

# 鲁棒联合稀疏不相关回归

李宗然<sup>1</sup> 陈秀宏<sup>1,2</sup> 陆 贇<sup>1</sup> 邵政毅<sup>1</sup>

1 江南大学人工智能与计算机学院 江苏 无锡 214122

2 江苏省媒体设计与软件技术重点实验室 江苏 无锡 214122

(18851656843@163.com)

**摘 要** 常见的无监督特征选择方法考虑的只是选择具有判别性的特征,而忽略了特征的冗余性,并且没有考虑到小类问题,故而影响分类性能。基于此背景,提出鲁棒不相关回归算法。首先,对不相关回归进行研究,使用不相关正交约束,以便找出不相关但具有判别性的特征,不相关约束使得数据结构保持在 Stiefel 流形中,使模型具有封闭解,避免了传统的岭回归模型引发的可能的平凡解。其次,损失函数与正则化项使用  $L_{2,1}$  范数,保证模型的鲁棒性,得到具有稀疏性的投影矩阵;同时将小类问题考虑进去,使投影矩阵数量不受类别数的限制,得到足够多的投影矩阵,从而提升模型的性能。理论分析和多个数据集上的实验结果表明,所提出的方法比其他特征选择方法具有更好的性能。

**关键词:** 特征选择;鲁棒;联合;不相关;回归;小类问题

**中图法分类号** TP391.4

## Robust Joint Sparse Uncorrelated Regression

LI Zong-ran<sup>1</sup>, CHEN XIU-Hong<sup>1,2</sup>, LU Yun<sup>1</sup> and SHAO Zheng-yi<sup>1</sup>

1 School of Artificial Intelligence and Computer Science, Jiangnan University, Wuxi, Jiangsu 214122, China

2 Jiangsu Key Laboratory of Media Design and Software Technology, Wuxi, Jiangsu 214122, China

**Abstract** Common unsupervised feature selection methods only consider the selection of discriminative features, while ignoring the redundancy of features and failing to consider the problem of small classes, which affect the classification performance. Based on this background, a robust uncorrelated regression algorithm is proposed. First, research on uncorrelated regression, use uncorrelated orthogonal constraints to find irrelevant but discriminative features. Uncorrelated constraints keep the data structure in the Stiefel manifold, making the model have a closed solution, avoiding the possible trivial solutions caused by the traditional ridge regression model. Secondly, the loss function and the regularization term use the  $L_{2,1}$  norm to ensure the robustness of the model and obtain a sparse projection matrix. At the same time, the small class problem is taken into account, so that the number of projection matrices is not limited by the number of classes, and the result is enough projection matrices to improve the classification performance of the model. Theoretical analysis and experimental results on multiple data sets show that the proposed method has better performance than other feature selection methods.

**Keywords** Feature selection, Robust, Joint, Uncorrelated, Regression, Small-class

## 1 引言

高维数据已经广泛出现在模式识别、数据挖掘和计算机视觉等众多研究领域和应用中。但是,高维数据中存在着一一些不相关的冗余的特征,这导致学习任务的效率较低,分类性能也较差<sup>[1]</sup>。针对这些问题,特征选择<sup>[2]</sup>成为降维的主要方式之一。特征选择的目的是从高维数据中选择特征子集。另外,特征选择不改变数据原有的表示形式,并保持了数据变量的物理意义。

常见的有监督特征选择方法有:Fisher Score<sup>[3]</sup>、线性判别分析(Linear Discriminant Analysis, LDA)<sup>[4]</sup>、岭回归(Ridge Regression, RR)<sup>[5]</sup>、Nie 等对岭回归进行改进所提出的鲁棒

特征选择(Robust Feature Selection, RFS)<sup>[6]</sup>、Chen 等提出的用于二维监督特征选择的鲁棒图正则化稀疏矩阵回归方法<sup>[7]</sup>,以及 Mo 等提出的 RJSR<sup>[8]</sup>等,另外 Hou 等于 2017 年提出了稀疏矩阵回归(Sparse Matrix Regression, SMR)<sup>[9]</sup>方法,以解决现有大多数特征选择算法都需要先将样本数据向量化再进行学习所带来的问题,其性能也较优秀。为了扩大不同类别之间的余量并保留变换空间中样本的固有几何结构,Chen 等<sup>[10]</sup>针对二维监督特征选择提出动态图正则化和基于标签松弛的 SMR(Dynamic Graph Regularization and Label-based Relaxation, DGRLR-SMR)方法。考虑到矩阵数据中每个原始特征的相关性及其对回归模型最终估计的影响,Yu 等<sup>[11]</sup>引入了未知权重矩阵对矩阵数据中特征的相关性进行

编码,并提出了特征加权和基于图的矩阵回归(FWGMR)模型,用于图像监督特征选择。Yang 等<sup>[12]</sup>提出了基于自编码器和流形正则的结构保持无监督特征选择,适用于高维度数据。但是,由于标签的获取在很多情况下都非常困难,因此无监督的特征选择方法在现实中更加实用。常见的无监督特征选择方法有:将判别分析和范数最小化纳入无监督特征选择的联合框架中的无监督判别特征选择(UDFS)<sup>[13]</sup>、联合嵌入学习与稀疏回归(JELSR)<sup>[14]</sup>、Liu 等提出的全局和局部结构保存特征选择方法 GLSPFS<sup>[15]</sup>、Yuan 等提出的用于二维无监督特征选择的联合稀疏矩阵回归和非负谱分析(JSMRNS)<sup>[16]</sup>模型,以及 Li 等在 2019 年提出的广义不相关回归(Generalized Uncorrelated Regression, GURM)<sup>[17]</sup>。另外,很多特征选择方法都致力于选择更具有判别性的特征,而忽略了样本存在的冗余特征。除此之外,现有的很多无监督特征选择方法都忽略了小类问题<sup>[18]</sup>,所学习到的投影数量都受到类别数的限制,这在实际应用中会影响分类效果。

基于以上问题,本文提出了鲁棒不相关回归(Robust Uncorrelated Regression, RJUR)模型,该模型对 GURM 模型进行改进,对损失函数和正则化项使用联合  $L_{2,1}$  范数,使得在面对离群值以及噪声数据时,模型的鲁棒性能得到保证;另外 RJUR 模型还将小类问题考虑进去,使得投影矩阵数量不受类别数的限制,从而学习到足够的投影矩阵,提升模型的性能。在 6 个不同数据集上的实验效果表明, RJUR 方法优于大多数无监督特征选择方法。

## 2 相关工作

给定训练样本为  $\mathbf{X}=[\mathbf{x}_1; \mathbf{x}_2; \dots; \mathbf{x}_n] \in R^{n \times d}$ , 其中,  $n$  为样本个数,  $d$  为样本特征维数。

### 2.1 岭回归模型

经典的岭回归模型如下所示:

$$\min_{\mathbf{P}} \|\mathbf{XW} - \mathbf{Y}\|_F^2 + \lambda \|\mathbf{W}\|_F^2 \quad (1)$$

其中,  $\mathbf{W}$  为投影矩阵。

### 2.2 基于广义正交学习下的鲁棒联合稀疏回归(RJSR)

将联合稀疏性、鲁棒性和局部性保持特性与投影矩阵上的广义正交性约束相结合,并且考虑到小类问题, RJSR 的目标函数如下:

$$\min_{\mathbf{W}, \mathbf{F}, \mathbf{B}} \|\mathbf{XBA}^T + \mathbf{1h}^T - \mathbf{Y}\|_{2,1}^2 + \lambda \|\mathbf{B}\|_{2,1} \quad (2)$$

s. t.  $\mathbf{A}^T \mathbf{A} = \mathbf{I}, \mathbf{B}^T \mathbf{X}^T \mathbf{W} \mathbf{X} \mathbf{B} = \mathbf{I}$

其中,  $\mathbf{B} \in R^{d \times k}$  为投影矩阵,  $\mathbf{A} \in R^{c \times k}$  为辅助矩阵,  $\mathbf{W}$  是亲和矩阵,  $\mathbf{W}$  中的元素定义为:

$$W_{i,j} = \begin{cases} 1, & \text{若 } \mathbf{x}_i \text{ 是 } \mathbf{x}_j \text{ 的 } k \text{ 近邻或 } \mathbf{x}_j \text{ 是 } \mathbf{x}_i \text{ 的 } k \text{ 近邻} \\ 0, & \text{否则} \end{cases} \quad (3)$$

### 2.3 广义不相关回归模型(GURM)

岭回归模型很难直接应用到无监督学习中,但是 RJSR 模型能够应用到有监督学习中。当标签矩阵  $\mathbf{Y}$  未知时,它作为一个变量,需要在模型(1)中进行优化。在这种情况下,当  $\mathbf{W} = \mathbf{0}, \mathbf{h} = [1, 0, \dots, 0]^T$ , 并且  $\mathbf{Y} = [1, 0, \dots, 0]$  时,可能会触发潜在的平凡解。为了避免这种情况,并把回归模型应用于无监督特征选择中, Li 等提出了 GURM 模型,表示如下:

$$\min_{\mathbf{W}, \mathbf{F}, \mathbf{B}} \|\mathbf{XW} + \mathbf{1h}^T - \mathbf{F}\|_F^2 + \lambda \|\mathbf{W}\|_{2,1} \quad (4)$$

s. t.  $\mathbf{W} \in R_c^{d \times c}, \mathbf{F}^T \mathbf{F} = \mathbf{I}$

其中,  $\mathbf{W}$  是列满秩变换矩阵,  $\mathbf{F} = \{\mathbf{f}^1, \mathbf{f}^2, \dots, \mathbf{f}^n\} \in R^{n \times c}$  表示嵌入在输入数据空间的指标矩阵, 其中  $\mathbf{f}^c$  为  $c$  维流形中第  $i$  个样本的指标向量。对变换矩阵施加不相关约束  $\mathbf{W}^T \mathbf{S}_i \mathbf{W} = \mathbf{I}$ , 避免了平凡解, 其中  $\mathbf{S}_i = \mathbf{X}^T \mathbf{H} \mathbf{X}$  为全局散度矩阵,  $\mathbf{H} = \mathbf{I} - (1/n)\mathbf{1}\mathbf{1}^T$  为中心矩阵。

## 3 鲁棒联合不相关回归模型(RJUR)

### 3.1 模型建立

由于 GURM 模型的损失函数是基于 F 范数的, 这会导致模型对离群值以及噪声比较敏感; 另外, RR 以及 GURM 所学习到的投影数量受到类别数的限制, 由于学习后的投影矩阵  $\mathbf{W}$  的大小为  $d \times c$ , 这说明在学习的各个步骤中, 只有  $c$  个投影 ( $c$  为样本类别数) 可以用来进行特征选择。在实际应用中,  $c$  个投影并不能获得最佳的性能, 因此需要更多的投影来提高有效性, 这就是小类别问题。针对以上问题, 本文提出鲁棒联合不相关回归模型(RJUR), 表示如下:

$$\min_{\mathbf{A}, \mathbf{B}, \mathbf{F}, \mathbf{h}} \|\mathbf{XBA}^T + \mathbf{1h}^T - \mathbf{F}\|_{2,1} + \alpha \|\mathbf{B}\|_{2,1} \quad (5)$$

s. t.  $\mathbf{A}^T \mathbf{A} = \mathbf{I}, \mathbf{F}^T \mathbf{F} = \mathbf{I}, \mathbf{B}^T \mathbf{X}^T \mathbf{H} \mathbf{X} \mathbf{B} = \mathbf{I}$

其中,  $\mathbf{B} \in R^{d \times k}$  为投影矩阵,  $\mathbf{A} \in R^{c \times k}$  为辅助矩阵,  $k$  为目标特征维数。

RJUR 模型在损失函数与正则化项上使用联合  $L_{2,1}$  范数, 以保证模型对离群值以及噪声的鲁棒性, 同时为特征选择学习到一个联合稀疏投影, 改善了 GURM 模型对噪声数据的不稳定性; 另外 RJUR 模型考虑了小类问题, 使得所学习到的投影数量不受类别数的限制。

### 3.2 模型求解

对于不相关约束  $\mathbf{B}^T \mathbf{X}^T \mathbf{H} \mathbf{X} \mathbf{B} = \mathbf{I}$ , 令  $\bar{\mathbf{B}} = \sqrt{\mathbf{H}} \mathbf{X} \mathbf{B}$ ,  $\mathbf{C} = (\sqrt{\mathbf{H}} \mathbf{X})^{-1}$ , 有  $\bar{\mathbf{B}}^T \bar{\mathbf{B}} = \mathbf{I}, \mathbf{B} = \bar{\mathbf{C}} \bar{\mathbf{B}}$ , 因此原模型可以写为:

$$\min_{\mathbf{A}, \mathbf{B}, \mathbf{F}, \mathbf{h}} \|\mathbf{XCBA}^T + \mathbf{1h}^T - \mathbf{F}\|_{2,1} + \alpha \|\bar{\mathbf{C}} \bar{\mathbf{B}}\|_{2,1} \quad (6)$$

s. t.  $\mathbf{A}^T \mathbf{A} = \mathbf{I}, \mathbf{F}^T \mathbf{F} = \mathbf{I}, \bar{\mathbf{B}}^T \bar{\mathbf{B}} = \mathbf{I}$

根据  $L_{2,1}$  范数的定义, 有:

$$\mathbf{D}_i = \frac{1}{2 \|\mathbf{XCBA}^T + \mathbf{1h}^T - \mathbf{F}\|_2^i} \quad (7)$$

其中,  $(\mathbf{XCBA}^T + \mathbf{1h}^T - \mathbf{F})^i$  表示矩阵  $\mathbf{XCBA}^T + \mathbf{1h}^T - \mathbf{F}$  的第  $i$  行。  $\bar{\mathbf{D}}_i = \frac{1}{2 \|\bar{\mathbf{C}} \bar{\mathbf{B}}\|_2^i}$ , 其中,  $(\bar{\mathbf{C}} \bar{\mathbf{B}})^i$  表示矩阵  $\bar{\mathbf{C}} \bar{\mathbf{B}}$  的第  $i$  行。

因此

$$\begin{aligned} & \|\mathbf{XCBA}^T + \mathbf{1h}^T - \mathbf{F}\|_{2,1} + \alpha \|\bar{\mathbf{C}} \bar{\mathbf{B}}\|_{2,1} \\ &= \text{tr}((\mathbf{XCBA}^T + \mathbf{1h}^T - \mathbf{F})^T \mathbf{D} (\mathbf{XCBA}^T + \mathbf{1h}^T - \mathbf{F})) + \\ & \quad \alpha \text{tr}(\bar{\mathbf{B}}^T \bar{\mathbf{C}}^T \bar{\mathbf{D}} \bar{\mathbf{C}} \bar{\mathbf{B}}) \\ &= \text{tr}(\bar{\mathbf{B}}^T \bar{\mathbf{C}}^T \mathbf{X}^T \mathbf{D} \mathbf{X} \bar{\mathbf{C}} \bar{\mathbf{B}} + 2 \bar{\mathbf{B}}^T \bar{\mathbf{C}}^T \mathbf{X}^T \mathbf{D} \mathbf{1} \mathbf{h}^T \mathbf{A} - 2 \bar{\mathbf{B}}^T \bar{\mathbf{C}}^T \mathbf{X}^T \mathbf{D} \mathbf{F} \mathbf{A} + \mathbf{h}^T \mathbf{h} \mathbf{1}^T \mathbf{D} \mathbf{1} - 2 \mathbf{h}^T \mathbf{F}^T \mathbf{D} \mathbf{1} + \mathbf{F}^T \mathbf{D} \mathbf{F} + \alpha \bar{\mathbf{B}}^T \bar{\mathbf{C}}^T \bar{\mathbf{D}} \bar{\mathbf{C}} \bar{\mathbf{B}}) \end{aligned} \quad (8)$$

(1) 固定  $\mathbf{A}, \mathbf{B}, \mathbf{F}$ , 更新  $\mathbf{h}$

对式(8)求导令其等于 0, 可以得到:

$$\mathbf{h} = \frac{1}{s} (\mathbf{F}^T - \mathbf{A} \bar{\mathbf{B}}^T \bar{\mathbf{C}}^T \mathbf{X}^T) \mathbf{D} \mathbf{1} \quad (9)$$

其中,  $s = \mathbf{1}^T \mathbf{D} \mathbf{1}$ 。

(2) 固定  $\mathbf{B}, \mathbf{F}, \mathbf{h}$ , 更新  $\mathbf{A}$

$$\min_{\mathbf{A}} [\mathbf{A}^T (\mathbf{h} \mathbf{1}^T - \mathbf{F}^T) \mathbf{D} \mathbf{X} \bar{\mathbf{C}} \bar{\mathbf{B}}] \quad \text{s. t. } \mathbf{A}^T \mathbf{A} = \mathbf{I} \quad (10)$$

**定理 1**<sup>[18]</sup> 给定一个正交矩阵  $G \in R^{c \times k}$ , 秩为  $k$  的矩阵  $Q \in R^{c \times k}$ , 考虑以下优化问题:

$$\hat{G} = \arg \min_G \text{tr}(G^T Q) \quad \text{s. t. } G^T G = I_k \quad (11)$$

对  $Q$  进行奇异值分解  $Q = \tilde{U} \tilde{D} \tilde{V}^T$ , 则  $\hat{G} = \tilde{U} \tilde{V}^T$ . 根据定理 1, 对于给定的  $h$  和  $B$ , 对  $(h^T - F^T)DXCB$  进行奇异值分解:

$$(h^T - F^T)DXCB = \tilde{U} \tilde{D} \tilde{V}^T \quad (12)$$

那么  $A = UV^T$ .

(3) 固定  $A, F, h$ , 更新  $B$

因为  $B = CB$ , 需要先得到  $\bar{B}$  的最优值, 然后才能得到变量  $B$  的最优值.

**定理 2**<sup>[8]</sup> 假设  $S$  是任意一个对称矩阵,  $S$  的奇异值分解为  $\hat{U} \hat{D} \hat{V}^T$ , 下列优化问题:

$$\hat{M} = \arg \min_M \text{tr}(M^T X^T S X M - 2 M^T X^T Y) \quad (13)$$

等价于:

$$\hat{M} = \arg \min_M \| ((\hat{D}^{1/2} \hat{U}^T)^T)^{-1} Y - (\hat{D}^{1/2} \hat{U}^T) X M \|_2^2 \quad (14)$$

对于式(5), 去掉与  $\bar{B}$  无关的项, 得到:

$$\text{tr}[\bar{B}^T C^T (X^T DX + \alpha \bar{D}) C \bar{B} - 2 \bar{B}^T C^T X^T D(F - 1h^T)A] \quad (15)$$

令  $T = X^T DX + \alpha \bar{D}$ ,  $Z = X^T D(F - 1h^T)A$ , 则式(15)变为:  $\text{tr}(\bar{B}^T C^T T C \bar{B} - 2 \bar{B}^T C^T Z)$ , 根据定理 2, 对  $T$  进行 SVD 得到  $\hat{U} \hat{D} \hat{U}^T$ , 有:

$$\min_{\bar{B}} \text{tr}(\bar{B}^T C^T T C \bar{B} - 2 \bar{B}^T C^T Z) \Leftrightarrow \min_{\bar{B}} \| ((\hat{D}^{1/2} \hat{U}^T)^T)^{-1} Z - (\hat{D}^{1/2} \hat{U}^T) C \bar{B} \|_2^2 \quad (16)$$

令  $Y^* = ((\hat{D}^{1/2} \hat{U}^T)^T)^{-1} Z$ ,  $X^* = \hat{D}^{1/2} \hat{U}^T C$ , 则式(16)可以写为:

$$\| ((\hat{D}^{1/2} \hat{U}^T)^T)^{-1} Z - (\hat{D}^{1/2} \hat{U}^T) C \bar{B} \|_2^2 = \| Y^* - X^* \bar{B} \|_2^2 \quad (17)$$

因此, 变量  $\bar{B}$  的最优值可以通过以下优化问题得到:

$$\hat{B} = \arg \min_{\bar{B}} \| Y_B^* - X^* \bar{B} \|_2^2 \quad \text{s. t. } \bar{B}^T \bar{B} = I \quad (18)$$

这个问题可以使用正交匹配追踪法(OMP)<sup>[19]</sup>求解, 见算法 1. 得到变量  $\bar{B}$  的最优值后, 即可得到变量  $B$  的最优值.

(4) 固定  $A, B, h$ , 更新  $F$

用式(9)代替  $h$ , 原模型可以写为:

$$\min_{A, B, F} \| H(XCBA^T - F) \|_{2,1} + \alpha \| CB \|_{2,1} \quad \text{s. t. } A^T A = I, F^T F = I, \bar{B}^T \bar{B} = I \quad (19)$$

去掉与  $F$  无关的项, 式(19)可写为:

$$\| H(XCBA^T - F) \|_{2,1} = \text{tr}(\bar{B}^T C^T X^T \tilde{H} \tilde{D} H X C \bar{B} - 2 F^T \tilde{D} H X C \bar{B} A^T + F^T \tilde{H} \tilde{D} H F) \quad (20)$$

即:

$$\min_F \text{tr}(F^T Q F - 2 F^T P) \quad \text{s. t. } F^T F = I \quad (21)$$

其中,  $Q = H \tilde{D} H$ ,  $P = H \tilde{D} H X C \bar{B} A^T$ . 由于式(21)与 Stiefel 流形上的二次型问题的标准形式一致, 因此可以用文献[20]中提出的一种高效算法来求解, 即广义动力迭代法. 上述具体算法见算法 2. 所有变量更新完毕后, RJUR 的具体

算法如算法 3 所示.

**算法 1** OMP 算法

输入:  $\Phi, \tilde{y}$ , 收敛准则  $\epsilon$

输出:  $\hat{x} = x^l = \arg \min_{z: \text{supp}(z) \subseteq \Lambda^l} \| \tilde{y} - \Phi z \|_2$

1. 初始化:  $r^0 = \tilde{y}$ ,  $x^0 = 0$ ,  $\Lambda^0 = \phi$ ,  $l = 0$

2. Repeat

2.1. 匹配  $h^l = \Phi^T r^l$

2.2. 认证:

$\Lambda^{l+1} = \Lambda^l \cup \{ \arg \max_j |h^l(j)| \}$ , 如果存在多个最大值, 只选择其中一个;

2.3. 更新  $x^{l+1} = \arg \min_{z: \text{supp}(z) \subseteq \Lambda^{l+1}} \| \tilde{y} - \Phi z \|_2$ ,  $r^{l+1} = \tilde{y} - \Phi x^{l+1}$ ,  $l = l + 1$ ;

Until 目标函数收敛

**算法 2** 模型(20)求解

输入:  $Q \in R^{n \times n}$ ,  $P \in R^{n \times c}$

输出:  $F \in R^{n \times c}$

1. 初始化: 满足  $F^T F = I$  的随机矩阵, 通过幂方法使  $\tilde{A} = vI - A \in R^{n \times n}$  为正定矩阵的  $v$ .

2. Repeat

2.1. 更新  $R \leftarrow 2\tilde{A}F + 2C$ ;

2.2. 根据引理 2 解决  $\max_{F^T F = I} \text{tr}(F^T R)$  来更新  $F$ ;

Until 目标函数收敛

**算法 3** RJUR 算法

输入: 样本矩阵  $X \in R^{n \times d}$ , 中心矩阵  $H \in R^{n \times n}$ , 类别数  $c$ , 目标维数  $k$  ( $k \leq d$ ), 参数  $\alpha$

输出: 投影矩阵  $B = C\bar{B}$

1. 随机初始化:  $A \in R^{c \times k}$ ,  $B \in R^{d \times k}$ ,  $\bar{B} \in R^{n \times k}$ ,  $D \in R^{n \times n}$ ,  $\bar{D} \in R^{d \times d}$ ,  $h \in R^{c \times 1}$ ; 元素全为 1 的  $1 \in R^{n \times 1}$ , 令  $C = (\sqrt{H}X)^{-1}$

2. Repeat

2.1. 根据式(7)更新对角矩阵  $D$ ;

2.2. 根据  $\bar{D}_n = \frac{1}{2 \| (CB)^i \|_2}$  更新对角矩阵  $\bar{D}$ ;

2.3. 根据式(9)更新  $h$ ;

2.4. 根据模型(12)更新  $A$ ;

2.5. 根据算法 1 更新  $\bar{B}$ ;

2.6. 根据算法 2 更新  $F$ ;

Until 目标函数收敛

## 4 实验

为了评估 RJUR 算法的有效性, 本文选择了 UDFS<sup>[12]</sup>, JELSR<sup>[13]</sup>, GLSPFS<sup>[14]</sup> 以及 GURM<sup>[16]</sup> 与 RJUR 进行对比, 并使用最近邻分类器(NN)对算法在每个数据集上的结果进行测试, 从而获得分类精度. 实验用到的环境是 Windows 10, 处理器是 Intel(R) Core(TM) i7, 用到的软件是 MATLAB R2018.

### 4.1 数据集

本文选择了 6 个数据集进行实验, 其中包括 4 个人脸数据集 ORL<sup>[21]</sup>, Yale<sup>[22]</sup>, BIO<sup>[23]</sup> 和 JAFFE<sup>[24]</sup>, 一个物品数据集 COIL20<sup>[25]</sup>, 以及一个高光谱数据集 Pavia University.

数据集的具体信息如表 1 所列. 为了测试 RJUR 的鲁棒性, 还对 UMIST 数据集分别添加了  $8 \times 8$  以及  $12 \times 12$  的白块噪声, 对 COIL20 数据集分别施加了 10% 以及 20% 的椒盐噪声. 每个数据集的部分图像如图 1—图 3 所示.

表 1 数据集具体信息值

Table 1 Specific information of data set

| Data set         | Number of Samples | Number of categories | Number of features |
|------------------|-------------------|----------------------|--------------------|
| ORL              | 400               | 40                   | 1 024              |
| Yale             | 165               | 15                   | 1 024              |
| BIO              | 1 460             | 22                   | 1 024              |
| JAFFE            | 213               | 10                   | 1 024              |
| COIL20           | 1 440             | 20                   | 1 024              |
| Pavia University | 42 776            | 9                    | 103                |



(a)ORL 数据集



(b)Yale 数据集



(c)BIO 数据集



(d)JAFFE 数据集



(e)COIL20 数据集

图 1 数据集部分图像

Fig. 1 Partial image of dataset

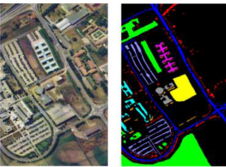
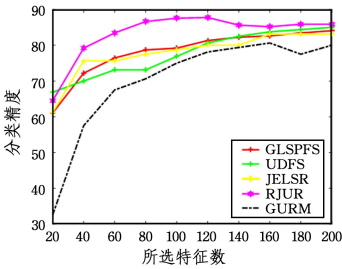
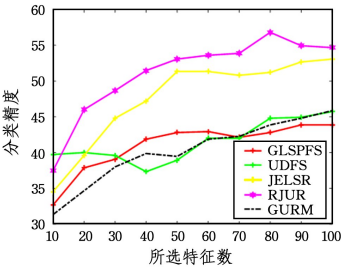


图 2 Pavia University 数据集

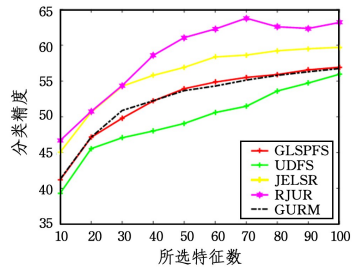
Fig. 2 Pavia University dataset



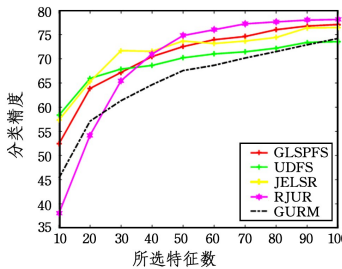
(a)ORL



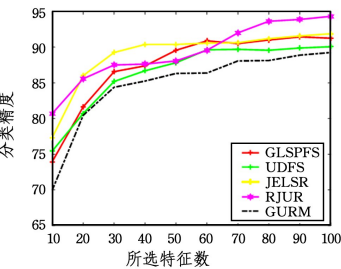
(b)Yale



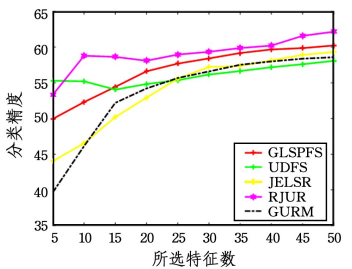
(c)BIO



(d)COIL



(e)JAFFE



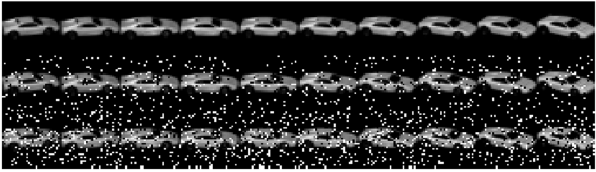
(f)Pavia University

图 4 不同算法在不同数据集上的分类精度

Fig. 4 Classification accuracies of different methods on six data sets with different numbers of selected features



(a)UMIST:噪声块大小分别为  $0\times0$ ,  $8\times8$  和  $12\times12$



(b)COIL-20:噪声密度分别为 0, 10% 和 20%

图 3 部分加噪图像

Fig. 3 Partially noisy images

4.2 实验分析

对于不同的数据集,所选取的训练集数量以及特征数量也是不同的。图 4 给出了每个算法在不同数据集、不同子空间维度上的分类精度。对于 ORL, Yale, BIO 数据集,每类图像随机选取 6 个样本组成训练集,其余的样本组成数据集;对于 COIL20 以及 JAFFE 数据集,每类图像随机选取 4 个样本组成训练集,剩下的样本组成测试集;对于高光谱数据集 Pavia University,每类图像随机选取 10 个样本组成训练集,剩下的样本组成测试集。

从图 4 可以看出,在大多数情况下, RJUR 方法相比其他对比特征选择方法能达到更好的性能,因为 RJUR 方法所学习到的投影矩阵是不相关的,并且所学习到的投影矩阵数量也不受数据集类别数的限制,可以得到足够的投影矩阵数量。

另外,为了说明 RJUR 的有效性,本节给出了不同算法在噪声数据集上的分类效果。

从图 5 可以看出, RJUR 在数据集存在噪声的情况下 也有很好的分类性能,这是因为 RJUR 采用了联合  $L_{2,1}$

范数,学习到的投影矩阵具有稀疏性,且面对离群值或者噪声数据时,RJUR 模型具有更好的鲁棒性。除此之外,为了说明不同样本数量对分类性能的影响,表 2 列出了不同训练样本数量情况下,不同方法在不同数据集上的分类效果。从表 2 可以看出,在不同训练样本数目下,RJUR 算法均优于其他 4 种算法,且随着训练样本数的增加,识别率也相应提高。另外,对于在样本数量较少的情况下,RJUR 方法的分类性能比其他对比方法更好。这是因为

RJUR 方法使用了学习系数投影矩阵,同时考虑到了小类问题,使得投影矩阵的数量不受类别数的限制。另外,RJUR 方法使用了不相关约束,所选出来的特征之间是不相关的。而其他对比方法并没有考虑到以上种种,比如 GURM 方法的损失函数是 F 范数,这使得 GURM 对噪声数据或者离群值比较敏感;UDFS,GLSPFS 等方法都没有考虑到小类问题,使得学习到的投影矩阵数量不够多,故而影响了分类性能。

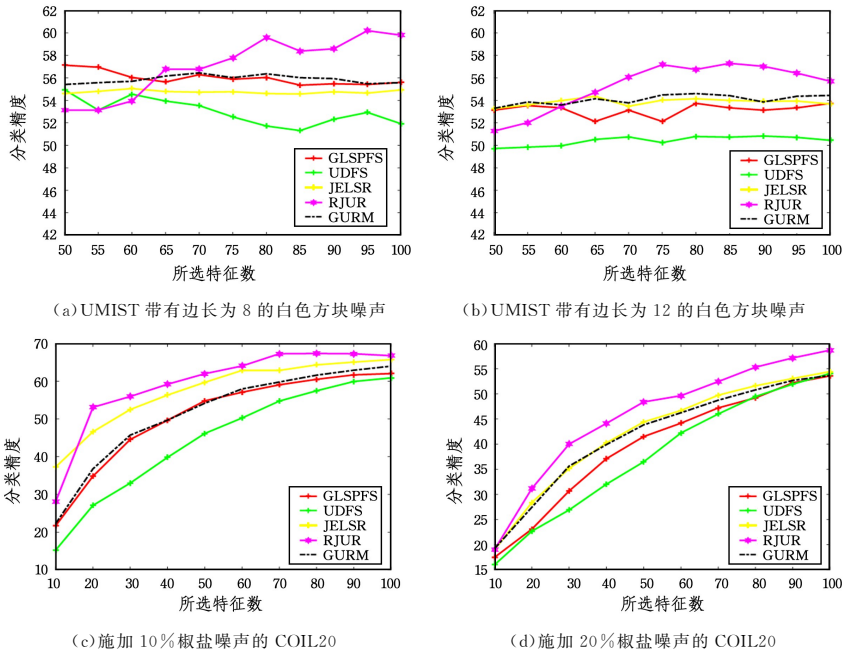


图 5 不同方法在两个噪声数据集上的分类精度

Fig. 5 Classification accuracies of different methods on two noisy data sets

表 2 不同训练样本数量对分类精度的影响

Table 2 Influence of number of training samples on the accuracy of classification

| Data set         | Number of training samples for each type | GURM    | GLSPFS  | UDFS    | JELSR   | RJUR    |
|------------------|------------------------------------------|---------|---------|---------|---------|---------|
| ORL              | 4                                        | 78.4167 | 75.6667 | 76.6667 | 76.2500 | 80.7917 |
|                  | 6                                        | 83.7500 | 84.0625 | 85.0000 | 83.1250 | 87.7500 |
|                  | 8                                        | 88.2500 | 88.0000 | 90.0000 | 90.0000 | 91.8750 |
| Yale             | 4                                        | 38.2857 | 40.2857 | 42.5714 | 45.6190 | 51.1429 |
|                  | 6                                        | 45.8667 | 43.8667 | 45.7333 | 53.0667 | 56.8000 |
|                  | 8                                        | 50.4444 | 47.1111 | 50.0000 | 54.8889 | 57.7778 |
| BIO              | 4                                        | 51.0423 | 51.8076 | 49.6137 | 53.3090 | 55.7872 |
|                  | 6                                        | 56.7395 | 56.9202 | 55.9413 | 59.7063 | 63.1777 |
|                  | 8                                        | 62.9439 | 62.4688 | 60.1012 | 65.3115 | 67.2118 |
| COIL20           | 4                                        | 76.7132 | 77.0956 | 73.5515 | 76.4706 | 78.1324 |
|                  | 6                                        | 80.0303 | 81.7045 | 74.9015 | 82.2045 | 83.1364 |
|                  | 8                                        | 84.4688 | 85.0938 | 80.6328 | 84.3750 | 86.9531 |
| JAFPE            | 4                                        | 89.2500 | 91.4375 | 90.0625 | 91.8750 | 94.3125 |
|                  | 6                                        | 93.8571 | 94.9286 | 93.3571 | 95.5000 | 96.2857 |
|                  | 8                                        | 96.4167 | 95.8333 | 94.5833 | 97.0833 | 97.8333 |
| Pavia University | 10                                       | 58.5817 | 60.2310 | 58.0774 | 59.3288 | 62.1586 |
|                  | 15                                       | 61.6433 | 63.2520 | 61.6046 | 63.3805 | 64.2293 |
|                  | 20                                       | 64.2992 | 63.7550 | 62.5676 | 64.8958 | 65.4491 |

4.3 参数设置

由 RJUR 算法的目标函数可知,正则化参数  $\alpha$  和特征维度  $k$  的取值对算法的分类性能有着很大影响。图 6 给出了 RJUR 算法在 3 个不同数据集上,不同正则化参数  $\alpha$  和特征维度  $k$  的取值对分类精度的影响。从图 6 可以看出,大多数情况下,

随着  $\alpha$  值与  $k$  值的增大,分类精度呈现先递增后递减并趋于稳定的变化趋势。对于 ORL 数据集,在  $\alpha=10^{-2}, k=70$  时, RJUR 算法的分类效果最好;在 Yale 数据集上,  $\alpha=10^{-1}, k=18$  时, RJUR 算法的分类精度达到最高;当  $\alpha=10^{-3}, k=30$  时, RJUR 算法在 BIO 数据集上的分类性能最好。

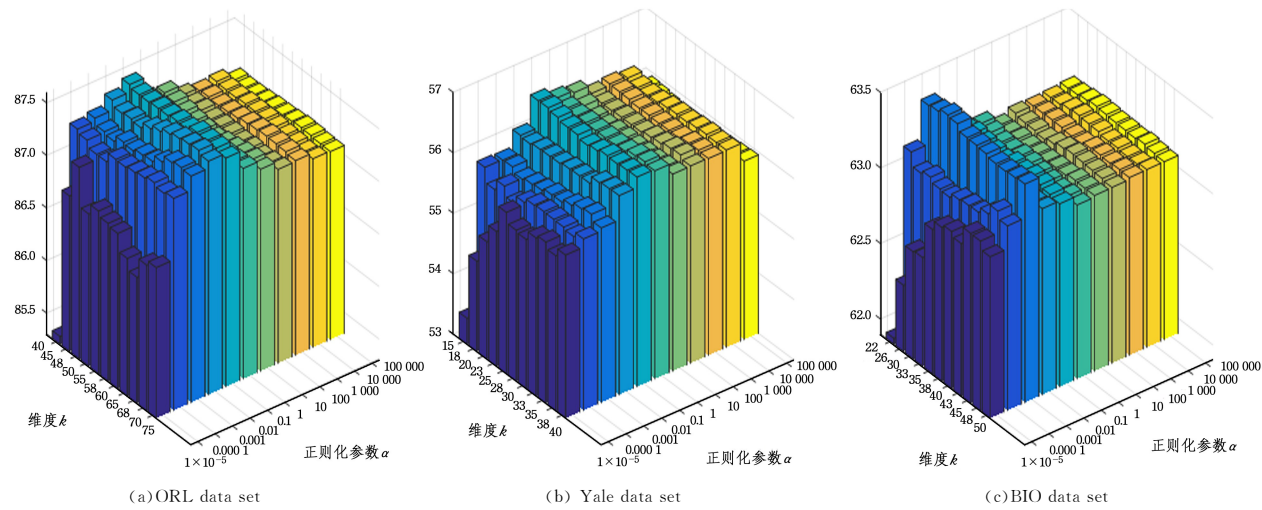


图 6 不同参数对不同数据集分类精度的影响

Fig. 6 Effects of different parameters on classification accuracy of different data sets

4.4 收敛性分析

图 7 给出了 RJUR 算法在数据集 ORL, Yale 以及 JAFFE 上的收敛性。3 个数据集的每类训练样本数量都为 6。从图 7 可以看出, RJUR 算法的收敛速度很快, 迭代次数为 4 左右就可以收敛到一个很小的值。

另外, 表 3 给出了不同算法在不同数据集上的运行时间。从表 3 可以看出, 相较对比算法而言, RJUR 的运行时间较长, 这是由于 OMP 算法稍微有点耗时。虽然 RJUR 算法的时间复杂度更高, 但总体来说算法的综合性能优秀。

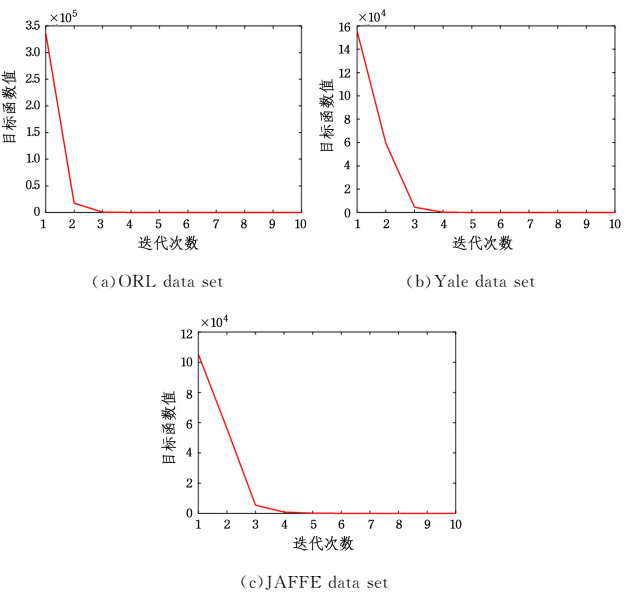


图 7 RJUR 在不同数据集上的收敛性

Fig. 7 Convergence of RJUR on different data sets

表 3 不同算法在不同数据集上的运行时间

Table 3 Running time of different algorithms on different data sets

| 数据集   | GURM | GLSPFS | UDFS | JELSR | RJUR |
|-------|------|--------|------|-------|------|
| Yale  | 1.63 | 0.03   | 2.09 | 0.89  | 6.06 |
| JAFFE | 1.64 | 0.02   | 2.02 | 0.86  | 6.72 |
| ORL   | 1.75 | 0.19   | 2.23 | 1.24  | 6.71 |

**结束语** 本文提出了一种鲁棒不相关回归的无监督特征选择方法(RJUR), 该方法对损失函数和正则化项使用了联合  $L_{2,1}$  范数, 在学习到一个稀疏投影矩阵的同时, 保证了模型的鲁棒性, 使得模型在面对噪声数据时依然能够保持不错的分类性能; 另外, RJUR 模型还考虑了实际应用中可能存在的小类问题, 使得学习到的投影矩阵数量不受类别数的限制, 能够得到足够多的投影矩阵数量, 从而提高算法的识别率。理论分析和实验结果都证明了 RJUR 算法的分类性能。

接下来的工作将会考虑对算法进行改进, 如何以矩阵数据直接作为输入, 从而避免小样本问题和维度过高带来的问题, 并将数据原始的位置关系考虑进去, 以进一步提升算法的性能。

参 考 文 献

[1] MAATEN L, POSTMA E, HERIK J. Dimensionality reduction: A comparative review[J]. Review Literature &. Arts of the Americas, 2009, 10(1): 158-165.

[2] JAIN A, ZONGKER D. Feature selection: evaluation, application, and small sample performance[J]. IEEE Trans. Pattern Anal. Mach. Intell, 1997, 19(2): 153-158.

[3] DUDA R O, HART P E, STORK D G. Pattern Classification [M]. Wiley, 2001: 63-69.

[4] BELHUMEUR P N, HESPANHA J P, KRIEGMAN D J. Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection[C]// European Conference on Computer Vision. Springer-Verlag, 1996: 15-28.

[5] HE J, DING L, LEI J, et al. Kernel ridge regression classification [C]// Proceedings of International Joint Conference on Neural Networks. IEEE, 2014: 137-146.

[6] NIE F, HUANG H, XIAO C, et al. Efficient and Robust Feature Selection via Joint  $\ell_{2,1}$ -Norms Minimization[C]// International Conference on Neural Information Processing Systems. Curran Associates Inc. , 2010: 1813-1821.

[7] CHEN X H, LU Y. Robust Graph Regularized Sparse Matrix Regression for Two-dimensional Supervised Feature Selection

- [J]. IET Image Processing, 2020, 14(9): 1740-1749.
- [8] MO D, LAI Z. Robust Jointly Sparse Regression with Generalized Orthogonal Learning for Image Feature Selection[J]. Pattern Recognition, 2019, 16(3): 964-985.
- [9] HOU C, JIAO Y, NIE F, et al. 2D Feature Selection by Sparse Matrix Regression[J]. IEEE Transactions on Image Processing, 2017, 26(9): 4255-4268.
- [10] CHEN X H, LU Y. Dynamic Graph Regularization and Label Relaxation- Based Sparse Matrix Regression for Two-dimensional Feature Selection [J]. IEEE Access, 2020, 8: 62855-62870.
- [11] YU Y, CHEN X. Joint feature weighting and adaptive graph-based matrix regression for image supervised feature Selection- ScienceDirect [J]. Signal Processing: Image Communication, 2020, 13(2): 9-16.
- [12] YANG L, JIANG A L, QIANG Y. Structure Preserving Unsupervised Feature Selection Based on Autoencoder and Manifold Regularization[J]. Computer Science, 2021, 48(8): 53-59.
- [13] YI Y, SHEN H T, MA Z, et al. L21-norm regularized discriminative feature selection for unsupervised learning[C]// Proceedings of the Twenty-Second international joint conference on Artificial Intelligence. AAAI Press, 2011: 56-59.
- [14] HOU C, NIE F, LI X, et al. Joint Embedding Learning and Sparse Regression: A Framework for Unsupervised Feature Selection[J]. IEEE Transactions on Cybernetics, 2013, 44(6): 793-804.
- [15] LIU X, WANG L, ZHANG J, et al. Global and Local Structure Preservation for Feature Selection[J]. IEEE Transactions on Neural Networks and Learning Systems, 2017, 25(6): 1083-1095.
- [16] YUAN H, LI J, LAI L L, et al. Joint sparse matrix regression and nonnegative spectral analysis for two-dimensional unsupervised feature selection[J]. Pattern Recognition, 2019, 66(3): 119-133.
- [17] LI X, ZHANG H, ZHANG R, et al. Generalized Uncorrelated Regression with Adaptive Graph for Unsupervised Feature Selection[J]. IEEE Transactions on Neural Networks and Learning Systems, 2018, 3(6): 1-9.
- [18] ZOU H, TIBSHIRANI H R. Sparse Principal Component Analysis[J]. Journal of Computational and Graphical Stats, 2006, 15(2): 1-30.
- [19] DAVENPORT M A, WAKIN M B. Analysis of Orthogonal Matching Pursuit using the Restricted Isometry Property[J]. IEEE Transactions on Information Theory, 2010, 56(9): 4395-4401.
- [20] NIE F, ZHANG R, LI X. A generalized power iteration method for solving quadratic problem on the Stiefel manifold[J]. Science China, 2017(11): 146-155.
- [21] BELHUMEUR P N, HESPAHNA J P, KRIEGMAN D J. Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1997, 19(7): 711-720.
- [22] GEORGHIADES A, BELHUMEUR P P, KRIEGMAN D. Yale face database [OL]. <http://cvc.yale.edu/projects/yalefaces/yalefaces.html>.
- [23] JESORSKY O. Robust Face Detection Using the Hausdorff Distance[J]. AVBPA, 2001, 32(10): 90-95.
- [24] LYONS M J, BUDYNEK J. Automatic classification of single facial images[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1999, 4(3): 93-104.
- [25] NENE S A. Columbia Object Image Library (COIL-20) [J]. Technical Report, 1996, 5(8): 63-68.



**LI Zong-Ran**, born in 1995, postgraduate. His main research interests include machine learning, pattern recognition and digital image processing.



**CHEN Xiu-hong**, born in 1964, Ph.D. His main research interests include digital image processing, pattern recognition, target detection and tracking, optimization theory and method and so on.

(责任编辑:李亚辉)