



JOURNAL OF IMAGE AND GRAPHICS

主办: 中国科学院空天信息创新研究院
中国图象图形学学会
北京应用物理与计算数学研究所

中国图象图形学报

2020

09

VOL.25

ISSN1006-8961
CN11-3758/TB



遥感边缘智能技术 P1719



第25卷第9期 (总第293期)

2020年9月16日

中国精品科技期刊
中国国际影响力优秀学术期刊
中国科技核心期刊
中文核心期刊

版权声明

凡向《中国图象图形学报》投稿, 均视为同意在本刊网站及CNKI等全文数据库出版, 所刊载论文已获得著作权人的授权。本刊所有图片均为非商业目的使用, 所有内容, 未经许可, 不得转载或以其他方式使用。

Copyright

All rights reserved by Journal of Image and Graphics, Institute of Remote Sensing and Digital Earth, CAS. The content (including but not limited text, photo, etc) published in this journal is for non-commercial use.

主管单位 中国科学院

主办单位 中国科学院空天信息创新研究院

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 顾行发

编辑出版 《中国图象图形学报》编辑出版委员会

通信地址 北京市海淀区北四环西路19号

邮 编 100190

电子信箱 jig@radi.ac.cn

电 话 010-58887035

网 址 www.cjig.cn

广告发布登记号 京朝工商广登字20170218号

总 发 行 北京报刊发行局

订 购 全国各地邮局

海外发行 中国国际图书贸易集团有限公司

(邮政信箱: 北京399信箱 邮编: 100048)

印刷装订 北京科信印刷有限公司

Journal of Image and Graphics

Title inscription: Song Jian

Monthly, Started in 1996

Superintended by Chinese Academy of Sciences

Sponsored by Aerospace Information Research Institute, CAS

China Society of Image and Graphics

Institute of Applied Physics and Computational Mathematics

Editor-in-Chief Gu Xingfa

Editor, Publisher Editorial and Publishing Board of Journal of
Image and Graphics

Address No. 19, North 4th Ring Road West, Haidian District,
Beijing, P. R. China

Zip code 100190

E-mail jig@radi.ac.cn

Telephone 010-58887035

Website www.cjig.cn

Distributed by Beijing Bureau for Distribution of Newspapers and Journals

Domestic All Local Post Offices in China

Overseas China International Book Trading Corporation
(P.O.Box 399, Beijing 100048, P.R.China)

Printed by Beijing Kexin Printing Co., Ltd.

CN 11-3758/TB

ISSN 1006-8961

CODEN ZTTXFZ

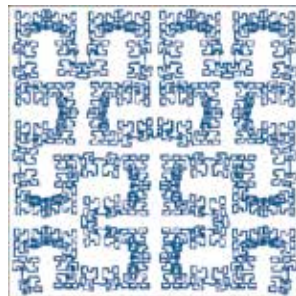
国外发行代号 M1406

国内邮发代号 82-831

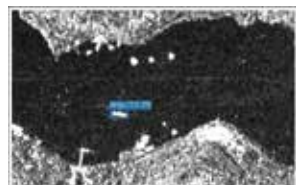
国内定价 60.00元



第一人称视角下的社会力优化多行人跟踪(第1869页)



分形曲线生成的频域方法(第1904页)



双向特征融合的数据自适应SAR图像舰船目标检测模型(第1943页)

学者观点

遥感边缘智能技术研究进展及挑战

孙显, 梁伟, 刁文辉, 曹志颖, 冯瑛超, 王冰, 付琨 1719

综述

红外弱小目标检测算法综述

李俊宏, 张萍, 王晓玮, 黄世泽 1739

目标检测中的尺度变换应用综述

申奉臻, 张萍, 罗金, 刘松阳, 冯世杰 1754

图像处理和编码

分层特征融合注意力网络图像超分辨率重建

雷鹏程, 刘丛, 唐坚刚, 彭敦陆 1773

适用小样本的无参考水下视频质量评价方法

宋巍, 刘诗梦, 黄冬梅, 王文娟, 王建 1787

图像分析和识别

结合面料属性和触觉感测的织物识别

邢寅初, 刘骊, 付晓东, 刘利军, 黄青松 1800

结合GAN的轻量级模糊车牌识别算法

段宾, 符祥, 江毅, 曾接贤 1813

非贪婪的鲁棒性度量学习算法

曾凡霞, 张文生 1825

不规则像素簇显著性检测算法

李明旭, 翟东海 1837

全局区域相异度阈值驱动构建的稀疏尺度集模型

王浩宇, 沈占锋, 柯映明, 许泽宇, 李硕, 焦淑慧 1848

图像理解和计算机视觉

分裂合并运动分割的多运动视觉里程计方法

王晨捷, 张云, 赵青, 王伟, 尹露, 罗斌, 张良培 1859

第一人称视角下的社会力优化多行人跟踪

杨廷召, 刘骊, 付晓东, 刘利军, 黄青松 1869

引入辅助损失的多场景车道线检测

陈立潮, 徐秀芝, 曹建芳, 潘理虎 1882

计算机图形学

旋转骨架法在二值图像分形维数计算中的应用

纪佑军, 何杰, 韩海水, 程忠钊, 曾保全 1894

分形曲线生成的频域方法

陈伟, 乔洁雯, 周晨 1904

医学图像处理

融合残差注意力机制的UNet视盘分割

侯向丹, 赵一浩, 刘洪普, 郭鸿湧, 于习欣, 丁梦园 1915

深层聚合残差密集网络的超声图像左心室分割

吴宣言, 缙新科, 朱子重, 魏域林, 王凯 1930

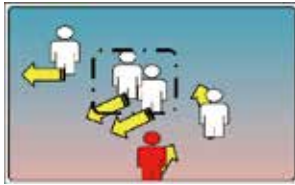
遥感图像处理

双向特征融合的数据自适应SAR图像舰船目标检测模型

张筱晗, 姚力波, 吕亚飞, 简涛, 赵志伟, 藏洁 1943

CONTENTS

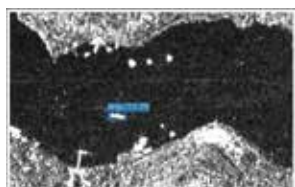
JOURNAL OF IMAGE AND GRAPHICS



Multi-pedestrian tracking optimized by social force model under first-person perspective(P1869)



Fractal curve generation method based on the frequency domain (P1904)



Data-adaptive single-shot ship detector with a bidirectional feature fusion module for SAR images (P1943)

Scholar View

Progress and challenges of remote sensing edge intelligence technology

Sun Xian, Liang Wei, Diao Wenhui, Cao Zhiying, Feng Yingchao, Wang Bing, Fu Kun 1719

Review

Infrared small-target detection algorithms: a survey

Li Junhong, Zhang Ping, Wang Xiaowei, Huang Shize 1739

Scale changing in general object detection: a survey

Shen Fengcan, Zhang Ping, Luo Jin, Liu Songyang, Feng Shijie 1754

Image Processing and Coding

Hierarchical feature fusion attention network for image super-resolution reconstruction

Lei Pengcheng, Liu Cong, Tang Jiangang, Peng Dunlu 1773

Non-reference underwater video quality assessment method for small size samples

Song Wei, Liu Shimeng, Huang Dongmei, Wang Wenjuan, Wang Jian 1787

Image Analysis and Recognition

Cloth recognition based on fabric properties and tactile sensing

Xing Yinchu, Liu Li, Fu Xiaodong, Liu Lijun, Huang Qingsong 1800

Lightweight blurred car plate recognition method combined with generated images

Duan Bin, Fu Xiang, Jiang Yi, Zeng Jiexian 1813

Robust distance metric learning algorithm based on nongreedy solution

Zeng Fanxia, Zhang Wensheng 1825

Significance detection method with irregular pixel clusters

Li Mingxu, Zhai Donghai 1837

Sparse scale set model based on global regional dissimilarity threshold

Wang Haoyu, Shen Zhanfeng, Ke Yingming, Xu Zeyu, Li Shuo, Jiao Shuhui 1848

Image Understanding and Computer Vision

Multi-motion visual odometry based on split-merged motion segmentation

Wang Chenjie, Zhang Yun, Zhao Qing, Wang Wei, Yin Lu, Luo Bin, Zhang Liangpei 1859

Multi-pedestrian tracking optimized by social force model under first-person perspective

Yang Tingzhao, Liu Li, Fu Xiaodong, Liu Lijun, Huang Qingsong 1869

Multi-scenario lane line detection with auxiliary loss

Chen Lichao, Xu Xiuzhi, Cao Jianfang, Pan Lihu 1882

Computer Graphics

Application of skeleton rotating method in fractal dimension calculate of binary image

Ji Youjun, He Jie, Han Haishui, Cheng Zhongzhao, Zeng Baoquan 1894

Fractal curve generation method based on the frequency domain

Chen Wei, Qiao Jiewen, Zhou Chen 1904

Medical Image Processing

Optic disk segmentation by combining UNet and residual attention mechanism

Hou Xiangdan, Zhao Yihao, Liu Hongpu, Guo Hongyong, Yu Xixin, Ding Mengyuan 1915

Left ventricular segmentation on ultrasound images using deep layer aggregation for residual dense networks

Wu Xuanyan, Gou Xinke, Zhu Zizhong, Wei Yulin, Wang Kai 1930

Remote Sensing Image Processing

Data-adaptive single-shot ship detector with a bidirectional feature fusion module for SAR images

Zhang Xiaohan, Yao Libo, Lyu Yafei, Jian Tao, Zhao Zhiwei, Zang Jie 1943

中图法分类号: TP391.4 文献标识码: A 文章编号: 1006-8961(2020)09-1825-12

论文引用格式: Zeng F X and Zhang W S. 2020. Robust distance metric learning algorithm based on nongreedy solution. Journal of Image and Graphics, 25(09): 1825-1836 (曾凡霞, 张文生. 2020. 非贪婪的鲁棒性度量学习算法. 中国图象图形学报, 25(09): 1825-1836) [DOI: 10.11834/jig.190652]

非贪婪的鲁棒性度量学习算法

曾凡霞, 张文生

1. 中国科学院自动化研究所, 北京 100190; 2. 中国科学院大学, 北京 100049

摘要: 目的 度量学习是机器学习与图像处理中依赖于任务的基础研究问题。由于实际应用背景复杂, 在大量不可避免的噪声环境下, 度量学习方法的性能受到一定影响。为了降低噪声影响, 现有方法常用 L1 距离取代 L2 距离, 这种方式可以同时减小相似样本和不相似样本的损失尺度, 却忽略了噪声对类内和类间样本的不同影响。为此, 本文提出了一种非贪婪的鲁棒性度量学习算法——基于 L2/L1 损失的边缘费歇尔分析 (marginal Fisher analysis based on L2/L1 loss, MFA-L2/L1), 采用更具判别性的损失, 可提升噪声环境下的识别性能。**方法** 在边缘费歇尔分析 (marginal Fisher analysis, MFA) 方法的基础上, 所提模型采用 L2 距离刻画相似样本损失、L1 距离刻画不相似样本损失, 同时加大对两类样本的惩罚程度以提升方法的判别性。首先, 针对模型非凸带来的求解困难, 将目标函数转为迭代两个凸函数之差便于求解; 然后, 受 DCA (difference of convex functions algorithm) 思想启发, 推导出非贪婪的迭代求解算法, 求得最终度量矩阵; 最后, 算法的理论证明保证了迭代算法的收敛性。**结果** 在 5 个 UCI (University of California Irvine) 数据集和 7 个人脸数据集上进行对比实验: 1) 在不同程度噪声的 5 个 UCI 数据集上, MFA-L2/L1 算法最优, 且具有较好的抗噪性, 尤其在 30% 噪声程度的 Seeds 和 Wine 数据集上, 与次优方法 LDA-Ng (non-greedy L1-norm linear discriminant analysis) 相比, MFA-L2/L1 的准确率高出 9%; 2) 在不同维度的 AR 和 FEI 人脸数据集上的实验, 验证了模型采用 L1 损失、采用 L2 损失提升了模型的判别性; 3) 在 Senthil、Yale、ORL、Caltech 和 UMIST 人脸数据集的仿真实验中, MFA-L2/L1 算法呈现出较强鲁棒性, 性能排名第 1。**结论** 本文提出了一种基于 L2/L1 损失的鲁棒性度量学习模型, 并推导了一种便捷有效的非贪婪式求解算法, 进行了算法收敛性的理论分析。在不同数据集的不同噪声情况下的实验结果表明, 所提算法具有较好的识别率和鲁棒性。

关键词: 距离度量学习; 鲁棒性; 非贪婪算法; 边缘费歇尔分析 (MFA); 分类识别; L2/L1 损失

Robust distance metric learning algorithm based on nongreedy solution

Zeng Fanxia, Zhang Wensheng

1. Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China;

2. University of Chinese Academy of Sciences, Beijing 100049, China

Abstract: **Objective** Distance metric learning, which is task dependent, is an essential research issue in machine learning and image processing. As a usual preprocessing step for classification and recognition, distance metric learning method enhances the performance of machine learning methods, such as clustering and k-nearest neighbor. This kind of method aims to learn the underlying structure of distribution given the semantic similarity or labels of samples. The optimization goal of distance metric learning is to shrink the distance between similar pairs and expand the distance between dissimilar pairs.

收稿日期: 2019-12-19; 修回日期: 2020-04-14; 预印本日期: 2020-04-21

基金项目: 国家自然科学基金项目 (61702518)

Supported by: National Natural Science Foundation of China (61702518)

With the development of machine learning, many distance metric learning methods have been proposed and applied successfully. These methods use L2 distance to measure loss. However, the L2 distance amplifies the influence of noise with the square operation, because of the unavoidable noise resulting from the complex application background. This situation may make the observed dissimilar and similar pairs respectively appear to be further apart and closer than in fact, which misleads the learning. Thus, the existence of noise leads to the poor performance of distance metric learning in machine learning and pattern recognition. Existing robust distance metric learning methods usually adopt L1 distance rather than L2 distance, aiming to diminish the distance scale of pairs simultaneously and thus reduce the influence of noise. However, by adopting a kind of loss for similar and dissimilar pairs, these methods neglect the different optimization directions between intra- and interclasses. Therefore, a robust distance metric learning algorithm based on L2/L1 loss with nongreedy solution is proposed. This method adopts more discriminative loss and improves the recognition performance in the presence of feature noise.

Method Based on marginal Fisher analysis (MFA) method, the proposed model adopts L2 and L1 distance to measure the loss of similar pairs and dissimilar pairs, respectively. In details, it considers the influence of noise on intra-class and interclass scatters separately. Affected by noise, the observed similar pairs may be closer than the truth, and the observed dissimilar pairs may be further apart than the truth, which may leads that the punishments on pairs is not enough. Adopting L2/L1 loss, L2 distance gives more punishments on similar samples than L1 distance does, and L1 distance gives more punishments on dissimilar samples than L2 distance does. However, the model is nonconvex, which is involved with both minimization of L2 norm and maximization of the L1 norm. Hence, the existing solution for trace ratio problems are not suitable for the objective function. This makes the object difficult to solve, thus a nongreedy solution is derived. First, the objective function, which is a ratio, is transformed into a difference of two convex functions during each iteration. This process makes the object easy to solve. Then, inspired by the idea of the difference of convex function algorithms (DCA), an iterative algorithm with nongreedy solution is derived to solve the object, and a projection matrix is learned. Analyzing the algorithm, in fact, an auxiliary function of the object is generated during the iteration, which makes sure that the value of objective function decreases. Lastly, a theoretical analysis of the iterative algorithm is performed to ensure the convergence of the proposal.

Result Synthetic experiments are conducted on five UCI (University of California Irvine) datasets and seven face datasets with noise. The nearest neighbor classifier based on the learned matrix is used to compare the performance of related methods. The proposed method outperforms other methods in terms of accuracy. First, when 5%, 15%, 25%, and 30% feature noise is added on five UCI datasets, MFA-L2/L1 achieves the best performance and exhibits robustness against noise. In addition, the accuracy of MFA-L2/L1 is 9% higher than the second-best method. Second, during the accuracy comparison with varying dimension of projection based on AR dataset, the proposal beats MFA, which validates that using the L1 loss for dissimilar pairs makes the proposal more discriminative. During the accuracy comparison with varying dimensions of projection based on FEI dataset, the proposal beats LDA-NgL1 (non-greedy L1-norm liner discriminant analysis), which validates that using the L2 loss for similar pairs makes the proposal more discriminative. Lastly, based on five face datasets with noise, MFA-L2/L1 behaves better than other methods in terms of robustness.

Conclusion This study proposes a robust distance metric learning method based on L2/L1 loss, which adopts different losses for similar pairs and dissimilar pairs respectively. To solve a nonconvex object, the object is transformed into a difference of two convex functions during each iteration. By virtue of the idea of DCA, an iterative algorithm is derived, which generates an auxiliary function decreasing the object. The convergence of the algorithm is guaranteed by a theoretical proof. Based on the public datasets, several series of experiments with different levels and types of synthetic noise are executed. Results on different datasets show that the proposal performs well and is robust to noise based on the experiment datasets. Future research will focus on local distance metric learning in the presence of label noise, because the label noise makes the observed distribution deviate from the true distribution more than the feature noise does.

Key words: distance metric learning; robustness; non-greedy algorithm; marginal Fisher analysis (MFA); classification and recognition; L2/L1 loss

0 引言

度量学习是机器学习与图像处理中的基础研究问题(Bellet 等, 2014; Moutafis 等, 2017), 具有重要的研究价值与意义。从度量函数类型角度分析, 度量学习方法可分为线性方法和非线性方法, 均旨在学习一个线性马氏矩阵(可分解为投影矩阵)或非线性变换, 使得相似样本距离更近、不相似样本距离更远。该方向的研究已广泛应用到图像检索(Dai 等, 2018)、人脸识别(Hu 等, 2014; Guillaumin 等, 2009)、行人再识别(Liao 等, 2015; 张晶和赵旭, 2017)、跨模态检索(Liong 等, 2017)、跟踪(Hu 等, 2016; 霍其润 等, 2017; Zhang 等, 2017)等领域。

在给定的样本相似性或标签的前提下, 传统的线性方法通过定义损失来学习一个马氏矩阵 $\mathbf{M}$ 或投影矩阵 $\mathbf{W}$ ($\mathbf{M} = \mathbf{W}^T \mathbf{W}$) 作为输出。当投影矩阵不满秩时, 度量学习方法可看做一种降维方法, 因此度量学习与诸如流形学习一类的降维方法有着密切的联系。大部分方法常基于二元组或三元组样本相似性建立损失。其中, Xing 等人(2002)提出的 MMC (Mahalanobis metric for clustering)、Hoi 等人(2006)提出的 DCA (discriminative component analysis) 和 Xiang 等人(2008)提出的针对聚类和分类的马氏度量方法是早期基于二元组关系的经典代表性方法。基于三元组关系, 代表性方法包括 LMNN (large margin nearest neighbor) (Weinberger 和 Saul, 2009) 以及后续的 SCML (sparse compositional metric learning) (Shi 等, 2014) 等。受线性方法的启发, 非线性方法常将数据扩展到核空间或利用深度神经网络学习一个非线性度量(Lu 等, 2017)。但这些工作都是基于 L2 距离建立损失, 易受实际应用的噪声干扰, 其性能受到一定影响(Ke 和 Kanade, 2005)。

如何针对噪声环境构建一个鲁棒的度量学习算法, 是提升该情境下分类和识别性能的关键点(Wang 等, 2014)。传统方法常假设类内和类间样本均服从两个不同的高斯分布, 因而模型采用 L2 损失(Xiang 等, 2008)。但实际应用中, 噪声的存在往往使得观测分布偏离真实分布, 采用 L2 距离反而会同时扩大样本距离的尺度, 加剧观测分布与真实分布的偏离程度。因此, 在考虑噪声的情况下, 一部分工作转而采用 L1 距离作为损失(Kwak, 2008; Nie

等, 2011), 其本质是假设类内样本和类间样本均服从两个不同的拉普拉斯分布(Zhang 等, 2016a)。由于 L1 损失可减小样本距离尺度以降低噪声对方法的影响, 很多学者基于 L1 损失提出了一系列卓有成效的方法。例如, 非监督情况下, PCA-L1 (principal component analysis based on L1-norm maximization) (Kwak, 2008) 采用 L1 损失取代 L2 损失, 最大化目标方差, 提升了算法的鲁棒性。有监督情况下, 基于 L1 损失的线性判别分析(L1-norm-based linear discriminant analysis, LDA-L1) (Zhong 和 Zhang, 2013) 方法和 Wang 等人(2014)利用 L1 损失提出的度量学习方法, 取得了一定效果。但这类方法均采用逐向量的贪婪式求解算法, 未对投影矩阵进行整体求解, 其优化效果并不太理想。Liu 等人(2017)在 LDA-L1 的基础上, 推导了一种非贪婪的矩阵整体求解算法, 提升了 LDA-L1 的性能。

然而, 上述方法存在两方面的问题。一方面, 这些方法本质上并未从判别性的角度对度量进行提升。L1 损失同时减小类内和类间距离尺度, 并未在减小类内散度的同时增大类间散度。因此, 要提升方法的判别性, 应在扩大类内样本距离尺度的同时减小类间样本距离尺度。这样可同时扩大两类损失的惩罚程度, 以更好地提升算法的鲁棒性和判别性。另一方面, 类内样本和类间样本应考虑不同的损失(Ye 等, 2018)。

鉴于上述分析, 基于边缘费歇尔分析(marginal Fisher analysis, MFA) (Xu 等, 2007; Xiong 等, 2014) 方法, 本文提出了一种采用 L2/L1 损失的非贪婪的鲁棒性度量学习方法, 以同时最小化类内样本的 L2 距离和最大化类间样本的 L1 距离为目标, 加大对两类样本的惩罚程度, 提升算法的判别性。针对目标函数非凸带来的求解困难, 将目标函数转化为迭代优化两个凸函数之差进行求解。随后, 受 DCA (difference of convex functions algorithm) (An 和 Tao, 2005) 思想启发, 推导了一种非贪婪的迭代求解算法 MFA-L2/L1, 从理论上分析了迭代过程的收敛性。随后在公开数据集上进行了大量仿真实验, 结果表明, 本文 MFA-L2/L1 方法具有一定的优越性和鲁棒性。

1 相关方法

给定训练样本集 $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N] \in \mathbf{R}^{d \times N}$,

$\mathbf{x}_i$ 是第 i 个含有 d 维特征的训练样本。对于样本对, 有 $(\mathbf{x}_i, \mathbf{x}_j) \in \mathbf{S}$ 或 $(\mathbf{x}_i, \mathbf{x}_j) \in \mathbf{D}$, 其中, $\mathbf{S}$ 是相似样本集, $\mathbf{D}$ 是不相似样本集。记待学习的投影矩阵为 $\mathbf{W} \in \mathbf{R}^{d \times m}$, 其中, m 表示投影向量的个数。

1.1 基于 L1 范数的线性判别分析

考虑到 L2 损失容易受噪声影响 (Ke 和 Kanade, 2005), Zhong 和 Zhang (2013) 提出了基于 L1 损失的线性判别分析 (LDA-L1)。该方法致力于在目标函数下学习一个投影矩阵 $\mathbf{W}$, 具体为

$$\begin{aligned} \max_{\mathbf{W}} \quad & \frac{\sum_{i=1}^c n_i \|(\bar{\mathbf{x}}_i - \bar{\mathbf{x}})^T \mathbf{W}\|_1}{\sum_{i=1}^c \sum_{\mathbf{x}_j \in c_i} \|(\mathbf{x}_j - \bar{\mathbf{x}}_i)^T \mathbf{W}\|_1} \\ \text{s. t.} \quad & \mathbf{W}^T \mathbf{W} = \mathbf{I} \end{aligned} \quad (1)$$

式中, $\bar{\mathbf{x}}$ 是总体样本均值, $\bar{\mathbf{x}}_i$ 是第 i 类样本的均值, n_i 表示对应第 i 个类别的样本个数, c 表示总类别数, $\mathbf{x}_j \in c_i$ 表示第 j 个样本 $\mathbf{x}_j$ 属于第 i 个类别类 c_i , $(\mathbf{x}_j - \bar{\mathbf{x}}_i)^T$ 表示列向量 $\mathbf{x}_j - \bar{\mathbf{x}}_i$ 的转置。与传统线性判别分析 (linear discriminant analysis, LDA) 不同的是, 该目标函数采用 L1 损失刻画类内散度和类间散度。由于目标函数同时涉及 L1 范数最大化和最小化, 求解较为困难, LDA-L1 通过贪婪的逐向量优化算法求得最终投影矩阵。为了改善贪婪式求解算法在模型式 (1) 中的不足, Liu 等人 (2017) 提出了 non-greedy L1-norm LDA (LDA-NgL1) 算法, 该方法对投影矩阵进行整体求解, 相对贪婪算法有一定的性能提升。

1.2 边缘费歇尔分析 (MFA)

线性判别分析假设类内样本和类间样本均服从高斯分布, 但真实数据成因复杂, 难以满足高斯假设 (Xu 等, 2007)。针对这一限制, MFA (Xu 等, 2007) 根据样本之间的近邻关系重新定义了类内散度与类间散度, 基于该目标函数学习投影矩阵 $\mathbf{W}$, 具体为

$$\begin{aligned} \min_{\mathbf{W}} \quad & \frac{\sum_{\mathbf{x}_{ij} \in \mathbf{S}_M} \|\mathbf{x}_{ij}^T \mathbf{W}\|_2^2}{\sum_{\mathbf{x}_{ij} \in \mathbf{D}_M} \|\mathbf{x}_{ij}^T \mathbf{W}\|_2^2} \\ \text{s. t.} \quad & \mathbf{W}^T \mathbf{W} = \mathbf{I} \end{aligned} \quad (2)$$

式中, $\mathbf{x}_{ij}$ 是关于相似或不相似样本对 $(\mathbf{x}_i, \mathbf{x}_j)$ 的差向量, $\mathbf{x}_{ij} = \mathbf{x}_i - \mathbf{x}_j$, 行向量 $\mathbf{x}_{ij}^T$ 是列向量 $\mathbf{x}_{ij}$ 的转置。此处采用样本对关系形式表示 MFA, $\mathbf{S}_M$ 和 $\mathbf{D}_M$ 分别是基于内部图 $\mathbf{G}_S = \{\mathbf{X}, \mathbf{A}_S\}$ 的相似样本集和基于惩罚图 $\mathbf{G}_D = \{\mathbf{X}, \mathbf{A}_D\}$ 的不相似样本集。两个图 $\mathbf{A}_S$

和 $\mathbf{A}_D$ 的连接权定义为

$$A_S(i, j) = \begin{cases} 1 & i \in N_{k_S}^+(j) \text{ 或 } j \in N_{k_S}^+(i) \\ 0 & \text{其他} \end{cases} \quad (3)$$

$$A_D(i, j) = \begin{cases} 1 & i \in N_{k_D}^-(j) \text{ 或 } j \in N_{k_D}^-(i) \\ 0 & \text{其他} \end{cases} \quad (4)$$

式中, $N_{k_S}^+(i)$ 表示与第 i 个样本具有相同标签的 k_S 近邻集, $N_{k_D}^-(i)$ 表示与第 i 个样本具有不同标签的 k_D 近邻集。该目标函数可直接通过特征向量分解实现求解。然而, 由于 MFA 采用 L2 距离度量损失, 该方法受噪声影响较大。

2 方法

2.1 基于 L2/L1 损失的鲁棒性度量学习模型

由于实际应用中不可避免的噪声, 数据的观测分布易偏离真实分布, 特征噪声下的观测样本与真实样本分布可能如图 1 所示。

真实样本: ● 类别1 ▲ 类别2 观测样本: ● 类别1 ▲ 类别2

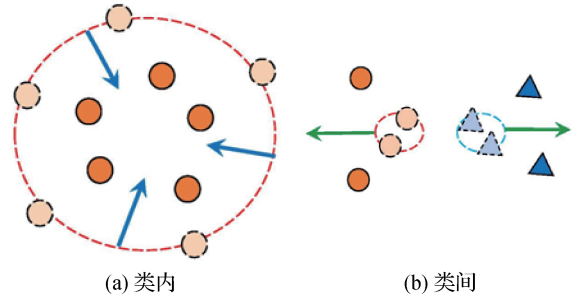


图 1 特征噪声下的观测样本与真实样本分布图

Fig. 1 Distribution of the observed samples and true samples in the presence of feature noise ((a) intraclass; (b) interclass)

考虑 L2 距离和 L1 距离具有不同的尺度性, 基于图 1 进行分析, 特征噪声对度量学习的负面影响可能表现在类内和类间两个方面。一方面, 如图 1 所示的类内情况, 噪声可能使得相似样本的真实距离比观测距离更远。类内样本应该采用尺度更大的距离损失, 加大对该类样本的惩罚, 使得同类样本间尽可能压缩、距离更近。因此, 该情况下采用尺度较小的 L1 距离定义类内损失, 可能使得惩罚程度不够; 采用尺度较大的 L2 距离定义类内损失可加大对类内样本的惩罚程度。另一方面, 如图 1 所示的类间情况, 噪声可能使得不相似样本的真实距离比观测距离更近。与类内情况相反, 类间样本应该采用

尺度较小的损失,加大对类间样本的惩罚,使得类间样本尽可能分离、距离更远。因此,该情况下采用尺度较小的 L1 距离定义类间损失,可加大对类间样本的惩罚程度。综合噪声对两类样本不同优化方向的影响,采用 L2 距离定义类内样本损失、L1 距离定义类间样本损失,同时加大对两类样本的惩罚程度,可提升投影矩阵的判别性。

利用 MFA 方法无需高斯假设的优势,本文提出一种基于 L2/L1 损失的边缘费歇尔分析 (marginal Fisher analysis based on L2/L1 loss, MFA-L2/L1) 模型,具体为

$$\min_{\mathbf{W}} \frac{\sum_{\mathbf{x}_{ij} \in \mathbf{S}_M} \|\mathbf{x}_{ij}^T \mathbf{W}\|_2^2}{\sum_{\mathbf{x}_{ij} \in \mathbf{D}_M} \|\mathbf{x}_{ij}^T \mathbf{W}\|_1} = \frac{\text{tr}(\mathbf{W}^T \mathbf{A}^T \mathbf{A} \mathbf{W})}{\|\mathbf{B} \mathbf{W}\|_1} \quad (5)$$

s. t. $\mathbf{W}^T \mathbf{W} = \mathbf{I}$

式中, $\mathbf{x}_{ij}$ 与 MFA 在式(2)中的定义相同,是来自相似样本集 $\mathbf{S}_M$ 或不相似样本 $\mathbf{D}_M$ 的差向量; $\text{tr}(\cdot)$ 表示对矩阵取迹,所有来自相似样本集 $\mathbf{S}_M$ 的行向量 $\mathbf{x}_{ij}^T$ 构成矩阵 $\mathbf{A}$,所有来自不相似样本集 $\mathbf{D}_M$ 的行向量 $\mathbf{x}_{ij}^T$ 构成矩阵 $\mathbf{B}$ 。为了方便表示,对任意大小为 $n_1 \times n_2$ 的矩阵 $\mathbf{C}$,定义其矩阵范数 $\|\mathbf{C}\|_1 = \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} |\mathbf{C}_{i,j}|$ 。由于该目标函数非凸,且同时涉及最小化 L2 范数和最大化 L1 范数, MFA 和 LDA-NgL1 的求解算法都将不再适合该模型。为了解决该模型的求解问题,本文将推导一种非贪婪的优化算法,实现逐矩阵求解,无需逐向量求解。

2.2 求解算法

与 MFA 和 LDA-L1 模型相比, MFA-L2/L1 的目标函数同样是两个凸函数的商形式。受已有工作 (Wang 等, 2014) 中定理 1 和定理 2 启发,目标函数式(5)可转化为

$$\mathbf{W}_t = \arg \min_{\mathbf{W}^T \mathbf{W} = \mathbf{I}} G(\mathbf{W}) - \lambda_{t-1} L(\mathbf{W}) \quad (6)$$

式中, $G(\mathbf{W}) = \text{tr}(\mathbf{W}^T \mathbf{A}^T \mathbf{A} \mathbf{W})$, $L(\mathbf{W}) = \|\mathbf{B} \mathbf{W}\|_1$, λ_{t-1} 是由第 $t-1$ 次投影矩阵 $\mathbf{W}_{t-1}$ 计算所得的目标函数值,具体为

$$\lambda_{t-1} = \frac{\text{tr}(\mathbf{W}_{t-1}^T \mathbf{A}^T \mathbf{A} \mathbf{W}_{t-1})}{\|\mathbf{B} \mathbf{W}_{t-1}\|_1} \quad (7)$$

用 $F(\mathbf{W})$ 表示 $G(\mathbf{W}) - \lambda_{t-1} L(\mathbf{W})$, 则

$$\arg \min_{\mathbf{W}^T \mathbf{W} = \mathbf{I}} F(\mathbf{W}) = \arg \min_{\mathbf{W}^T \mathbf{W} = \mathbf{I}} G(\mathbf{W}) - \lambda_{t-1} L(\mathbf{W}) \quad (8)$$

观察目标函数式(8)可知, λ_{t-1} 在迭代求解 $\mathbf{W}_t$ 时是常数,不影响推导分析。因此,优化 $F(\mathbf{W})$ 的实质为优化 $G(\mathbf{W})$ 和 $L(\mathbf{W})$ 两个凸函数的差。受 DCA (difference of convex functions algorithm) (An 和 Tao, 2005) 思想启发,在第 t 次迭代求解 $\mathbf{W}_t$ 时,固定 λ_{t-1} ,用上标 k 表示求解 $\mathbf{W}_t$ 时的内部迭代次数,与外部迭代次数 t 区分,可生成两个变量序列,即中间变量序列 $\mathbf{Y}^k$ 和求解变量序列 $\mathbf{W}^k$,具体为

$$\begin{cases} \mathbf{Y}^k \in \partial L(\mathbf{W}^k) \\ \mathbf{W}^{k+1} = \arg \min_{\mathbf{W}} G(\mathbf{W}) - \lambda_{t-1} [L(\mathbf{W}^k) + \text{tr}((\mathbf{Y}^k)^T (\mathbf{W} - \mathbf{W}^k))] \end{cases} \quad (9)$$

式中, $\mathbf{Y}^k$ 是 $L(\mathbf{W})$ 在 $\mathbf{W}^k$ 处的次梯度。

记 $H(\mathbf{W}, \mathbf{W}^k) = \text{tr}((\mathbf{Y}^k)^T (\mathbf{W} - \mathbf{W}^k))$, 则 $L(\mathbf{W}^k) + \text{tr}((\mathbf{Y}^k)^T (\mathbf{W} - \mathbf{W}^k)) = L(\mathbf{W}^k) + H(\mathbf{W}, \mathbf{W}^k)$ 是 $L(\mathbf{W})$ 在 $\mathbf{W}^k$ 处的一阶 Taylor 展开式。对于序列式(9),有如下定理:

定理 1 序列式(9)中的 $\mathbf{W}^{k+1} \in \mathbf{R}^{d \times m}$ 使得 $F(\mathbf{W}^{k+1}) \leq F(\mathbf{W}^k)$ 成立。

证明:由于 $\mathbf{Y}^k$ 是 $L(\mathbf{W})$ 在 $\mathbf{W}^k$ 处的次梯度,有 $L(\mathbf{W}) \geq L(\mathbf{W}^k) + \lambda_{t-1} H(\mathbf{W}, \mathbf{W}^k)$ 成立。所以,有

$$\begin{aligned} F(\mathbf{W}^{k+1}) &= G(\mathbf{W}^{k+1}) - \lambda_{t-1} L(\mathbf{W}^{k+1}) \leq \\ &G(\mathbf{W}^{k+1}) - \lambda_{t-1} [L(\mathbf{W}^k) + H(\mathbf{W}^{k+1}, \mathbf{W}^k)] \leq \\ &G(\mathbf{W}^k) - \lambda_{t-1} [L(\mathbf{W}^k) + H(\mathbf{W}^k, \mathbf{W}^k)] = \\ &G(\mathbf{W}^k) - \lambda_{t-1} L(\mathbf{W}^k) = F(\mathbf{W}^k) \end{aligned}$$

证毕。

从定理 1 可知,由序列式(9)可得出一系列下降的目标函数值 $F(\mathbf{W}^k)$ 。

由序列式(9)推导对应目标函数式(8)的具体计算为

$$\begin{cases} \mathbf{Y}^k = \mathbf{B}^T \text{sign}(\mathbf{B} \mathbf{W}^k) \\ \mathbf{W}^{k+1} = \arg \min_{\mathbf{W}} \text{tr}(\mathbf{W}^T \mathbf{A}^T \mathbf{A} \mathbf{W}) - \lambda_{t-1} H(\mathbf{W}, \mathbf{W}^k) \end{cases} \quad (10)$$

式中, $\text{sign}(\cdot)$ 是关于矩阵逐元素的函数,具体为

$$\text{sign}(x) = \begin{cases} 1 & x \geq 0 \\ -1 & x < 0 \end{cases}$$

在序列式(10)中,关于 $\mathbf{W}^k$ 的常数项 $L(\mathbf{W}^k)$ 已忽略。考虑序列式(10)中的求解变量序列 $\mathbf{W}^k$,对目标函数关于变量 $\mathbf{W}$ 求导,并令导数为零,有 $2\mathbf{A}^T \mathbf{A} \mathbf{W} - \lambda_{t-1} \mathbf{Y}^k = 0$,求解可得

$$\mathbf{W}^{k+1} = \frac{\lambda_{t-1}}{2} (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{Y}^k \quad (11)$$

根据以上推导,可得到关于 $\mathbf{Y}^k$ 和 $\mathbf{W}^k$ 两个序列的迭代计算,直至序列式(10)迭代终止。迭代终止时的最终 $\mathbf{W}^{k+1}$ 则为第 t 次迭代所求的 $\mathbf{W}_t$, 然后代入 $\mathbf{W}_t$ 可求得 λ_t 。

进一步,考虑矩阵 $\mathbf{W}$ 的正交性,约束 $\mathbf{W}^T \mathbf{W} = \mathbf{I}$ 可通过正交强迫一致问题 (orthogonal procrustes problem) (Green, 1952) $\mathbf{W} = \arg \min_{\mathbf{W}^T \mathbf{W} = \mathbf{I}} \|\mathbf{W} - \mathbf{W}_t\|_F$ 来解决,其解为 $\mathbf{W} = \mathbf{U} \mathbf{V}^T$ 。其中, $\mathbf{U}$ 和 $\mathbf{V}^T$ 是大小分别为 $d \times k_1$ 和 $k_1 \times m$ 的正交矩阵,均来自 $\mathbf{W}_t$ 的奇异值分解 (singular value decomposition, SVD), k_1 是非零奇异值的个数。总结以上分析, MFA-L2/L1 求解的算法 1 具体步骤如下:

输入: $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N] \in \mathbf{R}^{d \times N}$, 投影矩阵维度 m , 总迭代次数 T , 固定 λ_{t-1} 时的内部迭代次数 l 和终止误差 ε 。

输出: 投影矩阵 $\mathbf{W}$ 。

初始化: 随机初始化投影矩阵 $\mathbf{W}_0$, 使其满足正交性: $\mathbf{W}_0^T \mathbf{W}_0 = \mathbf{I}$; 令 $t = 1$ 。

1) 按式(7)更新 λ_{t-1} ;

2) 令 $k = 1$, $\mathbf{W}^1 = \mathbf{W}_{t-1}$;

3) 按照式(10)中第 1 个序列更新 $\mathbf{Y}^k$;

4) 按照式(11)更新 $\mathbf{W}^{k+1}$, $k = k + 1$;

5) 检查内部序列式(10)是否满足迭代终止条件; 若满足则跳出序列迭代, 得到固定 λ_{t-1} 时所求的 $\mathbf{W}_t$; 若不满足返回步骤 3);

6) 计算关于 $\mathbf{W}_t$ 的奇异值分解, 更新第 t 次迭代投影矩阵 $\mathbf{W}_t = \mathbf{U} \mathbf{V}^T$;

7) $t = t + 1$;

8) 检查是否满足收敛或满足迭代次数 $t > T$, 若不满足则返回步骤 1); 若满足则结束迭代, 输出投影矩阵 $\mathbf{W}$ 。

从迭代流程分析, 序列 $\mathbf{Y}^k$ 的作用是构造一个关于目标函数式(8)的辅助函数, 即序列式(10)中关于 $\mathbf{W}$ 的目标函数, 使得在迭代过程中目标函数式(8)下降。整个算法过程始终对 $\mathbf{W}$ 进行整体求解, 属于非贪婪式算法, 因此该算法比逐向量的贪婪式算法更为便捷。

2.3 算法收敛性分析

上节讨论了求解 $\mathbf{W}_t$ 时序列的单调性, 本节将联合 $\mathbf{W}_t$ 和 λ_t 两个变量分析算法的整体收敛性。

定理 2 算法 1 使得目标函数式(5)下降。

证明: 用投影向量 $(\mathbf{w}_l)_t$ 表示对矩阵 $\mathbf{W}_t$ 的第 l 个投影向量, 用 $\mathbf{b}^i$ 表示矩阵 $\mathbf{B}$ 的第 i 个行向量, 假设矩阵 $\mathbf{B}$ 的行向量个数为 n_b 。

用 $\frac{\mathbf{b}^i \mathbf{w}_l}{|\mathbf{b}^i \mathbf{w}_l|}$ 表示 $\text{sign}(\mathbf{b}^i \mathbf{w}_l)$ (当分母为 0 时, 令其为 1), 由前面的分析可知, $\mathbf{W}_t$ 使得目标 $F(\mathbf{W})$ 下降, 则

$$\begin{aligned} & \sum_{l=1}^m (\mathbf{w}_l)_t^T \mathbf{A}^T \mathbf{A} (\mathbf{w}_l)_t - \\ & \lambda_{t-1} \sum_{l=1}^m \sum_{i=1}^{n_b} \frac{(\mathbf{w}_l)_t^T (\mathbf{b}^i)^T \mathbf{b}^i (\mathbf{w}_l)_{t-1}}{|\mathbf{b}^i (\mathbf{w}_l)_{t-1}|} \leq \\ & \sum_{l=1}^m (\mathbf{w}_l)_{t-1}^T \mathbf{A}^T \mathbf{A} (\mathbf{w}_l)_{t-1} - \\ & \lambda_{t-1} \sum_{l=1}^m \sum_{i=1}^{n_b} \frac{(\mathbf{w}_l)_{t-1}^T (\mathbf{b}^i)^T \mathbf{b}^i (\mathbf{w}_l)_{t-1}}{|\mathbf{b}^i (\mathbf{w}_l)_{t-1}|} \quad (12) \end{aligned}$$

根据柯西—施瓦兹不等式, 在第 t 次迭代与第 $t-1$ 次迭代之间, 对投影矩阵 $\mathbf{W}$ 的第 l 个投影向量, 有

$$(\mathbf{b}^i (\mathbf{w}_l)_t)^T \mathbf{b}^i (\mathbf{w}_l)_{t-1} \leq |\mathbf{b}^i (\mathbf{w}_l)_t| |\mathbf{b}^i (\mathbf{w}_l)_{t-1}| \quad (13)$$

将矩阵 $\mathbf{B}$ 的 n_b 个行向量叠加, 则有

$$\begin{aligned} & \sum_i \frac{(\mathbf{b}^i (\mathbf{w}_l)_t)^T \mathbf{b}^i (\mathbf{w}_l)_{t-1}}{|\mathbf{b}^i (\mathbf{w}_l)_{t-1}|} - \sum_i |\mathbf{b}^i (\mathbf{w}_l)_t| \leq \\ & \sum_i \frac{(\mathbf{b}^i (\mathbf{w}_l)_{t-1})^T \mathbf{b}^i (\mathbf{w}_l)_{t-1}}{|\mathbf{b}^i (\mathbf{w}_l)_{t-1}|} - \sum_i |\mathbf{b}^i (\mathbf{w}_l)_{t-1}| = 0 \quad (14) \end{aligned}$$

对于投影矩阵 $\mathbf{W}$ 的 m 个投影向量, 将其所对应的公式(14)代入式(12), 得到

$$\begin{aligned} & \sum_{l=1}^m (\mathbf{w}_l)_t^T \mathbf{A}^T \mathbf{A} (\mathbf{w}_l)_t - \lambda_{t-1} \sum_{l=1}^m \|\mathbf{B}(\mathbf{w}_l)_t\|_1 \leq \\ & \sum_{l=1}^m (\mathbf{w}_l)_{t-1}^T \mathbf{A}^T \mathbf{A} (\mathbf{w}_l)_{t-1} - \lambda_{t-1} \sum_{l=1}^m \|\mathbf{B}(\mathbf{w}_l)_{t-1}\|_1 = 0 \quad (15) \end{aligned}$$

最后, 得出

$$\frac{\sum_{l=1}^m \|\mathbf{A}(\mathbf{w}_l)_t\|_2^2}{\sum_{l=1}^m \|\mathbf{B}(\mathbf{w}_l)_t\|_1} \leq \lambda_{t-1} = \frac{\sum_{l=1}^m \|\mathbf{A}(\mathbf{w}_l)_{t-1}\|_2^2}{\sum_{l=1}^m \|\mathbf{B}(\mathbf{w}_l)_{t-1}\|_1} \quad (16)$$

即

$$\frac{\text{tr}(\mathbf{W}_t^T \mathbf{A}^T \mathbf{A} \mathbf{W}_t)}{\|\mathbf{B} \mathbf{W}_t\|_1} = \lambda_t \leq \lambda_{t-1} = \frac{\text{tr}(\mathbf{W}_{t-1}^T \mathbf{A}^T \mathbf{A} \mathbf{W}_{t-1})}{\|\mathbf{B} \mathbf{W}_{t-1}\|_1} \quad (17)$$

因此, 目标函数式(5)在迭代过程中下降, 证毕。

3 实验

为了验证本文算法的鲁棒性和有效性,实验分别在5个UCI(University of California Irvine)数据集(Dua和Graff,2019)和7个人脸数据集上添加不同程度和类型的噪声,与相关的代表性算法进行对比。为确保实验的公正性,所有实验均在训练样本中学习一个度量矩阵或变换矩阵,之后均基于所学矩阵采用最近邻算法对测试样本进行分类。

3.1 UCI数据集上的实验

为分析本文算法在不同程度噪声下的性能,本组实验基于Statlog(Heart)、Parkinsons(Little等,2007)、Seeds、Sonar和Wine等5个UCI数据集(Dua和Graff,2019)进行方法对比。UCI数据集信息如表1所示。为避免特征尺度与噪声尺度不一致影响实验评价,将数据集每个特征维度的模标准化为1,参考椒盐噪声的定义,随机选取部分样本的部分特征置为0或1作为特征噪声,分别加入比例为5%、15%、25%和30%左右的噪声,按噪声比的平方根对样本数和特征数取整,随机选取对应数量的样本和维度加入噪声。训练集和测试集的比例为7:3,该划分重复100次,取平均值作为结果。

表1 UCI数据集信息

Table 1 Characteristics of UCI datasets

数据集	样本数	特征数	类别数
Statlog (Heart)	270	13	2
Parkinsons	195	22	2
Seeds	210	7	3
Sonar	208	60	2
Wine	178	13	3

选用LDA-L1(Zhong和Zhang,2013)、LDA-NgL1(Liu等,2017)、MFA-NgL1、MFA(Xiong等,2014)和DNS(discriminative null space)(Zhang等,2016b)作为对比方法。对于MFA-NgL1模型,在MFA的样本关系下采用与LDA-NgL1相同的L1损失与解法。LDA-L1、LDA-NgL1、MFA-NgL1和本文MFA-L2/L1属于迭代解法,MFA和DNS属于非迭代解法。LDA-L1梯度步长设为0.01,其他设置均遵循原方法的设置。将矩阵第 t 次迭代时MFA-L2/L1

序列式(10)的迭代次数设置为 $l=1$ 。由于LDA-NgL1和MFA-NgL1矩阵的第 t 次迭代属于梯度法,需要较高的迭代次数,因此梯度迭代次数设置为100。当矩阵总迭代次数 $t>100$ 或误差满足

$$\varepsilon = \frac{\|W_{t+1} - W_t\|_F}{\|W_t\|_F} < e^{-11}$$

时,算法终止。

在100次样本划分过程中均采用了不同的随机种子,各算法在同一随机种子下保持同一初始矩阵。实验事先采用主成分分析(principal component analysis, PCA)计算出95%贡献率的成分个数 p ,投影矩阵维度设置为 $p+1$ 。各算法的排名及性能结果分别如表2和图2所示。

表2 各算法在UCI数据集上的实验排名
Table 2 Ranking of different algorithms based on UCI datasets experiment

算法	排名
LDA-NgL1(Liu等,2017)	2.75
MFA-NgL1	3.15
LDA-L1(Zhong和Zhang,2013)	4.60
MFA(Xiong等,2014)	3.70
DNS(Zhang等,2016b)	5.40
本文(MFA-L2/L1)	1.40

注:加粗字体为最优结果。

从表2和图2可以看出,在不同程度的噪声下,MFA-L2/L1均表现出较好的准确性,在Parkinson、Seeds和Wine数据集上的性能始终超越其他对比方法。且在Wine数据集上,本文方法的准确率与次优方法LDA-NgL1的差异尤为明显,当噪声程度处于15%~30%时,所提方法的准确率比LDA-NgL1高出9%。在图2(a)中,本文方法虽然在Heart和Sonar数据集上的性能略低于LDA-NgL1、MFA-NgL1和LDA-L1,但图2(b)~(d)显示,随噪声程度升高,MFA-L2/L1的性能逐渐超过其他方法,与次优方法间的差异也逐渐增大。观察4个子图之间的变换,随噪声程度增加,MFA-L2/L1具有较好的抗噪性。综合实验分析表明,MFA-L2/L1在本次实验的加噪数据集上表现出一定的鲁棒性。

3.2 AR数据集上的实验

为了具体地比较MFA-L2/L1和MFA两个相近

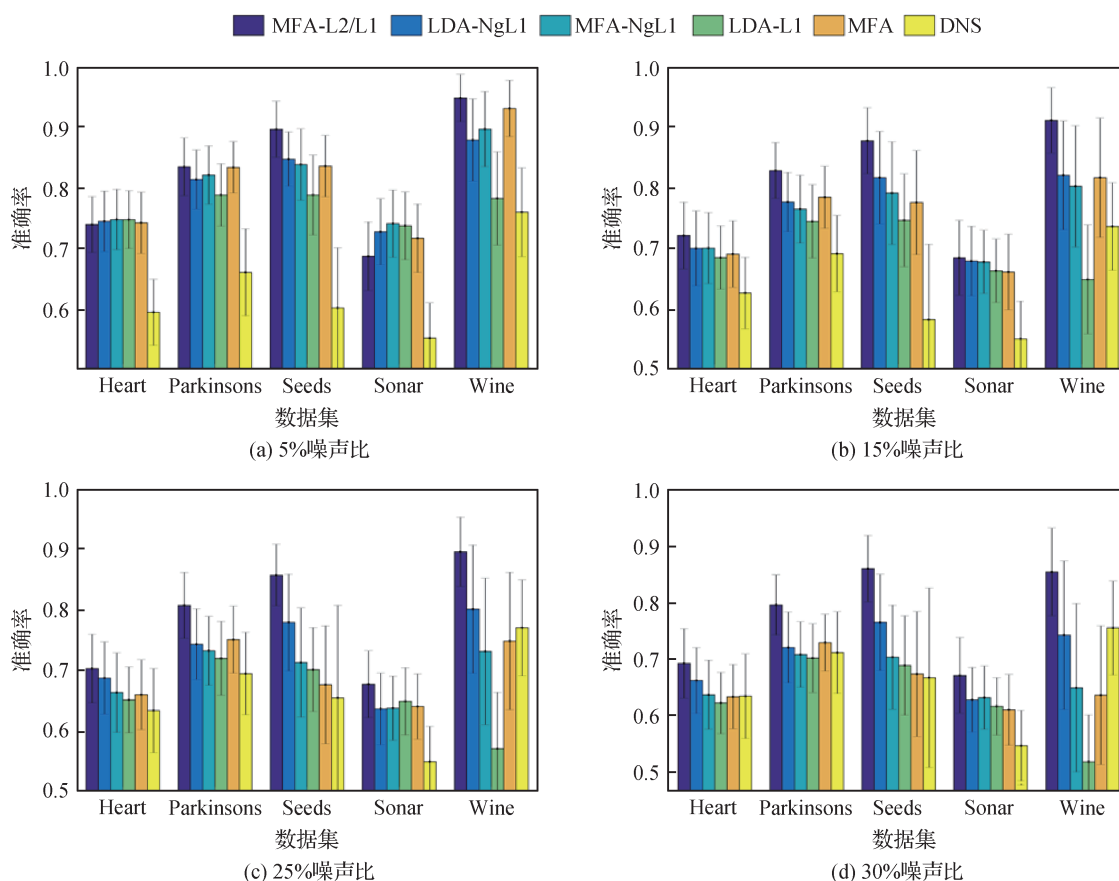


图2 各UCI数据集在不同程度噪声下的准确率结果

Fig. 2 Accuracy on UCI dataset under different levels of noise

((a) noise rate = 5%; (b) noise rate = 15%; (c) noise rate = 25%; (d) noise rate = 30%)

方法的差异,实验在AR数据集(Martinez和Kak, 2001)上对两种方法进行对比,该数据集为经典的人脸识别数据集,实验选取AR数据集裁剪后的2600幅图像,共100人(男女各50人),每人为一类,每类26幅,包括光照、表情和是否佩戴眼镜、围巾等变化的图像。

实验将图像转换为 50×40 像素的灰度图像,对每个类别随机选取6幅图像,每幅图像分别加入

15%比例的随机块和椒盐噪声两种噪声。从每个类别中随机选取13幅图像作为训练样本,剩余13幅图像作为测试样本。实验事先采用PCA保留95%的贡献率,对所有样本进行前处理。各方法的实验设置与UCI数据集实验一致。为避免实验结果的随机性,实验过程重复10次,对比了投影矩阵在80~225维之间的性能变化,平均结果如图3所示。

由图3可知,在高斯块噪声情况对应的子图3(a)

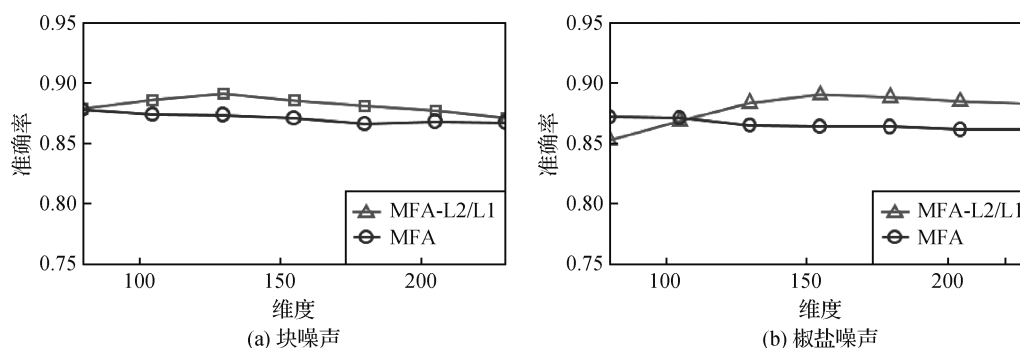


图3 在AR数据集上MFA-L2/L1与MFA算法的准确率对比结果

Fig. 3 Comparison of accuracy between MFA-L2/L1 and MFA on AR dataset ((a) patch noise; (b) salt and pepper noise)

中, MFA-L2/L1 的性能好于 MFA。在椒盐噪声情况对应的子图 3(b) 中, MFA 的性能随维度增加发生了微弱下降, 之后保持稳定; MFA-L2/L1 的性能随维度增加逐渐上升, 超过 MFA 之后保持稳定。如前文所述, 两种方法模型均采用 L2 损失定义类内散度, 差异在于类间散度的定义不同。相对而言, MFA-L2/L1 采用 L1 损失定义类间散度, 具有更好的判别性。因此, 结合结果得知, 由于对类间散度采用 L1 损失, MFA-L2/L1 对类间样本赋予了更大的惩罚, 在本次实验不同维度和噪声类型下的性能优于 MFA 方法。

3.3 FEI 数据集上的实验

基于 MFA-L2/L1 与 LDA-NgL1 两种非贪婪算法在 UCI 数据集上的稳定性, 在公开的 FEI 人脸识别数据集 (Thomaz 和 Giral di, 2010) 上对两种方法进行比较。FEI 人脸数据集包括 200 个人 (男女各 100 人) 在不同角度和光照环境下的各 14 幅 RGB 图像。实验固定选取 6 个场景共 1 200 幅图像 (均为正面人脸) 进行训练和测试。

实验将图像转换为 32×32 像素的灰度图像。前面的实验已考察了不同噪声程度和噪声类型的情况, 此处只针对一种噪声情况下的性能进行考察。参考 Chen 等人 (2018) 的方式, 对每个类别随机选取 2 幅图像, 在图像中随机选取 25% 的像素点, 将其从原始值 a 翻转至 $255 - a$ 作为像素噪声。每个类别随机选取 3 幅干净图像作为测试样本, 其余图像作为训练样本。基于前面的实验, MFA-L2/L1 和 LDA-NgL1 均设置相同的矩阵初始化方式和迭代终止条件。为避免实验结果的随机性, 训练集和测试集的划分过程重复 10 次, 对比投影矩阵在 100 ~ 250 维之间的性能变化, 平均结果如图 4 所示。

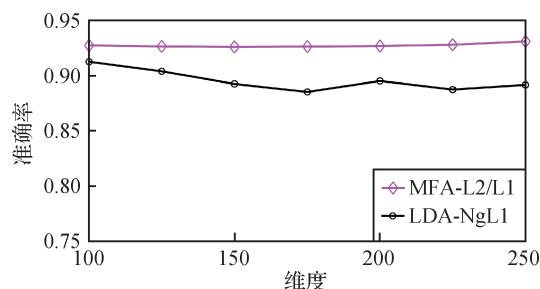


图4 在 FEI 数据集上 MFA-L2/L1 与 LDA-NgL1 算法的准确率对比结果

Fig. 4 Comparison of accuracy between MFA-L2/L1 and LDA-NgL1 algorithms on FEI dataset

由图 4 可知, LDA-NgL1 的性能随维度增加发生了一定波动; MFA-L2/L1 的性能随维度增加呈现上升趋势, 在不同维度下始终高于 LDA-NgL1。由于两种非贪婪算法的区别主要在于对类内散度的定义不同, 相对 LDA-NgL1 而言, MFA-L2/L1 采用 L2 损失定义类内散度, 在本次实验中提升了算法的判别性。

3.4 5 个人脸数据集上的综合实验

为了进一步全面对比 MFA-L2/L1 与同类方法的性能, 除了选取 LDA-NgL1、MFA-NgL1、LDA-L1、MFA 和 DNS 以外, 还选取了经典的度量学习方法 KISSME (keep it simple and straightforward metric learning) (Köstinger 等, 2012)。其中, DNS 主要通过计算类内样本的零空间来学习变换矩阵, KISSME 主要依赖类内样本和类间样本的两个协方差矩阵来计算度量矩阵, 两种方法的投影矩阵维数固定, 无需设定。因此, 实验均只在固定维度下进行对比。实验设置的初始化和终止条件与前面一致。为避免实验结果的随机性, 训练集和测试集的划分过程依然重复 10 次。由于第 3.3 节实验未进行 PCA 处理, 算法性能并无较大差异, 实验依然采用 PCA 保留 95% 的贡献率, 对所有样本进行前处理。5 个人脸数据集的信息和加噪方式如下:

1) Senthil 数据集。该数据集由 5 个人的 80 幅人脸图像构成, 每人各 16 幅。实验时, 先将每幅图像转换为 32×32 像素的灰度图像, 然后对每个类别随机选取 4 幅图像作为噪声图像, 并按第 3.3 节的方式加入 20% 的像素噪声。在每个类别中随机选取 6 幅干净图像和 2 幅噪声图像作为训练样本, 剩下的作为测试样本。投影矩阵的维度设置为 25。

2) Yale 数据集 (Belhumeur 等, 1997)。该数据集采集了 15 个人的 165 幅人脸图像, 每人各 11 幅。实验时, 先将每幅图像转换为 32×32 像素的灰度图像, 然后对每个类别随机选取 3 幅图像, 加入面积为 15% 的随机块噪声。从每个类别中随机选取 5 幅干净图像作为测试样本, 剩下的噪声图像和干净图像作为训练样本。投影矩阵的维度设置为 40。

3) ORL (Olivetti Research Laboratory) 数据集 (Samaria 和 Harter, 1994)。该数据集采集了 40 个人的 400 幅人脸图像, 每人各 10 幅。实验时, 先将每幅图像转换为 33×28 像素的灰度图像, 然后对每个类别随机选取 3 幅图像, 加入 15% 的随机块噪声作为噪声样本。从每个类别中随机选取 5 幅干净图

像作为测试样本,剩下的噪声图像和干净图像作为训练样本。投影矩阵的维度设置为 35。

4) Caltech 数据集 (Weber, 1999)。该数据集采集了 31 个人的 450 幅人脸图像,其中 5 个类别只有 1 幅图像,其余每个类别包含 5 ~ 29 幅图像不等。实验时,先将每幅图像裁剪为 193×162 像素的灰度图像,然后转换为 32×27 像素。考虑数据集类别的不平衡性,将不超过 5 个样本的类别全部作为噪声样本,在其余类别中随机选取 5 幅图像作为噪声样本,并按第 3.3 节的方式加入 25% 的像素噪声。从每个类别中随机选取一半图像作为训练样本,剩下一半作为测试样本。投影矩阵的维度设置为 60。

5) UMIST 数据集 (Graham 和 Allinson, 1998)。该数据集采集了 20 个人的 575 幅人脸图像,每个类别包含 19 ~ 48 幅不等。实验时,先将每幅图像转换为 32×32 像素的灰度图像,然后对每个类别随机选取 10 幅图像作为噪声样本,并按第 3.3 节的方式加入 20% 的像素噪声。从每个类别中随机选取占样本总数一半的干净图像作为测试样本,其余的

噪声图像和干净图像作为训练样本。投影矩阵的维度设置为 25。

5 个人脸数据集的实验结果如表 3 所示,在实验的噪声环境下, MFA-L2/L1 方法优于其他方法。在 MFA-L2/L1、LDA-NgL1、MFA-NgL1 和 LDA-L1 几个迭代算法中, LDA-L1 的性能较差,原因在于该方法属于贪婪式方法,采用逐向量求解算法使得该方法性能较不稳定。该现象从另一层面上验证了非贪婪方法的优势。对比 LDA-NgL1、MFA-NgL1 与 MFA 的差异,发现前两者并没有太大的优势,说明仅采用 L1 损失刻画类内散度和类间散度并未充分提升方法的判别性。DNS 方法性能不佳的原因在于数据集的零空间有限使得其表达能力有限。KISSME 方法由于在全局空间学习度量矩阵,未着重突出有效的低维结构,使得其性能也表现不佳,这一缺点也在 Liao 等人 (2015) 的文献中有所印证。对比不同的模型和方法,验证了 MFA-L2/L1 方法利用不同的损失刻画类内散度和类间散度,在实验的噪声数据上具有一定的鲁棒性和有效性。

表 3 在 5 个人脸数据集上不同方法的准确率对比结果
Table 3 Comparison of accuracy among different methods on five face datasets

方法	Senthil	Yale	ORL	Caltech	UMIST
	维度 = 25	维度 = 40	维度 = 35	维度 = 60	维度 = 25
DNS (Zhang 等, 2016b)	63.5	77.87	—	83.81	67.44
KISSME (Köstinger 等, 2012)	76.0	80.27	—	94.51	97.44
LDA-L1 (Zhong 和 Zhang, 2013)	77.25	79.07	90.5	91.53	78.18
LDA-NgL1 (Liu 等, 2017)	83.0	86.67	93.2	96.65	96.56
MFA-NgL1	82.0	88.67	93.3	96.56	96.74
MFA (Xiong 等, 2014)	83.0	87.87	92.4	96.65	97.89
本文 (MFA-L2/L1)	84.0	90.27	94.1	97.02	98.11

注:加粗字体表示每列最优结果;“—”表示准确率低于 60%,不予报告。

4 结 论

针对特征噪声对度量学习方法的影响,本文提出了一种基于 L2/L1 损失的非贪婪的鲁棒性度量学习算法,并针对非凸目标函数推导了高效的求解算法。本文算法分别考虑噪声对类内样本和类间样本不同优化方向的影响,使所学投影矩阵具有较好

的判别性。针对优化目标非凸带来的求解困难,推导了一种迭代的非贪婪求解算法。在 5 个 UCI 数据集和 7 个人脸数据集上进行了仿真加噪实验,结果表明,本文算法相对其他算法具有较好的抗噪性和判别性。

在监督和弱监督学习中,样本主要由特征和标签相关信息组成。相对特征噪声,标签噪声下的观测分布偏离真实分布更为严重。而基于样本三元组

的局部方法已成为度量学习方法中最为重要的一类。因此,下一步主要针对标签噪声下的度量学习进行研究,针对标签噪声设计局部的鲁棒性度量学习方法,以提升机器学习方法的性能。

参考文献 (References)

- An L T H and Tao P D. 2005. The DC (difference of convex functions) programming and DCA revisited with DC models of real world non-convex optimization problems. *Annals of Operations Research*, 133(1/4): 23-46 [DOI: 10.1007/s10479-004-5022-1]
- Belhumeur P N, Hespanha J P and Kriegman D J. 1997. Eigenfaces vs. fisherfaces: recognition using class specific linear projection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7): 711-720 [DOI: 10.1109/34.598228]
- Bellet A, Habrard A and Sebban M. 2014. A survey on metric learning for feature vectors and structured data [EB/OL]. [2019-12-13]. <https://arxiv.org/pdf/1306.6709.pdf>
- Chen W J, Li C N, Shao Y H, Zhang J and Deng N Y. 2018. Robust L_1 -norm multi-weight vector projection support vector machine with efficient algorithm. *Neurocomputing*, 315: 345-361 [DOI: 10.1016/j.neucom.2018.04.083]
- Dai G X, Xie J and Fang Y. 2018. Deep correlated holistic metric learning for sketch-based 3D shape retrieval. *IEEE Transactions on Image Processing*, 27(7): 3374-3386 [DOI: 10.1109/TIP.2018.2817042]
- Dua D and Graff C. 2019. UCI machine learning repository [DB/OL]. [2019-12-13]. <http://archive.ics.uci.edu/ml>
- Graham D B and Allinson N M. 1998. Characterising virtual eigensignatures for general purpose face recognition//Wechsler H, Phillips P J, Bruce V, Soulié F F and Huang T S, eds. *Face Recognition*. Berlin Heidelberg: Springer: 446-456 [DOI: 10.1007/978-3-642-72201-1_25]
- Green B F. 1952. The orthogonal approximation of an oblique structure in factor analysis. *Psychometrika*, 17(4): 429-440 [DOI: 10.1007/BF02288918]
- Guillaumin M, Verbeek J and Schmid C. 2009. Is that you? Metric learning approaches for face identification//Proceedings of the 12th IEEE International Conference on Computer Vision. Kyoto: IEEE: 498-505 [DOI: 10.1109/ICCV.2009.5459197]
- Hoi S C H, Liu W, Lyu M R and Ma W Y. 2006. Learning distance metrics with contextual constraints for image retrieval//Proceedings of 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. New York: IEEE: 2072-2078 [DOI: 10.1109/CVPR.2006.167]
- Hu J L, Lu J W and Tan Y P. 2014. Discriminative deep metric learning for face verification in the wild//Proceedings of 2014 IEEE Conference on Computer Vision and Pattern Recognition. Columbus: IEEE: 1875-1882 [DOI: 10.1109/CVPR.2014.242]
- Hu J L, Lu J W and Tan Y P. 2016. Deep metric learning for visual tracking. *IEEE Transactions on Circuits and Systems for Video Technology*, 26(11): 2056-2068 [DOI: 10.1109/TCSVT.2015.2477936]
- Huo Q R, Lu Y, Liu Y and Chao J B. 2017. Metric learning for tracking utilizing phase congruency. *Journal of Image and Graphics*, 22(3): 297-304 (霍其润, 陆耀, 刘羽, 巢进波. 2017. 结合相位一致性的度量学习跟踪. *中国图象图形学报*, 22(3): 297-304) [DOI: 10.11834/jig.20170303]
- Ke Q F and Kanade T. 2005. Robust L_1 norm factorization in the presence of outliers and missing data by alternative convex programming//Proceedings of 2005 IEEE Computer Society Conference on Computer vision and Pattern Recognition. San Diego: IEEE: 739-746 [DOI: 10.1109/CVPR.2005.309]
- Köstinger M, Hirzer M, Wohlhart P, Roth P M and Bischof H. 2012. Large scale metric learning from equivalence constraints//Proceedings of 2012 IEEE Conference on Computer vision and Pattern Recognition. Providence: IEEE: 2288-2295 [DOI: 10.1109/CVPR.2012.6247939]
- Kwak N. 2008. Principal component analysis based on L_1 -norm maximization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(9): 1672-1680 [DOI: 10.1109/TPAMI.2008.114]
- Liao S C, Hu Y, Zhu X Y and Li S Z. 2015. Person re-identification by local maximal occurrence representation and metric learning//Proceedings of 2015 IEEE Conference on Computer vision and Pattern Recognition. Boston: IEEE: 2197-2206 [DOI: 10.1109/CVPR.2015.7298832]
- Liong V E, Lu J W, Tan Y P and Zhou J. 2017. Deep coupled metric learning for cross-modal matching. *IEEE Transactions on Multimedia*, 19(6): 1234-1244 [DOI: 10.1109/TMM.2016.2646180]
- Little M A, McSharry P E, Roberts S J, Costello D A and Moroz I M. 2007. Exploiting nonlinear recurrence and fractal scaling properties for voice disorder detection. *Biomedical Engineering Online*, 6: #23 [DOI: 10.1186/1475-925X-6-23]
- Liu Y, Gao Q X, Miao S, Gao X B, Nie F P and Li Y S. 2017. A non-greedy algorithm for L_1 -norm LDA. *IEEE Transactions on Image Processing*, 26(2): 684-695 [DOI: 10.1109/TIP.2016.2621667]
- Lu J W, Hu J L and Zhou J. 2017. Deep metric learning for visual understanding: an overview of recent advances. *IEEE Signal Processing Magazine*, 34(6): 76-84 [DOI: 10.1109/MSP.2017.2732900]
- Martinez A M and Kak A C. 2001. PCA versus LDA. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(2): 228-233 [DOI: 10.1109/34.908974]
- Moutafis P, Leng M J and Kakadiaris I A. 2017. An overview and empirical comparison of distance metric learning methods. *IEEE Transactions on Cybernetics*, 47(3): 612-625 [DOI: 10.1109/TCYB.

2016. 2521767]
- Nie F P, Huang H, Ding C, Luo D J and Wang H. 2011. Robust principal component analysis with non-greedy L1-norm maximization//Proceedings of the 22nd International Joint Conference on Artificial Intelligence. Barcelona, Spain; AAAI: 1433-1438
- Samaria F S and Harter A C. 1994. Parameterisation of a stochastic model for human face identification//Proceedings of 1994 IEEE Workshop on Applications of Computer Vision. Sarasota; IEEE: 138-142 [DOI: 10.1109/ACV.1994.341300]
- Shi Y, Bellet A and Sha F. 2014. Sparse compositional metric learning//Proceedings of the 28th AAAI Conference on Artificial Intelligence. Québec; AAAI: 2078-2084
- Thomaz C E and Giraldo G A. 2010. A new ranking method for principal components analysis and its application to face image analysis. Image and Vision Computing, 28(6): 902-913 [DOI: 10.1016/j.imavis.2009.11.005]
- Wang H, Nie F P and Huang H. 2014. Robust distance metric learning via simultaneous L1-norm minimization and maximization//Proceedings of the 31st International Conference on Machine Learning. Beijing; JMLR: 1836-1844
- Weber M. 1999. Frontal face dataset [DB/OL]. [2019-12-13]. <https://wenku.baidu.com/view/f1fc60ec81c758f5f61f672a.html>
- Weinberger K Q and Saul L K. 2009. Distance metric learning for large margin nearest neighbor classification. Journal of Machine Learning Research, 10: 207-244
- Xiang S M, Nie F P and Zhang C S. 2008. Learning a Mahalanobis distance metric for data clustering and classification. Pattern Recognition, 41(12): 3600-3612 [DOI: 10.1016/j.patcog.2008.05.018]
- Xing E P, Ng A Y, Jordan M I and Russell S. 2002. Distance metric learning with application to clustering with side-information//Proceedings of Advances in Neural Information Processing Systems. Vancouver; MIT Press: 521-528
- Xiong F, Gou M R, Camps O and Szaier M. 2014. Person re-identification using kernel-based metric learning methods//Proceedings of the 13th European Conference on Computer Vision. Zurich; Springer: 1-16 [DOI: 10.1007/978-3-319-10584-0_1]
- Xu D, Yan S C, Tao D C, Lin S and Zhang H J. 2007. Marginal fisher analysis and its variants for human gait recognition and content-based image retrieval. IEEE Transactions on Image Processing, 16(11): 2811-2821 [DOI: 10.1109/TIP.2007.906769]
- Ye Q L, Fu L Y, Zhang Z, Zhao H H and Naiem M. 2018. Lp- and Ls-norm distance based robust linear discriminant analysis. Neural Networks, 105: 393-404 [DOI: 10.1016/j.neunet.2018.05.020]
- Zhang J and Zhao X. 2017. Global-local metric learning for person re-identification. Journal of Image and Graphics, 22(4): 472-481 (张晶, 赵旭. 2017. 整合全局-局部度量学习的人体目标再识别. 中国图象图形学报, 22(4): 472-481) [DOI: 10.11834/jig.20170407]
- Zhang L, Wei W, Tian C N, Li F and Zhang Y N. 2016a. Exploring structured sparsity by a reweighted Laplace prior for hyperspectral compressive sensing. IEEE Transactions on Image Processing, 25(10): 4974-4988 [DOI: 10.1109/TIP.2016.2598652]
- Zhang L, Xiang T and Gong S G. 2016b. Learning a discriminative null space for person re-identification//Proceedings of 2016 IEEE Conference on Computer Vision and Pattern recognition. Las Vegas; IEEE: 1239-1248 [DOI: 10.1109/CVPR.2016.139]
- Zhang S P, Qi Y K, Jiang F, Lan X Y, Yuen P C and Zhou H Y. 2018. Point-to-set distance metric learning on deep representations for visual tracking. IEEE Transactions on Intelligent Transportation Systems, 19(1): 187-198 [DOI: 10.1109/TITS.2017.2766093]
- Zhong F J and Zhang J S. 2013. Linear discriminant analysis based on L1-norm maximization. IEEE Transactions on Image Processing, 22(8): 3018-3027 [DOI: 10.1109/TIP.2013.2253476]

作者简介



曾凡霞, 1988年生, 女, 博士研究生, 主要研究方向为机器学习、度量学习、重采样。

E-mail: zengfanxia2014@ia.ac.cn



张文生, 通信作者, 男, 研究员, 博士生导师, 主要研究方向为人工智能、机器学习、大数据知识挖掘。

E-mail: zhangwenshengia@hotmail.com