◎模式识别与人工智能◎

基于稀疏学习的低秩属性选择算法

胡荣耀[1,2],刘星毅[2,3],程德波[1,2],何 威[1,2]
HU Rongyao[1,2], LIU Xingyi[2,3], CHENG Debo[1,2], HE Wei[1,2]

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.广西区域多源信息集成与智能处理协同创新中心,广西 桂林 541004
3.广西钦州学院,广西 钦州 535000
1.Guangxi Key Lab of Multi-source Information Mining & Security, Guangxi Normal University, Guilin, Guangxi 541004, China
2.Guangxi Collaborative Innovation Center of Multi-source Information Integration and Intelligent Processing, Guilin, Guangxi 541004, China
3.Qinzhou University, Qinzhou, Guangxi 535000, China

HU Rongyao, LIU Xingyi, CHENG Debo, et al. Low-rank feature selection algorithm based on sparse learning. Computer Engineering and Applications, 2017, 53(10):132-138.

Abstact: The traditional regression model does not ensure to output good performance since it conducts feature selection without considering the correlation between labels. To address this issue, this paper proposes a novel robust low-rank feature selection method. Specifically, this paper considers the correlation between labels into a low-rank regression model and then employs an $l_{2,p}-$ norm regularization term to conduct feature selection. Meanwhile, this paper also considers subspace learning method(i.e., Linear Discriminant Analysis(LDA)) into the proposed feature selection model to adjust the result of feature selection. The iteration between feature selection and LDA enables to output optimal features until the algorithm converges. The experimental results on six public datasets show that the proposed feature selection method outperformed four comparison methods.

Key words: linear regression; linear discriminant analysis; feature selection; subspace learning; sparse learning

摘 要:针对回归模型在进行属性选择未考虑类标签之间关系从而导致回归效果不理想,提出了一种新的具有鲁棒性的低秩属性选择算法。具体为,在线性回归的模型框架下,通过低秩约束来考虑类标签间的相关性和通过稀疏学习理论中的$l_{2,p}-$范数来考虑属性间的关联结构,以此去除不相关的冗余属性的影响;算法通过嵌入子空间学习方法(线性判别分析(LDA))来调整属性选择结果。经实验验证,提出的属性选择算法在六个公开数据集上的效果均优于四种对比算法。

关键词:线性回归;线性判别分析;属性选择;子空间学习;稀疏学习
文献标志码:A **中图分类号:**TP181 doi:10.3778/j.issn.1002-8331.1512-0056

基金项目:国家自然科学基金(No.61170131, No.61450001, No.61263035, No.61363009, No.61573270);广西自然科学基金(No.2012GXNSFGA060004, No.2015GXNSFCB139011);广西研究生教育创新计划项目(No.YCSZ2016046, No.XYCSZ2017064)。

作者简介:胡荣耀(1992—),男,硕士,主要研究领域为数据挖掘、机器学习,E-mail:610273948@qq.com;刘星毅(1972—),男,通讯作者,副教授,主要研究领域为数据挖掘;程德波(1990—),男,硕士,主要研究领域为数据挖掘、机器学习;何威(1989—),男,硕士,主要研究领域为数据挖掘、机器学习。

收稿日期:2015-12-03 **修回日期:**2016-03-04 **文章编号:**1002-8331(2017)10-0132-07
CNKI**网络优先出版:**2016-03-25, http://www.cnki.net/kcms/detail/11.2127.TP.20160325.2025.074.html

1 引言

在模式识别,计算机视觉,数据挖掘等诸多领域,高维度数据随处可见。对高维数据的处理大大增加了数据处理的时间和空间复杂度[1]。因此,在高维数据中进行属性约简以减少数据的维度,寻找到一个规模较小且具有代表性的属性子集成为机器学习的一个重要研究领域[2-4]。属性约简不仅降低了处理时间,而且能够得到更紧凑和更具有泛化能力的学习模型。

现有属性约简方法通常包括属性选择和子空间学习两种[5-7]方法。其中,属性选择是通过预设标准对每个属性进行积分排序,然后输出部分能代表整个属性集的子集,常见的属性选择方法有t-test检验法[8]和稀疏逻辑回归方法[9]等;而子空间学习[10]是通过投影实现高维属性从高维空间到低维空间的映射,来保持数据间的关联结构,常见方法如:线性判别分析(Linear Discriminant Analysis,LDA)[11],典型关联分析(Canonical Correlation Analysis,CCA)[12]算法等。近年来,有学者提出利用低秩回归模型进行属性选择。低秩回归模型通过嵌入一个线性等式在预测和响应两者之间来观测数据,使得属性选择效果[13]更佳,而且已经广泛应用到能够获取默认类在相关的模式中。但低秩回归方法直接在现实应用中易出现以下问题:首先,当输入的数据具有超大的属性时(如:文档数据、人脸数据等),传统的回归模型对于分析如此的高维数据具有很低的性能;其次,一般线性回归模型并不注重在不同的响应之间的相关性,最具代表的是最小二乘回归[14],该方法仅仅是对每个预测的数据分别产生一个响应。

针对以上问题,提出一种结合属性选择和子空间学习(即LDA)的属性约简方法——基于稀疏学习的低秩属性选择算法(Low-rank feature selection algorithm based on sparse learning,SLR_FS)。首先,在线性回归的模型框架中直接进行低秩属性选择,其中,低秩属性选择运用了两种方法:稀疏方法(利用 $l_{2,p}-$ 范数灵活地考虑样本和类标签之间的关系)和低秩方法(通过假设关系矩阵具有低秩,以此来考虑类标签之间的关联性);其次,为了使得提取出的属性更具有代表性,在模型中有效地嵌入经典子空间学习方法——LDA算法,用来对低秩属性选择的结果进行微调;最后,提出一种新的优化方法,即对目标函数按顺序迭代执行低秩属性选择和子空间学习方法的优化求解算法,不断交替的迭代使得结果达到最优,最终取得全局最优解。本文提出的SLR_FS算法结合属性选择和子空间学习各自的优点,经实验验证,该算法在分类准确率上能够达到较好的效果。

2 稀疏学习简介

稀疏学习(Sparse Learning)理论最初应用于图像视觉,由于具有强大的应用价值及内在理论,使得稀疏学习[15]目前得到了迅速的发展,并已经在机器学习与模式识别等领域得到广泛的应用。

在稀疏学习的基本理论中,通过对模型的参数向量 $\boldsymbol{w}\in R^n$ 进行稀疏假设,实现稀疏正则化,再用训练样本对参数 $\boldsymbol{w}$ 进行拟合,主要实现的目标为:

$$\min_{w} g(\boldsymbol{w})=f(\boldsymbol{w})+\lambda\phi(\boldsymbol{w}) \tag{1}$$

其中, $f(\boldsymbol{w})$ 是损失函数, $\phi(\boldsymbol{w})$ 是正则化项, λ 为控制损失函数和正则化项之间的平衡参数,通过改变 λ 的值调节 $\boldsymbol{w}$ 的稀疏性。即 λ 越大,权值矩阵 $\boldsymbol{w}$ 越稀疏,反之亦然。

当稀疏学习应用于属性选择算法中时,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,通过对输入数据进行稀疏约束,使之变得稀疏,这样可以去除冗余属性和不相关属性,使得重要属性能够被保存下来,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,更能够保证本文提出的模型求得唯一的全局最优解[16]。

目标函数(1)中的 $f(\boldsymbol{w})$ 表示为损失函数,常见的损失函数包括:对数损失函数,据对损失函数等;同时, $\phi(\boldsymbol{w})$ 表示为正则化项,常见的正则化项包括 $l_{2,1}-$ 范数, l_1- 范数等,但 $l_{2,p}-$ 范数更具灵活性,通过参数 p 来控制属性间关联结构的度,进而更好地进行属性选择。

3 算法描述及求解

3.1 算法描述

传统的线性回归模型,对于分类问题的处理解决如下:

$$\min_{Z}\|\boldsymbol{Y}-\boldsymbol{X}^{\mathrm{T}}\boldsymbol{Z}\|_F^2 \tag{2}$$

其中, $\boldsymbol{X}=[x_1,x_2,\cdots,x_n]\in\boldsymbol{R}^{m\times n}$ 表示训练样本矩阵, n 和 m 分别表示样本数和属性数, $\boldsymbol{Y}\in\boldsymbol{R}^{n\times k}$ 表示类标签矩阵, n 和 k 分别表示样本数和类别数,比如: $\boldsymbol{Y}_{i,j}=\dfrac{1}{\sqrt{n_j}}$ 代表第 i 个数据点属于第 j 个类别, n_j 表示第 j 个类的样本数。最后模型输出的系数矩阵 $\boldsymbol{Z}\in\boldsymbol{R}^{m\times k}$ 表示训练样本和测试样本间重建的系数矩阵。

通常,大多数事物之间都隐含某种联系,在多类分类问题中,这种关系可能会更加明显[6,10]。因此,为了探测这种隐藏结构,同时利用这种特性来增强已经学习好的模型。为此,算法引入低秩结构来约束系数矩阵 $\boldsymbol{Z}$,即使用 $r<\min(n,k)$ 约束。一般地,秩最小化问题是一个非凸状且NP难问题,但是秩约束的目标却是可解的,即该全局解能够得到的[17]。因此算法在公式(2)的线性回归中,对系数矩阵 $\boldsymbol{Z}$ 使用了低秩约束,所以问题转化为解决如下的问题:

$$\min_{Z}\|\boldsymbol{Y}-\boldsymbol{X}^{\mathrm{T}}\boldsymbol{Z}\|_F^2 \quad \text{s.t.}\ \ rank(\boldsymbol{Z})\leqslant r \tag{3}$$

由于低秩结构对数据子集选择及数据间的结构选择难以保证高效率,文献[18]指出,低秩的线性回归等同于在LDA的子空间中做线性回归。因此,本文算法把公式(3)投影到LDA空间中,且为了对低秩结构更好地优化和有效地考虑类别间的关联结构。算法令 $Z=AB$ 拥有一个低秩 r , $r<\min(n,k)$ 具体模型如下:

$$\min_{A,B}\|Y-X^{\mathrm{T}}AB\|_F^2 \tag{4}$$

其中, $A\in R^{m\times r}$, $B\in R^{r\times k}$ 。因此,重构的系数矩阵 Z ,矩阵 A 和 B 同时具有一个低秩结构。因此,上述的目标函数和公式(3)是等价的,为了具有更加清楚的解释判别投影,公式(4)重写如下:

$$\min_{A,B}\|Y-(A^{\mathrm{T}}X)^{\mathrm{T}}B\|_F^2 \tag{5}$$

上述目标函数表明 A 也能够被作为一个映射使用。而为了使模型更好地应用于属性选择,同时让式(4)模型在选取重要属性时能有效避免维度灾难,在模型中有效的添加 $l_{2,p}-$ 范数来作为正则化项,使得所有的属性既能够保持不同类的低秩结构,并通过 $p(0<p<2)$ 的控制来控制所有属性中不同类之间的关联结构,进一步对属性进行稀疏处理,具体的新模型如下:

$$\min_{A,B}\|Y-X^{\mathrm{T}}AB\|_F^2+\lambda\|AB\|_{2,p} \tag{6}$$

其中, $X\in R^{m\times n}$ (n 和 m 分别表示样本数和属性数)为训练样本, $Y\in R^{n\times k}$ (n 和 k 分别表示样本数和类别数)为类标签或响应矩阵,λ 是调整参数, AB 表示重构系数矩阵,而 $A\in R^{m\times r}$ 和 $B\in R^{r\times k}$,低秩结构意味着 $rank(AB)=r<\min(n,k)$,稀疏结构化表示令矩阵 AB 的大多数的行收缩为零,而属性中对应的非零系数则被作为有代表的属性子集,以此帮助更好的属性选择。

本文算法的伪代码如下:

算法 1 SLR_FS算法伪代码

输入:1.训练数据集 $X\in R^{n\times d}$;

2.正则化参数 λ ;

3.SVM分类器中参数 $(c,g)\in[-10{:}2{:}10]$,分类器类型选择线性模式;

输出:分类准确率;

1.通过训练样本得出类指示矩阵 $Y=[y_{i,k}]$;

2.依据所选择的模型式(6):调用算法2求解全局最优解得到自表征矩阵 $Z^*=A^*B^*$;

5.利用最优解 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;

6.对新的属性集构成的样本采用SVM分类。

3.2 算法优化

首先,把公式(6)化简如下:

$$\min_{A,B}\|Y-X^{\mathrm{T}}AB\|_F^2+\lambda\mathrm{Tr}(B^{\mathrm{T}}A^{\mathrm{T}}DAB) \tag{7}$$

其中, $D\in R^{m\times m}$ 是一个对角矩阵, D 中的每个元素定义如下:

$$d_{ii}=\frac{1}{(2/p)(\|g^i\|_2)^{2-p}},\text{s.t. } i=1,2,\cdots,m,0<p<2 \tag{8}$$

其中, g^i 表示矩阵 $G^*=A^*B^*$ 的第 i 行, A^* 和 B^* 分别为 A 和 B 的最优解,而 G^* 为重构系数矩阵 Z 的最优解。令公式(6)等于 $J(A,B)$,并对公式(6)中的 B 求偏导数,令所得偏导数等于零:

$$\frac{\partial J(A,B)}{\partial B}=-2A^{\mathrm{T}}XY+2A^{\mathrm{T}}XX^{\mathrm{T}}AB+2\lambda A^{\mathrm{T}}DAB=0 \tag{9}$$

从上面的等式求解可以得到 B :

$$B=(A^{\mathrm{T}}(XX^{\mathrm{T}}+\lambda D)A)^{-1}A^{\mathrm{T}}XY \tag{10}$$

然后,把得到的矩阵 B 的表达式代入到公式(7)中得到:

$$\max_{A}\mathrm{Tr}((A^{\mathrm{T}}(XX^{\mathrm{T}}+\lambda D)A)^{-1}A^{\mathrm{T}}XYY^{\mathrm{T}}X^{\mathrm{T}}A) \tag{11}$$

注意到:

$$S_t=XX^{\mathrm{T}},S_b=XYY^{\mathrm{T}}X^{\mathrm{T}} \tag{12}$$

其中, S_t 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,通过求解公式(11)可以得到 A 如下:

$$A=\arg\max_{A}\{\mathrm{Tr}((A^{\mathrm{T}}(S_t+\lambda D)A)^{-1}A^{\mathrm{T}}S_bA)\} \tag{13}$$

公式(13)的求解问题明显是LDA求解问题,因此,上式对应的全局最优解是:求 $S_t^{-1}S_b$ 中非零属性值所对应的属性向量。如果 S_t 是奇异的,则计算 $S_t^{+}S_b$ 中非零属性值对应的属性向量,而 S_t^{+} 表示 S_t 的伪逆矩阵。由于系数矩阵的最优解 $G^*=A^*B^*$ 的列空间和A的列空间相同,所以提出的新的方法同样与特殊的正则化LDA具有同样的列空间。

本文提出的SLR_FS算法具有如下优点:

(1)线性回归模型有效结合 $l_{2,p}-$ 范数来考虑属性及其对应响应变量间的相关性,来对原始属性集通过稀疏处理进行属性选择(其中,稀疏处理可以保留重要属性以达到选择的目的,以此消除输入数据中的冗余属性和不相关属性),同时,利用低秩约束来考虑响应变量间的相关性以进行子空间选择。

(2)在新的回归模型中有效地融入子空间学习中的线性判别分析(LDA)算法来调整子空间属性选择的结果,以此保证得到的属性选择结果达到最优。因此,使得提出的算法比单一的子空间学习或者属性选择方法具有更强的分类能力。

(3)子空间学习在迭代中被设计为适应低秩属性选择后的结果,使得选择后的属性对应的响应矩阵具有一致性,在选出代表的属性和其对应的响应变量之间进行非线性学习,因此可以扩展算法的应用范围。

(4)对本文的目标函数提出了一种区别于交替方向乘子法的求解方法,即固定低秩属性选择的结果,以此提高子空间学习的性能,接着,固定子空间学习的结果,确保低秩属性选择能输出更具代表性的属性集。该优化算法能保证目标值在每次迭代过程中逐步收敛趋近

与全局最优解,最终取得全局最优解。

算法2 优化求解式(6)的伪代码

输入:

1. 训练数据集 $X\in R^{m\times n}$;
2. 类标签数据集 $Y\in R^{n\times k}$;
3. 低秩参数 r;
4. 属性中控制关联结构度的参数 p;

输出:矩阵 $A\in R^{m\times r}$ 和 $B\in R^{r\times k}$;

1. 初始化 $t=0$;
2. 通过随机初始化矩阵 $A^{(t)}$ 和 $B^{(t)}$,得到初始化 $Z^{(t)}=A^{(t)}B^{(t)}$;
3. 初始化 $D^{(t)}=I\in R^{m\times m}$;
4. 重复:
5. 通过等式(13)计算 $A^{(t+1)}$;
6. 通过等式(10)计算 $B^{(t+1)}$;
7. 更新对角矩阵 $D^{(t+1)}\in R^{m\times m}$,第 i 个对角元素由公式(8)计算得出,

其中 $g^i=[A(t+1)B(t+1)]^i$;

8. 更新 $t=t+1$;
9. 直到公式(6)收敛;
10. 结束;

4 优化算法收敛证明

由于公式(6)是非平滑的,且有两个变量 A 和 B 需要优化,而正则化项也是非平滑的,使得该目标函数求解更加难以解决。但本文提出使用一个简单的有效的算法来求解问题,下文给出算法1收敛证明的具体步骤。

具体步骤如下:

从算法2中的步骤4至步骤10,能够得到第 t 次迭代对应的结果:

$$<A^{(t+1)},B^{(t+1)}>=\arg\min_{A,B}\|Y-X^{\mathrm{T}}AB\|_F^2+\lambda\mathrm{Tr}(B^TA^{\mathrm{T}}D^{(t)}AB) \quad (14)$$

由此得到:

$$\|Y-X^{\mathrm{T}}A^{(t+1)}B^{(t+1)}\|_F^2+\lambda\mathrm{Tr}(B^{(t+1)^T}A^{(t+1)^T}D^{(t)}A^{(t+1)}B^{(t+1)})\leqslant \|Y-XA^{(t)}B^{(t)}\|_F^2+\lambda\mathrm{Tr}(B^{(t)^T}A^{(t)^T}D^{(t)}A^{(t)}B^{(t)}) \quad (15)$$

其中,注意 $G^{(t)}=A^{(t)}B^{(t)}$,$G^{(t+1)}=A^{(t+1)}B^{(t+1)}$,由公式(8)组成的对角矩阵 D,代入不等式(15)可以得到:

$$\|Y-X^{\mathrm{T}}G^{(t+1)}\|_F^2+\lambda\sum_{i=1}^{m}\frac{(\|g^{i(t+1)}\|_2)^{2(2-p)}}{(2/p)(\|g^{i(t)}\|_2)^{2-p}}\leqslant \|Y-X^{\mathrm{T}}G^{(t)}\|_F^2+\lambda\sum_{i=1}^{m}\frac{(\|g^{i(t)}\|_2)^{2(2-p)}}{(2/p)(\|g^{i(t)}\|_2)^{2-p}} \quad (16)$$

其中,$g^{i(t)}$ 和 $g^{i(t+1)}$ 分别是 $G^{(t)}$ 和 $G^{(t+1)}$ 的第 i 行。而对于每个i,可以得到以下不等式:

$$(\|g^{i(t+1)}\|_2)^{2-p}-\frac{(\|g^{i(t+1)}\|_2)^{2(2-p)}}{(2/p)(\|g^{i(t)}\|_2)^{2-p}}\leqslant (\|g^{i(t)}\|_2)^{2-p}-\frac{(\|g^{i(t)}\|_2)^{2(2-p)}}{(2/p)(\|g^{i(t)}\|_2)^{2-p}} \quad (17)$$

然后,通过上面的不等式乘以正则化参数 λ 得到累加的总和后,就可以获得:

$$\lambda\sum_{i=1}^{m}(\|g^{i(t+1)}\|_2)^{2-p}-\frac{(\|g^{i(t+1)}\|_2)^{2(2-p)}}{(2/p)(\|g^{i(t)}\|_2)^{2-p}}\leqslant \lambda\sum_{i=1}^{m}(\|g^{i(t)}\|_2)^{2-p}-\frac{(\|g^{i(t)}\|_2)^{2(2-p)}}{(2/p)(\|g^{i(t)}\|_2)^{2-p}} \quad (18)$$

最后,联合不等式(16)和不等式(18)就可以得到:

$$\|Y-X^{\mathrm{T}}G^{(t+1)}\|_F^2+\lambda\sum_{i=1}^{m}(\|g^{i(t+1)}\|_2)^{2(2-p)}\leqslant \|Y-X^{\mathrm{T}}G^{(t)}\|_F^2+\lambda\sum_{i=1}^{m}(\|g^{i(t)}\|_2)^{2(2-p)} \quad (19)$$

综合上述,得到了如下结果:

$$\|Y-X^{\mathrm{T}}G^{(t+1)}\|_F^2+\lambda\|G^{(t+1)}\|_{2,p}\leqslant \|Y-X^{\mathrm{T}}G^{(t)}\|_F^2+\lambda\|G^{(t)}\|_{2,p} \quad (20)$$

上述不等式表明算法1在每次迭代中,目标函数都是单调递减的,因此,算法1可以达到收敛的结果。

5 实验与结果分析

5.1 实验数据集和对比算法

本文用六个数据集测试提出的SLR_FS属性选择算法性能,其中数据集DBWorld,madelon,isolets来自UCI[19],PCMAC来自属性选择数据集[20],Yale-32和ORL都是人脸数据集,分别来自文献[21]和[22]。

表1 数据集信息统计

数据集	样本数	属性数	类数	类型
DBWorld	64	4 702	2	Text
madelon	2 000	500	2	Multivariate
PCMAC	1 943	3 289	2	Text
Yale-32	165	1 024	15	Image,Face
isolets	1 559	617	26	Multivariate
ORL	400	1 024	40	Image,Face

实验运行在Win7系统上,使用Matlab2014a软件进行编程实现和实验。实验选用七种对比算法与本文提出的新的算法进行比较。对比算法详细介绍如下:

NFS方法(Non Feature Selection):对原始数据不做任何处理,直接使用LIBSVM工具箱[23]进行SVM分类,相比SLR_FS等属性选择算法,不仅数据处理量大,而且容易受到噪声数据和冗余数据的影响,时间复杂度为 $O(n^3)$。

LDA(Linear Discriminant Analysis):在对数据处理过程中,仅考虑数据之间的全局结构,且LDA算法引用在本文提出的算法中对选择的最终属性进行微调,相比SLR_FS方法不但数据处理量大,且易受噪声等冗余数据的影响,时间复杂度为 $O(n^3)$。

LR(Linear Regression):利用最小平方函数对一个或多个自变量和因变量之间进行建模的一种回归分析,由

于未考虑 l_F- 范数对噪声数据的敏感性，相比 SLR_FS 方法，受到冗余数据和离群数据影响较大，时间复杂度为 $O(n^2)$；

LGR（Logistic Regression）：是通过放弃最小二乘法的无偏性，专用于共线性数据分析的有偏估计的回归方法，相比 SLR_FS 方法，虽然利用 l_F- 范数先对原始数据进行预处理，但该算法没有考虑数据与数据之间的关联结构，对于多类问题，不同类别的数据之间关联结构是非常重要地，时间复杂度为 $O(n^2)$。

RSR 方法[24]：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，同时，系数矩阵中的系数大小即表示对应属性重要性的强弱，相比 SLR_FS 方法，虽然对数据进行了相应的稀疏处理，但是，由于缺乏类标签的有效辅助，同时，未能对数据之间相关性进行考虑，故效果不是很理想，时间复杂度为 $O(n^2)$。

MI 方法[25]：先通过对原始数据进行对数压缩，然后选取一个映射函数的方法对原始数据进行映射处理，并按得分高低进行属性重要性的选取，相比 SLR_FS 方法，该方法既对数据进行了预处理，还充分考虑数据与类标签之间的依赖性，但是，该方法未能充分利用不同类别数据之间的相关结构，时间复杂度为 $O(n)$。

SD 方法[25]：通过一个映射函数对原始数据进行处理，得到的分数高低对应属性重要性强弱，最后对排序靠前的属性进行提取，相比 SLR_FS 方法，该方法没有对数据进行预处理去除噪声和冗余数据，同时，没有对不同类别数据与数据之间的关联结构进行处理分析，时间复杂度为 $O(n)$。

最后，对于 SLR_FS 方法，由于需要进行特征值分解，故时间复杂度为 $O(n^2)$。

对于所有对比算法，采用 10-折交叉验证方法选择参数，且 SVM 分类器中参数 $(c,g)\in[-10{:}2{:}10]$。NFS、LDA、LR、LGR 和 RSR 五种对比算法中的正则化参数 λ 设置为 $\lambda\in[10^{-3},10^{-2},10^{-1},1,10^{1},10^{2},10^{3}]$，MI 和 SD 两种对比算法分层参数根据原文提示使用默认值 12；本文 SLR_FS 算法中 1 与上述五种对比算法中的设置一致，且低秩参数 r 设置为 $r<\min(n,k)$。

5.2　实验结果和分析

本文采用十折交叉验证方法将原始数据分成训练集和测试集，再利用 SVM 进行分类，同时，把分类准确率作为评价指标，即分类准确率越高表示算法效果越好。所有算法均保证在同一实验环境下进行，最后提取 10 次运行的实验结果的均值加减均方差来评估各算法的性能，在每个数据集上运行 10 次以减少实验误差，以此来增加实验的稳定性。各算法在 6 个数据集上实验结果对比图见图 1~图 6，具体实验数据结果见表 2。

如图 1~图 6 所示，可以清楚地看到 SLR_FS 算法在 6 个数据集上的分类准确率，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 次实验结果比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。

对于多类的数据集，比如：isolets，Yale-32 和 ORL 数

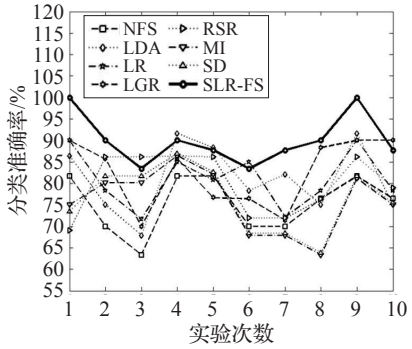

图 1　DBWorld **数据集**

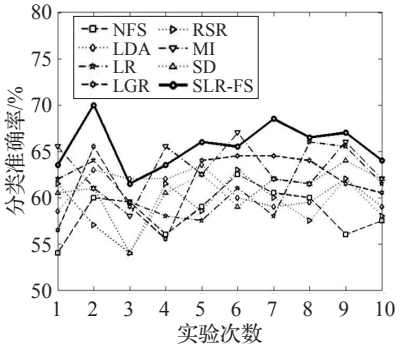

图 2　madelon **数据集**

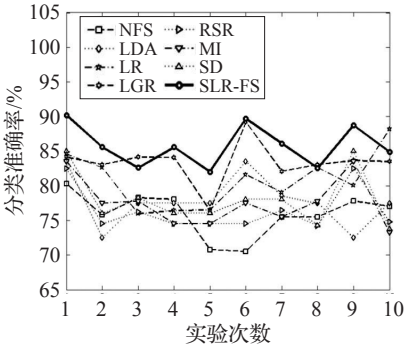

图 3　PCMAC **数据集**

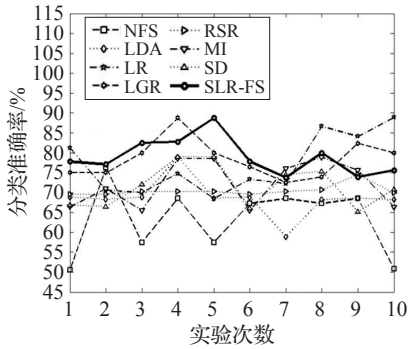

图 4　Yale-32 **数据集**

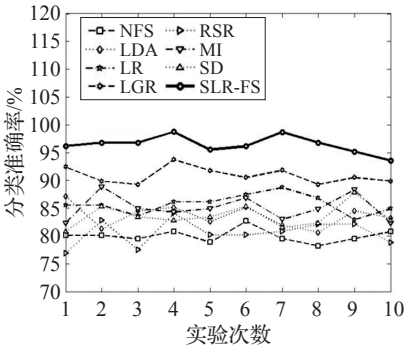

图 5　isolets **数据集**

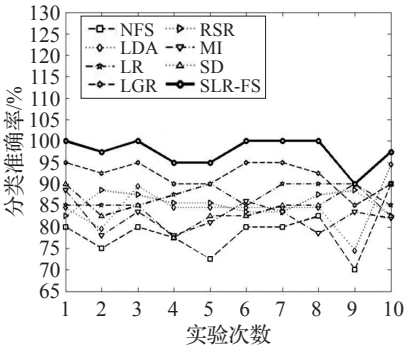

图 6　ORL **数据集**

表2 准确率(均值±方差)的统计结果 %

数据集	NFS	LDA	LR	LGR	RSR	MI	SD	SLR_FS
DBWorld	75.29±6.60	81.14±8.16	80.95±6.51	82.43±7.98	79.90±7.15	75.90±7.47	76.40±7.67	90.00±5.79
madelon	58.50±3.64	60.85±2.90	61.30±3.11	61.55±3.57	59.30±2.79	63.10±2.81	60.70±2.71	65.60±2.56
PCMAC	75.95±3.16	77.79±3.69	80.76±3.16	83.33±3.15	76.46±3.27	77.54±3.53	78.04±3.98	85.80±2.96
Yale-32	63.27±6.12	68.63±4.94	77.02±5.14	78.40±4.85	70.47±5.04	72.21±5.86	71.51±5.32	79.00±4.64
isolets	80.01±2.25	83.60±2.02	85.77±1.71	90.89±1.49	80.53±2.32	85.05±2.32	83.55±2.08	96.44±1.49
ORL	78.75±5.56	84.50±5.27	87.25±2.49	92.00±3.29	85.50±2.40	82.75±3.49	84.25±3.74	97.50±2.36
平均值	71.96	76.09	78.84	81.43	75.36	76.09	75.74	85.72

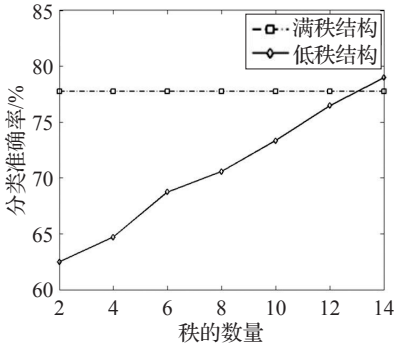

图7 Yale-32数据集

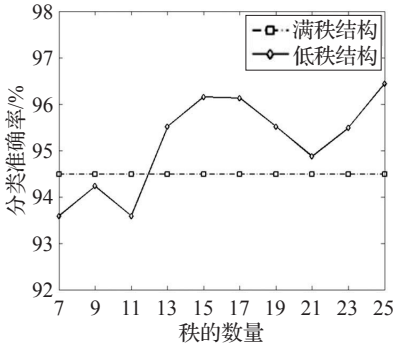

图8 isolets数据集

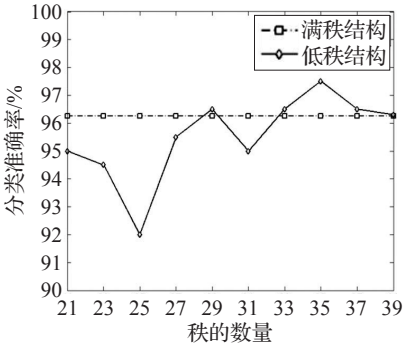

图9 ORL数据集

据集,可以通过控制低秩的数量来得到不同的分类结果,通过图7~图9所示,具有低秩结构的SLR_FS算法比满秩的效果更佳,分类准确率更高。

通过表2分析得出,SLR_FS算法与其他四种对比算法比较均取得最高的分类准确率,具体地,与NFS方法比较平均提高了13.76%,与LGR算法对比平均提高了4.29%。由于madelon数据集中的数据全部是正整数且数据之间较为接近,故五种算法所得准确率比较接近,SLR_FS算法最好提高了7.1%。关于同类型分析,除了在文本类型两个数据集结果较为接近,只相差4.2%的准确率,而在多元类型上的两个数据集准确率相差较大。在ORL数据集上,SLR_FS算法效果最佳,不但与NFS相比提高了18.75%,而且比LGR算法提高了5.5%。这是由于SLR_FS利用低秩属性选择方法,包括用 $l_{2,p}-$ 范数考虑样本和类标签之间关系,还用低秩约束来考虑类标签间的关联性,最后利用线性判别分析调整属性选择的结果,保证数据的几何结构,故能显著提高模型的性能。

由于NFS方法没有预先对数据进行建模预处理,所以没有对应的建模时间。对于其他七种算法的建模时间对比如表3所示。

表3 每次建模运行时间的统计结果 s

数据集	LDA	LR	LGR	RSR	MI	SD	SLR_FS
DBWorld	133.627	18.937	14.064	3.605	0.380	0.3222	46.745
madelon	0.190	0.329	0.218	3.460	0.465	0.4222	0.661
PCMAC	35.638	24.045	3.151	14.468	0.465	0.4222	18.154
Yale-32	8.145	1.777	1.677	1.134	0.271	0.224	3.318
isolets	2.738	0.551	0.384	4.171	0.533	0.479	0.733
ORL	9.199	2.245	1.526	2.310	0.453	0.363	3.200

由于不同数据集有不同的数据分布,且含有不同的干扰因素。实验结果显示,本文提出的SLR_FS算法在每个数据集上均比对比算法好,相对所有对比算法,本文SLR_FS算法具有最强的鲁棒性。同时,跟子空间学习方法比较,本文算法不仅保持提取后重要的属性,还利用子空间学习对重要属性进一步微调,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。

6 结束语

本文提出一种全新的属性选择算法——SLR_FS算法。该算法是基于线性回归的框架,通过低秩属性选择,包括低秩方法和稀疏方法中 $l_{2,p}-$ 范数共同去除冗余属性,同时,利用线性判别分析(LDA)对属性选择的结果进行微调,保证得到的结果达到最优。因此,算法有效地融合子空间学习和低秩属性选择在一个回归模型中。既扩展了子空间学习在回归模型上应用范畴,也扩展了低秩属性选择在保持数据几何结构方面的不足。而且实验显示本文算法能够在分类准确率和稳定性上取得很好的分类效果。在将来的工作中,将尝试在半监督属性选择方面拓展本文提出算法的应用范畴,并尝试使用自表征、图稀疏等技术来改进算法。

参考文献:

[1] Zhu X, Huang Z, Shen H T, et al.Dimensionality reduction by mixed kernel canonical correlation analysis[J].Pattern Recognition, 2012, 45(8):3003-3016.

[2] Zhu X, Zhang S C, Jin Z, et al.Missing value estimation

for mixed-attribute data sets[J].IEEE Transactions on Knowledge & Data Engineering,2010,23(1):110-121.

[3] Zhu X,Huang Z,Yang Y,et al.Self-taught dimensionality reduction on the high-dimensional small-sized data[J].Pattern Recognition,2013,46(1):215-229.

[4] Zhu X,Suk H I,Shen D.Matrix-similarity based loss function and feature selection for alzheimer's disease diagnosis[C]// 2014 IEEE Conference on Computer Vision and Pattern Recognition(CVPR).IEEE,2014:3089-3096.

[5] Zhu X,Huang Z,Shen H G,et al.Linear cross-modal hashing for efficient multimedia search[C]//Proceedings of the 21st ACM International Conference on Multimedia,2013:143-152.

[6] Zhu X,Huang Z,Cui J,et al.Video-to-shot tag propagation by graph sparse group lasso[J].IEEE Transactions on Multimedia,2013,15(3):633-646.

[7] Zhu X,Zhang L,Huang Z.A sparse embedding and least variance encoding approach to hashing[J].IEEE Transactions on Image Processing,2014,23(9):3737-3750.

[8] Konietschke F,Pauly M.Bootstrapping and permuting paired *t*-test type statistics[J].Statistics & Computing,2013,24(3):283-296.

[9] Liimatainen K,Heikkila R,Yli-Harja O.Sparse logistic regression and polynomial modelling for detection of artificial drainage networks[J].Remote Sensing Letters,2015,6:311-320.

[10] Zhu X,Huang Z,Cheng H,et al.Sparse hashing for fast multimedia search[J].Acm Transactions on Information Systems,2013,31(2):595-605.

[11] Fan Z,Xu Y,Zhang D.Local linear discriminant analysis framework using sample neighbors[J].IEEE Transactions on Neural Networks,2011,22(7):1119-1132.

[12] Sun L,Ji S,Ye J.Canonical correlation analysis for multilabel classification:a least-squares formulation,extensions,and analysis[J].IEEE Transactions on Pattern Analysis & Machine Intelligence,2011,33(1):194-200.

[13] Zhu X,Suk H I,Shen D G.A novel matrix-similarity based loss function for joint regression and classification in AD diagnosis[J].NeuroImage,2014,100:91-105.

[14] Sun H,Wu Q.Least square regression with indefinite kernels and coefficient regularization[J].Applied & Computational Harmonic Analysis,2011,30(1):96-109.

[15] Lai H,Pan Y,Liu C,et al.Sparse learning-to-rank via an efficient primal-dual algorithm[J].IEEE Transactions on Computers,2013,62(6):1221-1233.

[16] Adel A A.Theory and methodology on the global optimal solution to a general reverse logistics inventory model for deteriorating items and time-varying rates[J].Computers & Industrial Engineering,2011,60(2):236-247.

[17] Xiang S,Zhu Y,Shen X,et al.Optimal exact least squares rank minimization[C]//Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining,2012:480-488.

[18] Cai X,Ding C,Nie F,et al.On the equivalent of low-rank linear regressions and linear discriminant analysis based regressions[C]//Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.ACM,2013:1124-1132.

[19] UCI repository of machine learning datasets[EB/OL].[2015-04-10].http://archive.ics.uci.edu/ml/.

[20] Feature selection datasets[EB/OL].[2015-04-10].http://featureselection.asu.edu/datasets.php.

[21] Song Y,Morency L P,Davis R.Learning a sparse codebook of facial and body microexpressions for emotion recognition[J].Association for Computing Machinery,2013:237-244.

[22] Chen Y,Chiang J.Face recognition using combined multiple feature extraction based on Fourier-Mellin approach for single example image per person[J].Pattern Recognition Letters,2010,31(13):1833-1841.

[23] LIBSVM—A Library for Support Vector Machines[EB/OL].[2015-04-10].http://www.csie.ntu.edu.tw/~cjlin/libsvm.

[24] Zhu P,Zuo W,Zhang L,et al.Unsupervised feature selection by regularized self-representation[J].Pattern Recognition,2015,48(2):438-446.

[25] Pohjalainen J,Räsänen O,Kadioglu S.Feature selection methods and their combinations in high-dimensional classification of speaker likability,intelligibility and personality traits[J].Computer Speech & Language,2013,29(1):145-171.