

# 基于超图稀疏的低秩属性选择算法用于多回归分析\*

罗 葵<sup>1</sup>, 苏毅娟<sup>2†</sup>, 雷 聪<sup>1</sup>, 胡荣耀<sup>1</sup>, 杨利锋<sup>1</sup>, 李永钢<sup>1</sup>

(1. 广西师范大学 广西多源信息挖掘与安全重点实验室, 广西 桂林 541004; 2. 广西师范学院 计算机与信息工程学院, 南宁 530023)

**摘要:** 针对已有的多回归算法未能更好地考虑不同列类别之间的关联性, 提出了一种基于超图稀疏的属性选择方法。利用稀疏理论中的  $\ell_{2,p}$ -范数来改进线性回归模型, 结合低秩结构和超图表示来分别考虑同一类别间的局部结构和不同类别间的全局结构, 通过子空间学习中的线性判别分析(LDA)来进一步调整重构系数矩阵的结果。经实验验证, 提出的算法在六个公开数据集上相比四种对比算法, 在多回归分析中均能取得更好的结果。

**关键词:** 多回归分析; 超图表示; 子空间学习; 稀疏学习

中图分类号: TP301.6 文献标志码: A 文章编号: 1001-3695(2018)09-2671-05

doi: 10.3969/j.issn.1001-3695.2018.09.025

## Multi-output regression analysis algorithm based on hypergraph and low-rank

Luo Yan<sup>1</sup>, Su Yijuan<sup>2†</sup>, Lei Cong<sup>1</sup>, Hu Rongyao<sup>1</sup>, Yang Lifeng<sup>1</sup>, Li Yonggang<sup>1</sup>

(1. Guangxi Key Laboratory of Multi-source Information Mining & Security, Guangxi Normal University, Guilin Guangxi 541004, China; 2. School of Computer & Information Engineering, Guangxi Teachers Education University, Nanning 530023, China)

**Abstract:** Since the traditional regression model does not take advantage of the correlation among class labels, this paper proposed a multi-output regression algorithm by considering hypergraph and low-rank. Firstly, this paper employed an  $\ell_{2,p}$ -norm regularization term to improve the model of linear regression. Then, this paper combined low rank structure with hypergraph representation to consider both the local structure of same categories and global structure of different categories. Finally, the model further adjusted the results of reconstruction coefficient matrix by linear discriminant analysis (LDA). Experimental results show that the proposed method obtains the best results, compares to four competing methods on six public datasets.

**Key words:** multiple regression analysis; hypergraph representation; subspace learning; sparse learning

## 0 引言

在大数据时代背景下, 计算机视觉、数据挖掘等领域, 高维数据急剧地增加<sup>[1]</sup>。同时, 高维数据的维数高和数量大等特点给相关研究与应用带来了巨大挑战, 即极度扩增了数据的存储空间和增大了时间复杂度, 使得高维数据不能直接在实际应用中应用。因此, 在数据的处理过程中, 如何有效且高效地利用高维数据具有非常重大的意义<sup>[2]</sup>。实际中, 高维数据可由多个低维子空间结构表示, 对高维数据进行属性约简即通过寻找其低维子空间结构, 能够利用样本之间的关联以及样本属性之间的关联来降低数据的维度<sup>[3]</sup>。属性约简的目标是从数据的属性集中选择一个重要且具有代表性的子集作为数据的新属性集, 并且新属性集能够维持原高维数据的数据结构, 甚至可以提高分类效果, 因而属性约简成为机器学习的一个重要领域<sup>[4]</sup>。

属性约简方法根据类标签的使用情况可以分为三类, 即无监督方法、有监督方法和半监督方法<sup>[5-6]</sup>。其中, 无监督方法一般是根据属性之间的相关性或者数据的图结构来提取属性。经典的无监督属性约简方法有主成分分析(principal component analysis, PCA)法<sup>[7]</sup>、局部保留投影(locality preserving pro-

jection, LPP)法<sup>[8]</sup>、局部线性嵌入(locally linear embedding, LLE)法<sup>[9]</sup>等。有监督方法通常是根据类标签来搜索重要的属性, 并且根据已知的类标签来评估算法的性能。常见的有监督属性约简方法是线性判别分析(linear discriminant analysis, LDA)法<sup>[10]</sup>, 它已经被广泛地应用于医疗、生物等行业。而半监督属性方法能够充分利用有标签和无标签的样本信息来选择最具判别力的属性。目前比较经典的半监督降维方法有半监督判别分析(semi-supervised discriminant analysis, SDA)法<sup>[11]</sup>、半监督维数约减(semi-supervised dimensionality reduction, SDR)法<sup>[12]</sup>等。将属性约简方法应用于线性回归模型时, 通常能取得较好的效果。但是在分析高维数据时, 传统的回归模型表现出很低的性能; 其次, 一般线性回归模型在不同的响应之间不会考虑其相关性, 其典型的代表是最小二乘回归<sup>[13]</sup>, 此方法只是对每个预测的数据分别产生一个响应。

针对以上问题, 本文提出一种结合属性选择、低秩超图和子空间学习(即LDA)的属性约简方法用于回归分析——基于超图和低秩的多回归分析算法(multi-output regression analysis algorithm based on hypergraph and low-rank, HLMR-FS)。本文首先结合线性回归和稀疏学习(利用 $\ell_{2,p}$ -范数更灵活地考虑类标签与样本之间的联系)应用于回归分析; 接着为了使得模型选择出来的属性能够更好地保留数据的局部信息, 在模型中嵌

收稿日期: 2017-04-21; 修回日期: 2017-05-31 基金项目: 国家自然科学基金资助项目(61450001, 61672177, 61573270); 国家“973”计划资助项目(2013CB329404); 中国博士后科学基金资助项目(2015M570837); 广西自然科学基金资助项目(2012GXNSFGA060004, 2015GXNSFCB139011, 2015GXNSFAA139306); 广西研究生教育创新计划项目(XYCSZ2017064, YCSW2017065)

作者简介: 罗葵(1989-), 男, 安徽安庆人, 硕士研究生, 主要研究方向为数据挖掘、机器学习; 苏毅娟(1976-), 女(通信作者), 广西桂林人, 副教授, 主要研究方向为机器学习、数据挖掘(519736662@qq.com); 雷聪(1991-), 男, 湖北大冶人, 硕士研究生, 主要研究方向为数据挖掘、机器学习; 胡荣耀(1992-), 男, 江西景德镇人, 硕士研究生, 主要研究方向为数据挖掘、机器学习; 杨利锋(1989-), 男, 广西平乐人, 硕士研究生, 主要研究方向为数据挖掘、机器学习; 李永钢(1989-), 男, 河北保定人, 硕士研究生, 主要研究方向为数据挖掘、机器学习。

入一个基于超图的拉普拉斯矩阵,用于保持各属性之间的更深层次的关系;然后在此回归模型上运用低秩表示(假设系数矩阵具有低秩效果,从而考虑到类标签的相关性)来提高效率和准确率;最后探索出一种新的算法优化方法,即对目标函数按顺序执行低秩属性选择和子空间学习方法,并不断交替迭代地执行此过程,使得结果达到最优,最终取得全局最优解。本文提出的 HLMR-FS 算法结合属性选择、低秩超图和子空间学习各自优点用于回归分析。

### 1 相关理论背景及简介

#### 1.1 稀疏学习简介

稀疏学习(sparse learning)<sup>[14]</sup>理论最早主要应用于图形、图像视觉等领域<sup>[15]</sup>。由于具有强大的应用价值及理论支撑,所以稀疏学习得到了迅速发展,并且已经在机器学习与模式识别等领域得到了广泛的应用。设  $w \in R^n$  是模型的参数向量,  $n$  是样本数,其中  $w_1, w_2, \dots, w_n$  表示  $n$  个样本参数,且  $w = [w_1; w_2; \dots; w_n]$ 。本质上,数据的特点是由少量关键特征决定的,属性选择的目标就是找出一个有代表性的特征子集,所有属性都可以通过相关数据来重建。鉴于稀疏学习可实现特征的自动选择,为此本文算法采用稀疏学习的方法来自动选择重要的特征子集。在稀疏学习的基本理论中,通过对模型的参数向量  $w \in R^n$  进行稀疏假设,实现稀疏正则化,再用训练样本对参数  $w$  进行拟合。主要实现的目标函数为

$$\min_w g(w) = f(w) + \lambda \varphi(w) \quad (1)$$

其中:  $f(w)$  是损失函数;  $\varphi(w)$  是正则化项;  $\lambda$  为控制损失函数与正则化项之间的平衡参数,通过改变  $\lambda$  的值来调节  $w$  的稀疏性,且  $\lambda$  越大,  $w$  越稀疏。将稀疏学习应用于属性选择算法中,能够将原始数据之间的系数权重作为重要的鉴别信息引入模型,通过稀疏约束来使得输入数据进行稀疏表示,这样可以去除冗余和不相关属性,同时保证重要属性能够被选择。鉴于稀疏学习的正则化因子通常选用能够凸优化求解的范数,如此便能够保证本文提出的模型求得唯一的全局最优解<sup>[16]</sup>。在稀疏学习中,  $\ell_0$ -范数是最有效的稀疏正则因子,但因其求解为 NP 难题,所以很多文献均采用近似正则项  $\ell_1$ -范数来替代  $\ell_0$ -范数;而  $\ell_{2,1}$ -范数能导致行稀疏,已经被证明比  $\ell_1$ -范数更适用于属性选择。但  $\ell_{2,p}$ -范数更具灵活性,通过参数  $p$  来寻找属性间本质的关联结构,能够使系数矩阵得到行稀疏,从而去除数据中噪声的干扰和冗余特征来提高回归模型的效果。因此本文采用  $\ell_{2,p}$ -范数作为稀疏正则化因子来对系数矩阵进行稀疏处理,以达到剔除冗余和不相关属性的目的,且能够有效地减少离群点对结果的影响。

#### 1.2 超图简介

给定一个超图  $G = (V, E, w)$ , 其中  $V$  是一个有限节点集合,  $E$  是超图的超边集(在超图的图形表示中,将属于同一条边的点由一条闭曲线包围在一起,这条闭曲线称为超图的一条超边)。任意一个超边  $e \in E$  是顶点集  $V$  的一个子集,  $\cup_{e \in E} e = V$ , 并被赋予一个非负权重  $w(e)$ 。对于任意顶点  $v \in V$ , 它的阶(或称为度, degree)可被定义为如下表达式:

$$d(v) = \sum_{e \in E | v \in e} w(e) \quad (2)$$

对于超边  $e \in E$ , 它的阶可以定义为边上所有顶点的数目:

$$\delta(e) = |e| \quad (3)$$

超图的结构可以通过一个  $|V| \times |E|$  维的点边关联矩阵  $H$  来进行描述。在该矩阵中,如果一个顶点  $v$  在一个超边  $e$  上,其第  $(v, e)$  个元素被赋值为 1,  $h(v, e) = 1$ ; 如果该顶点不在该边上,则设置相关元素的值为 0。根据关联矩阵的定义,式(2)和(3)可以进一步表达为

$$d(v) = \sum_{e \in E | v \in e} w(e) h(v, e) \quad (4)$$

$$\delta(e) = \sum_{v \in e} h(v, e) \quad (5)$$

超图样例与对应的关联矩阵如图 1 所示。

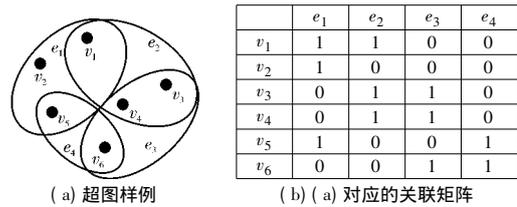


图 1 超图样例与对应的关联矩阵

归一化的超图拉普拉斯矩阵<sup>[17]</sup>定义如下:

$$L_H = I - D_e^{-1/2} J W_H D_e^{-1} J^T D_v^{-1/2} \quad (6)$$

其中:  $I$  表示  $|V| \times |V|$  的单位矩阵;  $D_v$  表示由  $d(v)$  构成的对角矩阵;  $J$  是关联矩阵,能够充分描述超图的特点;  $W_H$  表示由超边权重矩阵  $w$  构成的对角矩阵;  $D_e$  用来表示由  $\delta(e)$  组成的对角矩阵。超图模型拓展了经典图模型中两两顶点组建边的构图方式,以具有某种相似属性的数据子集构建超边,从而可以有效刻画数据间的高阶关系。将超图(其中每个超图顶点代表一个样本,而超边代表样本间的关系)融合到 LPP 算法中,使模型在进行属性选择时能更好地考虑数据的局部结构。

## 2 算法描述和优化

### 2.1 算法描述

对于回归问题,传统的方法一般是利用最小二乘线性回归建模处理,具体如下:

$$\min_W \|Y - X^T W\|_F \quad (7)$$

其中:  $\|\cdot\|_F$  表示 Frobenius 范数; 令训练数据表示为  $X = [x_1, x_2, \dots, x_n] \in R^{d \times n}$ ,  $n$  和  $d$  分别表示样本数和属性数; 类标签数据表示为  $Y \in R^{n \times c}$ ,  $c$  表示样本对应类别数。本文用  $\ell_F$  范数来估计余量,其中,  $\|X\|_F = \sqrt{\sum_i \|x_i\|_2^2} = \sqrt{\sum_j \|x_j\|_2^2}$ , 而且模型输出的重构系数矩阵  $W \in R^{d \times c}$  表示训练样本与测试样本间重建的系数矩阵,可以得到其最优的模型解为  $W = (X X^T)^{-1} X Y$ 。

通常数据中含有许多离群点,然而  $\ell_F$ -范数对离群数据比较敏感;同时,为了防止求  $W$  时过拟合,本文利用  $\ell_{2,p}$ -范数作为正则化项,即  $R_1(W) = \|W\|_{2,p} = (\sum_{i=1}^d (\sum_{j=1}^c W_{ij}^2)^{p/2})^{1/p}$ , 其中  $p(0 < p < 2)$  来寻找属性之间的关联结构,对属性进行稀疏处理,从而获得新的回归模型如下:

$$\min_W \|Y - X^T W\|_F^2 + \alpha \|W\|_{2,p} \quad (8)$$

一般来说,大多数事物之间都隐含某种联系,对于多回归分析问题中,这种关系可能会更加明显<sup>[1618]</sup>。因此,为了探测这些多输出之间的关联结构,同时利用这种特性来增强多输出回归模型,算法引入低秩条件来约束系数矩阵  $W$ ,即使用  $\text{rank}(W) = r \leq \min(d, c)$  约束<sup>[19]</sup>。其中:  $n$  是样本数;  $c$  是每个样本的输出变量数。对于秩最小化问题是一个非凸状且 NP 难问题,然而秩约束的目标却是可解的,即该全局解是可以得到的<sup>[20]</sup>。在式(8)的回归模型中,对系数矩阵  $W$  使用低秩约束,所以问题转换为解决如下的问题:

$$\begin{aligned} \min_W & \|Y - X^T W\|_F^2 + \alpha \|W\|_{2,p} \\ \text{s. t.} & \text{rank}(W) \leq \min(d, c) \end{aligned} \quad (9)$$

为进一步增加模型的泛化性能,本文借鉴 LPP 算法,并且将其中的普通图改成超图,即在模型式(8)的基础上嵌入一个超图拉普拉斯正则化项  $R_2(W) = \text{tr}(W^T X L_H X^T W)$ , 其中  $L_H$  是通过式(6)得到的。这样既能保证样本数据在空间投影变换后数据的近邻关系保持不变,又考虑到了样本间的关系。故可以得到以下目标函数:

$$\min_W \|Y - X^T W\|_F^2 + \alpha \|W\|_{2,p} + \beta \text{tr}(W^T X L_H X^T W)$$

s. t.  $\text{rank}(W) \leq \min(d, \rho)$  (10)

其中:  $\alpha, \beta$  表示为相应正则化项的控制参数。

文献 [21] 指出, 低秩线性回归模型等同于在 LDA 的子空间中作线性回归。因此, 为了更有效地求解低秩约束条件和有效地考虑多输出之间的关联结构, 算法令  $Z = AB$  拥有一个低秩  $r, r < \min(d, \rho)$ 。本文的最终目标模型如下:

$$\min_{A,B} \|Y - X^T AB\|_F^2 + \alpha \|AB\|_{2,p} + \beta \text{tr}(B^T A^T X L_H X^T AB)$$

s. t.  $\text{rank}(AB) \leq \min(d, \rho)$  (11)

其中:  $AB$  表示由  $A \in R^{d \times r}$  和  $B \in R^{r \times c}$  构成的系数矩阵, 并且  $\text{rank}(AB) \leq \min(d, \rho)$ 。数据有噪声或离群点会导致系数矩阵的秩变大。使用低秩约束能考虑属性间的关系, 而且低秩约束就是进行 LDA 的子空间学习, 从而本文提出的方法考虑到了数据的全局结构。稀疏表示使矩阵  $AB$  中大多数的行收缩为零, 而  $AB$  中非零系数所对应的属性则被作为有代表的属性子集, 从而进行属性选择。

### 2.2 模型总结

本文提出的 HLMR-FS 算法在属性选择时考虑到类标签之间的关系, 具有较好的鲁棒性和泛化性。通过使用  $\ell_{2,p}$ -范数稀疏正则化因子惩罚回归系数矩阵, 从而去除冗余和不相关的属性, 使该模型能够自动地选择重要的属性, 所以该算法相对于子空间学习方法具有较好的解释性。通过在回归模型中嵌入一个超图拉普拉斯正则化项来保持数据之间的局部结构, 同时利用子空间学习方法来保持数据的全局结构, 因此该算法相对于已有的属性选择算法和常见的子空间学习(只是单一地考虑一种结构, 即全局结构或者局部结构)具有更好的效果。通过在回归系数矩阵上使用低秩约束、多输出之间的关联, 从而提高模型对于高维数据分类预测的准确率; 同时提出了一种区别于交替方向乘子法的求解方法, 首先固定低秩属性选择的结果, 以此增强子空间学习的能力; 其次固定子空间学习的结果, 确保低秩属性选择能输出更具判别力的属性集。该优化算法能保证目标函数值在每次迭代中逐步收敛, 最终取得全局最优解。

综上所述, 新算法结合了超图稀疏和属性选择, 以此来选取既具代表性又具有局部相关的属性子集; 同时利用低秩约束来考虑不同类标签之间的相关结构, 使得选取后的属性子集保持了不同类全局之间的相关结构, 并能够更好地应用于多回归研究与分析。本文算法的伪代码如下。

#### 算法 1 HLMR-FS 算法伪代码

输入: 训练样本  $X \in R^{d \times n}$ , 控制参数  $\alpha, \beta, p$ 。

输出: aCC 和 aRMSE 的均值。

- a) 通过训练样本得出类指示矩阵;
- b) 建立超图得到样本与超边的关系矩阵  $J$ , 然后利用式 (6) 得到超图拉普拉斯矩阵  $L_H$ ;
- c) 通过算法 2 求解全局最优解, 得到系数矩阵  $Z^* = A^* B^*$ ;
- d) 通过  $Z^*$  对原始属性集  $X$  进行属性选择后得到的属性集作为样本新的属性集;
- e) 对新的属性集构成的样本采用 SVM 分类, 求出测试样本的多输出预测值;
- f) 计算 aCC 和 aRMSE 并获得均值。

### 2.3 算法优化

本文算法采用交替迭代方法优化。

- a) 定义一个对角矩阵:

$$D^i = \frac{1}{(2/p) \|g^i\|_2^{2-p}} \quad \text{s. t. } i = 1, 2, \dots, d; 0 < p < 2 \quad (12)$$

其中:  $g^i$  表示矩阵  $G^* = A^* B^*$  的第  $i$  行;  $A^*$  和  $B^*$  分别表示  $A$  和  $B$  的最优解; 而  $G^*$  则表示回归系数矩阵  $Z$  的最优解。然后将式 (12) 代入式 (11) 化简可得

$$\min_{A,B} \text{tr}((Y - X^T AB)^T (Y - X^T AB)) + \alpha \text{tr}(B^T A^T DAB) + \beta \text{tr}(B^T A^T X L_H X^T AB)$$

(13)

- b) 本文采用两步交替优化法:

(a) 固定  $A$ , 优化  $B$ 。当固定  $A$  后, 式 (13) 可以看成是关于  $B$  的函数, 令式 (13) 等于  $J(A, B)$ , 并对式 (13) 中的  $B$  求偏导数, 然后令所得偏导数等于零。

$$\frac{\partial J(A, B)}{\partial B} = -A^T X^T Y + A^T X^T XAB + 2\alpha A^T X^T L_H XAB + 2\beta A^T DAB = 0$$

(14)

式 (14) 中可以解出  $B$ :

$$B = (A^T (X^T X + 2\alpha X^T L_H X + 2\beta D) A)^{-1} A^T X^T Y \quad (15)$$

(b) 固定  $B$ , 优化  $A$ 。当固定  $B$  后, 把式 (15) 得到的矩阵  $B$  代入到式 (13) 中得到

$$\max_A \text{tr}((A^T (X^T X + 2\alpha X^T L_H X + 2\beta D) A)^{-1} A^T X^T Y Y^T X A) \quad (16)$$

注意到  $S_i = X^T X + 2\alpha X^T L_H X + 2\beta D$ ,  $S_b = X^T Y Y^T X$  (17) 其中:  $S_i$  和  $S_b$  分别表示 LDA 中的类内离散矩阵和类间离散度矩阵。因此, 通过求解式 (16) 可以得到

$$A = \arg \max_A \{ \text{tr}(A^T S_i A)^{-1} A^T S_b A \} \quad (18)$$

式 (18) 的求解问题明显是 LDA 求解问题。因此, 式 (18) 对应的全局最优解是: 求  $S_i^{-1} S_b$  中非零属性值所对应的属性向量。如果  $S_i$  是奇异的, 则计算  $S_i^+ S_b$  中非零属性值所对应的属性向量, 而  $S_i^+$  表示  $S_i$  的伪逆矩阵。由于系数矩阵的最优解的列空间与  $A$  的列空间相同, 所以提出的新方法同样与特殊的正则化 LDA 具有相同的列空间。

#### 算法 2 优化求解式 (11) 的伪代码

输入: 训练样本  $X \in R^{d \times n}$ , 控制参数  $\alpha, \beta$ , 低秩参数  $r$ , 属性中控制关联结构度的参数  $p$ 。

输出: 矩阵  $A \in R^{d \times r}$  和  $B \in R^{r \times c}$ 。

- a) 初始化  $t = 1$ ;
- b) 通过随机初始化矩阵  $A^{(t)}$  和  $B^{(t)}$ , 得到初始化矩阵  $Z^{(t)} = A^{(t)} B^{(t)}$ ;
- c) 初始化  $D^{(t)} = I \in R^{d \times d}$ ;
- d) 重复;
- e) 通过式 (15) 计算  $B^{(t+1)}$ ;
- f) 通过式 (18) 计算  $A^{(t+1)}$ ;
- g) 更新对角矩阵  $D^{(t+1)} \in R^{d \times d}$ , 第  $i$  个对角元素由式 (12) 计算得出

其中  $D^i = \frac{1}{(2/p) \|g^i\|_2^{2-p}}$ , s. t.  $i = 1, 2, \dots, d; 0 < p < 2$ ;

- h) 更新  $t = t + 1$ ;
- i) 直到式 (11) 收敛;
- j) 结束。

## 3 实验结果和分析

### 3.1 实验数据集和对比算法

本文在六个数据集上测试所提出的属性选择算法用于回归分析的性能, 其中数据集 EDM<sup>[22]</sup> 是火花机数据集; ATP1d 和 ATP7d 是航班票价数据集, 来自文献 [23]; OES97 和 OES10 是美国劳动统计局的数据, 来自文献 [23]; SF1 是太阳耀斑数据集, 源于文献 [24]。数据集详情如表 1 所示。

表 1 数据集信息统计

数据集	样本数	属性数	类数	数据集	样本数	属性数	类数
EDM	154	16	2	OES97	334	263	16
ATP1d	337	411	6	OES10	403	298	16
ATP7d	296	411	6	SF1	323	10	3

实验在 Windows 7 系统上运行, 使用 MATLAB 2014a 软件进行编程实现和实验。实验选用四种对比算法与本文提出的新算法进行比较。对比算法介绍如下:

- a) LSG21 方法<sup>[25]</sup>。这是一种具有稀疏学习能力的新图结

构稀疏算法 通过在系数矩阵上应用  $\ell_{2,1}$ -范数进行约束,但没有采用低秩约束去考虑响应变量间的相关性;时间复杂度为  $O(n^2)$ 。

b) CSFS 方法<sup>[26]</sup>。这是一种凸的半监督多标签属性选择算法,它能够应用  $\ell_{2,1}$ -范数惩罚系数矩阵来进行属性选择,但是忽略了样本之间的关联结构;时间复杂度为  $O(n^2)$ 。

c) SLRR 方法<sup>[21]</sup>。这是一种稀疏低秩回归模型,利用低秩约束寻求数据中的低秩结构和应用  $\ell_{2,1}$ -范数进行特征选择,但是没有在回归模型中嵌入普通图拉普拉斯正则化项,因此不能保持数据之间的局部结构,相比 HLMR-FS 方法而言,回归效果欠佳;时间复杂度为  $O(n^2)$ 。

d) RSR 方法<sup>[27]</sup>。通过自表征的方法选择一个最具代表性的响应矩阵,然后嵌入到稀疏学习模型中进行属性选择,系数矩阵中的系数大小即表示对应属性重要性的强弱,但是未能对样本之间的相关性进行考虑;时间复杂度为  $O(n^2)$ 。

最后,对于 HLMR-FS 方法,由于需要进行特征值分解,所以时间复杂度为  $O(n^3)$ 。为了比较对比算法与 HLMR-FS 算法的性能,本文用平均相关系数(average correlation coefficient, aCC)以及均方差误差(average root mean square error, aRMSE)作为评价指标。其中,多次实验结果的均方差误差的平均值可以反映算法的稳定性,均方差值越小说明算法的稳定性越好。而平均相关系数表示预测值与测试样本目标值之间的相关性,其值越大说明预测值越接近测试目标值,即算法预测的可信度越高。平均相关系数和均方差误差的计算公式<sup>[28]</sup>分别如下:

$$aCC = \frac{1}{d} \sum_{i=1}^d \frac{\sum_{j=1}^{ntest} (y_i^{(j)} - \bar{y}_i) (\hat{y}_i^{(j)} - \hat{\bar{y}}_i)}{\sqrt{\sum_{j=1}^{ntest} (y_i^{(j)} - \bar{y}_i)^2 \sum_{j=1}^{ntest} (\hat{y}_i^{(j)} - \hat{\bar{y}}_i)^2}}$$

$$aRMSE = \frac{1}{d} \sum_{i=1}^d \sqrt{\frac{\sum_{j=1}^{ntest} (y_i^{(j)} - \hat{y}_i^{(j)})^2}{ntest}}$$

其中: ntest 表示测试样本数;  $y_i$  为测试样本目标值;  $\hat{y}_i$  为预测值;  $\bar{y}_i$  为目标值均值;  $\hat{\bar{y}}_i$  为预测值均值。

对比算法和 HLMR-FS 算法的实验都采用 10-折交叉验证法。其首先将整个数据集随机分成均匀的 10 份,然后选择其中的 1 份作为测试集,其余的 9 份作为训练集。并且在 10-折交叉验证的每一折中再进行 5-折交叉验证,使用 LIBSVM 训练来学习回归模型。在模型的参数设置中,本文把控制稀疏的参数  $\alpha$  和控制子空间结构的参数  $\beta$  设置为  $(\alpha, \beta) \in [10^{-3}, 10^{-2}, 10^{-1}, 1, 10^1, 10^2, 10^3]$ ,低秩参数  $r$  设置为  $r < \min(n, \rho)$ ,  $\ell_{2,p}$ -范数中的参数  $p$  在本文实验中均设置为 1,以及 LIBSVM 工具箱中的参数  $(c, g) \in \{2^{-5}, 2^{-4}, \dots, 2^5\}$ 。

### 3.2 实验结果和分析

实验通过 10-折交叉验证的方法把原始数据划分为训练集和测试集,再运用 LIBSVM 内部的 5-折交叉验证并保留每次返回 5-折交叉验证的最好的 aCC 和 aRMSE 结果,所以 10-折交叉验证可获得 10 个 aCC 和 aRMSE 结果,然后分别取 10 个 aCC 和 aRMSE 结果的均值用于比较。对比算法和 HLMR-FS 算法均保证在同一实验环境下运行,其中 LSG21、CSFS、SLRR、RSR 和本文提出的 HLMR-FS 方法在六个数据集上实验结果对比如图 27 所示,具体的 aCC 结果如表 2 所示。

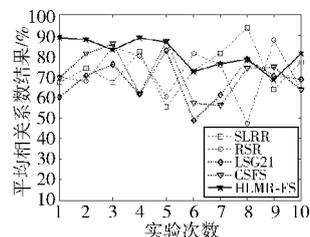


图2 数据集 EDM

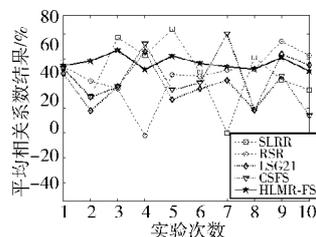


图3 数据集 SF1

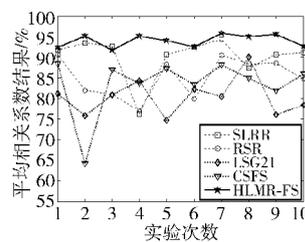


图4 数据集 ATP1d

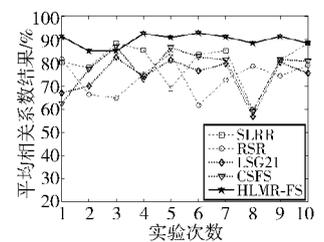


图5 数据集 ATP7d

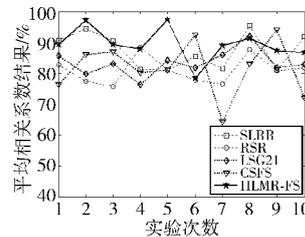


图6 数据集 OES97

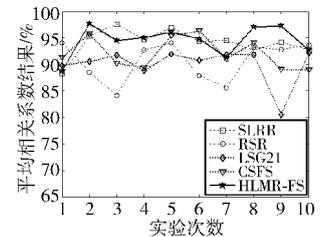


图7 数据集 OES10

表2 平均相关系数(aCC)统计结果

数据集	LSG21	CSFS	SLRR	RSR	HLMR-FS
EDM	72.22	71.05	73.44	73.75	81.19
SF1	50.75	52.95	54.92	53.57	56.57
ATP1d	84.89	83.43	90.05	87.06	94.01
ATP7d	78.74	77.00	79.91	80.77	89.67
OES97	87.85	81.57	87.29	82.98	89.34
OES10	94.50	92.17	94.15	92.68	95.13
平均	78.16	76.36	79.96	78.47	84.32

通过图 27 可以明显地看出,HLMR-FS 算法在六个数据集上的平均相关系数,由于 10-折交叉验证的随机性,并不能保证每次的结果都是最好的,但是每个数据集上 10 次实验结果对比算法大部分都较高,最终的平均分类准确率也是最高的。从表 2 数据可以看出,HLMR-FS 算法与其他四种对比算法在所有数据集上均取得了最高的平均相关系数。具体地,HLMR-FS 算法在 OES10 数据集上取得了最优的结果,其平均相关系数为 95.13%,相比 LSG21、CSFS、SLRR 和 RSR 算法分别高于 0.63%、2.96%、0.98%、2.45%。另外,在 ATP1d 数据集上,HLMR-FS 算法的 aCC 结果为 94.01%,远高于对比算法 LSG21、CSFS、SLRR 和 RSR 的 aCC,分别提高了 9.12%、10.58%、3.96%、9.95%。与 LSG21 算法比较平均提高了 6.16%;与 CSFS 算法比较平均提高了 7.96%;与 SLRR 算法比较平均提高了 4.36%;与 RSR 算法比较平均提高了 5.75%。

HLMR-FS 算法在 SF1 数据集上的效果不是很理想,但总体上依然取得了最好的效果,达到 90% 左右,并且平均能够达到 84.32%,超出对比算法平均值的 5% 左右。

在这六个数据集上的四种对比算法都获得了很好的 aCC 实验结果。其中 LSG21 算法利用图结构稀疏的方法获得多输出回归的结果,但是没有对样本进行选择,并且在构建图结构的过程中计算代价很大,不利于大量数据的多输出回归分析。CSFS 算法通过应用  $\ell_{2,1}$ -范数约束来进行属性选择,但是忽略了数据之间的关联结构,在回归模型上获得的 aCC 结果不是很理想。SLRR 算法通过低秩约束去构建低秩回归模型,并且利用  $\ell_{2,1}$ -范数正则化项惩罚回归系数矩阵来进行特征选择,但是该算法也没有对样本进行选择,数据中的噪声和离群点会干扰回归模型的学习,所以其获得的 aCC 结果比 HLMR-FS 的 aCC 结果差。RSR 算法通过自表征的方法选择一个最具代表性的响应矩阵,然后嵌入到稀疏学习模型中进行属性选择,系数矩阵的元素表示对应属性的重要性,虽然对数据进行了相应的稀疏处理,但是没有考虑类标签与数据之间的相关性,故 aCC 结果不是很理想。

另外,LSG21、CSFS、SLRR、RSR 和本文提出的 HLMR-FS 方法在上述六个数据集上实验的 aRMSE 结果如表 3 所示。

表3 均方差误差(aRMSE)结果

数据集	LSG21	CSFS	SLRR	RSR	HLMR-FS
EDM	0.044 4	0.045 5	0.046 0	0.046 1	0.044 0
SF1	0.050 5	0.042 7	0.041 4	0.021 9	0.020 0
ATP1d	0.015 5	0.015 3	0.008 4	0.009 3	0.006 7
ATP7d	0.014 7	0.015 3	0.009 0	0.008 8	0.007 0
OES97	0.015 7	0.017 4	0.014 8	0.016 3	0.014 6
OES10	0.012 7	0.014 4	0.013 2	0.012 6	0.012 6
平均	0.025 6	0.025 1	0.022 1	0.019 2	0.017 5

从表3的aRMSE实验结果可知,HLMR-FS算法在数据集EDM、SF1、ATP1d、ATP7d、OES97和OES10上获得的aRMSE结果分别为0.044 0、0.020 0、0.006 7、0.007 0、0.014 6和0.012 6,都拥有非常小的aRMSE值,并且均对比算法的aRMSE实验结果低,说明本文提出的HLMR-FS算法具有较强的稳定性。

不同数据集有不同的数据分布,且含有不同的干扰因素。实验结果表明,本文提出的HLMR-FS算法用于回归分析在每个数据集上都要比对比算法好些,对于对比算法来说,HLMR-FS算法具有最高的鲁棒性,同时也说明HLMR-FS算法输出了更具判别力的属性集。同时,与子空间学习方法相比,本文算法不仅保留了提取后的重要属性,还利用子空间学习对重要属性作进一步微调,相比单一的属性选择算法保证了数据自身的全局结构,因此使得本文算法具有更好的解释性。

#### 4 结束语

本文提出了一种综合性的属性选择算法(HLMR-FS算法)用于多回归分析。该算法是在线性回归基础上,为能够去除数据中的噪声和离群点的干扰,利用有效的样本学习模型,该方法在低秩回归上利用 $\ell_{2,1}$ -范数进行样本选择,在低秩方法和稀疏方法中用 $\ell_{2,p}$ -范数共同去除冗余属性,同时嵌入了一个超图拉普拉斯正则化项,弥补了普通图在表示多元关系方面的不足,然后利用LDA算法来对属性选择的结果进行进一步调整,确保得到最优的结果。因此,算法有效地在一个比较经典的回归模型中融合了子空间学习和低秩超图属性选择,既扩展了子空间学习在回归模型上的应用,也弥补了低秩属性选择在保持数据几何结构方面的不足。实验表明,本文算法能够在分类准确率和稳定性上有显著的提高。在以后的工作中,将尝试在半监督属性选择方面进一步扩展本文提出的算法。

#### 参考文献:

- [1] Zhang Shichao, Jin Zhi, Zhu Xiaofeng. Missing data imputation by utilizing information within incomplete instances [J]. *Journal of Systems and Software*, 2011, 84(3): 452-459.
- [2] Zhu Xiaofeng, Suk H I, Lee S W, et al. Subspace regularized sparse multitask learning for multiclass neurodegenerative disease identification [J]. *IEEE Trans on Biomedical Engineering*, 2016, 63(3): 607-618.
- [3] Zhu Xiaofeng, Huang Zi, Shen Hengtao, et al. Dimensionality reduction by mixed kernel canonical correlation analysis [J]. *Pattern Recognition*, 2012, 45(8): 3003-3016.
- [4] Zhu Xiaofeng, Zhang Shichao, Jin Zhi, et al. Missing value estimation for mixed-attribute datasets [J]. *IEEE Trans on Knowledge and Data Engineering*, 2011, 23(1): 110-121.
- [5] Zhang Shichao, Li Xuelong, Zong Ming, et al. Learning  $k$  for KNN classification [J]. *ACM Trans on Intelligent Systems and Technology*, 2017, 8(3): Article No. 43.
- [6] Zhu Xiaofeng, Suk H I, Wang Li, et al. A novel relational regularization feature selection method for joint regression and classification in AD diagnosis [J]. *Medical Image Analysis* 2015 75(6): 570-577.
- [7] Pyatykh S, Hesser J, Lei Zheng. Image noiselevel estimation by principal component analysis [J]. *IEEE Trans on Image Processing*, 2013, 22(2): 687-699.
- [8] Zhang Shichao, Li Xuelong, Zong Ming, et al. Efficient KNN classi-

- fication with different numbers of nearest neighbors [J]. *IEEE Trans on Neural Networks and Learning Systems*, 2017, 29(5): 1774-1785.
- [9] Zhu Xiaofeng, Li Xuelong, Zhang Shichao. Block-row sparse multi-view multilabel learning for image classification [J]. *IEEE Trans on Cybernetics*, 2016, 46(2): 450-461.
- [10] Fan Zizhu, Xu Yong, Zhang D. Local linear discriminant analysis framework using sample neighbors [J]. *IEEE Trans on Neural Networks*, 2011, 22(7): 1119-1132.
- [11] Liimatainen K, Heikkila R, Yli-Harja O. Sparse logistic regression and polynomial modelling for detection of artificial drainage networks [J]. *Remote Sensing Letters*, 2015, 6(4): 311-320.
- [12] Zhu Xiaofeng, Suk H I, Shen Dinggang. A novel matrix-similarity based loss function for joint regression and classification in AD diagnosis [J]. *NeuroImage*, 2014, 100(10): 91-105.
- [13] Sun Hongwei, Wu Qiang. Least square regression with indefinite kernels and coefficient regularization [J]. *Applied & Computational Harmonic Analysis*, 2011, 30(1): 96-109.
- [14] Lai Hanjiang, Pan Yan, Liu Cong, et al. Sparse learning-to-rank via an efficient primal-dual algorithm [J]. *IEEE Trans on Computers*, 2013, 62(6): 1221-1233.
- [15] Zhu Xiaofeng, He Wei, Li Yonggang, et al. One-step spectral clustering via dynamically learning affinity matrix and subspace [C]//Proc of the 31st AAAI Conference on Artificial Intelligence. Palo Alto, CA: AAAI Press, 2017: 2963-2969.
- [16] Zhu Xiaofeng, Huang Zi, Cheng Hong, et al. Sparse hashing for fast multimedia search [J]. *ACM Trans on Information Systems*, 2013, 31(2): 595-605.
- [17] Zhou Dengyong, Huang Jiayuan, Schölkopf B. Learning with hypergraphs: clustering, classification, and embedding [C]//Proc of the 19th International Conference on Neural Information Processing Systems. Cambridge, MA: MIT Press 2006: 1601-1608.
- [18] Zhu Xiaofeng, Huang Zi, Cui Jiangtao, et al. Video-to-shot tag propagation by graph sparse group lasso [J]. *IEEE Trans on Multimedia*, 2013, 15(3): 633-646.
- [19] Ling C X, Yang Qiang, Wang Jianning, et al. Decision trees with minimal costs [C]//Proc of the 21st International Conference on Machine Learning. New York: ACM Press, 2004: 69.
- [20] Xiang Shuo, Zhu Yunzhang, Shen Xiaotong, et al. Optimal exact least squares rank minimization [C]//Proc of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York: ACM Press, 2012: 480-488.
- [21] Cai Xiao, Ding C, Nie Feiping, et al. On the equivalent of low-rank linear regressions and linear discriminant analysis based regressions [C]//Proc of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York: ACM Press, 2013: 1124-1132.
- [22] Karalić A, Bratko I. First order regression [J]. *Machine Learning*, 1997, 26(2): 147-176.
- [23] Spyromitros-Xious E, Groves W, Tsoumakas G, et al. Multilabel classification methods for multi-target regression [EB/OL]. (2012). <https://arxiv.org/pdf/1211.6581v4.pdf>.
- [24] Bache K, Lichman M. UCI machine learning repository [EB/OL]. <http://archive.ics.uci.edu/ml>.
- [25] Cai Xiao, Nie Feiping, Cai Weidong, et al. New graph structured sparsity model for multi-label image annotations [C]//Proc of IEEE International Conference on Computer Vision. Washington DC: IEEE Computer Society, 2013: 801-808.
- [26] Chang Xiaojun, Nie Feiping, Yang Yi, et al. A convex formulation for semi-supervised multi-label feature selection [C]//Proc of the 28th AAAI Conference on Artificial Intelligence. Palo Alto, CA: AAAI Press, 2014: 1171-1177.
- [27] Miao Zhang, Ding C, Ya Zhang, et al. Feature selection at the discrete limit [C]//Proc of the 28th AAAI Conference on Artificial Intelligence. Palo Alto, CA: AAAI Press, 2014: 1355-1361.
- [28] Izenman A J. Reduced-rank regression for the multivariate linear model [J]. *Journal of Multivariate Analysis*, 1975, 5(2): 248-264.
- [29] Sun Liang, Ji Shuiwang, Ye Jieping. Canonical correlation analysis for multilabel classification: a least-squares formulation, extensions, and analysis [J]. *IEEE Trans on Pattern Analysis & Machine Intelligence*, 2011, 33(1): 194-200.