

Journal of Image
and Graphics

中国图象图形学报



ISSN1006-8961
CN11-3758/TB

2012 **12**
Vol.17 No.

中国科学院遥感应用研究所
中国图象图形学学会主办
北京应用物理与计算数学研究所

中国图象图形学报

Zhongguo Tuxiang Tuxing Xuebao

2012年12月 第17卷 第12期(总第200期)

目次

综述

高斯混合模型、求解算法及视觉应用综述..... 管涛,李玲玲(1461)

图像处理和编码

基于曲波变换和色度模型的彩色图像去噪 张云强,张培林,王国德,卓洪波(1472)

子矢量排序的渐进不相似度逼近算法 李阳,潘志斌,吴鑫鹏(1478)

稀疏梯度先验模型的正则化图像复原 刘伟豪,梅林,蔡烜(1485)

高精度亚像素全局运动估计的上采样梯度互相关算法 余应淮,王锦荣(1492)

图像分析和识别

点到弦距离累加的自适应角点检测 王万良,金亦挺,赵燕伟,胡峰俊(1500)

最大子分类间隔准则的核函数优化 张静,杨智勇,于红芸,孙晓燕(1509)

核典型相关性鉴别分析 陈伟琦,程强(1516)

小波变换和特征加权融合的人脸识别 赵焕利,王玉德,张学志,薛乃玉(1522)

具有身份标识的多目标跟踪 安国成,肖坦(1528)

图像理解和计算机视觉

用于圆锥体位姿测量的曲面标记..... 李蒙,谌德荣,周广铭,蒋玉萍,高翔霄(1534)

计算机图形学

基于特征识别的3维人脸动画模型自动构造…………… 徐明, 乔宁博, 文振焜, 曾新, 采振祥(1540)

3D GIS 环境下雨雪天气实时仿真…………… 李欢, 樊红, 冯浩(1548)

医学图像处理

结合马尔可夫随机场与模糊 C-均值聚类的脑 MRI 图像分割…………… 余黎煌, 钟华, 张石(1554)

遥感图像处理

面向遥感影像镶嵌的 SVR 色彩一致性处理…………… 吴炜, 骆剑承, 李均力, 杨海平, 沈占锋(1561)

遥感影像投影寻踪回归分类模型…………… 张正健, 李祚泳, 秦宁生, 刘志红, 巴桑(1568)

《中国图象图形学报》2012 年第 17 卷总目次…………… (1)

志谢…………… 封2

中国图象图形学报

刊名题字: 宋 健

月刊(1996 年创刊)

第 17 卷 第 12 期

2012 年 12 月 16 日出版

主管单位 中国科学院

主 办 中国科学院遥感应用研究所

中国图象图形学学会

北京应用物理与计算数学研究所

主 编 李小文

编辑出版 《中国图象图形学报》编辑出版委员会

北京 9718 信箱 邮编 100101

电子信箱:jig@irsa.ac.cn

电话:010-64807995 010-82614429

网 址:www.cjig.cn

印刷装订 北京北林印刷厂

广告经营许可证 京朝工商广字第 0346 号

总 发 行 北京报刊发行局

订 购 全国各地邮局

国外发行 中国国际图书贸易总公司

(中国国际书店)

(北京 399 信箱 邮编 100044)

Superintended by Chinese Academy of Sciences

Sponsored by Institute of Remote Sensing Application,

CAS China Society of Image and Graphics

Institute of Applied Physics and Computational

Mathematics

Chief editor LI Xiaowen

Editor, Publisher Editorial and Publishing Board

of Journal of Image and Graphics

(P. O. Box 9718, Beijing 100101, China)

E-mail:jig@irsa.ac.cn

Distributed by Beijing Bureau for Distribution of Newspapers

and Journals

Domestic All Local Post Offices in China

Foreign China International Book Trading Corporation

(P. O. Box 399, Beijing 100044, China)

Printed by Beijing Beilin Printing House

ISSN 1006-8961 CN11-3758/TB CODE ZTTXFZ 国内邮发代号: 82-831 国外发行代号: M1406 国内定价: 45.00 元

Journal of Image and Graphics

(Monthly, Started in 1996)

Vol. 17 No. 12 December 2012

Contents

Review

Overview of Gaussian mixture models, solving algorithms and visual applications Guan Tao, Li Lingling(1461)

Image Processing and Coding

Denoising method for color images based on chrominance model and curvelet transform
..... Zhang Yunqiang, Zhang Peilin, Wang Guode, Zhuo Hongbo(1472)

Improved incremental dissimilarity approximations algorithm using sub-vector sorting Li Yang, Pan Zhibin, Wu Xinpeng(1478)

Regularized image restoration algorithm on sparse gradient prior model Liu Weihao, Mei Lin, Cai Xuan(1485)

High accuracy sub-pixel global motion estimation based on upsampled gradient cross-correlation algorithm
..... Yu Yinghuai, Wang Jinrong(1492)

Image Analysis and Recognition

Adaptive corner detection based on chord-to-point distance accumulation
..... Wang Wanliang, Jin Yiting, Zhao Yanwei, Hu Fengjun(1500)

Kernel optimization approach based on maximum subclass margin criterion
..... Zhang Jing, Yang Zhiyong, Yu Hongyun, Sun Xiaoyan(1509)

Kernel discriminant analysis based on canonical correlation Chen Weiqi, Cheng Qiang(1516)

Face recognition based on wavelet transform and weighted fusion of face features
..... Zhao Huanli, Wang Yude, Zhang Xuezhi, Xue Naiyu(1522)

Multi-target tracking algorithms with identification label An Guocheng, Xiao Tan(1528)

Image Understanding and Computer Vision

Curved surface marker used for cone pose measurement
..... Li Meng, Chen Derong, Zhou Guangming, Jiang Yuping, Gao Xiangxiao(1534)

Computer Graphics

The approach to automatically construct animation models based on 3D facial geometry and texture features recognition
..... Xu Ming, Qiao Ningbo, Wen Zhenkun, Zeng Xin, Cai Zhenxiang(1540)

Real-time snow and rain rendering in 3D GIS environment Li Huan, Fan Hong, Feng Hao(1548)

Medical Image Processing

Fuzzy C-means clustering algorithm combined with markov random field for brain MR image segmentation
..... She Lihuang, Zhong Hua, Zhang Shi(1554)

Remote Sensing Image Processing

Support vector regression color normalization method for image mosaic
..... Wu Wei, Luo Jiancheng, Li Junli, Yang Haiping, Shen Zhanfeng(1561)

Remote sensing image classification model based on projection pursuit regression Zhang Zhengjian, Li Zuoyong, Qin Ningsheng, Liu Zhihong, Ba Sang(1568)

中图法分类号:TP391 文献标识码:A 文章编号:1006-8961(2012)12-1509-07

论文引用格式:张静,杨智勇,于红芸,孙晓燕. 最大子分类间隔准则的核函数优化[J]. 中国图象图形学报,2012,17(12):1509-1515.

最大子分类间隔准则的核函数优化

张静¹, 杨智勇², 于红芸³, 孙晓燕¹

1. 海军航空工程学院信号与信息处理重点实验室, 烟台 264001; 2. 海军航空工程学院7系, 烟台 264001;
3. 鲁东大学数学与信息学院, 烟台 264001

摘要: 在分析现有基于经验特征空间核函数优化方法局限性的基础上,提出一种基于最大子分类间隔准则的核函数优化方法。该方法首先建立最大子分类间隔准则,然后结合数据在经验特征空间中的特点给出样本数据的类间散布矩阵和类内散布矩阵的表达式,最后利用奇异值分解实现核函数参数的优化选取。本文利用UCI(University of California, Irvine)数据对算法进行仿真实验,仿真结果表明了本文方法的正确性和有效性。

关键词: 核函数;核优化;最大子分类间隔准则;目标识别

Kernel optimization approach based on maximum subclass margin criterion

Zhang Jing¹, Yang Zhiyong², Yu Hongyun³, Sun Xiaoyan¹

1. Shandong Provincial Key Laboratory of Signal & Information Processing, Naval Aeronautical and Astronautical University, Yantai 264001, China;
2. No. 7 Department, Naval Aeronautical and Astronautical University, Yantai 264001, China;
3. The School of the Information and Mathematics, Lu Dong University, Yantai 264001, China

Abstract: In order to deal with the kernel optimization, a new kernel data-dependent optimization kernel approach based on maximum subclass margin criterion is proposed. In this scheme, a maximum subclass margin function is created firstly. Then, the in-between-subclass and inter-subclass scatter matrix in the empirical feature space are defined. Finally, the optimal coefficients vector is solved by the selected optimization criterion. Experimental results based on UCI data show that it is effective and feasible.

Key words: kernel function; kernel optimization; maximum subclass margin criterion; target recognition

0 引言

核方法是一种用于学习的新方法^[1]。由于其优越的非线性处理能力,核方法在各个领域得到广泛应用,如图像处理、目标识别等。

核方法用于目标识别,其分类性能的好坏关

键取决于核空间的构造,这就涉及核函数的正确选取问题,即核优化问题。目前,核机器学习领域对核函数优化方面的研究主要分为两派:对核函数的优化^[2-6]和对核函数参数的优化^[7-9]。基于核矩阵的优化主要从核矩阵自身的特性出发,通过寻求集合上的凸性、对偶性完成核函数的优化。Cristianinni等人^[2]提出的核排列优化方法就是一

收稿日期:2011-07-27;修回日期:2012-04-08

基金项目:国家自然科学基金项目(61032001;61102167)

第一作者简介:张静(1976—),女,副教授,2008年于海军航空工程学院获通信与信息系统工学博士学位,主要研究方向为目标识别和图像信息处理。E-mail:zy913@sina.com

种简单的核矩阵优化方法,这种方法利用矩阵之间的排列尺度,衡量两个矩阵的相似性。Yeung 等人^[5]在半正定规划的基础上提出了两种核矩阵学习方法,即分别通过对核 A 和核 β 的半正定规划实现核函数的优化。除此之外, Bach 等人^[6]的多核学习方法、Ong 等人^[10]的超分辨核学习方法都是性能较好的核矩阵学习方法。从性能上看,基于核矩阵的优化方法设计出的核函数灵活,对数据具有较强的鲁棒性,但这些方法普遍相对复杂,存在优化复杂度高等缺陷。

相比于基于核矩阵优化方法,基于核参数的优化方法要更加简单、易行。这种方法主要针对核函数中的参数进行优化,通过预先选取一定的核函数,如高斯核函数、多项式核函数等,利用各种优化准则对参数进行优化以实现核函数的优化^[11-12]。基于经验特征空间的核函数优化方法就是一种形式上更为灵活的核参数优化方法^[13]。这种方法利用经验特征空间中数据之间的特点,利用各种核鉴别准则函数,如 Xiong 等人的核 Fisher 准则^[13] (KFC)、Chen 等人的局部核 Fisher 准则^[14-15] (LKFC) 等,实现数据在经验特征空间的线性可分。然而,这类方法普遍存在一定的问题:1) 算法采用了梯度下降法实现核函数参数的寻优,迭代过程自身比较耗时;2) 优化过程需要提前设置初始迭代参数,这些参数的设置对优化参数的选择起到决定作用,当参数设置不合理时将直接影响核函数参数的优化,甚至使核函数参数无法达到最优。

针对上述问题,提出一种基于最大子分类间隔准则 (MSMC) 的核函数优化方法,这种方法的优点在于:1) 避开核 Fisher 准则和梯度下降法,采用最大间隔准则作为优化准则函数,利用奇异值分解来选择最优核参数,在降低计算时间的同时能够获得较好的识别效果;2) 建立最大子分类间隔准则函数,适用于多模数据的分类。利用 UCI 数据对算法进行仿真实验。实验结果表明本文方法正确可行。

1 基于最大子分类间隔准则的核函数优化方法

1.1 基于数据的核函数

由文献^[13],假设训练数据 $(x_1, y_1), \dots, (x_n,$

$y_n) \in \mathbf{R}^d \times \{\pm 1\}$, 定义基于数据的核函数为

$$k(x, y) = q(x)q(y)k_0(x, y) \quad (1)$$

这里 $x, y \in \mathbf{R}^d$, $k_0(x, y)$ 表示基本核函数,一般选用高斯核函数、Poly 核函数等等。 $q(\cdot)$ 表示因子函数,定义为

$$q(x) = \alpha_0 + \sum_{r=1}^m \alpha_r k_1(x, \alpha_r) \quad (2)$$

式中, α_r 为联合系数, $k_1(x, \alpha_r) = e^{-\gamma \|x - \alpha_r\|^2}$, α_r 称为“经验核”,满足 $\{\alpha_r \in \mathbf{R}^d, r = 1, \dots, m\}$, 其值可以从训练数据中选取。定义 $k(x, y)$ 、 $k_0(x, y)$ 对应的核矩阵分别为 \mathbf{K} 和 \mathbf{K}_0 , 则

$$\mathbf{K} = [q(x_i)q(x_j)k_0(x_i, x_j)]_{n \times n} = \mathbf{Q}\mathbf{K}_0\mathbf{Q} \quad (3)$$

式中, \mathbf{Q} 为对角线矩阵, 其对角线上的元素为 $\{q(x_1), q(x_2), \dots, q(x_n)\}$ 。定义 $\mathbf{q} = (q(x_1), q(x_2), \dots, q(x_n))^T$ 、 $\boldsymbol{\alpha} = (\alpha_0, \alpha_1, \dots, \alpha_m)^T$, 则

$$\mathbf{q}(\boldsymbol{\alpha}) = \begin{bmatrix} 1 & K_1(x_1, a_1) & \cdots & K_1(x_1, a_m) \\ 1 & K_1(x_2, a_1) & \cdots & K_1(x_2, a_m) \\ \vdots & \vdots & & \vdots \\ 1 & K_1(x_n, a_1) & \cdots & K_1(x_n, a_m) \end{bmatrix} \begin{bmatrix} \alpha_0 \\ \alpha_1 \\ \vdots \\ \alpha_m \end{bmatrix} = \mathbf{K}_1 \boldsymbol{\alpha} \quad (4)$$

式中, \mathbf{K}_1 是一个 $n \times (m+1)$ 的矩阵。

1.2 基于最大子分类间隔准则的优化目标函数的建立

核函数的优化需要建立一定的优化准则之上,对于多模数据而言,因为同类数据间由多个子类构成,因此建立一种基于最大子分类间隔准则的优化目标函数,通过求解获取到式(4)中的最优核参数。

假设一组训练数据 $\{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$, 其中 $x_i \in \mathbf{R}^d (i = 1, \dots, n)$ 是数据空间中第 i 个样本, $y_i (i = 1, \dots, n)$ 是第 i 个样本对应的类标信息,且 $y_i \in \{1, 2, \dots, C\}$, 则两个子类别 c_{ij} 和 c_{kl} 之间的间隔可表示为

$$D = \frac{1}{2} \sum_{i=1}^{c-1} \sum_{j=1}^{h_i} \sum_{k=i+1}^c \sum_{l=1}^{h_k} p_{ij} p_{kl} d(c_{ij}, c_{kl}) \quad (5)$$

式中, p_{ij}, p_{kl} 表示分别表示第 i 类中第 j 个子类和第 k 类中第 l 个子类的先验知识,一般可以根据各个子类中的样本数目设置 $p_{ij} = \frac{n_{ij}}{N}, p_{kl} = \frac{n_{kl}}{N}$ 。然而,这种间隔并不能很好地反映出子类间数据的分类程度,因此还需要考虑子类的散布程度,重新定义间隔距离为

$$D = \sum_{i=1}^{c-1} \sum_{j=1}^{H_i} \sum_{k=i+1}^c \sum_{l=1}^{H_k} p_{ij} p_{kl} d(c_{ij}, c_{kl}) - \sum_{i=1}^c \sum_{j=1}^{H_i} \mathbf{S}(c_{ij}) \quad (6)$$

式中, $\mathbf{S}(c_{ij})$ 定义为 c_{ij} 子类的散射矩阵, 可表示为

$$\mathbf{S}(c_{ij}) = \text{tr}(\mathbf{S}_{ij}) \quad (7)$$

对式(6)进行化简

$$D = \sum_{i=1}^{c-1} \sum_{j=1}^{H_i} \sum_{k=i+1}^c \sum_{l=1}^{H_k} p_{ij} p_{kl} (\boldsymbol{\mu}_{ij} - \boldsymbol{\mu}_{kl})^T (\boldsymbol{\mu}_{ij} - \boldsymbol{\mu}_{kl}) - \sum_{i=1}^c \sum_{j=1}^{H_i} \text{tr}(\mathbf{S}_{ij}) = \text{tr}(\mathbf{S}_b) - \text{tr}(\mathbf{S}_\omega) \quad (8)$$

式中, $\boldsymbol{\mu}_{ij}$ 表示第 i 类中第 j 个子类的均值。 \mathbf{S}_b 表示类间散射矩阵, \mathbf{S}_ω 表示类内散射矩阵。因此, 求取两个子类间的最大间隔可由式(8)建立准则

$$J = \max(D) = \max[\text{tr}(\mathbf{S}_b - \mathbf{S}_\omega)] = \max[\text{tr}(2\mathbf{S}_b - \mathbf{S}_T)] = \max[\text{tr}(\boldsymbol{\alpha}^T (2\mathbf{S}_B - \mathbf{S}_T) \boldsymbol{\alpha})] \quad (9)$$

推导见附录 A。则经验特征空间中利用最大子分类间隔准则优化核函数就是令

$$\begin{cases} \max(\boldsymbol{\alpha}^T (2\mathbf{S}_B - \mathbf{S}_T) \boldsymbol{\alpha}) \\ \text{s. t. } \boldsymbol{\alpha}^T \boldsymbol{\alpha} - 1 = 0 \end{cases} \quad (10)$$

对式(10)进行 SVD 分解, 则最大特征值所对应的特征矢量即为最优 $\boldsymbol{\alpha}$, 其求解过程可参见文献[16-17]。

2 仿真实验及分析

利用 UCI 数据中的两类数据 (Monks1、Monks3、Ionosphere、Wisconsin Breast Cancer、Wisconsin Breast Cancer Cont) 和多类数据 (Iris、Wine) 进行实验。

仿真过程中, 将各个数据集随机分成互不重复的 3 组, 分别作为训练样本集、测试样本集和构建经验中心数据集。将优化后的核函数代入 KPCA 方法中提取用以识别的特征, 并利用传统的模板匹配法和高斯 SVM 法作为识别器对测试数据集进行识别。在利用 UCI 数据进行仿真实验时, 仿真过程中基本核函数选用高斯核函数, 参数设置为 $\sigma^2 = 5000$, 子类数设置为 2。图 1 以 Monks-3 和 Iris 为例给出各种优化方法处理后的特征空间分布图, 其中 KPCA 表示利用传统的、未经优化的 KPCA 处理后数据在特征空间的分布情况; KPCA + KFC 表示利用文

献[13]提出的优化方法处理后的结果; KPCA + LKFC 表示利用文献[14]提出的方法获得的结果; 而 MSMC 则表示利用本文所提出的优化方法处理后的特征空间分布图。表 1 给出利用不同优化方法获得的识别结果。表 2 给出不同优化方法耗时对比表。

由图 1、表 1 和表 2 可以得出以下结论:

1) 利用 MSMC 能够实现目标数据的识别, 该方法是一种行之有效的方法。由图 1 特征空间的分布图可以看出, 利用 MSMC 方法处理后的数据在特征空间可以获得较好的分类效果。

2) 由表 1 可以看出, 利用本文方法可以获得较好的识别率, 这充分说明了 MSMC 方法虽然采用了与 Chen 和 Xiong 方法不同的准则函数和优化方法, 但获得与后者同样的识别结果, 且优化过程中无需设置初始参数, 减少了优化算法对参数的依赖性。

3) 从不同分类器获得的识别结果来看, 无论哪种优化算法, 采用 Norm 分类器获得的识别率普遍低于采用 GSVM 分类器获得的识别率。

4) 由表 2 可以看出, 未经优化的 KPCA 耗时最小, 但识别效果最差; 其次是 Xiong 的核优化方法, 识别性能较差, 不适用于多模分布的数据, 耗时最长的是 Chen 的方法, 识别效果较好, 适用于多模分布的数据。对比 KPCA&LKFC 和 KPCA&MSMC 方法, 两种方法均能处理多模分布结构的数据, 而且均具有较好的分类效果, 但 KPCA&LKFC 方法的耗时长。因此, 从综合效果上来看, 本文 KPCA&MSMC 方法最好。

3 结 论

研究了一种新的基于最大子分类间隔准则的核参数优化方法。这种方法首先建立最大子分类间隔准则, 然后利用奇异值分解的方法获得经验特征空间中最优的核函数参数。这种方法的优点是: 无需提前设置优化算法的初始值, 减少了因初始值设置不当而引起的优化失衡; 这种优化算法可以在获得与其他优化算法相当的识别性能的基础上, 降低运行时间, 减少计算量。利用仿真 UCI 数据进行仿真实验, 验证了该方法具有良好的性能。

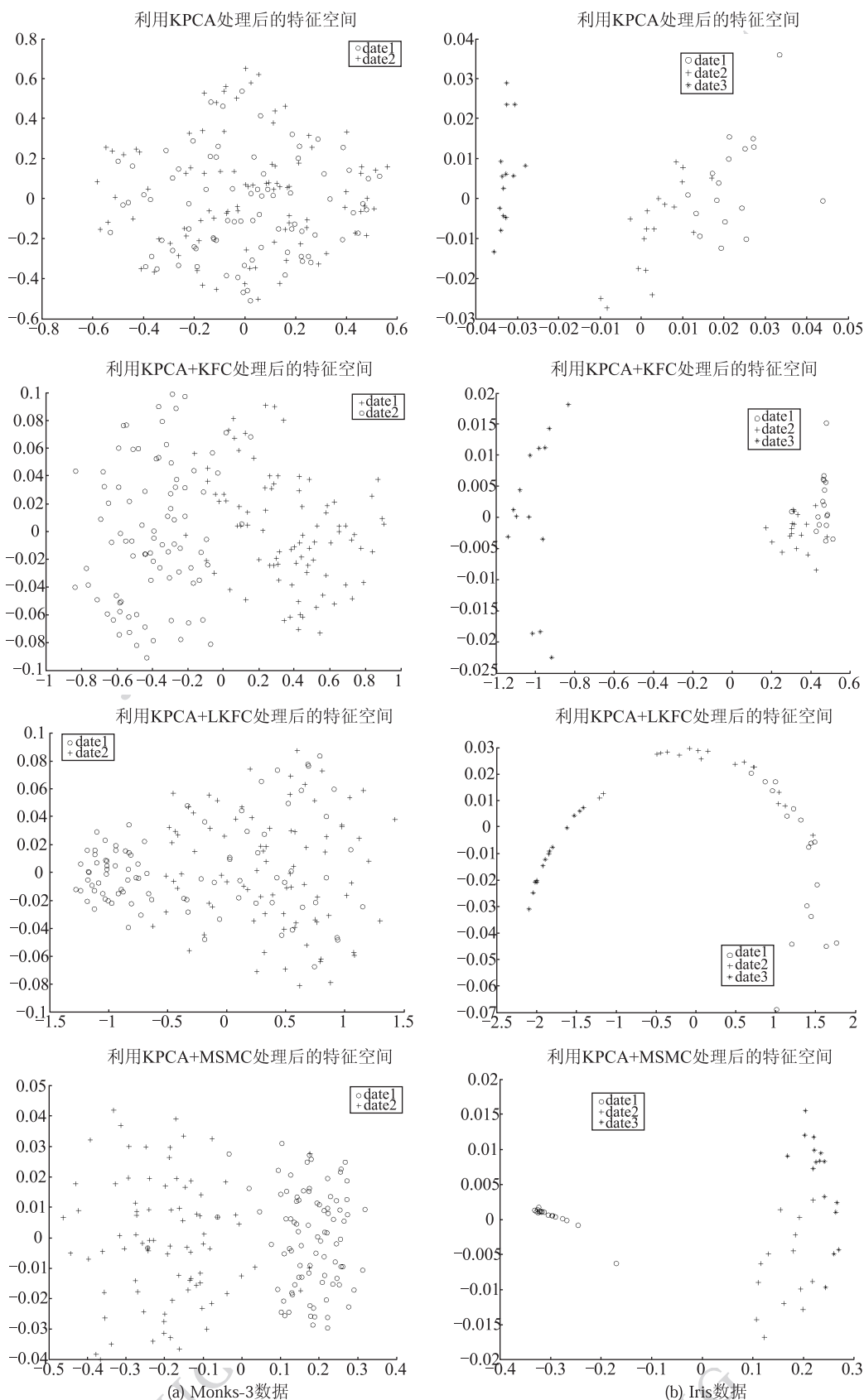


图 1 利用不同方法处理后数据在特征空间中的分布

Fig. 1 Two-dimensional projection of data set in the empirical space after different kernel optimization methods.

表 1 利用不同优化方法获得的识别结果

Table 1 Test classification performance comparison with different kernel optimization methods

优化方法	分类器	Monks-1	Monks-3	Ionosphere	Wisconsin Breast Cancer	Breast cancer Wisconsin cont	Iris	Wine
KPCA	Norm	75.44	80.34	86.32	95.88	95.51	92.31	93.23
	GSVM	79.32	81.23	89.47	96.01	96.63	93.53	95.61
KPCA & KFC	Norm	78.72	93.07	93.41	94.45	95.38	91.24	95.42
	GSVM	82.95	93.07	93.53	95.03	95.07	95.12	96.01
KPCA & LKFC	Norm	95.32	93.78	90.62	95.26	95.94	87.63	96.22
	GSVM	96.24	94.25	94.77	95.95	96.88	96.23	96.57
KPCA & MSMC	Norm	95.25	93.10	90.26	95.24	95.26	89.23	95.23
	GSVM	96.50	94.35	93.42	97.12	96.10	95.52	97.12

表 2 不同优化方法的耗时对比

Table 2 The average CPU time for different kernel optimization methods

	优化方法			
	KPCA	KPCA & KFC	KPCA & LKFC	KPCA & MSMC
耗时/s	0.133 3	0.305 739	9.469 8	1.105 2

参考文献 (References)

[1] Kong R, Shi Z S, Guo L. Improving performance of kernel principal component analysis using combination kernel functions[J]. Journal of Image and Graphics, 2004,9(1):40-45. [孔锐,施泽生,郭立. 利用组合核函数提高核主成分分析的性能[J]. 中国图象图形学报,2004,9(1):40-45.]

[2] Cristianinni N, Kandola J, Elisseeff A. On kernel target alignment[J]. Journal of Machine Learning Research, 2002, 4(1): 367-373.

[3] Amari S, Wu S. Improving support vector machine classifier by modifying kernel function[J]. Neural Network, 1999, 12(6): 738-789.

[4] Yang Y H. Adaptive credit scoring with kernel learning methods [J]. European Journal of Operational Research, 2007, 183(3):1521-1536.

[5] Yeung D Y, Chang H. A kernel approach for semi-supervised metric learning [J]. IEEE Trans. on Neural Network, 2007, 18(1):141-149.

[6] Bach F R, Lanckriet G R, Jordan M I. Multiple kernel learning, conic duality, and the SMO algorithm [C]//Proceedings of the 21st International Conference on Machine Learning. Banff, Canada: ACM, Press, 2004, 41-48

[7] Yeung D Y, Chang H, Dai G. Learning the kernel matrix by maximizing a KFD-based class separability criterion[J]. Pattern Recognition, 2007, 40(7):2021-2028.

[8] Chen B, Liu H W, Bao Z, et al. A kernel optimization algorithm based on fusion kernel for high-resolution range profiles recognition[J]. ACTA Electronica Sinica, 2006,34(6):1146-1151. [陈渤,刘宏伟,保铮,等. 一种针对雷达高分辨距离像的融合核优化算法[J]. 电子学报2006,34(6):1146-1151.]

[9] Zhang D Q, Chen S C, Zhou Z H. Learning the kernel parameters in kernel minimum distance classifier [J]. Pattern Recognition, 2006, 39(1):133-135.

[10] Ong C S, Smola A J, Williamson R C. Learning the kernel with hyper-kernels[J]. J. Mach. Learn, 2005, 6(12):1001-1030.

[11] Sollich P. Bayesian methods for support vector machines: Evidence and predictive class probabilities [J]. Machine Learning, 2002, 46(1-3):21-52.

[12] Lee M M S, Keerthi S S, Ong C J. An efficient method for computing leave-one-out error in support vector machines with Gaussian kernels [J]. IEEE Trans. on Neural Networks, 2004, 15(3):133-135.

[13] Xiong H L, Swamy M N S, Ahmad M O. Optimizing the kernel in the empirical feature space[J]. IEEE Trans. on Neural Networks, 2005, 16(2):460-472.

[14] Chen B, Liu H W, Bo Z. A kernel optimization method based on the localized kernel Fisher criterion [J]. Pattern Recognition, 2007,41(3):1098-1109.

[15] Chen B, Liu H W, Bo Z. Optimizing the data-dependent kernel under a unified kernel optimization framework[J]. Pattern Recognition, 2008,41(6):2107-2119.

[16] Li J B, Pan J S, Lu Z M. Kernel optimization-based discriminant analysis[J]. Neural Comput & Applic, 2009,18(6):603-612.

[17] Li H F, Jiang T. Efficient and robust feature extraction by maximum margin criterion[J]. IEEE Trans. on Neural Netw, 2006, 17(1):157-164.

附录 A

由于在经验特征空间中,任意一个样本数据可由 $Y = KPA^{-\frac{1}{2}}$, 这里 K 是式 (1) 定义的基于数据的核函数, 且 $K = PAP^T$, 则可在经验特征空间中求取 $\text{tr}(S_b)$ 和 $\text{tr}(S_i)$ 。

$$S_b = \sum_{i=1}^{C-1} \sum_{j=1}^{H_i} \sum_{k=i+1}^C \sum_{l=1}^{H_k} p_{kl} (\mu_{ij} - \mu_{kl}) (\mu_{ij} - \mu_{kl})^T = \frac{S_{b1} - S_{b2}}{2} \quad (A1)$$

式中,

$$S_{b1} = \sum_{i=1}^C \sum_{j=1}^{H_i} \sum_{k=1}^C \sum_{l=1}^{H_k} p_{ij} p_{kl} (\mu_{ij} - \mu_{kl}) (\mu_{ij} - \mu_{kl})^T \quad (A2)$$

$$S_{b2} = \sum_{i=1}^C \sum_{j=1}^{H_i} \sum_{k=1}^{H_k} p_{ij} p_{ik} (\mu_{ij} - \mu_{kl}) (\mu_{ij} - \mu_{kl})^T \quad (A3)$$

由此可得

$$\text{tr}(S_b) = \text{tr}\left(\frac{S_{b1} - S_{b2}}{2}\right) \quad (A4)$$

下面对式 (A4) 进行推导。

$$\begin{aligned} \text{tr}(S_{b1}) &= \sum_{i=1}^C \sum_{j=1}^{H_i} \sum_{k=1}^C \sum_{l=1}^{H_k} p_{ij} p_{kl} (\mu_{ij} - \mu_{kl}) (\mu_{ij} - \mu_{kl})^T = \\ &= 2 \sum_{i=1}^C \sum_{j=1}^{H_i} \frac{n_{ij}}{n} ((\mu_{ij})^T (\mu_{ij})) - \\ &= 2 \sum_{i=1}^C \sum_{j=1}^{H_i} \sum_{k=1}^C \sum_{l=1}^{H_k} \frac{n_{ij}}{n} \frac{n_{kl}}{n} ((\mu_{ij})^T (\mu_{kl})) \end{aligned} \quad (A5)$$

式中, n_{ij} 表示第 i 类中第 j 个子类包含样本的个数, n 代表样本总数。为了推导方便, 可以写出

$$\mu_{ij} = \frac{1}{n_{ij}} \sum_{m=1}^{n_{ij}} y_{ij}(m) = \frac{1}{n_{ij}} Y_{ij}^T \mathbf{1}_{n_{ij}} \quad (A6)$$

$$\mu_{kl} = \frac{1}{n_{kl}} \sum_{w=1}^{n_{kl}} y_{kl}(m) = \frac{1}{n_{kl}} Y_{kl}^T \mathbf{1}_{n_{kl}} \quad (A7)$$

将式 (A6) (A7) 代入 (A5) 可得

$$\begin{aligned} 2 \sum_{i=1}^C \sum_{j=1}^{H_i} \frac{n_{ij}}{n} ((\mu_{ij})^T (\mu_{ij})) &= \\ 2 [\mathbf{1}_{n_{11}}^T, \mathbf{1}_{n_{12}}^T, \dots, \mathbf{1}_{n_{CH_C}}^T] A Y Y^T A^T \begin{bmatrix} \mathbf{1}_{n_{11}} \\ \mathbf{1}_{n_{12}} \\ \vdots \\ \mathbf{1}_{n_{CH_C}} \end{bmatrix} &= \\ 2 \cdot \mathbf{1}_n^T A Y Y^T A^T \mathbf{1}_n & \end{aligned} \quad (A8)$$

$$2 \sum_{i=1}^C \sum_{j=1}^{H_i} \sum_{k=1}^C \sum_{l=1}^{H_k} \frac{n_{ij}}{n} \frac{n_{kl}}{n} ((\mu_{ij})^T (\mu_{kl})) =$$

$$2 \frac{1}{n^2} \begin{bmatrix} \mathbf{1}_{n_{11}}^T \\ \mathbf{1}_{n_{12}}^T \\ \vdots \\ \mathbf{1}_{n_{H_1}}^T \\ \mathbf{1}_{n_{21}}^T \\ \vdots \\ \mathbf{1}_{n_{CH_C}}^T \end{bmatrix}^T Y Y^T \begin{bmatrix} \mathbf{1}_{n_{11}} \\ \mathbf{1}_{n_{12}} \\ \vdots \\ \mathbf{1}_{n_{H_1}} \\ \mathbf{1}_{n_{21}} \\ \vdots \\ \mathbf{1}_{n_{CH_C}} \end{bmatrix} = 2 \cdot \mathbf{1}_n^T B Y Y^T B^T \mathbf{1}_n \quad (A9)$$

$$A = \begin{bmatrix} \left(\frac{1}{\sqrt{n \cdot n_{11}}}\right)_{n_{11} \times n_{11}} & \dots & 0 \\ \vdots & & \vdots \\ 0 & \dots & \left(\frac{1}{\sqrt{n \cdot n_{CH_C}}}\right)_{n_{CH_C} \times n_{CH_C}} \end{bmatrix}$$

$$B = \begin{bmatrix} \frac{1}{n} & \dots & 0 \\ \vdots & & \vdots \\ 0 & \dots & \frac{1}{n} \end{bmatrix}$$

$$\begin{aligned} \text{tr}(S_{b2}) &= \sum_{i=1}^C \sum_{j=1}^{H_i} \sum_{k=1}^{H_k} p_{ij} p_{ik} (\mu_{ij} - \mu_{ik}) (\mu_{ij} - \mu_{ik})^T = \\ &= 2 \sum_{i=1}^C \sum_{j=1}^{H_i} \frac{n_{ij}}{n} \frac{n_i}{n} ((\mu_{ij})^T (\mu_{ij})) - \\ &= 2 \sum_{i=1}^C \sum_{j=1}^{H_i} \sum_{k=1}^{H_k} \frac{n_{ij}}{n} \frac{n_{ik}}{n} (\mu_{ij} \mu_{ik}) \end{aligned} \quad (A10)$$

$$\begin{aligned} 2 \sum_{i=1}^C \sum_{j=1}^{H_i} \frac{n_{ij}}{n} \frac{n_i}{n} ((\mu_{ij})^T (\mu_{ij})) &= \\ 2 \sum_{i=1}^C \sum_{j=1}^{H_i} \frac{n_i}{n^2 \cdot n_{ij}} (\mathbf{1}_{n_{ij}}^T Y_{ij} Y_{ij}^T \mathbf{1}_{n_{ij}}) &= \\ 2 \cdot \mathbf{1}_n^T C Y Y^T C^T \mathbf{1}_n & \end{aligned} \quad (A11)$$

$$\begin{aligned} 2 \sum_{i=1}^C \sum_{j=1}^{H_i} \sum_{k=1}^{H_k} \frac{n_{ij}}{n} \frac{n_{ik}}{n} (\mu_{ij} \mu_{ik}) &= \\ \sum_{i=1}^C \sum_{j=1}^{H_i} \sum_{k=1}^{H_k} \frac{1}{n} \frac{1}{n} (\mathbf{1}_{n_{ij}}^T Y_{ij} Y_{ik}^T \mathbf{1}_{n_{ik}}) &= \\ 2 \mathbf{1}_n^T D Y Y^T D^T \mathbf{1}_n & \end{aligned} \quad (A12)$$

式中

$$C = \begin{bmatrix} \left(\frac{\sqrt{n_1}}{n \cdot \sqrt{n_{11}}}\right)_{n_{11} \times n_{11}} & \dots & 0 \\ \vdots & & \vdots \\ 0 & \dots & \left(\frac{\sqrt{n_C}}{n \cdot \sqrt{n_{CH_C}}}\right)_{n_{CH_C} \times n_{CH_C}} \end{bmatrix}$$

$$D = \begin{bmatrix} \left(\frac{1}{n}\right)_{n_1 \times n_1} & \cdots & 0 \\ \vdots & & \vdots \\ 0 & \cdots & \left(\frac{1}{n}\right)_{n_C \times n_C} \end{bmatrix}$$

由上述推导可得

$$\begin{aligned} \text{tr}(S_b) &= \mathbf{1}_n^T \mathbf{M} \mathbf{Y} \mathbf{Y}^T \mathbf{M}^T \mathbf{1}_n = \\ \mathbf{1}_n^T \mathbf{M} \mathbf{K} \mathbf{M}^T \mathbf{1}_n &= \mathbf{1}_n^T \mathbf{M} \mathbf{Q} \mathbf{K}_0 \mathbf{Q}^T \mathbf{M}^T \mathbf{1}_n = \\ &= \mathbf{1}_n^T \mathbf{Q} \mathbf{M} \mathbf{K}_0 \mathbf{M}^T \mathbf{Q}^T \mathbf{1}_n \end{aligned} \quad (\text{A12})$$

式中, $\mathbf{M} = (\mathbf{A} - \mathbf{B} - \mathbf{C} + \mathbf{D})$ 为 $n \times n$ 的对角阵; $\mathbf{K}_0 = \mathbf{Y}_0^T \mathbf{Y}_0$, 为基本核。 $\mathbf{Q} = \text{diag}(f(y_1), f(y_2), \dots, f(y_n))$,

而 $f(y) = \alpha_0 + \sum_{i=1}^m \alpha_i e(y, \bar{y}_i)$, 故 $\mathbf{1}_n^T \mathbf{Q} = (\mathbf{E}\boldsymbol{\alpha})^T$ 。

因此

$$\begin{aligned} \text{tr}(S_b) &= \\ (\mathbf{E}\boldsymbol{\alpha})^T (\mathbf{M} \mathbf{Y}_0) (\mathbf{M} \mathbf{Y}_0)^T \mathbf{E}\boldsymbol{\alpha} &= \\ \boldsymbol{\alpha}^T \mathbf{E}^T (\mathbf{M} \mathbf{Y}_0) (\mathbf{M} \mathbf{Y}_0)^T \mathbf{E}\boldsymbol{\alpha} &= \\ \boldsymbol{\alpha}^T (\mathbf{E}^T \mathbf{M} \mathbf{Y}_0) (\mathbf{E}^T \mathbf{M} \mathbf{Y}_0)^T \boldsymbol{\alpha} \end{aligned} \quad (\text{A14})$$

令 $\mathbf{X}_B = \mathbf{E}^T \mathbf{M} \mathbf{Y}_0$, 则

$$\text{tr}(S_b) = \boldsymbol{\alpha}^T \mathbf{X}_B \mathbf{X}_B^T \boldsymbol{\alpha} \quad (\text{A15})$$

下面推导 S_t 如下:

$$\begin{aligned} \text{tr}(S_t) &= \frac{1}{n} \sum_{i=1}^n (\mathbf{y}_i - \boldsymbol{\mu})^T (\mathbf{y}_i - \boldsymbol{\mu}) = \\ \frac{1}{n} \sum_{i=1}^n \mathbf{y}_i^T \mathbf{y}_i - \boldsymbol{\mu}^T \boldsymbol{\mu} &= \mathbf{1}_n^T (\mathbf{F} - \mathbf{G}) \mathbf{Y} \mathbf{Y}^T (\mathbf{F} - \mathbf{G})^T \mathbf{1}_n = \\ \mathbf{1}_n^T \mathbf{W} \mathbf{Y} \mathbf{Y}^T \mathbf{W}^T \mathbf{1}_n &= \boldsymbol{\alpha}^T \mathbf{E}^T \mathbf{W} \mathbf{Y}_0 \mathbf{Y}_0^T \mathbf{W}^T \mathbf{E}\boldsymbol{\alpha} \end{aligned} \quad (\text{A16})$$

式中

$$\mathbf{F} = \begin{bmatrix} \left(\frac{1}{\sqrt{n}}\right)_{1 \times 1} & \cdots & 0 \\ \vdots & & \vdots \\ 0 & \cdots & \left(\frac{1}{\sqrt{n}}\right)_{n \times n} \end{bmatrix}$$

$$\mathbf{G} = \begin{bmatrix} \left(\frac{1}{n}\right)_{1 \times 1} & \cdots & 0 \\ \vdots & & \vdots \\ 0 & \cdots & \left(\frac{1}{n}\right)_{n \times n} \end{bmatrix}$$

令 $\mathbf{X}_T = \mathbf{E}^T \mathbf{W} \mathbf{Y}_0$, 则

$$\text{tr}(S_t) = \boldsymbol{\alpha}^T \mathbf{X}_T \mathbf{X}_T^T \boldsymbol{\alpha} \quad (\text{A17})$$