是最典型的基于聚类结构粒度的降维算法.SDA 假设数据的分布满足混合高斯模型,采用 NN 聚类方法(nearest neighbor-based)将每类样本聚成若干个聚类,然后将聚类信息注入类间协方差矩阵 S_B:

$$S_B=\sum_{i=1}^{c-1}\sum_{j=1}^{H_i}\sum_{k=i+1}^{c}\sum_{l=1}^{H_k}p_{ij}p_{kl}(\mu_{ij}-\mu_{kl})(\mu_{ij}-\mu_{kl})^T \tag{1}$$

其中,c 表示样本的类别数,$H_i(i=1,2,\ldots,c)$表示第 i 类的聚类个数,$p_{ij}=N_{ij}/N_i$ 和μ_{ij} 分别表示第 i 类中第 j 个聚类的先验概率和均值,N_i 表示第 i 类的样本数,N_{ij} 表示第 i 类中第 j 个聚类的样本数.

SDA 仅将聚类信息注入类间协方差矩阵 S_B,故而 Yang 等人提出了将聚类信息同时引入类内和类间协方差矩阵的子类判别分析 SWBDA.在 SWBDA 中,注入聚类信息的类间协方差矩阵 S_B 的定义如下:

$$S_B=\sum_{i=1}^{c}N_i(\mu_i-\mu)(\mu_i-\mu)^T \tag{2}$$

$$\mu_i=\frac{1}{H_i}\sum_j\mu_{ij},\ j=1,2,\ldots,H_i \tag{3}$$

其中,μ_i 和μ分别表示第 i 类的样本均值和整体样本的均值.在 SWBDA 中,类内协方差矩阵 S_W 定义如下:

$$S_W=\sum_i S_{W_i},\ i=1,2,\ldots,c \tag{4}$$

© 中国科学院软件研究所 http://www.jos.org.cn

$$S_{W_i}=\frac{1}{H_i}\sum_j\frac{N_{ij}}{N_i}S_{W_{ij}},\ j=1,2,...,H_i \tag{5}$$

$$S_{W_{ij}}=\sum_{l=1}^{N_{ij}}(x_l-\mu_{ij})(x_l-\mu_{ij})^T \tag{6}$$

2 引入成对代价的子类判别分析 PCSCDA

SDA 将聚类信息注入类间协方差矩阵 S_B,却没有注入类内协方差矩阵 S_W.SWBDA 试图将聚类信息注入 S_B 和 S_W,但本文对 SWBDA 算法分析后发现:(a) SWBDA 中第 i 类的均值μ_I(见公式(3))由各个聚类的均值 μ_{ij}(j=1,2,…,H_i)等贡献所构成,如果其中一个聚类为离群类,其均值中心远离其他聚类,那么μ_i 将受该离群聚类的影响,偏向离群聚类;(b) 如果第 i 类中所有聚类的样本个数均相同,即 $N_{i1}=N_{i2}=...=N_{iH_i}$, 则公式(3)定义的$\mu_i$ 和原始类均值的定义一致,导致在由公式(2)定义的类间协方差矩阵 S_B 中无法体现聚类的结构信息.此外,SDA 和 SWBDA 都是代价不敏感的降维算法,没有考虑实际应用中的不同错分损失问题.基于以上分析,下面本文首先分析人脸识别门禁系统,然后将聚类信息和错分代价注入判别分析框架,设计代价敏感的子类判别分析算法.

2.1 基于人脸识别的门禁系统是子类学习问题

人脸识别门禁系统的任务是将入侵者拒之门外,而允许合法者进入.在门禁系统中,假设存在 $c-1$ 个合法者类:$G_1,\ldots,G_{c-1}$ 和入侵者类 I.对于这个 c 类的识别问题,存在 3 种类型的错分:(1) False rejection:将合法者错分成入侵者;(2) False acceptance:将入侵者错分成合法者;(3) False identification:将某一合法者错分成其他合法者.由于这 3 种错分的损失不同,Zhang 等人首次将人脸识别门禁系统看成代价敏感的学习问题[4],对不同的错分赋予不同的错分代价:(1) C_{GI}:将合法者错分成入侵者的代价;(2) C_{IG}:将入侵者错分成合法者的代价;(3) C_{GG}:将合法者错分成其他合法者的代价.最终可得错分代价矩阵 C,见表 1.

Table 1 Misclassification cost matrix C

表 1 错分代价矩阵 C

	G_1	…	G_{c-1}	I
G_1	0	…	C_{GG}	C_{GI}
…	…	…	…	…
G_{c-1}	C_{GG}	…	0	C_{GI}
I	C_{IG}	…	C_{IG}	0

一般而言,在基于人脸识别的门禁系统中,入侵者类 I 是由多个入侵者组成的混合类.图 1 直观地表示了门禁系统中各类的数据分布,其中,入侵者类由 K 个入侵者组成.从图 1 容易观察到,入侵者类的数据分布是一个天然的混合高斯分布,其中,单个高斯分布的个数即为入侵者的个数.因此,$c-1$ 个合法者类和入侵者类的聚类个数分别为 $H_{G_1}=H_{G_2}=...=H_{G_{c-1}}=1, H_I=K$,K 为入侵者的个数.

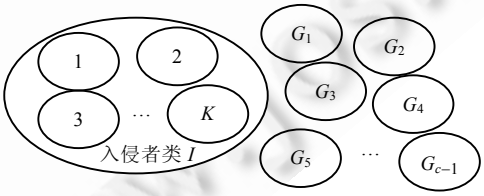


Fig.1 Door-Locker system based on face recognition

图 1 基于人脸识别的门禁系统

2.2 注入聚类信息的判别分析算法

为了在判别分析的类内和类间协方差矩阵中注入聚类信息,PCSCDA 算法的类间协方差矩阵 S_B 定义如下:

© 中国科学院软件研究所 http://www.jos.org.cn

$$S_B=\sum_{i=1}^{c-1}\sum_{j=1}^{H_i}\sum_{k=i+1}^{c}\sum_{l=1}^{H_k}p_{ij}p_{kl}(\mu_{ij}-\mu_{kl})(\mu_{ij}-\mu_{kl})^T \tag{7}$$

其中,相关符号的定义见公式(1).PCSCDA 算法的类内协方差矩阵 S_W 定义如下:

$$S_W=\sum_{i=1}^{c}S_{W_i} \tag{8}$$

$$S_{W_i}=\sum_{j=1}^{H_i}\sum_{l=1}^{N_{ij}}(x_l-\mu_{ij})(x_l-\mu_{ij})^T \tag{9}$$

其中,x_l 表示第 l 个样本.通过定义公式(7)~公式(9),PCSCDA 将聚类信息同时注入了判别分析的 S_B 和 S_W 中.

2.3 注入成对代价信息的判别分析算法

传统的判别型降维算法[1,27,29−31]寻找一个投影方向 $W_{D\times d}$,$D>>d$,将数据从 D 维空间投影到 d 维空间,使得在 d 维空间上等价或者近似于最小贝叶斯错误率准则,例如 LDA(linear discriminant analysis)算法[31].如果解决的是两类问题,类数据分布满足高斯分布且类协方差矩阵相等,那么 LDA 和最小贝叶斯错误率准则的决策平面一致[32];如果解决的是多类问题,直接设计满足最小贝叶斯错误率准则的线性判别分析算法就比较困难,因此,Loog 等人通过在 LDA 模型中引入加权函数,提出了近似于最小成对贝叶斯错误率准则的 aPAC(approximate pairwise accuracy criterion)算法[29,30].本文 PCSCDA 解决的是一个多类的子类学习问题,同时考虑子类间的不同错分损失问题和离群类问题.为此,PCSCDA 采用与 aPAC 相似的思路,在判别分析框架中引入加权函数 ω,使得在投影后的 d 维空间上近似于成对贝叶斯风险准则.

在总类别数为 c 类的子类学习问题中,成对贝叶斯风险准则定义如下:

$$\max_{W_{D\times d}}:J(W)=\sum_{i=1}^{c-1}\sum_{j=1}^{H_i}\sum_{k=i+1}^{c}\sum_{l=1}^{H_k}p_{ij}p_{kl}A_{ij,kl}(W) \tag{10}$$

其中,$W_{D\times d}$ 是投影方向;p_{ij} 表示第 i 类中第 j 个子类的先验概率;$A_{ij,kl}(W)$表示在投影后的 d 维空间上,第 i 类中第 j 个聚类和第 k 类中第 l 个聚类的样本考虑了不同错分损失问题的贝叶斯准确率.

显然,PCSCDA 首先需要计算出公式(10)定义的成对贝叶斯风险,然后才能给出加权函数 ω,使得 PCSCDA 近似于成对贝叶斯风险.但是,直接计算 d 维下的成对贝叶斯风险比较困难,因此,PCSCDA 首先在 d=1 维空间下计算成对贝叶斯风险,然后在 d=1 维空间下定义加权函数 ω,最后将其推广到 d(d>1)维空间,定义近似于成对贝叶斯风险的 PCSCDA 模型.

2.3.1 一维空间上的成对贝叶斯风险

如果降维后的维数为 d=1,一个 c 类的子类学习问题的成对贝叶斯风险准则定义如下:

$$\max_{W_{D\times 1}}:J(W)=\sum_{i=1}^{c-1}\sum_{j=1}^{H_i}\sum_{k=i+1}^{c}\sum_{l=1}^{H_k}p_{ij}p_{kl}A_{ij,kl}(W) \tag{11}$$

这里,PCSCDA 假设每个子类的数据分布满足高斯分布,各子类的协方差矩阵相同:

$$\Sigma_{11}=\Sigma_{12}=\ldots=\Sigma_{1H_1}=\Sigma_{21}=\ldots=\Sigma_{2H_2}=\ldots=\Sigma_{cH_c}=I,$$

且各子类的类先验概率相等 $p_{11}=p_{12}=\ldots=p_{1H_1}=p_{21}=\ldots=p_{2H_2}=\ldots=p_{cH_c}$. 本文认为:如果通过聚类可以将混合高斯分布中多个单高斯分布的数据聚成相应的聚类,那么在单个聚类上假设数据满足单高斯分布则是合理的.为了考虑子类间的不同错分损失问题,PCSCDA 对不同的错分赋予不同的错分代价.例如任意的两个子类 ij 和 kl,PCSCDA 假设 $C_{ik}\neq C_{ki}$.根据最小贝叶斯风险决策准则[32],可以得到:

$$\text{if } R(ij|x)<R(kl|x),\ \text{then } x\in\text{subclass } j \text{ of class } i \tag{12}$$

其中,$R(ij|x)=C_{ki}P(kl|x)$,$R(kl|x)=C_{ik}P(ij|x)$.根据贝叶斯公式,公式(12)可以改写成

$$\text{if } C_{ki}P(x|kl)p_{kl}<C_{ik}P(x|ij)p_{ij},\ \text{then } x\in\text{subclass } j \text{ of class } i \tag{13}$$

其中,$P(x|ij)$和 $P(x|kl)$分别表示第 i 类中第 j 个子类和第 k 类中第 l 个子类的类分布密度函数.令 $P(ijk)=C_{ik}p_{ij}$,$P(kli)=C_{ki}p_{kl}$,可得:

$$\text{if } P(kli)P(x|kl)<P(ijk)P(x|ij),\ \text{then } x\in\text{subclass } j \text{ of class } i \tag{14}$$

© 中国科学院软件研究所 http://www.jos.org.cn

值得注意的是,在公式(14)中,$P(ijk)$和$P(kli)$一般不相等.这是因为在人脸识别门禁系统中,假设i类为某一合法者类,k 类为入侵者类,将合法者错分成入侵者的损失要小于将入侵者错分成合法者的损失:$C_{ik}<C_{ki}$;先前假设各子类的类先验概率相等: $p_{11}=p_{12}=\ldots=p_{1H_1}=p_{21}=\ldots=p_{2H_2}=\ldots=p_{cH_c}$. 因此,显然 $P(ijk)$和 $P(kli)$不等.

假设 $P(ijk)<P(kli)$,在 d=1 维下,子类 ij 和 kl 的贝叶斯决策平面如图 2 所示,其中,$x=T$ 是子类 ij 和子类 kl 的最优贝叶斯决策平面.

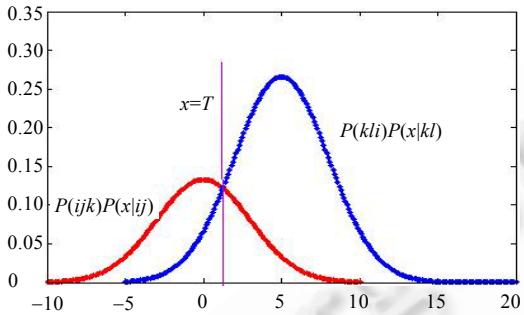


Fig.2 Optimal Bayes decision-making of subclass ij and kl

图 2 子类 ij 和 kl 的最优贝叶斯决策

根据图 2,经过简单推导,可以求得:

$$A_{ij,kl}(W)=P(kli)\left[\frac{1}{2}+\frac{1}{2}\mathrm{erf}\left(\frac{\ln\frac{P(kli)}{P(ijk)}}{\sqrt{2}}\frac{1}{\tilde{\mu}_{kl}-\tilde{\mu}_{ij}}+\frac{\tilde{\mu}_{kl}-\tilde{\mu}_{ij}}{2\sqrt{2}}\right)\right]+P(ijk)\left[\frac{1}{2}+\frac{1}{2}\mathrm{erf}\left(\frac{\ln\frac{P(ijk)}{P(kli)}}{\sqrt{2}}\frac{1}{\tilde{\mu}_{kl}-\tilde{\mu}_{ij}}+\frac{\tilde{\mu}_{kl}-\tilde{\mu}_{ij}}{2\sqrt{2}}\right)\right] \tag{15}$$

其中, $\tilde{\mu}_{ij}=W^T\mu_{ij}$ 和 $\tilde{\mu}_{kl}=W^T\mu_{kl}$ 分别表示在 d=1 维下,子类 ij 和 kl 的样本均值, $\mathrm{erf}(x)=\frac{2}{\sqrt{\pi}}\int_0^x\exp(-t^2)\mathrm{d}t$.

2.3.2 定义一维空间上的加权函数ω

公式(15)求得的 $A_{ij,kl}(W)$比较复杂,将其代入公式(11),直接最优化成对贝叶斯风险准则比较困难.为此,本文将 $A_{ij,kl}(W)$退化为 $\tilde{A}_{ij,kl}(W)$,然后代入公式(11),进行优化求解,具体证明过程见附录. $\tilde{A}_{ij,kl}(W)$ 的具体定义如下:

$$\tilde{A}_{ij,kl}(W)=(P(ijk)+P(kli))\left(\frac{1}{2}+\frac{1}{2}\mathrm{erf}\left(\frac{\tilde{\mu}_{kl}-\tilde{\mu}_{ij}}{2\sqrt{2}}\right)\right) \tag{16}$$

因而,可得退化后的成对贝叶斯风险准则:

$$\max_W: J(W)=\sum_{i=1}^{c-1}\sum_{j=1}^{H_i}\sum_{k=i+1}^{c}\sum_{l=1}^{H_k}p_{ij}p_{kl}A_{ij,kl}(W) \tag{17}$$

其中, $A_{ij,kl}(W)=(P(ijk)+P(kli))\left(\frac{1}{2}+\frac{1}{2}\mathrm{erf}\left(\frac{W^T\mu_{kl}-W^T\mu_{ij}}{2\sqrt{2}}\right)\right)$.

这里,本文采用与 aPAC 相似的策略,定义加权函数$\omega(\Delta_{ij,kl})$,并将其引入判别分析框架,得到 PCSCDA 模型:

$$\max_W: \sum_{i=1}^{c-1}\sum_{j=1}^{H_i}\sum_{k=i+1}^{c}\sum_{l=1}^{H_k}p_{ij}p_{kl}\omega(\Delta_{ij,kl})tr((W^TS_WW)^{-1}(W^TS_{ij,kl}W)) \tag{18}$$

其中, $\omega(\Delta_{ij,kl})=(P(ijk)+P(kli))\left(\frac{1}{2}+\frac{1}{2\Delta_{ij,kl}^2}\mathrm{erf}\left(\frac{\Delta_{ij,kl}}{2\sqrt{2}}\right)\right)$, $\Delta_{ij,kl}=\sqrt{(\mu_{ij}-\mu_{kl})^TS_W^{-1}(\mu_{ij}-\mu_{kl})}$, $S_{ij,kl}=(\mu_{ij}-\mu_{kl})(\mu_{ij}-\mu_{kl})^T$.

令 $W=S_W^{-1/2}V_{D\times1}$, 修改公式(18)可得:

$$\max_{V_{D\times1}}: \sum_{i=1}^{c-1}\sum_{j=1}^{H_i}\sum_{k=i+1}^{c}\sum_{l=1}^{H_k}p_{ij}p_{kl}\omega(\Delta_{ij,kl})tr(V^TS_W^{-1/2}S_{ij,kl}S_W^{-1/2}V) \tag{19}$$

© 中国科学院软件研究所 http://www.jos.org.cn

本文定义的加权函数$\omega(\Delta_{ij,kl})$有如下两个性质:

(a) 通过在判别分析框架中引入加权函数$\omega(\Delta_{ij,kl})$,PCSCDA 模型近似于成对贝叶斯风险准则,并且两者在最优解处等价,即公式(21)和公式(23)在最优解处完全等价[30].

(b) 加权函数$\omega(\Delta_{ij,kl})$有效地抑制了离群子类对投影方向的影响.$\omega(\Delta_{ij,kl})$是关于$\Delta_{ij,kl}$的减函数,离群子类的类中心远离其他子类的类中心,$\Delta_{ij,kl}$较大,故离群子类的加权值$\omega(\Delta_{ij,kl})$较小,有效地抑制了离群子类对投影方向的贡献.

2.3.3 推广至 d 维空间

在 d 维空间上,PCSCDA 模型定义如下:

$$\max_{V}: J(V)=\sum_{m=1}^{d}\sum_{i=1}^{c-1}\sum_{j=1}^{H_i}\sum_{k=i+1}^{c}\sum_{l=1}^{H_k}p_{ij}p_{kl}\omega(\Delta_{ij,kl})tr(e_m^T\tilde{S}_{ij,kl}e_m) \tag{20}$$

其中,$V=(e_1,e_2,...,e_d)$,$\tilde{S}_{ij,kl}=S_W^{-1/2}S_{ij,kl}S_W^{-1/2}$. 最终的投影矩阵为$W_{D\times d}=S_W^{-1/2}V$.

观察公式(20),PCSCDA 希望寻找 d 个正交的投影向量,使得 PCSCDA 近似 d 维空间上的贝叶斯风险累加和.由于$\sum_{i=1}^{c-1}\sum_{j=1}^{H_i}\sum_{k=i+1}^{c}\sum_{l=1}^{H_k}p_{ij}p_{kl}\omega(\Delta_{ij,kl})\tilde{S}_{ij,kl}$是对称矩阵,对其进行矩阵特征分解,求得的特征向量正交,互不影响,故 d 维的投影向量可直接由矩阵的特征分解得到.

2.4 不同的类分布密度问题对S_W的影响

类协方差矩阵体现了类分布密度:类 i 的类分布密度越高,该类的数据分布就越紧,对应的$tr(S_{W_i})$值就越小,反之亦然.如果某一类 i 的类分布密度明显低于其他类,那么整体的类协方差矩阵$S_W=\sum_{i=1}^{c}S_{W_i}$将主要由$S_{W_i}$决定,其他类的类协方差矩阵将得不到体现,因此需要在 S_W 中考虑不同的类分布密度问题.另外,由于不同种类的错分,其损失不同,因此还需要在 S_W 中引入错分代价.

综合考虑不同的类分布密度问题以及不同的错分损失问题,PCSCDA 的类协方差矩阵 S_W 定义如下:

$$S_W=\sum_{i=1}^{c}f(i)S_{W_i} \tag{21}$$

其中,$f(i)$是重要性函数,用于度量各类样本的重要性,定义如下:

$$\left.\begin{aligned}&f(i)=C_{i*}/tr(S_{W_i})^{\alpha},\ i=1,2,...,c\\&C_{i*}=\sum_{k=1,k\neq i}^{c}C_{ik}\end{aligned}\right\} \tag{22}$$

在公式(22)中,C_{i*}表示将第 i 类错分成其他类的错分代价[33];参数α用于平衡不同的类分布密度对类内协方差矩阵 S_W 的影响,α值越大,对分布密度低的类的惩罚力度越大,当α=0 时,PCSCDA 不考虑类分布密度问题.

2.5 PCSCDA的模型描述

综上所述,PCSCDA 算法的模型描述如下:

$$\left.\begin{aligned}&\max_{V_{Dxd}}:\sum_{i=1}^{c-1}\sum_{j=1}^{H_i}\sum_{k=i+1}^{c}\sum_{l=1}^{H_k}p_{ij}p_{kl}\omega(\Delta_{ij,kl})tr(V^TS_W^{-1/2}(\mu_{ij}-\mu_{kl})(\mu_{ij}-\mu_{kl})^TS_W^{-1/2}V)\\&S_W=\sum_{i=1}^{c}f(i)S_{W_i},\ S_{W_i}=\sum_{j=1}^{H_i}\sum_{l=1}^{N_{ij}}(x_l-\mu_{ij})(x_l-\mu_{ij})^T\\&f(i)=C_{i*}/tr(S_{W_i})^{\alpha},\ i=1,2,...,c,\ C_{i*}=\sum_{k=1,k\neq i}^{c}C_{ik}\end{aligned}\right\} \tag{23}$$

其中,c 是类别数;H_i 是第 i 类的聚类个数;p_{ij} 和 p_{kl} 分别表示第 i 类中第 j 个子类和第 k 类中第 l 个子类的先验概

© 中国科学院软件研究所 http://www.jos.org.cn

率; $\omega(\Delta_{ij,kl})=(P(ijk)+P(kli))\left(\frac{1}{2}+\frac{1}{2\Delta_{ij,kl}^2}\mathrm{erf}\left(\frac{\Delta_{ij,kl}}{2\sqrt{2}}\right)\right)$; $\Delta_{ij,kl}=\sqrt{(\mu_{ij}-\mu_{kl})^T S_W^{-1}(\mu_{ij}-\mu_{kl})}$; $P(ijk)=C_{ik}p_{ij}, P(kli)=C_{ki}p_{kl}$; $f(i)$是重要性函数,用于综合度量 S_W 中各类样本的重要性;C_{i*}表示将第 i 类错分成其他类的错分代价;α是参数;x_l 是第 l 个样本.最终的投影矩阵为 $W_{D\times d}=S_W^{-1/2}V$.

如果将 PCSCDA 应用于人脸识别门禁系统,假设前 $c-1$ 类为合法者类,第 c 类为入侵者类,那么聚类个数分别为 $H_1=H_2=\ldots=H_{c-1}=1, H_c=K$.根据错分代价矩阵 C 的定义(见表 1),重要性函数 $f(i)$可定义如下:

$$f(i)=\begin{cases}((c-2)C_{GG}+C_{GI})/tr(S_{W_i})^{\alpha}, & \text{if } i=1,2,\ldots,c-1\\ (c-1)C_{IG}/tr(S_{W_i})^{\alpha}, & \text{otherwise}\end{cases} \tag{24}$$

3 实 验

3.1 实验设置

3.1.1 数据集设置

为了检验 PCSCDA 算法的性能,本文选择 ORL[17]和 Extended Yale B[10]数据集进行验证.为了接近真实的人脸识别门禁系统,合法者的类别数分别设置为 $M_{\mathrm{ORL}}=30, M_{\text{Extended YaleB}}=30$,剩余样本为入侵者 N_I;错分代价与文献[4−8,16]一致:$C_{IG}=20, C_{GI}=2, C_{GG}=1$.在 ORL,Extended Yale B 数据集中,每类分别选择 5 张、30 张作为训练集,剩余图片作为测试集.分类器采用 KNN 和代价敏感的 McKNN[4],k 值取 3.实验重复 10 次,取平均.具体实验设置见表 2.

Table 2 Datasets settings of PCSCDA

表 2 PCSCDA 的数据集设置

数据集	合法者的类别数 M	每类合法者的样本数 N_G	入侵者的样本数 N_I	错分代价比 C_{IG}:C_{GI}:C_{GG}
ORL	30	5	50	20:2:1
Extended Yale B	30	30	240	20:2:1

3.1.2 对比方法与评价准则

为了展示算法进行性能,本文分别与基于类结构粒度的判别分析算法 LDA[1],aPAC[29,30],CSLDA[5−7],基于聚类结构粒度的判别分析算法 SDA[26],SWBDA[25]以及局部结构粒度的判别分析算法 LSDA(locality sensitive discriminant analysis)[34]进行比较.实验采用 5 种评价准则:

(1) 整体错分损失 $Cost$;

(2) 将入侵者错分成合法者的比例 Err_{IG};

(3) 将合法者错分成入侵者的比例 Err_{GI};

(4) 将合法者错分成其他合法者的比例 Err_{GG};

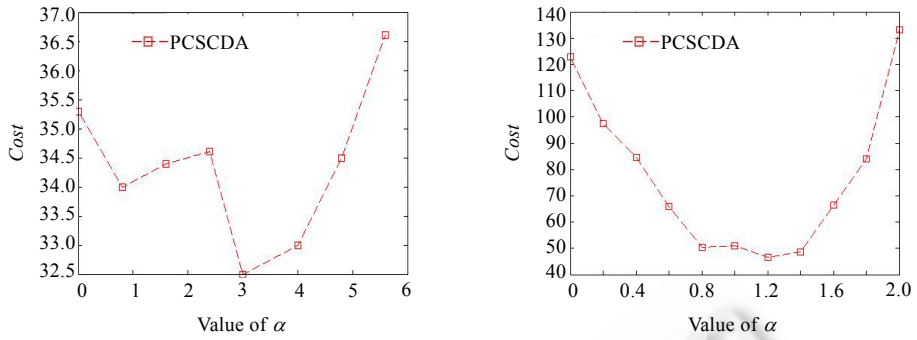
(5) 整体的识别错误率 Err.

3.2 实验结果与分析

3.2.1 参数α对 PCSCDA 算法的影响

在 PCSCDA 算法中,平衡参数α是比较重要的参数,用于平衡数据分布密度不同的类对类内协方差矩阵 S_W 的贡献.实验在 ORL 和 Extended Yale B 数据集上,通过变化参数α的值检验参数α对 PCSCDA 的影响.实验采用 KNN 分类器,实验结果如图 3 所示.从图 3 可知,当$\alpha=3$ 以及$\alpha\in[0.8,1.4]$时,PCSCDA 算法分别在 ORL 数据集和 Extended Yale B 数据集上取得最优的分类结果.

© 中国科学院软件研究所 http://www.jos.org.cn



(a) ORL (b) Extended Yale B

Fig.3 Performance of PCSCDA vs. α

图 3 PCSCDA 在不同α值下的性能

3.2.2 不同降维方法的识别性能比较

为了检验算法的性能,实验在表 2 的数据集设置下,采用第 3.1.2 节描述的 6 种对比方法和全部 5 种评价准则.由于各算法降维后的维数不同,这里选择算法的最优结果进行识别性能比较,结果见表 3.

Table 3 Classification performance of different dimensionality reduction methods (mean±std)

表 3 不同降维方法的分类性能(平均值±标准差)

		LDA	aPAC	CSLDA	SDA	SWBDA	LSDA	PCSCDA
ORL	*Cost* (最佳维数)	54.5±23 (30)	50.3±23 (30)	46.6±22 (30)	52.4±29 (38)	50.3±33 (37)	69.8±52 (45)	32.5±24 (39)
	Err_{IG} (%)	0.65±1.1	0.6±1	0.5±0.9	1±1.4	0.9±1.6	1.2±2.6	0.65±1.1
	Err_{GI} (%)	6.3±3.8	5.75±3.6	6±4.3	1.25±1.5	2.1±2.4	1±1.2	0.2±0.6
	Err_{GG} (%)	1.75±2.15	1.65±1.4	1.2±0.8	3.7±1.9	2.9±2.2	7.8±2.9	2.85±2
	Err (%)	8.65±4	8±4	7.7±5	5.9±3	6±3.7	10±4	3.7±3
Extended Yale B	*Cost* (最佳维数)	110±34 (27)	112±29 (30)	284±82 (30)	72±32 (35)	98.6±40 (37)	180.7±66 (93)	46.6±21 (37)
	Err_{IG} (%)	0.21±1.8	0.22±1.6	0.95±4	0.13±1.5	0.24±2.1	0.49±3.5	0.09±0.9
	Err_{GI} (%)	1.7±2.1	1.78±3	1±3.3	0.9±2.4	0.93±2	1.35±4.8	0.36±1.4
	Err_{GG} (%)	0.98±5	0.8±4	1.2±5.3	1±4.4	1±4	1.5±2.1	1±4
	Err (%)	2.9±4	2.8±3.5	3.3±9	2.17±3.8	2.18±3.7	3.4±3.7	1.5±4.8

观察表 3 可知:

(1) SDA,SWBDA 以及 PCSCDA 的识别效果分别好于 LDA,aPAC.这说明本文对人脸识别门禁系统采用混合高斯分布的假设是合理的,可以将其视作一个子类学习问题.

(2) PCSCDA 的识别性能要优于其他的识别方法.这是因为 PCSCDA 模型近似于成对贝叶斯风险准则,所提取的特征更具有判别能力.

3.2.3 不同人脸识别方式的识别性能比较

在人脸识别问题中存在两种识别方式:

(1) 代价不敏感的降维算法(LDA,aPAC,SDA,SWBDA,LSDA)和代价敏感分类器(McKNN,McKLR[4])的识别组合;

(2) 代价敏感降维算法(PCSCDA)和代价不敏感的分类器(KNN,KLR[32])的识别组合.

为了比较这两种识别方式,本文在表 2 的数据集设置下,采用整体错分损失 *Cost* 作为评价准则.实验结果见表 4.由表 4 可知:基于 PCSCDA 降维算法的识别组合取得了最优的识别性能.这是因为:

(1) PCSCDA 模型近似于成对贝叶斯风险准则,提取的特征具有较强的判别能力;

(2) 在人脸识别的降维部分注入错分代价,有效地避免了代价信息在降维部分的丢失.

© 中国科学院软件研究所 http://www.jos.org.cn

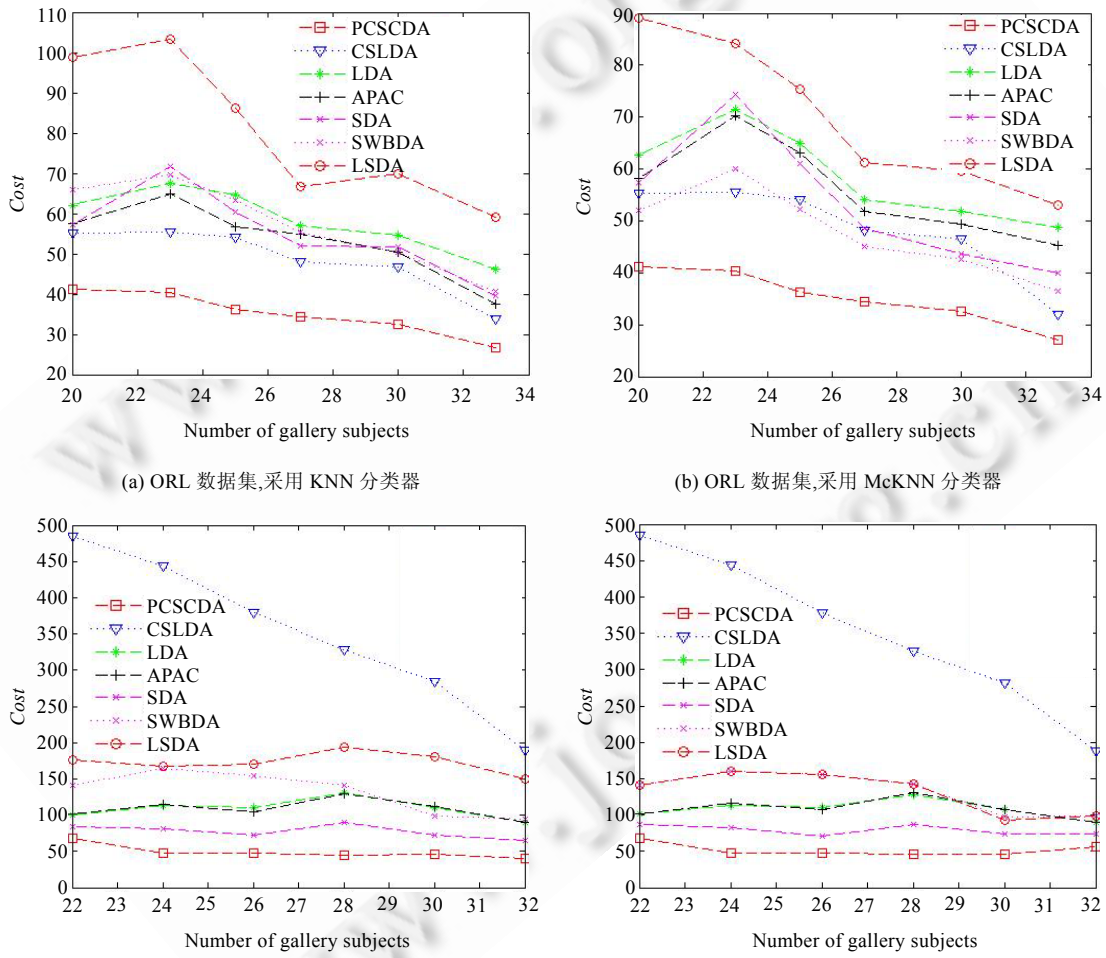
Table 4 Classification performance of different face recognition modes (mean±std)

表 4 不同人脸识别方式的分类性能(平均值±标准差)

	ORL	Extended Yale B		ORL	Extended Yale B
LDA+KNN	54.5±23(30)	110.1±34(27)	LDA+KLR	57±19(30)	2132.5±215(30)
LDA+McKNN	51.9±16(30)	108.1±30(27)	LDA+McKLR	90.7±11(29)	1967±198(30)
aPAC+McKNN	49.4±14(30)	107.4±30(27)	aPAC+McKLR	91.8±10(29)	1927±180(30)
SDA+McKNN	39.5±23(35)	74.3±32(35)	SDA+McKLR	84.3±6.6(35)	1894±151(35)
SWBDA+McKNN	42.6±20(37)	97.1±37(37)	SWBDA+McKLR	97.6±9.3(38)	1932±165(36)
LSDA+McKNN	59.6±23(36)	174.8±34(181)	LSDA+McKLR	70.4±11(88)	1946±182(100)
PCSCDA+KNN	32.5±24(39)	46.6±21(37)	PCSCDA+KLR	49.8±15(39)	1724±153(37)

3.2.4 算法在不同合法者类别数下的识别性能比较

为了检验 PCSCDA 在不同的合法者类别数下的性能,本文令 C_{IG}:C_{GI}:C_{GG}=20:2:1,通过改变合法者的类别数来检验算法的性能.实验中采用的评价指标为整体错分损失 *Cost*,选择 KNN 和 McKNN 作为分类器.ORL 和 Extended Yale B 数据集中合法者的类别数分别为 M_{ORL}=20,23,25,27,30,33,$M_{\mathrm{Extended\ YaleB}}$=22,24,26,28,30,32.实验结果如图 4 所示.由图 4 可知,PCSCDA 在不同的合法者类别数下都取得了较低的错分代价.



(a) ORL 数据集,采用 KNN 分类器

(b) ORL 数据集,采用 McKNN 分类器

(c) Extended Yale B 数据集,采用 KNN 分类器

(d) Extended Yale B 数据集,采用 McKNN 分类器

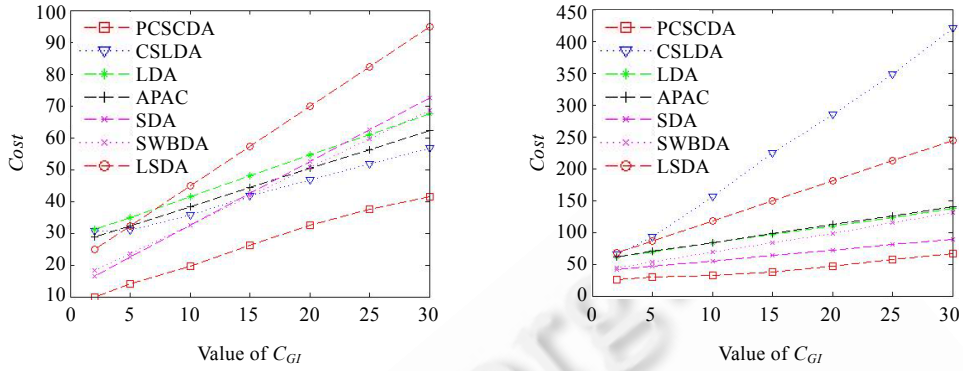
Fig.4 Classification performance under different gallery persons

图 4 不同合法者人数下的分类性能

© 中国科学院软件研究所 http://www.jos.org.cn

3.2.5 算法在不同错分代价比下的识别性能比较

为了检验 PCSCDA 在不同错分代价下的性能,本文令 C_{GI}=2,C_{GG}=1,通过改变 C_{IG} 的值,观察算法的性能.实验中采用的评价指标为整体错分损失 *Cost*,采用 KNN 分类器.C_{IG} 的值设置为 C_{IG}=2,5,10,15,20,25,30.实验结果如图 5 所示.由图 5 可知,PCSCDA 算法在不同的错分代价下都取得了较低的错分代价.



(a) ORL 数据集 (b) Extended Yale B 数据集

Fig.5 Classification performance under different misclassification costs

图 5 不同错分代价下的分类性能

3.2.6 算法的执行时间比较

实验采用 Matlab R2009 软件,使用的电脑配置如下:2.83GHz CPU,2G RAM.在表 2 的数据集设置下,采用 KNN 分类器,算法的训练时间和测试时间见表 5.表 5 表明:PCSCDA 的训练时间长于其他 6 种算法,测试时间与 SDA,SWBDA,LSDA 算法可比较.SDA,SWBDA,LSDA 以及 PCSCDA 测试时间较长,是因为这些算法可以提取更多的有用特征,它们的最佳维数一般高于基于类结构粒度的算法(LDA,aPAC,CSLDA).

Table 5 Comparison of the running time (mean±std) (s)

表 5 执行时间的比较(平均值±标准差) (秒)

		LDA	aPAC	CSLDA	SDA	SWBDA	LSDA	PCSCDA
Extended Yale B	Train	7.9±0.7	19.7±1.1	8.09±1.1	12.9±0.6	5.2±0.6	3.8±0.8	22.1±0.8
	Test	17.3±0.5	17.6±0.9	17.2±1.1	23.5±1	25.3±1	108±7.8	25±1
ORL	Train	7.53±0.1	9.59±0.5	9.3±0.73	12.1±0.4	4.77±0.4	2.59±0.1	16±1.1
	Test	0.58±0.03	0.56±0.02	0.57±0.03	0.75±0.03	0.83±0.1	0.96±0.03	0.74±0.02

4 结束语

本文基于数据的聚类结构粒度,提出了一种引入成对代价的子类判别分析算法 PCSCDA.不仅将聚类信息和错分代价同时注入判别分析算法,而且通过分析人脸识别的门禁系统,将其视为一个子类学习问题,并将 PCSCDA 算法成功地应用到其中.人脸数据集 ORL 和 Extended Yale B 上的实验结果表明了 PCSCDA 算法的有效性.

在真实的人脸识别问题中,错分代价很难得到,因此,对错分代价的学习[4,35]是未来的一个主要研究方向.

References:

[1] Belhumeur PN, Hespanha J. Eigenfaces vs. fisherfaces: Recognition using specific linear projection. IEEE Trans. on Pattern Analysis and Machine Intelligence, 1997,19(7):711−720. [doi: 10.1109/34.598228]

[2] Qiao LS, Chen SC, Tan XY. Sparsity preserving projections with applications to face recognition. Pattern Recognition, 2010,43(1): 331−341. [doi: 10.1016/j.patcog.2009.05.005]

© 中国科学院软件研究所 http://www.jos.org.cn

[3] Wright J, Yang A, Sastry S, Ma Y. Robust face recognition via sparse representation. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2009,31(2):210−227. [doi: 10.1109/TPAMI.2008.79]

[4] Zhang Y, Zhou ZH. Cost-Sensitive face recognition. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2010,32(10): 1758−1769. [doi: 10.1109/TPAMI.2009.195]

[5] Lu JW, Tan YP. Cost-Sensitive subspace learning for face recognition. In: Mortensen EN, Yang MH, eds. Proc. of the IEEE Int'l Conf. on Computer Vision and Pattern Recognition. San Francisco: IEEE Computer Society, 2010. 2661−2666. [doi: 10.1109/CVPR.2010.5539983]

[6] Lu JW, Tan YP. Cost-Sensitive subspace learning for human age estimation. In: Proc. of the IEEE 17th Int'l Conf. on Image Processing. IEEE Computer Society, 2010. 1593−1596. [doi: 10.1109/ICIP.2010.5650873]

[7] Lu JW, Zhou XZ, Tan YP. Cost-Sensitive semi-supervised discriminant analysis for face recognition. IEEE Trans. on Information Forensics and Security, 2012,7(3):944−953. [doi: 10.1109/TIFS.2012.2188389]

[8] Man JY, Jing XY, Zhang D, Lan C. Sparse cost-sensitive classifier with application to face recognition. In: Proc. of the IEEE 18th Int'l Conf. on Image Processing. IEEE Computer Society, 2011. 1773−1776. [doi: 10.1109/ICIP.2011.6115804]

[9] Ahonen T, Hadid A. Face description with local binary patterns. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2006, 28(12):2014−2037. [doi: 10.1109/TPAMI.2006.244]

[10] Georghiades AS, Belhumeur PN, Kriegman D. From few to many: Illumination cone models for face recognition under variable lighting and pose. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2001,23(6):643−660. [doi: 10.1109/34.927464]

[11] He XF, Yan SC, Hu YX, Niyogi P, Zhang HJ. Face recognition using laplacianfaces. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2005,27(3):328−340. [doi: 10.1109/TPAMI.2005.55]

[12] Jain K, Ross A, Prabhakar S. An introduction to biometric recognition. IEEE Trans. on Circuits and Systems for Video Technology, 2004,14(1):4−20. [doi: 10.1109/TCSVT.2003.818349]

[13] Li F, Wechsler H. Open set face recognition using transduction. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2005, 27(11):1686−1697. [doi: 10.1109/TPAMI.2005.224]

[14] Turk M, Pentland A. Eigenfaces for recognition. Journal of Cognitive Neuroscience, 1991,3(1):71−86. [doi: 10.1162/jocn.1991.3.1.71]

[15] Zhao W, Chellappa R, Phillips PJ. Face recognition: A literature survey. ACM Computing Surveys, 2003,35(1):399−458. [doi: 10.1145/954339.954342]

[16] Wan JW, Yang M, Gao Y, Chen YJ. A weighted cost sensitive locality preserving projection for face recognition. Ruan Jian Xue Bao/Journal of Software, 2013,24(5):1155−1164 (in Chinese with English abstract). http://www.jos.org.cn/1000-9825/4263.htm [doi: 10.3724/SP.J.1001.2013.04263]

[17] Samaria F, Harter A. Parameterisation of a stochastic model for human face identification. In: Proc. of the 2nd IEEE Workshop Applications Computer Vision. IEEE Computer Society, 1994. 138−142. [doi: 10.1109/ACV.1994.341300]

[18] Lee Y, Lin Y, Wahba G. Multicategory support vector machines: Theory and application to the classification of microarray data and satellite radiance data. Journal of American Statistical Association, 2004,99(465):67−81. [doi: 10.1198/016214504000000098]

[19] Li YF, Kwok J, Zhou ZH. Cost-Sensitive semi-supervised support vector machine. In: Proc. of the 24th AAAI Conf. on Artificial Intelligence (AAAI 2010). 2010. 500−505.

[20] Raudys S, Raudys A. Pairwise costs in multiclass perceptrons. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2010, 32(7):1324−1328. [doi: 10.1109/TPAMI.2010.72]

[21] Vandera S. CSNL: A cost-sensitive non-linear decision tree algorithm. ACM Trans. on Knowledge Discovery from Data, 2010,4(2): 6−25. [doi: 10.1145/1754428.1754429]

[22] Shirazi HM, Vasconcelos N. Cost-Sensitive boosting. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2011,33(2): 294−309. [doi: 10.1109/TPAMI.2010.71]

[23] Morik K, Brochhausen P. Combining statistical learning with a knowledge-based approach: A case study in intensive care monitoring. In: Proc. of the 16th Int'l Conf. on Machine Learning. Morgan Kaufmann Publishers, 1999. 268−277.

[24] Huang R, Liu Q, Lu H, Ma S. Solving the small sample size problem of LDA. In: Proc. of the 16th Int'l Conf. on Pattern Recognition, Vol.3. IEEE Computer Society, 2002. 29−32. [doi: 10.1109/ICPR.2002.1047787]

© 中国科学院软件研究所 http://www.jos.org.cn

[25] Yang B, Chen SC, Wu XD. A structurally motivated framework for discriminant analysis. Pattern Analysis and Applications, 2011, 14:349−367. [doi: 10.1007/s10044-011-0228-8]

[26] Zhu M, Martinez AM. Subclass discriminant analysis. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2006,28(8): 1274−1286. [doi: 10.1109/TPAMI.2006.172]

[27] You D, Martinez AM. Bayes optimal kernel discriminant analysis. IEEE Int'l Conf. on Computer Vision and Pattern Recognition. San Francisco: IEEE Computer Society, 2010. 3533−3538. [doi: 10.1109/CVPR.2010.5539952]

[28] Blake C. UCI repository of machine learning databases. 1998. http://www.ics.uci.edu/~mlearn/MLRepository.html

[29] Loog M, Duin RPW, Umbach RH. Multiclass linear dimension reduction by weighted pairwise fisher criteria. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2001,23(7):762−766. [doi: 10.1109/34.935849]

[30] Loog M. Approximate pairwise accuracy criteria for multiclass linear dimension reduction: Generalisations of the fisher criteria [Ph.D. Thesis]. Delft: University of Delft Press, 1999.

[31] Fisher RA. The use of multiple measurements in taxonomic problems. Annals of Eugenics, 1936,7(2):179−188.

[32] Hastie T, Tibshirani P, Friedman J. The Elements of Statistical Learning. 2nd ed., New York: Springer-Verlag, 2001.

[33] Ting KM. An instance-weighting method to induce cost-sensitive trees. IEEE Trans. on Knowledge and Data Engineering, 2002, 14(3):659−665. [doi: 10.1109/TKDE.2002.1000348]

[34] Cai D, He XF, Zhou K. Locality sensitive discriminant analysis. In: Proc. of the Int'l Joint Conf. on Artificial Intelligence. Morgan Kaufmann Publishers, 2007. 1713−1726.

[35] Liu XY, Zhou ZH. Learning with cost intervals. In: Proc. of the 16th ACM SIGKDD Conf. on Knowledge Discovery and Data Ming. 2010. 494−505. [doi: 10.1145/1835804.1835857]

附中文参考文献:

[16] 万建武,杨明,吉根林,陈银娟.一种面向人脸识别的加权代价敏感局部保持投影.软件学报,2013,24(5):1155−1164. http://www.jos.org.cn/1000-9825/4263.htm [doi: 10.3724/SP.J.1001.2013.04263]

附录

定理 1. 如果子类 ij 和 kl 满足图 2 所示的分布, $\tilde{\mu}_{ij} < \tilde{\mu}_{kl}$,$P(ijk)<P(kli)$,那么 $\max_W : A_{ij,kl}(W) = \max_W : \tilde{A}_{ij,kl}(W)$.

证明:已知 $\mathrm{erf}(x)$是一个单调递增函数,并且满足 $\lim_{x\to+\infty}\mathrm{erf}(x)=1, \lim_{x\to-\infty}\mathrm{erf}(x)=-1$. 这里把 $A_{ij,kl}(W)$分成两个部分: $P(kli)\left[\frac{1}{2}+\frac{1}{2}\mathrm{erf}\left(\frac{\ln\frac{P(kli)}{P(ijk)}}{\sqrt{2}}\frac{1}{\tilde{\mu}_{kl}-\tilde{\mu}_{ij}}+\frac{\tilde{\mu}_{kl}-\tilde{\mu}_{ij}}{2\sqrt{2}}\right)\right]$ 和 $P(ijk)\left[\frac{1}{2}+\frac{1}{2}\mathrm{erf}\left(\frac{\ln\frac{P(ijk)}{P(kli)}}{\sqrt{2}}\frac{1}{\tilde{\mu}_{kl}-\tilde{\mu}_{ij}}+\frac{\tilde{\mu}_{kl}-\tilde{\mu}_{ij}}{2\sqrt{2}}\right)\right]$.

首先,求 $\frac{\ln\frac{P(ijk)}{P(kli)}}{\sqrt{2}}\frac{1}{\tilde{\mu}_{kl}-\tilde{\mu}_{ij}}+\frac{\tilde{\mu}_{kl}-\tilde{\mu}_{ij}}{2\sqrt{2}}$ 关于 $\tilde{\mu}_{kl}-\tilde{\mu}_{ij}$ 的一阶导数,得到:

$$\left(\frac{\ln\frac{P(ijk)}{P(kli)}}{\sqrt{2}}\frac{1}{\tilde{\mu}_{kl}-\tilde{\mu}_{ij}}+\frac{\tilde{\mu}_{kl}-\tilde{\mu}_{ij}}{2\sqrt{2}}\right)'_{\tilde{\mu}_{kl}-\tilde{\mu}_{ij}} = -\frac{\ln\frac{P(ijk)}{P(kli)}}{\sqrt{2}}\frac{1}{(\tilde{\mu}_{kl}-\tilde{\mu}_{ij})^2}+\frac{1}{2\sqrt{2}}>0 \tag{25}$$

根据复合函数的单调性,如果满足 $u=g(x)$单调增,$f(u)$单调增,那么 $f(g(x))$关于 x 单调增,可以得到 $P(ijk)\left[\frac{1}{2}+\frac{1}{2}\mathrm{erf}\left(\frac{\ln\frac{P(ijk)}{P(kli)}}{\sqrt{2}}\frac{1}{\tilde{\mu}_{kl}-\tilde{\mu}_{ij}}+\frac{\tilde{\mu}_{kl}-\tilde{\mu}_{ij}}{2\sqrt{2}}\right)\right]$ 关于 $\tilde{\mu}_{kl}-\tilde{\mu}_{ij}$ 在$[0,+\infty]$上单调增.这里,根据已知条件 $\tilde{\mu}_{kl}-\tilde{\mu}_{ij}>0$, 故 $\tilde{\mu}_{kl}-\tilde{\mu}_{ij}\in[0,+\infty]$.

© 中国科学院软件研究所 http://www.jos.org.cn

然后,求 $\frac{\ln\frac{P(kli)}{P(ijk)}}{\sqrt{2}}\frac{1}{\tilde{\mu}_{kl}-\tilde{\mu}_{ij}}+\frac{\tilde{\mu}_{kl}-\tilde{\mu}_{ij}}{2\sqrt{2}}$ 关于 $\tilde{\mu}_{kl}-\tilde{\mu}_{ij}$ 的二阶导数,得到:

$$\left(\frac{\ln\frac{P(kli)}{P(ijk)}}{\sqrt{2}}\frac{1}{\tilde{\mu}_{kl}-\tilde{\mu}_{ij}}+\frac{\tilde{\mu}_{kl}-\tilde{\mu}_{ij}}{2\sqrt{2}}\right)''_{\tilde{\mu}_{kl}-\tilde{\mu}_{ij}}=\sqrt{2}\ln\frac{P(kli)}{P(ijk)}(\tilde{\mu}_{kl}-\tilde{\mu}_{ij})^{-3}>0 \tag{26}$$

根据凸函数性质,$\frac{\ln\frac{P(kli)}{P(ijk)}}{\sqrt{2}}\frac{1}{\tilde{\mu}_{kl}-\tilde{\mu}_{ij}}+\frac{\tilde{\mu}_{kl}-\tilde{\mu}_{ij}}{2\sqrt{2}}$ 为凸函数.

又由于 $\mathrm{erf}(x)$是一个单调递增函数,故 $P(ijk)\left[\frac{1}{2}+\frac{1}{2}\mathrm{erf}\left(\frac{\ln\frac{P(ijk)}{P(kli)}}{\sqrt{2}}\frac{1}{\tilde{\mu}_{kl}-\tilde{\mu}_{ij}}+\frac{\tilde{\mu}_{kl}-\tilde{\mu}_{ij}}{2\sqrt{2}}\right)\right]$在两个端点处:0 或者+∞,取得最大值.

综上,$A_{ij,kl}(W)$的最大值只可能存在于端点处:0 或者+∞.令 $\tilde{\mu}_{kl}-\tilde{\mu}_{ij}=0$, 得到:

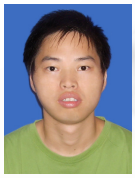
$$A_{ij,kl}(W)|_0=P(kli)+0=P(kli) \tag{27}$$

令 $\tilde{\mu}_{kl}-\tilde{\mu}_{ij}=+\infty$, 得到:

$$A_{ij,kl}(W)|_{+\infty}=P(kli)+P(ijk) \tag{28}$$

故,$A_{ij,kl}(W)$在 $\tilde{\mu}_{kl}-\tilde{\mu}_{ij}=+\infty$ 处取得最大值,最大值为 $P(kli)+P(ijk)$.

同理,可以证明 $\tilde{A}_{ij,kl}(W)$ 在 $\tilde{\mu}_{kl}-\tilde{\mu}_{ij}=+\infty$ 处取得最大值,最大值为 $P(kli)+P(ijk)$.定理得证. □



万建武(1986－),男,江苏常州人,博士,主要研究领域为数据挖掘,机器学习,模式识别.

E-mail: jianwuwan@gmail.com

杨明(1964－),男,博士,教授,博士生导师,主要研究领域为数据挖掘,机器学习,模式识别.

E-mail: m.yang@njnu.edu.cn

© 中国科学院软件研究所 http://www.jos.org.cn