

核偏最小二乘回归在面部表情识别中的应用

朱 炯

(中国矿业大学计算机科学与技术学院, 江苏 徐州 221000)

[摘要] 面部表情识别作为一个分类问题, 由于其在人机交互中扮演着重要角色, 已成为人脸识别领域的前沿研究方向. 本文分析了(核)偏最小二乘回归并获得了一种新方法来解决这些问题. 此外还发现核偏最小二乘回归的第一阶段等价于特征提取的广义判别分析并证明了核偏最小二乘回归可以通过定义虚拟矩阵直接应用于分类问题. 通过实验发现, 提出的线性降维改进算法在大多数情况下优于其他常用算法; 基于面部表情识别的核偏最小二乘回归算法在人脸数据集上也取得了良好的分类效果.

[关键词] 人脸识别, 面部表情, 特征提取, 核偏最小二乘回归

[中图分类号] TP183 [文献标志码] A [文章编号] 1001-4616(2018)03-0014-08

Kernel Partial Least Squares Regression with Applications to Facial Expression Recognition

Zhu Jiong

(School of Computer Science and Technology, China University of Mining and Technology, Xuzhou 221000, China)

Abstract: As a classification problem, facial expression recognition has been playing an important role in human-machine interaction, and has become a frontier research direction in face recognition field. This paper analyzes (kernel) partial least squares regression and obtains a new method to solve them. Moreover, it is of interest to note that the first stage of kernel partial least squares regression is equivalent to generalized discriminant analysis for feature extraction. At last, this paper also demonstrates that the kernel partial least squares regression can be directly applied to the classification problems by defining the dummy matrix. In conclusion, experimental results show that the proposed algorithm outperforms other linear dimensionality reduction algorithm commonly used in most cases; the kernel partial least squares regression algorithm based on facial expression recognition also achieves good classification performance.

Key words: face recognition, facial expression, feature extraction, kernel partial least squares regression

人的面部表情含有丰富的人体行为信息, 在人与人的交往中, 面部表情是除了声音以外的非常重要的通信方式, 作为信息的载体, 表情能够传达很多语音所不能传达的信息, 因此在语言学中、医疗中、商业等方面都有着十分重要的应用. 在计算机科学领域的研究, 面部表情的识别对于实现自然的人机交互以及自动图像理解也具有重大的意义. 面部表情自动识别 (Automatic Facial Expression Analysis) 对人脸的表情信息进行特征提取分析, 按照人的认识和思维方式加以归类和理解, 利用人类所具有的情感信息方面的先验知识, 使计算机进行联想、思考及推理, 进而从人脸信息中去分析理解人的情绪^[1-2]. 在人脸表情识别中, 经常是提供人脸表情的类信息的^[3-4]. 然而, 在某些情况下, 由于面部表情的多样性和复杂性, 使用语义代表面部图像是有效的, 语义毕竟是信息表示的一种形式. 虽然很多根据类信息的面部表情识别算法已被提出, 但在语义表征信息的面部表情识别方面做的工作还是很少. 在某种意义上, 面部表情识别可以被视为语义信息方面的回归问题^[5-6]. 因此, 偏最小二乘回归算法 (partial least square, PLS) 适用于语义信息方面的面部表情识别.

本文分析了(核)偏最小二乘回归(kernel partial least square, KPLS)并获得了一种新方法来解决这些

收稿日期: 2018-04-16.

基金项目: 国家自然科学基金(61672522).

通讯联系人: 朱炯, 硕士研究生, 研究方向: 机器学习. E-mail: 804194762@qq.com

问题,即权重可以通过求解一个广义特征值问题来得到. 我们发现了核偏最小二乘回归的第一阶段等价于特征提取的广义判别分析(generalized discriminant analysis, GDA)并通过定义虚拟矩阵的方式使其应用于分类问题,实现了回归方程在面部表情识别分类上的应用.

1 偏最小二乘算法概述及新的模型算法

1.1 偏最小二乘算法概述

假设 $x_i \in \mathbf{R}^N (i=1, \dots, n)$ 是 n 个数据点, $y_i \in \mathbf{R}^L (i=1, \dots, n)$ 是对应的响应值. 令 $\mathbf{X} = [x_1, \dots, x_n]^T, \mathbf{Y} = [y_1, \dots, y_n]^T$. 因此, \mathbf{X} 是一个 $n \times N$ 维的矩阵, \mathbf{Y} 是 $n \times L$ 的矩阵. 在 PLS 的框架中, 假设 \mathbf{X} 和 \mathbf{Y} 不失广义性的集中. 事实上, 这是很容易通过进行数据转换获得的. PLS 回归的目的是从 \mathbf{X} 中预测 \mathbf{Y} 并且描述 \mathbf{X} 和 \mathbf{Y} 中的共同结构. 具体来说, PLS 需要将 \mathbf{X} 和 \mathbf{Y} 分解成如下形式:

$$\mathbf{X} = \mathbf{T}\mathbf{P}^T + \mathbf{E}, \quad (1)$$

$$\mathbf{Y} = \mathbf{U}\mathbf{Q}^T + \mathbf{F}, \quad (2)$$

其中, \mathbf{T} 和 \mathbf{U} 是提取出来的 k 个得分向量的 $n \times k$ 的矩阵, \mathbf{P} 和 \mathbf{Q} 是载荷矩阵, \mathbf{E} 和 \mathbf{F} 是误差矩阵. 在非线性交互偏最小二乘算法(NIPALS)中, 权值向量 \mathbf{w} 和 \mathbf{c} 可通过下式得到:

$$[\text{cov}(\mathbf{t}, \mathbf{u})]^2 = [\text{cov}(\mathbf{X}\mathbf{w}, \mathbf{Y}\mathbf{c})]^2 = \max_{\|\mathbf{r}\| = \|\mathbf{s}\| = 1} [\text{cov}(\mathbf{X}\mathbf{r}, \mathbf{Y}\mathbf{s})]^2, \quad (3)$$

其中, $\text{cov}(\mathbf{t}, \mathbf{u}) = \frac{\mathbf{t}^T \mathbf{u}}{n}$ 表示得分向量 \mathbf{t} 和 \mathbf{u} 之间的样本协方差. \mathbf{t} 和 \mathbf{u} 分别是 \mathbf{T} 和 \mathbf{U} 的列向量. NIPALS 算法通过下述步骤继续从矩阵 \mathbf{X} 和 \mathbf{Y} 中提取向量 \mathbf{t}, \mathbf{u} 以及权重向量 \mathbf{w} 和 \mathbf{c} .

(1) 初始化 \mathbf{u} ; (2) $\mathbf{w} = \mathbf{X}^T \mathbf{u} / (\mathbf{u}^T \mathbf{u})$, $\mathbf{w} \leftarrow \mathbf{w} / \|\mathbf{w}\|$; (3) $\mathbf{t} = \mathbf{X}\mathbf{w}$; (4) $\mathbf{c} = \mathbf{Y}^T \mathbf{t} / (\mathbf{t}^T \mathbf{t})$, $\mathbf{c} \leftarrow \mathbf{c} / \|\mathbf{c}\|$; (5) $\mathbf{u} = \mathbf{Y}\mathbf{c}$; (6) 重复步骤 2~5 直到收敛; (7) 收敛 \mathbf{X} 和 \mathbf{Y} , $\mathbf{X} \leftarrow \mathbf{X} - \mathbf{t}\mathbf{t}^T \mathbf{X} / (\mathbf{t}^T \mathbf{t})$, $\mathbf{Y} \leftarrow \mathbf{Y} - \mathbf{u}\mathbf{u}^T \mathbf{Y} / (\mathbf{u}^T \mathbf{u})$.

显然, NIPALS 算法在提取出得分向量 \mathbf{t} 和 \mathbf{u} 之后再次开始使用了收敛矩阵 \mathbf{X} 和 \mathbf{Y} . 此外有几种不同的为 PLS 回归修改的 NIPALS 算法. 其中一种叫做 Lewi 方法, 它经常被应用于 PLS 回归, 算法步骤如下:

(1) 初始化 \mathbf{u} ; (2) $\mathbf{w} = \mathbf{X}^T \mathbf{u}$; (3) $\mathbf{t} = \mathbf{X}\mathbf{w}$, $\mathbf{t} \leftarrow \mathbf{t} / \|\mathbf{t}\|$; (4) $\mathbf{c} = \mathbf{Y}^T \mathbf{t}$; (5) $\mathbf{u} = \mathbf{Y}\mathbf{c}$, $\mathbf{u} \leftarrow \mathbf{u} / \|\mathbf{u}\|$; (6) 重复步骤 2~5 直到收敛; (7) 收敛 \mathbf{X} 和 \mathbf{Y} , $\mathbf{X} \leftarrow \mathbf{X} - \mathbf{t}\mathbf{t}^T \mathbf{X}$, $\mathbf{Y} \leftarrow \mathbf{Y} - \mathbf{t}\mathbf{t}^T \mathbf{Y}$.

1.2 偏最小二乘回归的新模型及其算法

我们能发现在 Lewi 方法和 NIPALS 之间有一些不同点. 在 Lewi 方法中, 得分向量 \mathbf{t} 和 \mathbf{u} 是归一化的, 然而在 NIPALS 算法中, 权重向量 \mathbf{w} 和 \mathbf{c} 是归一化的. 此外, 在这两种方法中, 收敛矩阵 \mathbf{X} 和 \mathbf{Y} 的形式也是不同的. 在 PLS 回归中, 经常假设在得分向量 \mathbf{t} 和 \mathbf{u} 之间存在线性关系, 即在得分矩阵之间存在如下关系:

$$\mathbf{U} = \mathbf{T}\mathbf{D} + \mathbf{H}, \quad (4)$$

此处, \mathbf{D} 是对角阵, \mathbf{H} 表示误差矩阵. 把式(4)代入式(2)得到:

$$\mathbf{Y} = (\mathbf{T}\mathbf{D} + \mathbf{H})\mathbf{Q}^T + \mathbf{F} = \mathbf{T}\mathbf{D}\mathbf{Q}^T + (\mathbf{H}\mathbf{Q}^T + \mathbf{F}). \quad (5)$$

由式(5), 如下的线性 PLS 回归模型可以被定义:

$$\mathbf{Y} = \mathbf{T}\mathbf{C}^T + \mathbf{F}^*, \quad (6)$$

其中, $\mathbf{C}^T = \mathbf{D}\mathbf{Q}^T$ 表示矩阵回归系数, $\mathbf{F}^* = \mathbf{H}\mathbf{Q}^T + \mathbf{F}$ 是 \mathbf{Y} 的误差矩阵. 为了模拟 \mathbf{X} 和 \mathbf{Y} 之间的关系, 可以使用下式:

$$\mathbf{T} = \mathbf{X}\mathbf{W}(\mathbf{P}^T \mathbf{W})^{-1}, \quad (7)$$

其中, \mathbf{W} 是权重矩阵, 它的列包含了 Lewi 算法里的权重向量 \mathbf{w} , \mathbf{P} 是在式(1)里定义的载荷向量组成的矩阵. 将式(7)代入式(6), 得到:

$$\mathbf{Y} = \mathbf{X}\mathbf{B} + \mathbf{F}^*, \quad (8)$$

式中, $\mathbf{B} = \mathbf{W}(\mathbf{P}^T \mathbf{W})^{-1} \mathbf{C}^T$ 表示矩阵的回归参数. 进一步而言, $\mathbf{T}, \mathbf{P}, \mathbf{W}$ 和 \mathbf{U} 之间存在如下关系:

$$\mathbf{P} = \mathbf{X}^T \mathbf{T} (\mathbf{T}^T \mathbf{T})^{-1}, \quad (9)$$

$$\mathbf{C} = \mathbf{Y}^T \mathbf{T} (\mathbf{T}^T \mathbf{T})^{-1}, \quad (10)$$

$$\mathbf{W} = \mathbf{X}^T \mathbf{U}. \quad (11)$$

值得注意的是, Lewi 方法的收敛方案保证所提取的分数向量的相互正交性及 \mathbf{T} 的分数向量的归一

化. 在这种情况下,把式(9)~(11)代入矩阵 B 得到:

$$B = X^T U (T^T X X^T U)^{-1} T^T Y. \quad (12)$$

在式(12)中可以看到两个参数矩阵 U 和 T . 为获得回归方程,我们需要将 U 和 T 的列向量通过上述提到的 Lewi 方法而得到. 因此,PLS 回归的估计值能利用式(12)中的矩阵 B 得到.

值得注意的是,我们需要通过实现 Lewi 方法来得到矩阵 B . 我们给出一种新的求解矩阵 B 的方法. 首先将式(7),(11)代入式(12)得到:

$$B = W (W^T X^T X W)^{-1} W^T X^T Y. \quad (13)$$

比较式(12)与式(13),发现式(13)只包含了未知矩阵 W ,当矩阵 W 已知时,我们就能得到回归方程.

下面阐述如何得到矩阵 W . 假设当足够多的分量被提取时, Y 用 XB 近似构造,可以得到下式:

$$\hat{Y} = XB = XW (W^T X^T X W)^{-1} W^T X^T Y. \quad (14)$$

将式(14)右乘 $Y^T X W$,左乘 X^T ,可得:

$$X^T \hat{Y} Y^T X W = X^T X W (W^T X^T X W)^{-1} W^T X^T Y Y^T X W. \quad (15)$$

令 $\Gamma = (W^T X^T X W)^{-1} W^T X^T Y Y^T X W$,即得:

$$X^T \hat{Y} Y^T X W = X^T X W \Gamma. \quad (16)$$

很显然式(16)是一个广义特征值问题. 通过解决式(16)可得到矩阵 W . 式(16)中的 \hat{Y} 是未知的. 为了解决这个问题,在实际应用中 \hat{Y} 采用了响应矩阵 Y 并应用联合对角化. 明确的说,找到矩阵 W 就是同时使矩阵 $X^T X$ 和 $X^T Y Y^T X$ 对角化,例如 $W^T X^T X W = I$ 和 $W^T X^T Y Y^T X W = \Gamma$,在这里 I 是单位矩阵, Γ 是元素按降序排列的对角阵. 能观察到式(16)的结果关于 $X^T X$ 是共轭正交的. 然而,在一些 PLS 回归问题中需要矩阵 W 是列正交的. 为此,可以通过改进 Gram-Schmidt 方法来得到正交矩阵 \bar{W} . 此外,发现当样本数量小于样本维度时 $X^T X$ 是奇异的. 幸运的是,一些有效克服广义特征方程中奇异性问题的算法在近些年已经被提出^[7]. 这里我们采用广义奇异值分解^[8]来解决式(16).

1.3 核偏最小二乘回归的新模型及其算法

给定一个非线性映射 ϕ ,输入数据空间 \mathbf{R}^N 被映射到特征空间 F :

$$\phi: \mathbf{R}^N \rightarrow F, x \mapsto \phi(x). \quad (17)$$

结果在原始空间的一种模式被映射到特征空间中的更高维向量. 值得注意的是,特征空间可能会有任意大,可能无限的维度. 因此在式(17)中数据 x_1, \dots, x_n 通过非线性映射 ϕ 映射到 $\phi(x_1), \dots, \phi(x_n)$. 通过定义内部运算可以得到一个核. 这个核是函数 k ,它满足 $k(x_i, x_j) = \langle \phi(x_i), \phi(x_j) \rangle, x_i, x_j \in \mathbf{R}^N$,此处的 ϕ 就是在式(17)中所定义的. 广泛运用的核函数是多项式核函数 $k(x_i, x_j) = (\langle x_i, x_j \rangle)^d$ 以及高斯核 $k(x_i, x_j) = \exp(-\|x_i - x_j\|^2 / \sigma)$, d 是多项式次数, σ 是高斯核的参数. 在特征空间中,式(18)有如下形式:

$$Y = \phi(X) B + F^*, \quad (18)$$

其中, $\phi(X) = [\phi(x_1), \dots, \phi(x_n)]^T$. 我们定义核矩阵 $K = \phi(X) \phi(X)^T$, Rosipal^[9] 基于 Lewi 方法及重构核理论为核 PLS 模型推导出如下算法:

(1) 初始化 u ; (2) $t = Ku, t \leftarrow t / \|t\|$; (3) $c = Y^T t / t^T t$; (4) $u = Yc, u \leftarrow u / \|u\|$; (5) 重复步骤 2~4 直到收敛; (6) 收敛 K 和 $Y, K \leftarrow (I - tt^T) K (I - tt^T), Y \leftarrow Y - tt^T Y$.

通过上述核 PLS 算法求得得分矩阵 U 和 T ,从而回归参数矩阵 B 也能解得,进一步的数据预测也可以得到. 类似于线性 PLS 回归,我们将给出一个新的方法在不使用得分矩阵 U 和 T 的情况下获得数据预测. 注意到式(16)可以表示成如下式子:

$$\phi(X)^T Y Y^T \phi(X) W = \phi(X)^T \phi(X) W \Gamma. \quad (19)$$

式(16)中的 \hat{Y} 用相应矩阵 Y 近似表示. 根据重构核理论^[10-11],每个 $W_j (j=1, \dots, k)$ 可以线性扩展为

$$W_j = \sum_{i=1}^n \alpha_i^j \phi(x_i), \quad (20)$$

式中, W_j 是矩阵 W 的第 j 列. 将式(20)代入式(19)得到:

$$\phi(X)^T Y Y^T \phi(X) \phi(X)^T A = \phi(X)^T \phi(X) \phi(X)^T A \Gamma, \quad (21)$$

其中, $A = [a_1, \dots, a_k], a_i = (\alpha_i^1, \dots, \alpha_i^n)^T$. 将式(21)左乘 $\phi(X)$,可得:

$$\phi(X) \phi(X)^T Y Y^T \phi(X) \phi(X)^T A = \phi(X) \phi(X)^T \phi(X) \phi(X)^T A \Gamma. \quad (22)$$

利用核矩阵 $\mathbf{K} = \boldsymbol{\phi}(\mathbf{X})\boldsymbol{\phi}(\mathbf{X})^T$, 可以进一步得到如下式子:

$$\mathbf{K}\mathbf{Y}\mathbf{Y}^T\mathbf{K}\mathbf{A} = \mathbf{K}\mathbf{K}\mathbf{A}\mathbf{\Gamma}. \quad (23)$$

式(23)等价于解决如下的广义瑞利商问题:

$$\mathbf{A}^* = \arg \max \text{trace} \left(\frac{\mathbf{A}^T \mathbf{K}\mathbf{Y}\mathbf{Y}^T \mathbf{K}\mathbf{A}}{\mathbf{A}^T \mathbf{K}\mathbf{K}\mathbf{A}} \right). \quad (24)$$

值得注意的是, 由于集中化的数据, \mathbf{K} 是一个奇异矩阵. 在此情况下, 可以使用广义奇异值分解技术来解决式(24). 因为矩阵 \mathbf{A} 通过解决式(24)而求得, 所以它不可能列正交, 可以使用修改后的 Gram-Schmidt 方法来使得 \mathbf{A} 正交化. 此外, 要注意在式(23)、(24)中计算 \mathbf{K} 时需要集中化的数据. 与核 PCA 相似, 在核 PLS 回归算法中也要考虑集中的数据^[12-13]. 因此, 可以在特征空间中应用下式来使得数据集中心化:

$$\mathbf{K}_{ce}^{tr} = \left(\mathbf{I} - \frac{1}{n} \mathbf{e}_n \mathbf{e}_n^T \right) \mathbf{K}^{tr} \left(\mathbf{I} - \frac{1}{n} \mathbf{e}_n \mathbf{e}_n^T \right), \quad (25)$$

$$\mathbf{K}_{ce}^{te} = \left(\mathbf{K}^{te} - \frac{1}{n} \mathbf{e} \mathbf{e}^T \mathbf{K}^{tr} \right) \left(\mathbf{I} - \frac{1}{n} \mathbf{e}_n \mathbf{e}_n^T \right), \quad (26)$$

式中, \mathbf{e}_n 是 $n \times 1$ 的元素全为 1 的向量, \mathbf{e} 是元素全为 1 且维度等于测试样本数量的向量, \mathbf{I} 是一个 $n \times n$ 的单位阵, \mathbf{K}^{tr} 是训练样本对应的核矩阵, \mathbf{K}^{te} 是根据训练样本和测试样本得到的核矩阵.

因此, 测试集的回归方程可以表示为:

$$\mathbf{Y}^{te} = \mathbf{K}_{ce}^{te} \mathbf{A} (\mathbf{A}^T \mathbf{K}_{ce}^{tr} \mathbf{K}_{ce}^{tr} \mathbf{A})^{-1} \mathbf{A}^T \mathbf{K}_{ce}^{tr} \mathbf{Y}. \quad (27)$$

具体而言, 从式(27)能得到测试样本的回归值. 由前面的分析将核 PLS 回归方法总结如下:

(1) 从式(25)中获得 \mathbf{K}^{tr} 和 \mathbf{K}_{ce}^{tr} ; (2) 求解广义特征方程(23)并求得 \mathbf{A} ; (3) 如果需要的話用修改后的 Gram-Schmidt 方法来正交化 \mathbf{A} ; (4) 为测试集从式(26)中求得 \mathbf{K}_{ce}^{te} ; (5) 将 \mathbf{A} , \mathbf{K}_{ce}^{te} 和 \mathbf{K}_{ce}^{tr} 代入式(27)求得 \mathbf{Y}^{te} .

从上述算法中看到我们提出的 KPLS 回归是一种非迭代方法. 为了得到回归方程, 通过解决广义特征值问题来求得矩阵 \mathbf{W} , 然而, 矩阵 \mathbf{U} 和 \mathbf{T} 是通过迭代过程获得的.

2 核偏最小二乘回归与核判别分析之间的关系

本节进一步分析核广义判别分析与核偏最小二乘回归之间的关系. 为简单起见, 在核广义判别分析中也同样考虑集中化数据. Baudat 和 Anouar 推导出了下式:

$$\lambda = \frac{\mathbf{a}^T \mathbf{K}\mathbf{G}\mathbf{K}\mathbf{a}}{\mathbf{a}^T \mathbf{K}\mathbf{K}\mathbf{a}}, \quad (28)$$

式中, \mathbf{a} 是一个判别向量, $\mathbf{G} = (\mathbf{G}_i)_{i=1,2,\dots,l}$ 是一个 $n \times n$ 的块对角阵, 它的子块 \mathbf{G}_i 是一个 $n_i \times n_i$ 且每一项都等于 $1/n_i$ 的矩阵, l 是类的数量, n_i 是类 i 中样本的数量.

对比式(24)与式(28), 可以发现这两个式有着相同的形式. 注意到式(28)只包含一个判别向量, 式(24)与式(28)最主要的区别之一是前者采用了回归矩阵 \mathbf{Y} , 而后者采用了矩阵 \mathbf{G} . 实际上, 当(核)偏最小二乘回归模型被应用到分类问题时, 我们需要定义一个虚拟矩阵. 虚拟矩阵一般定义如下 $\mathbf{Y} = [\mathbf{Y}_1, \dots, \mathbf{Y}_l]$, 此处 $\mathbf{Y}_i = [\underbrace{0, \dots, 0}_{n_1 + \dots + n_{i-1}}, \underbrace{1, \dots, 1}_{n_i}, \underbrace{0, \dots, 0}_{n_{i+1} + \dots + n_l}]^T$ 是一个 $n \times 1$ 向量. 因此, \mathbf{Y} 是一个 $n \times l$ 矩阵. 上述对于虚拟矩阵简单的定义没有包含每类样本的样本数量. 下面给出虚拟矩阵的另外一种定义. 具体而言, 令 $\mathbf{Y} = \hat{\mathbf{G}} = [\mathbf{g}_1, \dots, \mathbf{g}_i, \dots, \mathbf{g}_l]$, 此处 $\mathbf{g}_i = [\underbrace{0, \dots, 0}_{n_1 + \dots + n_{i-1}}, \underbrace{1/\sqrt{n_i}, \dots, 1/\sqrt{n_i}}_{n_i}, \underbrace{0, \dots, 0}_{n_{i+1} + \dots + n_l}]^T$ 是一个 $n \times 1$ 向量. 然后 $\hat{\mathbf{G}}(\hat{\mathbf{G}})^T = \mathbf{G}$

和式(24)在考虑投影向量的假设下等价于广义判别分析的投影向量采用核方法. 因此, 在判别过程中需要根据上述组合来选择 \mathbf{Y} 而定义一个合理的虚拟矩阵. 在本节, 我们通过广义奇异值分解来解决式(24). 它表明在此能得到 m 个特征向量, 而 m 是矩阵 $\mathbf{K} \times \mathbf{K}$ 的秩. 因此, \mathbf{A} 的列数至多是 m . 在我们的方法中, 通过广义特征值分解来得到 \mathbf{A} , 然而 Rosipal^[14] 通过解决一个特征值问题来求得矩阵 \mathbf{T} . 此外, Rosipal 还给出了 KCCA 与 KPLS 之间的关系. 正如该文中所指出的, 通过 KCCA 提取出来的分量数量受限

于矩阵 \mathbf{Y} 的秩,然而它在 KPLS 中受限于核矩阵 \mathbf{K} 的秩. 正如以上讨论的,在我们方法中因为通过广义奇异值分解来解决式(24),所以提取的分量数量受限于核矩阵 \mathbf{K} 的秩.

当 KPLS 回归为分类和判别问题设计时,测试样本的回归向量通过式(27)得到. 在很多情况下,测试样本的类归属于一个索引,该索引对应的值是向量中最大的,如下式:

$$\text{index} = \arg \max_j (\mathbf{Y}_j^{(i)}), \quad (29)$$

其中, $\mathbf{Y}_j^{(i)}$ 是式(27)中 \mathbf{Y}^{te} 第 i 列中的第 j 个分量.

下面分析在实际应用中的分类问题上为什么选式(29)作为标准. 首先分析在线性情况下的 PLS. 当测试样本给出后,回归变量能通过下式获得:

$$\mathbf{y} = \mathbf{x}^{te} \mathbf{W} (\mathbf{W}^T \mathbf{X}^T \mathbf{X} \mathbf{W})^{-1} \mathbf{W}^T \mathbf{X}^T \mathbf{Y}. \quad (30)$$

注意到,我们通过广义奇异值分解来解决式(16). 式 $\mathbf{W}^T \mathbf{X}^T \mathbf{X} \mathbf{W} = \mathbf{I}$ 给出了矩阵 \mathbf{W} 没有被 Gram-Schmidt 方法正交化的情况. 事实上,在 Lewi 的方法中, \mathbf{W} 的列向量不需要互相正交化. 在这种情况下,令 $\mathbf{Y} = \hat{\mathbf{G}} = [\mathbf{g}_1, \dots, \mathbf{g}_i, \dots, \mathbf{g}_l]$ 约束于 $\mathbf{W}^T \mathbf{X}^T \mathbf{X} \mathbf{W} = \mathbf{I}$. 然后式(30)就被转换为下式:

$$\mathbf{y} = \mathbf{x}^{te} \mathbf{W} \mathbf{W}^T \bar{\mathbf{X}}, \quad (31)$$

其中, $\bar{\mathbf{X}} = (\bar{x}_1, \dots, \bar{x}_l)$, $\bar{x}_i = \sum_{j=1}^{n_i} x_j / \sqrt{n_i}$. 当每一类中训练样本数量相同时,例如 $n_1 = n_2 = \dots = n_c$,能得到:

$$\mathbf{y} = \sqrt{n_i} \mathbf{x}^{te} \mathbf{W} \mathbf{W}^T \hat{\mathbf{X}}, \quad (32)$$

其中, $\hat{\mathbf{X}} = (\hat{x}_1, \dots, \hat{x}_l)$, $\hat{x}_i = \sum_{j=1}^{n_i} x_j / n_i$.

由式(32)可知,向量 \mathbf{y} 表示一个测试样本的投影向量与每一类中样本均值的投影向量之间的相关性. 相关系数越大,测试样本属于该类的可能性越大. 在这种情况下,式(30)相当于一个最大相关分类器. 因此,在某种意义上这也揭示了为什么 PLS 回归在分类问题上能得到良好的性能. 基于上述原因,当合适的虚拟矩阵被定义时, KPLS 回归也能用在分类问题上. 在这种情况下, (核) 偏最小二乘回归的相关分类器有相似特征.

3 实验与结果

在本节,我们进行实验来评估核偏最小二乘回归在面部表情识别上的性能. 实验基于日本女性面部表情数据库和斯特林大学心理学系采集的人脸表情数据集.

3.1 基于日本女性面部表情数据库的实验

日本女性面部表情数据库包含 10 位日本女性的 213 幅图像. 每个人针对 7 种表情都有 2~4 幅图像,表情分别为:中性、快乐、悲伤、惊喜、生气、厌恶、中性以及恐惧. 每幅图像原始尺寸为 256×256 像素. 在我们的实验中,每幅人脸图像的中心区域都被手动的裁剪归一化为 32×32 像素. 在裁剪完人脸图像后,运用小波包变换来提取人脸特征并用其绝对值作为我们实验的特征^[15]. 图 1 表示一些裁剪的人脸以及其对应小波包变换的图像.

注意到,哈尔小波变换运用在图像分解而高斯核 $k(\mathbf{x}, \mathbf{y}) = \exp(-\|\mathbf{x} - \mathbf{y}\|^2 / \delta)$ 则运用在核方法. 对于



(a) 原始图像



(b) (a)中图像的小波包变换

从左到右依次是生气、厌恶、快乐、悲伤和惊喜(每种 2 幅图像)

图 1 一些人脸图像及其对应的尺度为 3 时的小波包变换

Fig. 1 Some face images and corresponding wavelet packet transform in scale 3

PLS, 我们有两种方法来得到矩阵 W . 一种是利用 Lewi 方法而另一种则是直接求解式 (16). 由于 W 的列向量能被视为投影向量, 我们用每个方法逐个检查 W 列向量的子空间的差别. 令 P_1 是用 Lewi 方法得到的矩阵 W 的正交基, P_2 是直接解决式 (16) 得到的矩阵 W 的正交基. 我们利用子空间距离来衡量两个子空间之间的距离, 表示为: $dist = \|P_1 P_1^T - P_2 P_2^T\|$, 这里 $\| \cdot \|$ 是 2 阶范数. 一般情况下, $dist$ 越大, 子空间之间的差异越大. 当且仅当 $dist = 0$ 时, 这两个子空间恒等.

在第一组实验中, 我们通过实验得出两种方法得到的矩阵 W 列向量间的距离. 由图 2 可以观察到, 随着提取特征数量的改变, Lewi 方法和我们方法中的投影子空间之间的距离几乎为 0, 这说明我们的方法和 Lewi 方法形成的子空间几乎是一样的. 在第二组实验中, 对比 IKPLS 与我们无核的方法的运行时间. 由图 3 可以看到我们方法的运行时间比 IKPLS 略短一些.

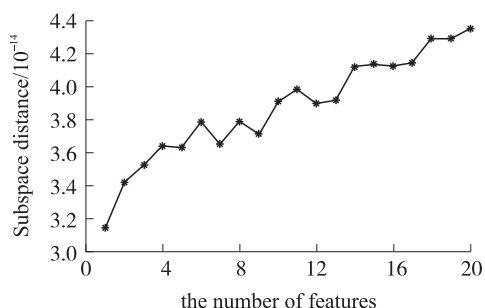


图 2 在不同特征数量的情况下由 Lewi 方法和本文方法得到的子空间之间的距离

Fig. 2 The distance between the subspaces spanned by Lewi's method and our method in different features

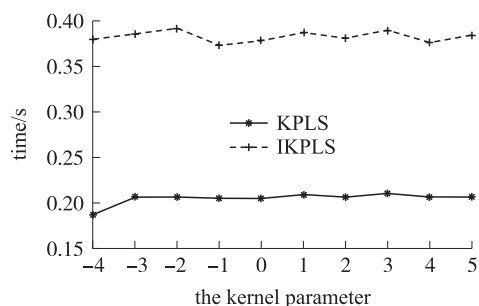


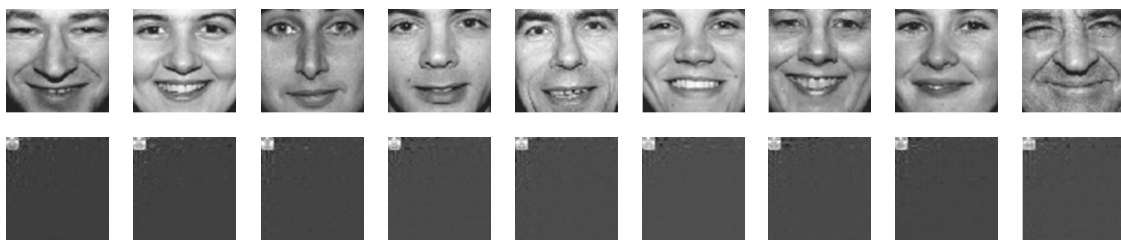
图 3 KPLS 和 IKPLS 在不同核参数下的运行时间

Fig. 3 The running time of KPLS and IKPLS in different kernel parameters

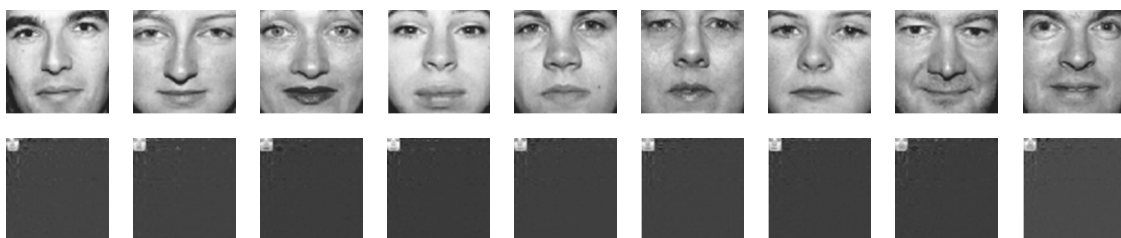
3.2 基于斯特林采集心理图像 (PICS) 的实验

在本实验中, 从斯特林大学心理学系收集人脸表情 (<http://pics.psych.stir.ac.uk/>), 只得到两类面部表情: 中性表情和微笑表情. 正如上节讨论的, 先将图像预处理, 每一幅图像归一化到 32×32 . 相类似的, 运用小波包变换 (哈尔小波) 来提取人脸特征并用其绝对值作为我们实验的特征. 图 4 表示一些归一化的人脸图像及其进行小波包变换后的图像.

我们分别对广义判别分析 (GDA), 使用核方法的广义判别分析 (KGDA), 偏最小二乘回归 (PLS) 以及迭代核偏最小二乘回归 (IKPLS) 做了测试. 注意到当判别分析应用于特征提取时, 有必要使用分类器, 我们使用十折交叉验证来评估分类性能. 在实验中, 高斯核里的最优参数从 $\{10^{-4}, 10^{-3}, \dots, 10^5\}$ 中选取, 在测试集上交叉验证来选取参数. 在 KPLS 和 IKPLS 上的核参数都是 10 000, 而在 KGDA 上的是 1 000.



(a) 在尺度3时微笑脸及其对应的小波包变换



(b) 在尺度3时中性脸及其对应的小波包变换

图 4 一些人脸图像及其对应的小波包变换

Fig. 4 Some face images and corresponding wavelet packet transform

图 5 表示在不同尺度下随着特征改变而得到的实验结果. 从图 5(a)、(b)、(c)中可以发现,一般情况下随着尺度增加分类性能也会提高. 在尺度为 1 时,KPLS 的性能是稳定的,KGDA 和 GDA 的结果随着特征改变而改变,IKPLS 的识别率高于其他方法. 当尺度为 3 时,KPLS 的性能不会随着特征的变化而变化,IKPLS 需要更多的特征来得到比 PLS 更优的性能.

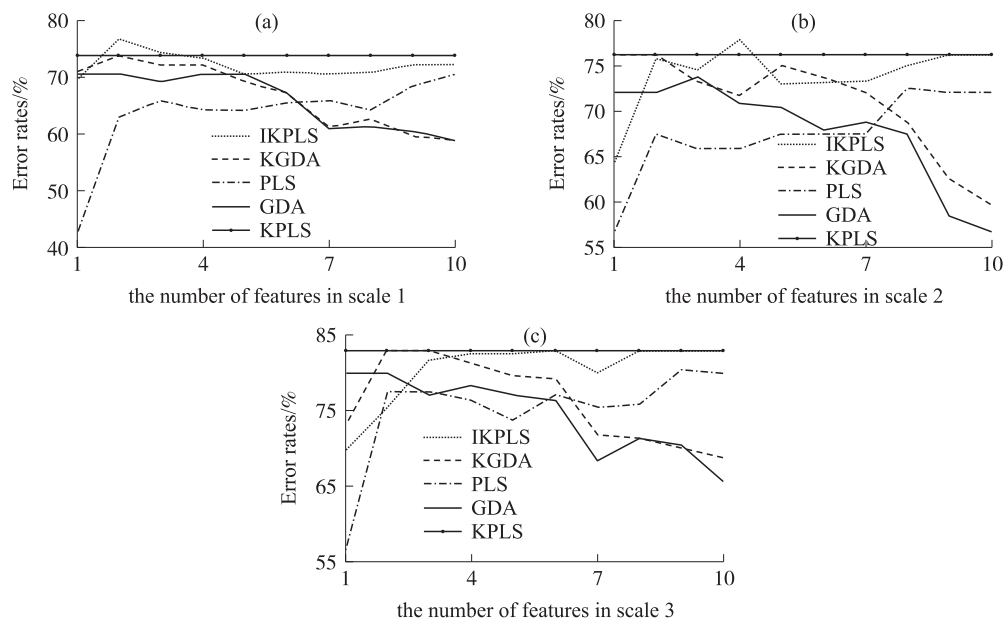


图 5 特征改变时的识别性能

Fig. 5 Recognition performance with the change of features

表 1 列出了在尺度为 3 时的最大平均识别率及标准差. 同样的,我们还采用线性支持向量机 (LSVM),结合高斯核与 adaboost 方法的支持向量机来进行测试. LSVM 的正规化参数是 0.1,SVM 的正规化参数是 1,SVM 的核参数是 10 000. 在表 1 中能够发现 KPLS,KGDA,GDA 和 IKPLS 获得了相同的并且优于 LSVM 及 SVM 的性能. 总的来说,实验结果表明,KPLS 回归即使在不知道语义的情况下也可以通过定义虚拟矩阵进行分类.

表 1 在 PICS 数据集上 KPLS,KGDA,IKPLS,PLS,GDA,LSVM,SVM 和 Adaboost 分别的最大平均识别率与标准差

Table 1 The maximal average recognition rates and standard deviations(std) of KPLS,KGDA,IKPLS,PLS,GDA,LSVM,SVM and Adaboost on PICS data sets

方法	KPLS	KGDA	IKPLS	PLS	GDA	LSVM	SVM	Adaboost
识别率/%	82.91	82.91	82.91	80.41	80.00	79.42	82.52	82.25
标准差	14.05	14.05	14.88	13.47	17.76	14.23	13.30	13.51
特征值	1	2	6	9	1	—	—	—

4 结语

本文证明了(核)偏最小二乘回归可以通过分析数据结构来降低数据维度,阐述了核判别分析与核偏最小二乘回归之间的关系. 通过定义虚拟矩阵的方式使核偏最小二乘回归等价于特征提取加一个分类器,并用我们的方法进行了面部表情识别. 从在日本女性面部表情的实验上可以看到语义表征能提高识别性能,这也证明了在一些情况下核偏最小二乘回归能得到比偏最小二乘回归更好的分类结果. 从在 PICS 上的实验看到,即使在语义未知时,也可以通过定义虚拟矩阵用核偏最小二乘回归来处理分类问题.

总的来说,在人脸识别中可以使用核偏最小二乘回归来得到性能改进. 然而,事实上由于面部表情的复杂性,只用一些简单的面部表情不足以代替在实际应用中的面部表情. 因此,为了更有效地识别和理解面部表情,在未来有必要研究更优的算法.

[参考文献]

- [1] 徐琳琳,张树美,赵俊莉. 基于图像的面部表情识别方法综述[J]. 计算机应用,2017,37(12):3509–3516,3546.
- [2] 宫玉娇. 面部表情识别方法研究综述[J]. 产业与科技论坛,2016,15(13):57–58.
- [3] 杜柏圣. 卷积网络的无监督特征提取对人脸识别的研究[J]. 计算机技术与发展,2018(6):17–20.
- [4] KAPIL J. Multiple feature descriptors based model for individual identification in group photos[J/OL]. [2018-04-06]. <http://doi.org/10.1016/j.jksuci.2017.02.002>.
- [5] IKHLAS A Q,ALI K J. A hybrid feature extraction framework for face recognition:HOG and compressive sensing[J]. International journal of handheld computing research,2017,8(1):1–13.
- [6] JIA Y,LUO Z. Summary of EMG feature extraction[J]. Chinese journal of electron devices,2007,30(1):326–330.
- [7] HOWLAND P,PARK H. Generalizing discriminant analysis using the generalized singular value decomposition[J]. IEEE transactions on pattern analysis and machine intelligence,2004,26(8):995–1006.
- [8] 尚晓琳,张澜. 基于奇异值分解的一类广义特征值反问题[J]. 内蒙古工业大学学报(自然科学版),2017,36(1):1–5.
- [9] ROSIPAL R,TREJO L J. Kernel partial least squares regression in reproducing kernel hilbert space[J]. Journal of machine learning research,2001,2(12):97–123.
- [10] HELANDER E,SILÉN H,VIRTANEN T,et al. Voice conversion using dynamic kernel partial least squares regression[J]. IEEE transactions on audio,speech,and language processing,2012,20(3):806–817.
- [11] SRINIVASAN B V,LUO Y,GARCIA R D,et al. A symmetric kernel partial least squares framework for speaker recognition[J]. IEEE transactions on audio,speech,and language processing,2013,21(7):1415–1423.
- [12] GUO G,MU G. Simultaneous dimensionality reduction and human age estimation via kernel partial least squares regression [C]//2011 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Colorado Springs, CO, USA:IEEE, 2011:657–664.
- [13] DHANJAL C,GUNN S R,SHAW T J. Efficient sparse kernel feature extraction based on partial least squares[J]. IEEE transactions on pattern analysis and machine intelligence,2009,31(8):1347–1361.
- [14] ROSIPAL R,TREJO L J,MATTHEWS B. Kernel PLS-SVC for linear and nonlinear classification[C]//Twentieth International Conference on Machine Learning. Washington DC,USA:AAAI,2003:640–647.
- [15] BAUDAT G,ANOUAR F. Generalized discriminant analysis using a kernel approach[J]. Neural computation,2000,12(10):2385–2404.

[责任编辑:丁 蓉]