

# 基于局部结构保持的自适应有序回归学习

王 芃, 吕 静, 沈华乐

(南京师范大学计算机科学与技术学院, 江苏 南京 210046)

[摘要] 有序回归学习是一种在训练模型过程中保持数据间序关系的机器学习方法, 在图像分类等领域有着广泛的应用. 现有的有序回归模型通过先验知识获得了更优的性能, 但是它们没有考虑数据内的局部结构信息. 本文在有序回归学习的同时保持局部结构信息, 并嵌入图像空间距离度量信息, 提出了一种基于局部结构保持的自适应有序回归方法 (SaLSP-LDLOR). 通过对局部保持矩阵进行模糊自适应处理, 获得了更好的鲁棒性. 实验结果表明, SaLSP-LDLOR 在有序图像分类的场景下具有更优的性能和良好的鲁棒性.

[关键词] 图像分类, 有序回归, 结构保持, 模糊自适应

[中图分类号] TP391 [文献标志码] A [文章编号] 1001-4616(2019)02-0009-08

## Improved Adaptive Ordinal Regression Learning Based on Locality Structure Preserving

Wang Peng, Lü Jing, Shen Huale

(School of Computer Science and Technology, Nanjing Normal University, Nanjing 210046, China)

**Abstract:** The ordinal regression learning is a kind of machine learning which preserves the order relations between data. It is widely used in image classification and other fields. Usually, there is some prior knowledge in the ordinal regression model, but the local structure information is not considered. Exploring such information can help to improve the effectiveness of classifiers. In this paper, we propose an improved adaptive ordinal regression method which is based on locality preserving structure (SaLSP-LDLOR), and the newly developed method considers embedding the spatial distance measurement information of the image. Experimental results with the standard data sets verify the effectiveness and the robustness of the proposed method.

**Key words:** image classification, ordinal regression, structure preserving, fuzzy adaptive

有序回归<sup>[1-2]</sup>指的是利用标签之间的自然有序构建的一种分类模式. 在实际应用中, 有序类标不仅有等级差异而且有一定相关性. 例如学生对老师的评价调查可分为 {差, 中, 良好, 好, 优秀} 5 个等级, 学生根据老师的课堂讲解、板书、作业批改、仪态举止等方面, 对老师进行评价. 通过 5 个等级的区分, 既反映了老师的受欢迎程度, 也反映了老师授课过程的差异度, 因此结合有序和回归特点的有序回归学习方法, 可以构建出更为精准的模型. 目前, 有序回归已经在医疗<sup>[3-4]</sup>, 年龄估计<sup>[5]</sup>, 信用评分<sup>[6]</sup>, 人脸识别<sup>[7]</sup>等众多领域得到了广泛应用.

目前, 已有众多有序回归方法相继被提出, 根据 Gutiérrez 等人<sup>[8]</sup>的研究, 有序回归方法可分为朴素方法、有序二分类方法和阈值方法. 朴素方法也称直接方法, 它将有序回归问题看作标准的分类或回归问题, 但忽略了类间有序关系, 从而影响分类准确性. Kramer 等人<sup>[9]</sup>将目标变量转换成实值连续变量, 利用回归树解决有序分类问题, 但由于类间距离未知, 难以找到合适的映射函数. Kotsiantis 等人<sup>[10]</sup>通过构建代价矩阵, 考虑代价敏感的有序回归, 但在没有先验知识的情况下, 难以选取合适的代价矩阵. 有序二分类方法是有序分类问题分解为一组二分类问题. 通常有两种分解策略, 分别是为每个子问题分别构建

收稿日期: 2018-11-23.

基金项目: 国家自然科学基金项目 (61876087)、赛尔网络下一代互联网技术创新项目资助 (NGII20170524)、江苏省高校自然科学研究项目 (18KJB520027).

通讯联系人: 吕静, 博士研究生, 讲师, 研究方向: 机器学习、模式识别. E-mail: jinglv@njnu.edu.cn

一个单输出的模型 (Multiple Model) 或将所有子问题集成一个多输出单模型 (Multiple-Output Single Model) [11-13]. Frank 和 Hall [14] 利用 C4.5 作为二分类器, 根据类标概率, 将多个独立的二分类器进行组合, 但可能出现负概率估计. Waegeman 等人 [15] 在此基础上给每个二分类器加上权重, 提高分类性能, 但可能导致某些测试类难以判别. 在文 [11] 中, 将类标分成 3 类, 为每个类构建分类器, 提高分类性能. 神经网络的出现, 可以以无缝方式处理多个响应. Deng 等人 [16] 将 ELMs 用于有序回归, 得到了好的泛化效果. 阈值方法是基于朴素方法, 通过训练阶段估计类标间距离, 将训练类划分成有序间隔. 通过训练学得一个映射  $f(\mathbf{x})$  和一系列阈值  $\mathbf{b} = (b_1, b_2, \dots, b_{Q-1}) \in R^{Q-1}$ , 其中  $b_1 < b_2 < \dots < b_{Q-1}$  在预测阶段, 将测试样本的输出值  $f(\mathbf{x}')$  与阈值逐一比较, 确定落在哪个区间, 得到相应的类别. SVM 模型被广泛应用于有序回归, 最早 Herbrich 等人 [17] 在有序回归方法中使用 SVM, 提出 Ranking SVM, Chu 等人 [18] 通过对阈值加上有序约束, 扩展 SVM, 提出两种不同算法, 处理有序分类问题, 但 SVM 方法通常计算复杂度高. Tian 等 [19] 结合图像空间信息提出新的有序回归方法, 由于阈值之间相对独立的优化过程无法保证全局最优解, 又提出有序最近中心投影回归方法 (OrNCP) [20].

针对上述所提有序回归方法的不足, 本文提出一种基于局部保持的有序回归方法 (LSP-LDLOR), 并将其扩展到基于局部保持的自适应有序回归方法 (SaLSP-LDLOR), 应用于图像分类. 本文的主要贡献有 3 个方面:

- (1) 通过拉普拉斯矩阵将数据的局部结构信息加入有序回归学习中, 以此增强投影向量的判别能力.
- (2) 对局部保持矩阵进行模糊自适应处理, 增强有序学习的鲁棒性.
- (3) 嵌入图像空间距离度量, 利用先验信息提升有序图像场景下的分类性能.

## 1 相关工作

### 1.1 图像欧氏距离嵌入的有序回归线性判别学习

Tian 等 [21] 将图像欧式距离度量 IMED 嵌入有序回归线性判别模型 (LDLOR) 中, 提出了一种基于图像欧氏距离嵌入的有序回归线性判别方法 (IMED-LDLOR) 应用于图像分类. 其中, IMED 是一种图像欧式距离度量方法. LDLOR 方法的核心是在对数据进行投影的同时最大程度地保持它们之间的序关系, 即最小化各类的类内离散度, 同时最大化序相邻的类间距离.

令  $(\mathbf{x}_i, y_i) \in R^d \times R, i=1, 2, \dots, N$  表示有  $N$  个维度为  $d$  的样本的训练数据集, 其中  $\mathbf{x}_i \in R^d$  表示第  $i$  个样本,  $y_i \in \{1, 2, \dots, K\}$  表示其对应的类别标记, 一共有  $K$  个有序类别, IMED-LDLOR 方法的目标函数如下:

$$\begin{aligned} \min_{\mathbf{w}, \rho} \quad & \mathbf{w}^T \mathbf{S}_w^{\text{IMED}} \mathbf{w} - C\rho \\ \text{s.t.} \quad & \mathbf{w}^T (\mathbf{m}_{k+1}^{\text{IMED}} - \mathbf{m}_k^{\text{IMED}}) \geq \rho, \quad k=1, 2, \dots, K-1, \end{aligned} \quad (1)$$

式中,  $\mathbf{w} \in R^d$  表示投影方向向量,  $\rho$  表示序相邻的两个类中样本均值之间的间隔, 当  $\rho > 0$  时, 类别的序关系得以保持,  $C$  表示惩罚系数,  $\mathbf{S}_w^{\text{IMED}}$  表示基于 IMED 度量的类内离散度矩阵, 定义如下:

$$\mathbf{S}_w^{\text{IMED}} = \frac{1}{N} \sum_{k=1}^K \sum_{\mathbf{x} \in X_k} (G^{1/2} \mathbf{x} - G^{1/2} \mathbf{m}_k) (G^{1/2} \mathbf{x} - G^{1/2} \mathbf{m}_k)^T \quad (2)$$

$\mathbf{m}_k^{\text{IMED}}$  表示基于 IMED 度量的第  $k$  类样本的中心,  $N_k$  表示第  $k$  类样本的规模,  $K$  表示类别数量.

$$\mathbf{m}_k^{\text{IMED}} = \frac{1}{N_k} \sum_{\mathbf{x} \in X_k} G^{1/2} \mathbf{x}, \quad k=1, 2, \dots, K-1 \quad (3)$$

式(1)分别实现了最小化同类数据的方差以及最大化序相邻的类中心之间的距离. 然而, IMED-LDLOR 只考虑了图像的空间结构信息, 并未考虑数据的局部结构信息, 所以存在一定的局限性. 因此文本提出了局部保持的有序回归学习方法.

### 1.2 局部保持投影方法

局部保持投影 (LPP) 算法 [22] 是将训练样本投影到一个低维空间的同时保存原始样本点的局部信息. 给定数据集  $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_m\}$ ,  $\mathbf{x}_i \in R^d$ , LPP 即寻找一个最合适的投影矩阵  $\mathbf{A}$ , 将  $d$  维数据点映射到  $f$  维空间, 得到相应的数据集  $\mathbf{Y} = \{\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_m\}$ ,  $\mathbf{y}_i \in R^f, \mathbf{y}_i = \mathbf{A} \mathbf{x}_i$ , 使得  $\frac{1}{2} \sum_{ij} \|\mathbf{y}_i - \mathbf{y}_j\|^2 w_{ij}$  最小. 常见局部

保持投影方法主要包含以下步骤,(1)构建邻接图:通过定义有  $m$  个顶点的图  $G$ ,如果  $\mathbf{x}_i$  和  $\mathbf{x}_j$  是近邻的,则利用  $\varepsilon$  邻域或  $k$  近邻方法在节点  $i$  和节点  $j$  之间添加相连边.一旦得到这样的邻接矩阵,LPP 会在选择投影时最大程度地保留它.(2)构建权值矩阵:主要通过热核法或简单构造法构造  $\mathbf{W}$  矩阵.  $\mathbf{W}_{ij}$  表示顶点  $i$  和顶点  $j$  之间边的权值.(3)寻找投影矩阵:计算  $\mathbf{X}\mathbf{L}\mathbf{X}^T\mathbf{a}=\lambda\mathbf{X}\mathbf{D}\mathbf{X}^T\mathbf{a}$  广义特征向量问题.其中, $\mathbf{D}$  是对角矩阵, $\mathbf{D}_{jj}=\sum_i \mathbf{W}_{ij}$ , $\mathbf{L}$  是拉普拉斯矩阵, $\mathbf{L}=\mathbf{D}-\mathbf{W}$ .虽然 LPP 能够有效地保持数据的流形结构,但是近邻图的构建对算法的识别效果起着重要的作用,参数的选择对于识别效果会造成一定影响,因此本文提出了基于局部保持的自适应有序回归学习方法.

## 2 基于局部保持的有序回归学习(LSP-LDLOR)

### 2.1 LSP-LDLOR 模型构建

针对 IMED-LDLOR 方法的不足,本文提出一种基于局部保持的有序回归学习方法 LSP-LDLOR.该方法学习数据的局部结构信息,使得在高维空间相邻的数据点在降维之后也是保持相邻的,以此增强投影向量的判别能力.其原始模型定义如下:

$$\begin{aligned} \min_{\mathbf{w}, \rho} \quad & \mathbf{w}^T \mathbf{S}_w \mathbf{w} + \lambda_1 \sum_{i=1}^N \sum_{j=1}^N (\mathbf{w}^T \mathbf{x}_i - \mathbf{w}^T \mathbf{x}_j)^2 \mathbf{A}_{ij} - C\rho, \\ \text{s.t.} \quad & \mathbf{w}^T (\mathbf{m}_{k+1} - \mathbf{m}_k) \geq \rho, \quad k = 1, 2, \dots, K-1. \end{aligned} \quad (4)$$

目标函数的第二项为局部保持项, $\mathbf{A}$  表示数据集加权图中边的权值矩阵,距离越近权值越大,因此可以约束相邻点在映射后的空间依然相邻.  $\lambda_1$  表示调节局部保持的参数.根据简单的代数公式可以推导如下:

$$\frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N (\mathbf{w}^T \mathbf{x}_i - \mathbf{w}^T \mathbf{x}_j)^2 \mathbf{A}_{ij} = \sum_{i=1}^N \mathbf{w}^T \mathbf{x}_i \mathbf{D}_{ii} \mathbf{x}_i^T \mathbf{w} - \sum_{i=1}^N \sum_{j=1}^N \mathbf{w}^T \mathbf{x}_i \mathbf{A}_{ij} \mathbf{x}_j^T \mathbf{w} = \mathbf{w}^T \mathbf{X} (\mathbf{D} - \mathbf{A}) \mathbf{X}^T \mathbf{w} = \mathbf{w}^T \mathbf{X} \mathbf{L} \mathbf{X}^T \mathbf{w}, \quad (5)$$

式中, $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N\}$ ,  $\mathbf{D}$  是对角矩阵, $\mathbf{D}_{jj} = \sum_i \mathbf{A}_{ij}$ ,  $\mathbf{L}$  表示拉普拉斯矩阵, $\mathbf{L} = \mathbf{D} - \mathbf{A}$ ,于是,目标函数可以重写如下:

$$\begin{aligned} \min_{\mathbf{w}, \rho} \quad & \mathbf{w}^T \mathbf{S}_w \mathbf{w} + \lambda_1 \mathbf{w}^T \mathbf{X} \mathbf{L} \mathbf{X}^T \mathbf{w} - C\rho, \\ \text{s.t.} \quad & \mathbf{w}^T (\mathbf{m}_{k+1} - \mathbf{m}_k) \geq \rho, \quad k = 1, 2, \dots, K-1. \end{aligned} \quad (6)$$

### 2.2 LSP-LDLOR 模型求解

上述目标函数是含不等式约束条件的凸二次规划问题,可以定义拉格朗日等式如下:

$$L(\mathbf{w}, \rho, \alpha) = \mathbf{w}^T \mathbf{S}_w \mathbf{w} + \lambda_1 \mathbf{w}^T \mathbf{X} \mathbf{L} \mathbf{X}^T \mathbf{w} - C\rho - \sum_{k=1}^{K-1} \alpha_k \{ \mathbf{w}^T (\mathbf{m}_{k+1} - \mathbf{m}_k) - \rho \}, \quad (7)$$

式中,拉格朗日乘子  $\alpha_k \geq 0$ .

为了获得拉格朗日乘子,对式(7)求导如下:

$$\frac{\partial L}{\partial \mathbf{w}} = 2(\mathbf{S}_w + \lambda_1 \mathbf{X} \mathbf{L} \mathbf{X}^T) \mathbf{w} - \sum_{k=1}^{K-1} \alpha_k (\mathbf{m}_{k+1} - \mathbf{m}_k) = 0 \Rightarrow \mathbf{w} = \frac{1}{2} (\mathbf{S}_w + \lambda_1 \mathbf{X} \mathbf{L} \mathbf{X}^T)^{-1} \sum_{k=1}^{K-1} \alpha_k (\mathbf{m}_{k+1} - \mathbf{m}_k), \quad (8)$$

$$\frac{\partial L}{\partial \rho} = \sum_{k=1}^{K-1} \alpha_k - C = 0 \Rightarrow \sum_{k=1}^{K-1} \alpha_k = C, \quad (9)$$

优化问题变为:

$$\begin{aligned} \min_{\alpha} \quad & \sum_{k=1}^{K-1} \alpha_k (\mathbf{m}_{k+1} - \mathbf{m}_k)^T (\mathbf{S}_w + \lambda_1 \mathbf{X} \mathbf{L} \mathbf{X}^T)^{-1} \sum_{k=1}^{K-1} \alpha_k (\mathbf{m}_{k+1} - \mathbf{m}_k), \\ \text{s.t.} \quad & \alpha_k \geq 0, k = 1, 2, \dots, K-1, \\ & \sum_{k=1}^{K-1} \alpha_k = C. \end{aligned} \quad (10)$$

上述优化问题是一个含线性约束的凸二次规划(QP)问题,可以使用内点法、作用集法、共轭梯度法等方法解决.求解出  $\alpha$  后将其代入式(8)即可获得投影方向  $\mathbf{w}$ .

当样本维度大于样本数目的时候, $\mathbf{S}_w + \lambda_1 \mathbf{X} \mathbf{L} \mathbf{X}^T$  可能会出现奇异矩阵,此时可以使用正则化方法解决:

在矩阵  $S_w + \lambda_1 XLX^T$  的对角线上添加一个很小的常量  $\mu > 0$ , 得到  $S_w + \lambda_1 XLX^T = S_w + \lambda_1 XLX^T + \mu I$ , 其中  $I$  是一个单位矩阵<sup>[21-22]</sup>, 通过交叉验证的方法可以确定最优的  $\mu$  值.

### 2.3 LSP-LDLOR 算法

输入: 训练集  $X \in R^{N \times d}$ , 标签集  $Y \in R^N$ , 参数  $\lambda_1, C$ , 近邻数  $k$ ;

步骤 1 计算各个类别样本的均值  $m_k, k=1, 2, \dots, K-1$ , 类内离散度矩阵  $S_w$  以及拉普拉斯矩阵  $L$ ;

步骤 2 求解凸二次规划问题 (10), 得到拉格朗日乘子  $\alpha$ ;

步骤 3 代入式 (8) 解出投影向量  $w$ ;

输出: 投影向量  $w$ .

### 2.4 LSP-LDLOR 模型的分类

基于 2.3 节, 根据学习得到的投影方向  $w$ , 将 LSP-LDLOR 模型的分类方法定义如下:

$$f(x) = \min_{k \in \{1, \dots, K\}} \{k : w^T x - b_k < 0\},$$

式中,  $b_k$  定义如下:

$$b_k = \frac{1}{2} w^T (m_{k+1} + m_k), \quad (11)$$

或

$$b_k = \frac{w^T (N_{k+1} m_{k+1} + N_k m_k)}{N_{k+1} + N_k}, \quad (12)$$

$N_k$  表示第  $k$  类样本的数量.

## 3 基于自适应局部保持的有序回归学习 (SaLSP-LDLOR)

### 3.1 SaLSP-LDLOR 模型构建

为了进一步提高 LSP-LDLOR 方法的鲁棒性, 本文还对局部保持矩阵进行了模糊自适应处理, 得到 SaLSP-LDLOR 方法. 其模型定义如下:

$$\begin{aligned} \min_{w, \rho, S} & w^T S_w w + \lambda_1 w^T S w + \lambda_2 \|S - S_0\|_F^2 - C\rho, \\ \text{s.t.} & w^T (m_{k+1} - m_k) \geq \rho, \quad k=1, 2, \dots, K-1. \end{aligned} \quad (13)$$

目标函数的第三项表示允许局部保持矩阵  $S$  在  $S_0$  的周围自适应调整, 其中,  $S_0 = XLX^T$ ,  $\lambda_2$  表示控制局部保持矩阵模糊自适应的参数.  $\lambda_2$  越大, 约束越强, 即  $S$  的调整幅度越小.

### 3.2 SaLSP-LDLOR 模型求解

对于式 (13) 的目标问题, 我们可以采用交替迭代的方法获取最优解, 通过反复更新投影向量  $w$  和序类别间隔  $\rho$  以及更新局部保持矩阵  $S$  这两个过程, 直至收敛为止.

#### 3.2.1 更新投影向量 $w$ 和序类别间隔 $\rho$

固定  $S$ , 更新  $w$  和  $\rho$ , 初始时  $S = S_0$ . 将式 (13) 重新定义如下:

$$\begin{aligned} \min_{w, \rho} & w^T S_w w + \lambda_1 w^T S w + \lambda_2 \|S - S_0\|_F^2 - C\rho, \\ \text{s.t.} & w^T (m_{k+1} - m_k) \geq \rho, \quad k=1, 2, \dots, K-1, \end{aligned} \quad (14)$$

此时和 2.2 节优化过程类似, 可以得到

$$w = \frac{1}{2} (S_w + \lambda_1 S)^{-1} \sum_{k=1}^{K-1} \alpha_k (m_{k+1} - m_k), \quad (15)$$

最终优化问题变为:

$$\begin{aligned} \min f(\alpha) &= \sum_{k=1}^{K-1} \alpha_k (m_{k+1} - m_k)^T (S_w + \lambda_1 S)^{-1} \sum_{k=1}^{K-1} \alpha_k (m_{k+1} - m_k), \\ \text{s.t.} & \alpha_k \geq 0, k=1, 2, \dots, K-1, \\ & \sum_{k=1}^{K-1} \alpha_k = C. \end{aligned} \quad (16)$$

求解出  $\alpha$  后将其带入式 (15) 即可获得投影方向  $w$ , 易知  $\rho$  的取值与更新过程无关.



### 3.2.2 更新局部保持矩阵 $S$

固定  $w$  和  $\rho$ , 更新  $S$ . 将式(14)重新定义如下:

$$J(S) = w^T S_w w + \lambda_1 w^T S w + \lambda_2 \|S - S_0\|_F^2 - C_\rho. \quad (17)$$

上述问题是一个无约束凸优化问题, 根据 Fermat 定理, 直接找到使目标函数导数为 0 的解析解即可, 也可使用梯度下降或牛顿法等迭代的方法来使  $S$  沿负梯度方向逐步逼近极小值点.

$$\frac{\partial J}{\partial S} = \lambda_1 w w^T + 2\lambda_2 (S - S_0) = 0 \Rightarrow S = S_0 - \frac{\lambda_1}{\lambda_2} w w^T, \quad (18)$$

### 3.3 SaLSP-LDLOR 算法

输入: 训练集  $X \in \mathbb{R}^{N \times d}$ , 标签集  $Y \in \mathbb{R}^N$ , 参数  $\lambda_1, C$ , 近邻数  $k$ ;

步骤 1 计算各个类别样本的均值  $m_k, k=1, 2, \dots, K-1$ , 类内离散度矩阵  $S_w$  以及局部保持矩阵  $S_0$ , 初始化  $S=S_0$ ;

步骤 2 固定  $S$ , 更新  $w$  和  $\rho$ , 求解凸二次规划问题(16)得到拉格朗日乘子  $\alpha$ , 然后代入式(15)解出投影向量  $w$ ;

步骤 3 固定  $w$  和  $\rho$ , 更新  $S$ , 根据式(18)解出  $S$ ;

步骤 4 重复步骤 2 到步骤 3, 直到满足收敛条件或达到最大迭代次数为止;

输出: 投影向量  $w$ .

SaLSP-LDLOR 模型的分类方法与 2.3 节完全一致.

## 4 复杂度分析

LSP-LDLOR 算法的复杂度体现在计算一个  $N \times N$  大小矩阵的逆以及解决一个二次规划(QP)问题: 求矩阵的逆的复杂度为  $O(N^2 \log(N))$ ; QP 问题中的 Hessian 矩阵大小为  $(K-1) \times (K-1)$ , 因此求解复杂度为  $O((K-1)^3)$ . 综上, 由于  $K \ll N$ , 所以 LSP-LDLOR 算法的总体复杂度为  $O(N^2 \log(N))$ .

而 SaLSP-LDLOR 算法还需考虑迭代次数  $t$ , 所以 SaLSP-LDLOR 算法的复杂度为  $O(tN^2 \log(N))$ .

## 5 实验及结果分析

本文在两个有序图像数据集——人脸表情数据集 JAFFE 和人脸年龄数据集 FG-NET 上进行实验, 将图像样本统一归一化为  $16 \times 16$  大小, 以排除图像尺度不同对预测性能造成的影响, 然后提取图像的灰度值作为其特征.

实验采用分类正确率(Acc)作为评价指标来评估 LSP-LDLOR 和 SaLSP-LDLOR 模型的性能. 使用交叉验证的评估方法来确定模型的参数, 将数据集随机划分 20 次, 取运行结果的平均值作为最终实验结果.

其中, 邻接图的构建采用  $k$  近邻方法, 权值矩阵的构建采用余弦相似度度量. 参数设置采用网格搜索法, 模型 SaLSP-LDLOR 中的参数  $\lambda_1/\lambda_2$  比值应该设置得较小, 否则  $S$  的抖动较大, 难以收敛. 图 1 显示了在数据集 JAFFE 上当  $\lambda_1/\lambda_2$  比值较大时目标函数值的变化情况, 可以看到, 目标函数值一直在上下波动, 并以微弱的趋势逐渐减小.

### 5.1 JAFFE 人脸表情识别

本文的第一个数据集选用 JAFFE 数据集, 该数据集采集自 10 个日本女学生, 共 213 张人脸表情图像, 包含 7 种人脸表情, 分别是愤怒、厌恶、恐惧、高兴、中性脸、悲伤和惊奇. JAFFE 中的部分人脸表情图像如图 2 所示.

表 1 显示了基于 JAFFE 数据集的人脸表情识别实验结果. 其中, 标有加粗下划线的实验结果具有  $t$ -test 统计显著性( $p$ -value: 0.05). “#Nper”表示出自每类的训练样本数.

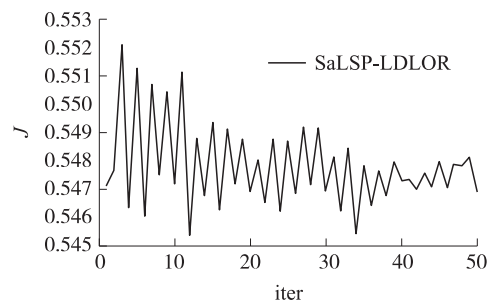


图 1 JAFFE 数据集  $\lambda_1/\lambda_2$  很大时目标函数值  
随迭代次数的变化

Fig. 1 The change of objective value with the increasing iterations when  $\lambda_1/\lambda_2$  in large enough of JAFFE

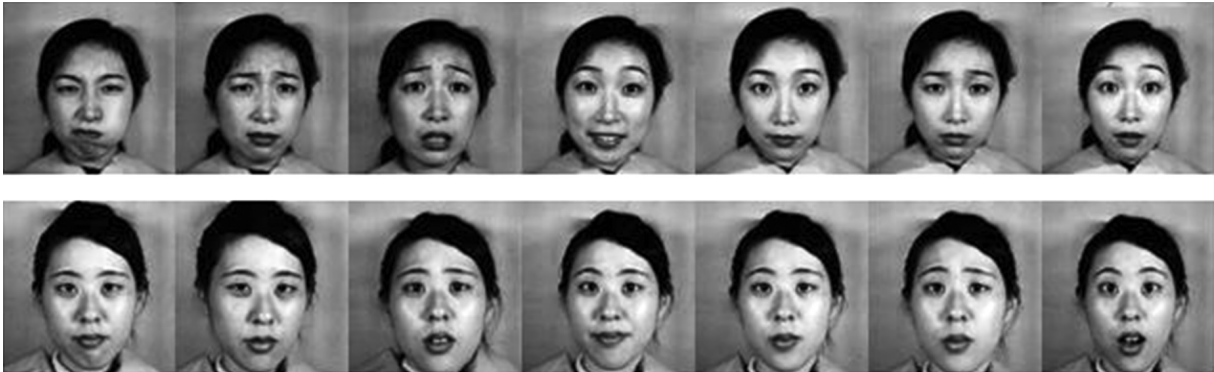


图 2 JAFFE 人脸表情样例

Fig. 2 Samples of facial expressions on JAFFE

表 1 不同 OR 方法在 JAFFE 上的 Acc 结果对比 (均值±标准差)

Table 1 Comparison of ACC results on JAFFE by different OR methods

#Nper	LDLOR	IMED-LDLOR	LSP-LDLOR	SaLSP-LDLOR
4	0.22±0.03	0.20±0.04	0.24±0.04	<b>0.25±0.03</b>
8	0.28±0.04	0.22±0.04	0.29±0.04	<b>0.30±0.04</b>
12	0.30±0.03	0.21±0.03	0.32±0.05	<b>0.32±0.04</b>
16	0.32±0.07	0.20±0.04	0.34±0.05	<b>0.34±0.04</b>
20	0.35±0.06	0.23±0.04	0.35±0.05	<b>0.35±0.05</b>
24	0.37±0.10	0.22±0.06	0.37±0.06	<b>0.37±0.05</b>

根据表 1,在每类的训练样本数#Nper 从 4 到 16 的情况下,LSP-LDLOR 方法和 SaLSP-LDLOR 方法较 LDLOR 方法和 IMED-LDLOR 方法在分类正确率上均能有大约两个百分点的提升,并且标准差保持不变或减少,只有在#Nper 为 12 的情况,标准差增大了一两个百分点,但是分类正确率也增大了一两个百分点,这表明数据的局部结构信息起到了作用. 然而,当#Nper 为 20 时,SaLSP-LDLOR 方法的分类准确率不变,标准差降低了 1 个百分点,当#Nper 为 24 时,SaLSP-LDLOR 方法的分类准确率不变,标准差降低了 5 个百分点,此时,SaLSP-LDLOR 方法的优越性并不明显,这与样本的随机划分情况有关. 此外,我们还可以发现 SaLSP-LDLOR 方法始终优于 LSP-LDLOR 方法,这是由于前者在后者的基础上增加了模糊自适应,因此增强了一定的鲁棒性.

图 3 是不同 OR 方法在数据集 JAFFE 上的分类准确率对比的直观表示. 可以看出,本文的方法 LSP-LDLOR 方法和 SaLSP-LDLOR 方法始终优于 LDLOR 方法和 IMED-LDLOR 方法,并且 SaLSP-LDLOR 方法性能更优.

5.2 FG-NET 人脸年龄识别

本文的第二个数据集选用 FG-NET 数据集,该数据集包含 1002 张人脸年龄图像,来自 82 个人,年龄跨度从 0 到 69 岁. 本文将图像样本划分为 8 个年龄段,分别是 0~1 岁、2~4 岁、5~8 岁、9~12 岁、13~16 岁、17~29 岁、30~43 岁、44~69 岁. FG-NET 中的部分人脸年龄图像如图 4 所示.

表 2 显示了基于 FG-NET 数据集的人脸年龄识别实验结果. 其中,标有下划线加粗的实验结果具有 *t*-test 统计显著性(*p*-value;0.05).

根据表 2,在每类的训练样本数#Nper 从 6 到 36 的情况下,LSP-LDLOR 方法和 SaLSP-LDLOR 方法较 LDLOR 方法和 IMED-LDLOR 方法在分类正确率上均能有显著提升,其中,在#Nper 为 6 到 24 时,提升了一两个百分点,效果并不明显,而在#Nper 为 30 和 36 时,分类正确率提升了四五个百分点. 这是因为在到

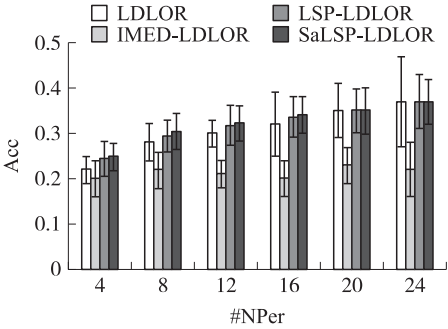


图 3 不同 OR 方法在 JAFFE 上的 Acc 结果对比

Fig. 3 Comparison of ACC results on JAFFE by different OR methods

达过拟合之前,随着训练样本数量的增加,分类器的学习性能会增强. 同样地,我们可以发现 SaLSP-LDLOR 方法的性能要优于 LSP-LDLOR 方法.



图 4 FG-NET 人脸年龄样例

Fig. 4 Samples of face age on FG-NET

表 2 不同 OR 方法在 FG-NET 上的 Acc 结果对比 (均值±标准差)

Table 2 Comparison of ACC results on FG-NET by different OR methods

#Nper	LDLOR	IMED-LDLOR	LSP-LDLOR	SaLSP-LDLOR
6	0.19±0.03	0.18±0.02	0.20±0.03	<b>0.20±0.02</b>
12	0.22±0.03	0.21±0.02	0.23±0.03	<b>0.24±0.03</b>
18	0.22±0.03	0.19±0.02	0.23±0.02	<b>0.24±0.03</b>
24	0.22±0.03	0.17±0.02	0.24±0.03	<b>0.25±0.03</b>
30	0.19±0.04	0.16±0.02	0.23±0.03	<b>0.24±0.03</b>
36	0.20±0.07	0.15±0.03	0.24±0.03	<b>0.25±0.03</b>

图 5 是不同 OR 方法在数据集 FG-NET 上的分类准确率对比的直观表示. 可以看出,本文的方法 LSP-LDLOR 方法和 SaLSP-LDLOR 方法始终优于 LDLOR 方法和 IMED-LDLOR 方法,并且 SaLSP-LDLOR 方法性能相对更优.

6 总结

本文提出了一种基于局部保持的有序回归方法 (LSP-LDLOR) 以及它的扩展自适应局部保持的有序回归方法 (SaLSP-LDLOR). 主要贡献包括:通过拉普拉斯矩阵将数据的局部结构信息加入有序回归学习中,以此增强投影向量的判别能力;对局部保持矩阵进行模糊自适应处理,增强了有序学习的鲁棒性;嵌入图像空间距离度量,利用先验信息提升针对有序图像的场景下的分类性能. 实验结果表明该方法在人脸表情识别和人脸年龄识别的图像分类任务中有很强的鲁棒性,尤其适合解决有序关系的图像分类问题.

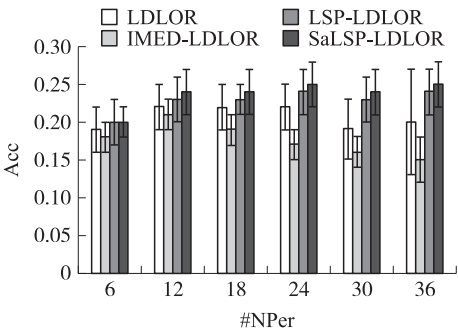


图 5 不同 OR 方法在 FG-NET 上的 Acc 结果对比

Fig. 5 Comparison of ACC results on FG-NET by different OR methods

[ 参考文献 ]

[ 1 ] MCCULLAGH P. Regression models for ordinal data[ J ]. Journal of the royal statistical society,1980,42(2):109–142.  
[ 2 ] NGUYEN B,CARLOS M,BERNARD DE B. Distance metric learning for ordinal classification based on triplet constraints[ J ]. Knowledge-based systems,2018,142:17–28.  
[ 3 ] BENDER R,ULRICH G. Ordinal logistic regression in medical research[ J ]. Journal of the royal college of physicians of london,1997,31(5):546–551.  
[ 4 ] PÉREZ O M,CRUZ R M,AYLLON T M D,et al. An organ allocation system for liver transplantation based on ordinal regres-

- sion[J]. Applied soft computing, 2014, 14(A): 88–98.
- [5] CHANG K Y, CHEN C S, HUNG Y P. Ordinal hyperplanes ranker with cost sensitivities for age estimation[C]//2011 IEEE conference on computer vision and pattern recognition, United States: Colorado springs, 2011.
- [6] DIKKERS H, LEON R. Support vector machines in ordinal classification: an application to corporate credit scoring[J]. Neural network world, 2005, 15(6): 491.
- [7] KIM M Y, VLADIMIR P. Structured output ordinal regression for dynamic facial emotion intensity prediction[M]. European conference on computer vision. Springer: Berlin, Heidelberg, 2010.
- [8] GUTIERREZ P, PEREZ O M, SANCHEZ M J, et al. Ordinal regression methods: survey and experimental study[J]. IEEE transactions on knowledge and data engineering, 2016, 28(1): 127–146.
- [9] KRAMER S, WIDMER G, PFAHRINGER B, et al. Prediction of ordinal classes using regression trees[J]. Fundamenta informaticae, 2001, 47(1/2): 1–13.
- [10] KOTSIANTIS S B, PANAGIOTIS P E. A cost sensitive technique for ordinal classification problems[C]//Third Hellenic conference on artificial intelligence. Greece: Samas, 2004.
- [11] PÉREZ O M, PEDRO A G, CÉSAR H M. Projection-based ensemble learning for ordinal regression[J]. IEEE transactions on cybernetics, 2014, 44(5): 681–694.
- [12] SÁNCHEZ M J, PEDRO A G, CESAR H M. Evolutionary ordinal extreme learning machine[C]//International Conference on Hybrid Artificial Intelligence Systems. Berlin, Heidelberg: Springer, 2013.
- [13] FERNÁNDEZ N F, ANNALISA R, SANTE C. Ordinal neural networks without iterative tuning[J]. IEEE transactions on neural networks and learning systems, 2014, 25(11): 2075–2085.
- [14] FRANK E, MARK H. A simple approach to ordinal classification[C]//European Conference on Machine Learning. Berlin, Heidelberg: Springer, 2001.
- [15] WAEGERMAN W, LUC B. An ensemble of weighted support vector machines for ordinal regression[J]. International journal of computer systems science and engineering, 2009, 3(1): 47–51.
- [16] DENG W Y, ZHENG Q H, LIAN S, et al. Ordinal extreme learning machine[J]. Neurocomputing, 2010, 74(1): 447–456.
- [17] HERBRICH R. Large margin rank boundaries for ordinal regression[J]. Advances in large margin classifiers, 2000: 115–132.
- [18] CHU W, SATHIYA K S. Support vector ordinal regression[J]. Neural computation, 2007, 19(3): 792–815.
- [19] TIAN Q, SONGCAN C. Comparative study among three strategies of incorporating spatial structures to ordinal image regression[J]. Neurocomputing, 2014, 136(20): 152–161.
- [20] TIAN Q, SONGCAN C. A novel ordinal learning strategy: ordinal nearest-centroid projection[J]. Knowledge-based systems, 2015, 88: 144–153.
- [21] WANG L W, YAN Z, JUFU F. On the euclidean distance of images[J]. IEEE transactions on pattern analysis and machine intelligence, 2005, 27(8): 1334–1339.
- [22] HE X F, PARTHA N. Locality preserving projections[J]. Advances in neural information processing systems, Cambridge: MIT Press, 2004, 16: 585–591.

[责任编辑: 陆炳新]