

基于属性自表达的低秩超图属性选择算法*

苏毅娟¹, 雷聪²⁺, 胡荣耀², 何威², 朱永华³

(1. 广西师范学院 计算机与信息工程学院, 南宁 530023; 2. 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; 3. 广西大学 计算机与电子信息学院, 南宁 530004)

摘要: 针对高维数据具有低秩形式和属性冗余等特点, 提出一种基于属性自表达的无监督超图属性选择算法。该算法首先利用属性自表达特点用其他属性稀疏地表达每个属性, 此自表达形式使用低秩假设寻找高维数据的低秩表示, 然后建立超图正则化因子保持高维数据的局部结构, 最后利用稀疏正则化因子进行属性选择。属性自表达特性确定属性的重要性, 低秩表示相当于考虑数据的全局信息进行子空间学习, 超图正则化因子考虑数据的局部结构对数据进行子空间学习。该算法实际上考虑数据全局和局部信息进行子空间学习, 更是一种嵌入了子空间学习的属性选择算法。实验结果表明, 该算法相比其他对比算法, 能更有效地选取属性, 并能取得很好的分类效果。

关键词: 属性选择; 属性自表达; 子空间学习; 超图; 低秩表示

中图分类号: TP391 文献标志码: A 文章编号: 1001-3695(2017)08-2294-05

doi:10.3969/j.issn.1001-3695.2017.08.012

Low rank hypergraph feature selection algorithm based on self-representation

Su Yijuan¹, Lei Cong²⁺, Hu Rongyao², He Wei², Zhu Yonghua³

(1. College of Computer & Information Engineering, Guangxi Teachers Education University, Nanning 530023, China; 2. Guangxi Key Laboratory of Multi-source Information Mining & Security, Guangxi Normal University, Guilin Guangxi 541004, China; 3. School of Computer, Electronics & Information, Guangxi University, Nanning 530004, China)

Abstract: Due to that high-dimensional data usually is low-rank and contains redundant features, this paper proposed a novel unsupervised hypergraph feature selection algorithm based on self-representation property of features. First, it considered the self-representation matrix to sparsely represent each feature by a linear combination of other features. Such self-representation property was then enforced a low-rank assumption to learn the low-rank representation of high-dimensional data, via considering the global structure of the data to conduct subspace learning. Second, it considered the local structure of the data by a hypergraph based regularizer. In this way, the proposed method integrated subspace learning into the framework of feature selection. Experimental results demonstrate that the proposed can select the best discriminative features and achieve the best classification performance, compared to the competing methods.

Key words: feature selection; self-representation; subspace learning; hypergraph; low-rank representation

0 引言

计算机视觉和数据挖掘等领域通常用高维特征表示数据^[1]。高维数据在能准确描述数据的同时也带来了很多问题, 如属性冗余、增大存储空间等。为了有效地利用这些高维数据, 且减少相应的处理时间, 对数据的预处理就显得尤为重要。因此, 对高维数据进行属性约简以缩小数据的维度^[2], 发现一个具有代表性且规模较小的子集成为一个重要的研究领域^[3]。

属性约简方法包括子空间学习和属性选择^[4,5]。子空间学习是通过投影矩阵将高维数据投影到低维空间, 以此来保持数据之间的关联结构。常见的方法有局部保留投影(locality preserving projection, LPP)^[6]、主成分分析(principal component analysis, PCA)^[7]、线性判别式分析(linear discriminant analysis,

LDA)算法等^[8]。而属性选择是通过一种搜索标准从原始的属性中找到一个最能表示原始特征的子集。它是一种很重要的技术, 被广泛地运用到模式识别和机器学习等领域^[9,10]。属性选择中比较常用的方法有 t-test 检验法^[11]和稀疏逻辑回归方法^[12]等。

研究表明, 属性选择方法更具有解释性, 而子空间学习方法通常比属性选择效果明显。为此, 本文首先结合属性选择和子空间学习进行属性约简以获取解释性和显著的效果。具体地说, 本文使用属性的自表达性和稀疏正则化因子对属性的重要性进行排序, 其中考虑数据的全局结构(通过低秩假设实现)和局部结构(通过超图正则化因子实现)进行子空间学习, 提出一种更加高效的属性选择算法, 称做基于属性自表达的低秩超图属性选择算法(low rank hypergraph feature selection algo-

收稿日期: 2016-05-27; 修回日期: 2016-07-05 基金项目: 国家自然科学基金资助项目(61450001, 61263035, 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(YCSZ2016045, YCSZ2016046, XYCSZ2017064, XYCSZ2017067, YCSW2017065)

作者简介: 苏毅娟(1976-), 女, 广西桂林人, 副教授, 主要研究方向为机器学习、数据挖掘; 雷聪(1991-), 男(通信作者), 湖北大冶人, 硕士, 主要研究方向为数据挖掘、机器学习(542832480@qq.com); 胡荣耀(1992-), 男, 江西景德镇人, 硕士, 主要研究方向为数据挖掘、机器学习; 何威(1989-), 男, 河南商丘人, 硕士, 主要研究方向为数据挖掘、机器学习; 朱永华(1994-), 男, 广西桂林人, 硕士, 主要研究方向为数据挖掘。

rithm based on the self-representation, BHSLR_FS)。

本文首先通过属性自表达和偏差项,得到一个自表达系数矩阵,从而解决了无监督属性选择方法中无类标签的特点;其次在线性回归的模型框架中进行低秩属性选择,其中,低秩属性选择运用了稀疏(利用 $l_{2,1}$ -范数来去除冗余的属性)和低秩(假设矩阵具有低秩,以此来排除噪声的干扰)两种方法,比子空间学习方法具有更好的解释性;接着嵌入一个基于超图的拉普拉斯矩阵来保持属性之间更深层次的关系,使模型能够更好地保留数据的局部信息,同时结合子空间学习方法——LDA 算法(考虑到数据的全局信息),可以对低秩属性选择的结果进行微调,进而帮助提高属性选择的效果;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,不断交替地迭代使得结果达到最优,最终取得全局最优解。由于本文同时考虑到了数据的全局结构和局部结构,所以比单一的子空间学习方法具有更好的效果。经实验验证,该算法在分类准确率上能够达到较好的效果。

1 相关理论背景及简介

1.1 稀疏学习简介

稀疏学习(sparse learning)^[13]最早主要应用于图形、图像视觉等领域^[14]。由于具有强大的应用价值及内在理论支撑,所以稀疏学习得到了迅速的发展。

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

$$\min_Z g(Z) = f(Z) + \lambda \phi(Z) \quad (1)$$

其中: $f(Z)$ 是损失函数; $\phi(Z)$ 是正则化项; λ 调节 Z 的稀疏性,且 λ 越大, Z 越稀疏,反之亦然。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^[15]。在稀疏学习中, l_0 -范数是最有效的稀疏正则因子,但因其求解为 NP 难,所以很多文献均采用近似正则项 l_1 -范数来替代 l_0 -范数;而 $l_{2,1}$ -范数能导致行稀疏,已经被证明比 l_1 -范数更适合于属性选择。因此本文采用 $l_{2,1}$ -范数作为稀疏正则化因子来对属性自表达进行组稀疏处理,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

一个超图被定义为一个三元组 $G_H = (V, E, w)$ 。其中: $V = \{1, \dots, n\}$ 是顶点的集合; E 是超边集合; w 是超边权重矩阵。每一个超边都有一个权重 $w(e_i)$, 这里设置为 1。如图 1(a) 所示,超边 e_1 是由顶点 v_1, v_2, v_3 组成,超边 e_2 是由顶点 v_3, v_4, v_5 组成,而超边 e_3 是由顶点 v_5, v_6 组成。超图 G_H 的连接关系表示为 $|V| \times |E|$ 的关联矩阵 J , 定义(图 1(b))为

$$J(v, e) = \begin{cases} 1 & \text{if } v \in e \\ 0 & \text{otherwise} \end{cases} \quad (2)$$

根据关联矩阵 J , 每一个顶点 v 的度表示为

$$d(v) = \sum_{e \in E} w(e) J(v, e) \quad (3)$$

每个超边 e 的度表示为

$$\delta(e) = \sum_{v \in V} J(v, e) \quad (4)$$

定义归一化的超图拉普拉斯矩阵^[16]:

$$L_H = I - D_v^{-\frac{1}{2}} J W_H D_e^{-1} J^T D_v^{-\frac{1}{2}} \quad (5)$$

其中: I 表示 $|V| \times |V|$ 的单位矩阵; D_e 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。

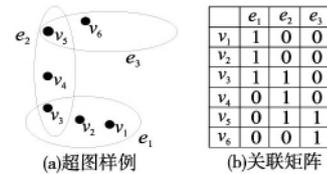


图1 超图样例和对应的关联矩阵

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

假设给定训练集 $X \in R^{n \times m}$, 其中: n 和 m 分别表示样本数和属性数。一般的属性选择方法通常是先计算样本间的相似性或流形结构,然后建立一个响应矩阵 Y ,使得属性选择问题转换为一个多输出的回归问题^[17]。但是要选取一个合适的响应矩阵是比较困难的。由于不同属性之间存在某种联系,所以本文算法用 X 作为响应矩阵,通过属性自表达来得到属性之间的关系:

$$X = XZ + eb \quad (6)$$

其中: $Z \in R^{m \times m}$ 表示自表达系数矩阵; $e \in R^{n \times 1}$ 表示元素全为 1 的列向量; $b \in R^{1 \times m}$ 表示偏差项。

为了使式(6)等号右边的式子尽量拟合 X , 本文用 l_F -范数来估计余量,即 $\min_{Z,b} \|X - XZ - eb\|_F^2$ 。同时,为了去除冗余和离群数据,添加一个正则化项 $R_1(Z)$ 。由于 $l_{2,1}$ -范数能够导致行稀疏结构,使 Z 中一些不重要的行的系数变小或者直接为 0,这样有利于排除无关的属性,使算法更加有效,所以本文采用 $l_{2,1}$ -范数作为正则化项惩罚目标项,即

$$R_1(Z) = \|Z\|_{2,1} = \sum_{i=1}^m \left(\sum_{j=1}^m Z_{ij}^2 \right)^{\frac{1}{2}}$$

为了进一步增加模型的泛化性能,本文借鉴了 LPP 算法,并且将其中的普通图改成超图,在之前提出的模型上嵌入一个超图拉普拉斯正则化项 $R_2(Z) = \lambda_2 \text{tr}(Z^T X^T L_H X Z)$, 其中 L_H 是通过式(5)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,又考虑到了样本间更深层次的关系。所以可得到以下函数:

$$\min_{Z,b} \|X - XZ - eb\|_F^2 + \lambda_1 \|Z\|_{2,1} + \lambda_2 \text{tr}(Z^T X^T L_H X Z) \quad (7)$$

其中: $\|\cdot\|_F$ 是 Frobenius 矩阵范数; $L_H \in R^{n \times n}$ 表示由数据构成的归一化超图拉普拉斯矩阵; λ_1 和 λ_2 为控制参数。

因噪声或离群点会增加自表达系数矩阵的秩,为了提高算法的效率,同时考虑到属性间的关系,对 Z 进行低秩限制^[18]: 假设 $\text{rank}(Z) = r, r \leq \min(m, n)$, 一个低秩限制的 Z 可以写成: $Z = AB$, 其中 $A \in R^{m \times r}$, 而 $B \in R^{r \times m}$ 。综上所述,最终得到以下目标函数:

$$\min_{A,B} \|X - XAB - eb\|_F^2 + \lambda_1 \|AB\|_{2,1} + \lambda_2 \text{tr}(\|B^T A^T L_H XAB\|) \quad (8)$$

s. t. $\text{rank}(AB) \leq \min(m, n)$

其中: m 表示属性数; n 表示样本数; AB 表示由 $A \in R^{m \times r}$ 和 $B \in R^{r \times m}$ 构成的自表达系数矩阵。低秩结构意味着 $\text{rank}(AB) = r \leq \min(m, n)$ 。数据有噪声或离群点会导致低秩。使用低秩限制能考虑属性间的关系, 而且还能证明实际上低秩限制就是进行 LDA 的子空间学习, 而 LDA 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, 以此帮助更好地进行属性选择。

本文提出的 BHSLR_FS 算法具有以下优点:

a) 由于算法采用了属性自表达来建模, 这种方式对噪声和离群点不敏感, 在建模时加入了一个偏差项以此来使模型更加准确, 所以该算法具有较好的鲁棒性和泛化性。然后在含有偏差项的属性自表达模型中加入 $l_{2,1}$ -范数稀疏正则化因子, 能够去除冗余和不相关的属性, 使该属性自表达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。

b) 考虑到数据之间更高层次的内在联系, 在新的属性自表达模型中嵌入了一个超图拉普拉斯正则化项, 以此来保持数据之间的局部结构; 同时, 利用子空间学习方法中的 LDA 算法来保持数据的全局结构且进一步调整属性选择的结果, 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。

c) 因为低秩的假设, 考虑了属性之间的关系, 融入到模型中能够有效地去除冗余的属性和噪声样本, 从而提高模型对于高维数据分类预测的准确率, 同时也能够提高算法的运行效率。

d) 对本文的目标函数提出了一种区别于交替方向乘法法的求解方法, 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值在每次迭代过程中逐步收敛趋近于全局最优解, 最终取得全局最优解。

本文算法的伪代码如下:

算法 1 BHSLR_FS 算法伪代码

输入: 训练样本 $X \in R^{n \times m}$, 控制参数 λ_1, λ_2 。

输出: 分类准确率。

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(5) 得到超图拉普拉斯矩阵 L_H ;

c) 依据所选择的模型调用算法 2 求解全局最优解, 得到自表达系数矩阵 $Z^* = A^* B^*$;

d) 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;

e) 对新的属性集构成的样本采用 SVM 分类。

2.2 算法优化

首先定义一个对角矩阵:

$$D^i = \frac{1}{2 \|g^i\|_2} \text{ s.t. } i = 1, 2, \dots, m \quad (9)$$

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^* 则表示自表达系数矩阵 Z 的最优解, 然后将等式(9) 代入式(8) 化简可得

$$\min_{A, B} \|X - XAB - eb\|_F^2 + \lambda_1 \text{tr}(B^T A^T DAB) + \lambda_2 \text{tr}(B^T A^T X^T L_H XAB) \quad (10)$$

这里采用两步交替优化法:

a) 固定 A, B , 优化 b 。

当固定 A, B 后, D 也是固定的。优化问题式(10) 将变为

$$\min_b \|X - XAB - eb\|_F^2 \quad (11)$$

式(11) 可以看成是关于 b 的函数, 所以对式(11) 求导, 并使导数为 0, 可以得到

$$2e^T eb + 2e^T XAB - 2e^T X = 0 \quad (12)$$

通过简单的数学转换便可以得到

$$b = \frac{1}{n} e^T X - \frac{1}{n} e^T XAB \quad (13)$$

b) 固定 b , 优化 A 和 B 。

当固定 b 之后, 将式(13) 代入式(10) 可以得到

$$\min_{A, B} \|X - XAB - e(\frac{1}{n} e^T X - \frac{1}{n} e^T XAB)\|_F^2 + \lambda_1 \text{tr}(B^T A^T DAB) + \lambda_2 \text{tr}(B^T A^T X^T L_H XAB) \quad (14)$$

让 $H = I_n - \frac{1}{n} ee^T \in R^{n \times n}$, 而 $I_n \in R^{n \times n}$ 是一个单位矩阵, 那么式(14) 可以改写成

$$\min_{A, B} \|HX - HXAB\|_F^2 + \lambda_1 \text{tr}(B^T A^T DAB) + \lambda_2 \text{tr}(B^T A^T X^T L_H XAB) \quad (15)$$

将式(15) 看成是 B 的函数, 并令其导数为 0, 便能够得到

$$\frac{\partial \text{tr}(HX - HXAB)^T (HX - HXAB)}{\partial B} + \lambda_1 \frac{\partial \text{tr}(B^T A^T DAB)}{\partial B} + \lambda_2 \frac{\partial \text{tr}(B^T A^T X^T L_H XAB)}{\partial B} = 0 \quad (16)$$

$$\Rightarrow 2A^T X^T H^T HXAB + 2\lambda_1 A^T DAB + 2\lambda_2 A^T X^T L_H XAB - 2A^T X^T H^T HX = 0$$

$$\Rightarrow B = (A^T (X^T H^T HX + \lambda_1 D + \lambda_2 X^T L_H X) A)^{-1} A^T X^T H^T HX$$

将式(16) 代入式(15) 可以得到

$$\max_A \text{tr}((A^T (X^T H^T HX + \lambda_1 D + \lambda_2 X^T L_H X) A)^{-1} A^T X^T H^T HX X X^T H^T HX A) \quad (17)$$

注意到:

$$S_i = X^T H^T HX + \lambda_1 D + \lambda_2 X^T L_H X$$

$$S_b = X^T H^T HX X^T H^T HX \quad (18)$$

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(17) 可以得到 A :

$$A = \arg \max_A \{ \text{tr}((A^T S_i A)^{-1} A^T S_b A) \} \quad (19)$$

式(19) 的求解问题明显是 LDA 求解问题, 因此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 $G^* = A^* B^*$ 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(8) 的伪代码

输入: 训练样本 $X \in R^{n \times m}$, 控制参数 λ_1, λ_2 , 低秩参数 r 。

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

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 $D^{(t)} = I \in R^{m \times m}$;

d) 重复:

e) 通过式(19) 计算 $A^{(t+1)}$;

f) 通过式(16) 计算 B^{t+1} ;

g) 通过式(13) 计算 b^{t+1} ;

h) 更新对角矩阵 $D^{t+1} \in R^{m \times m}$, 第 i 个对角元素由式(9) 计算得出, 其中 $g^i = [A((t+1) B((t+1))]^i$;

i) 更新 $t = t + 1$;

j) 直到式(8) 收敛;

k) 结束。

3 优化算法收敛性证明

由于式 (8) 是非平滑的, 而且有两个变量 A 和 B 需要优化, 并且正则化项也是非平滑的, 使得该目标函数求解更加困难。但本文提出使用一种简单有效的算法来求解该问题。下面将给出算法 1 收敛性证明的具体步骤。

从算法 2 中的步骤 d) ~ k), 能够得到第 t 次迭代对应的结果:

$$\langle A^{(t+1)}, B^{(t+1)} \rangle = \arg \min_{A, B} \|HX - HXAB\|_F^2 + \lambda_1 \text{tr}(B^T A^T D^{(t)} AB) + \lambda_2 \text{tr}(B^T A^T X^T L_H XAB) \quad (20)$$

由此可得

$$\begin{aligned} & \|HX - HXA^{(t+1)} B^{(t+1)}\|_F^2 + \lambda_1 \text{tr}(B^{(t+1)T} D^{(t)} A^{(t+1)} B^{(t+1)}) + \\ & \lambda_2 \text{tr}(B^{(t+1)T} A^{(t+1)T} X^T L_H X A^{(t+1)} B^{(t+1)}) \leq \\ & \|HX - HXA^{(t)} B^{(t)}\|_F^2 + \lambda_1 \text{tr}(B^{(t)T} A^{(t)T} D^{(t)} A^{(t)} B^{(t)}) + \\ & \lambda_2 \text{tr}(B^{(t)T} A^{(t)T} X^T L_H X A^{(t)} B^{(t)}) \end{aligned} \quad (21)$$

其中: 要看到 $G^{(t)} = A^{(t)} B^{(t)}$, $G^{(t+1)} = A^{(t+1)} B^{(t+1)}$ 。将由式 (9) 得到的对角矩阵 D 代入式 (21) 可以得到

$$\begin{aligned} & \|HX - HXG^{(t+1)}\|_F^2 + \lambda_1 \sum_{i=1}^m \frac{\|g^{i(t+1)}\|_2^2}{\|g^{i(t)}\|_2} + \\ & \lambda_2 \text{tr}(G^{(t+1)T} X^T L_H X G^{(t+1)}) \leq \\ & \|HX - HXG^{(t)}\|_F^2 + \lambda_1 \sum_{i=1}^m \frac{\|g^{i(t)}\|_2^2}{\|g^{i(t)}\|_2} + \\ & \lambda_2 \text{tr}(G^{(t)T} X^T L_H X G^{(t)}) \end{aligned} \quad (22)$$

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

$$\|g^{i(t+1)}\|_2 - \frac{\|g^{i(t+1)}\|_2^2}{2\|g^{i(t)}\|_2} \leq \|g^{i(t)}\|_2 - \frac{\|g^{i(t)}\|_2^2}{2\|g^{i(t)}\|_2} \quad (23)$$

接着对上面的不等式乘上控制参数 λ_1 且累加后可以得到

$$\lambda_1 \sum_{i=1}^m \|g^{i(t+1)}\|_2 - \frac{\|g^{i(t+1)}\|_2^2}{2\|g^{i(t)}\|_2} \leq \lambda_1 \sum_{i=1}^m \|g^{i(t)}\|_2 - \frac{\|g^{i(t)}\|_2^2}{2\|g^{i(t)}\|_2} \quad (24)$$

最后 结合式 (22) 和 (24) 便可以得到

$$\begin{aligned} & \|HX - HXG^{(t+1)}\|_F^2 + \lambda_1 \sum_{i=1}^m \|g^{i(t+1)}\|_2 + \\ & \lambda_2 \text{tr}(G^{(t+1)T} X^T L_H X G^{(t+1)}) \leq \\ & \|HX - HXG^{(t)}\|_F^2 + \lambda_1 \sum_{i=1}^m \|g^{i(t)}\|_2 + \\ & \lambda_2 \text{tr}(G^{(t)T} X^T L_H X G^{(t)}) \end{aligned} \quad (25)$$

综上所述 再结合式 (13) 和 $H = I_n - \frac{1}{n} ee^T$ 可知:

$$\begin{aligned} & \|X - XG^{(t+1)} - eb\|_2 + \lambda_1 \|G^{(t+1)}\|_{2,1} + \\ & \lambda_2 \text{tr}(G^{(t+1)T} X^T L_H X G^{(t+1)}) \leq \\ & \|X - XG^{(t)} - eb\|_2 + \lambda_1 \|G^{(t)}\|_{2,1} + \\ & \lambda_2 \text{tr}(G^{(t)T} X^T L_H X G^{(t)}) \end{aligned} \quad (26)$$

上述不等式表明算法 1 在每次的迭代中, 目标函数都是单调递减的, 所以本文提出的算法能最终达到全局收敛的结果。

4 实验结果和分析

4.1 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的性能,

其中数据集 ORL 是人脸数据集, 来自文献 [19], DBWorld、LungCancer、Soybean、isolets 来源于 UCI^[20], PCMAC 来自属性选择数据集^[21]。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
LungCancer	32	56	2	Soybean	47	35	4
DBWorld	64	4 702	2	isolets	1 560	617	26
PCMAC	1 943	3 289	2	ORL	400	1 024	40

所有实验均在 Windows 7 系统下运行, 使用 MATLAB 2014a 软件进行测试。本文实验选择四种对比算法来与本文提出的算法进行比较: NFS 方法 (non feature selection) 对原始数据不作任何处理, 直接使用 LIBSVM 工具箱^[22] 进行 SVM 分类, 对于所有算法, 均采用 5 折交叉验证方法选择参数, 且 SVM 分类器中参数, SVM 分类器类型统一选择线性模式; LDA (linear discriminant analysis)、LR (linear regression) 方法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; RSR 方法^[23] 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。

分析以上算法, NFS 方法直接对原始数据集进行 SVM 分类, 未对原始数据进行任何处理, 相比 BHSLR_FS 等属性选择算法, 不仅数据处理量大, 而且容易受到冗余数据和噪声数据的影响; LDA 方法只是单纯地考虑数据之间的整体结构; LR 方法未考虑范数对噪声数据的敏感性; RSR 没有考虑数据之间的相关性。

4.2 实验结果和分析

本文采用分类准确率作为评价指标, 分类准确率越高表示算法效果越好。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 SVM 进行分类得到分类准确率。所有的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。各算法在六个数据集上实验结果对比如图 2 ~ 7 所示, 具体数据结果如表 2 所示。

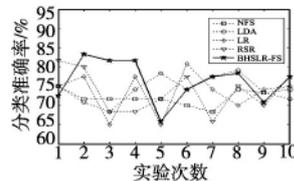


图2 数据集LungCancer

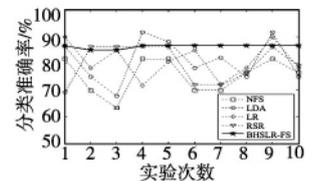


图3 数据集DBWorld

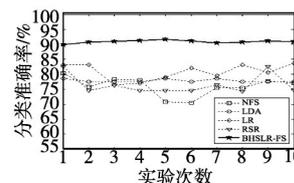


图4 数据集PCMAC

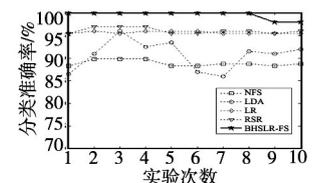


图5 数据集Soybean

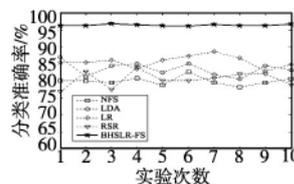


图6 数据集isolets

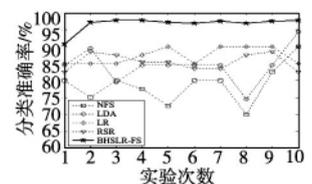


图7 数据集ORL

表 2 准确率(均值 ± 均方差)统计结果

数据集	NFS	LDA	LR	RSR	BHSLR_FS
LungCancer	72.17 ± 1.94	74.25 ± 3.26	73.42 ± 4.98	73.58 ± 5.20	76.33 ± 5.22
DBWorld	75.29 ± 6.60	81.14 ± 8.16	80.95 ± 6.51	79.90 ± 7.15	86.33 ± 0.67
PCMAC	75.95 ± 3.16	77.79 ± 3.69	80.76 ± 3.16	76.46 ± 3.27	90.84 ± 0.46
Soybean	88.90 ± 0.62	90.70 ± 3.08	95.70 ± 0.24	96.10 ± 0.62	99.60 ± 0.80
isolets	80.01 ± 2.25	83.60 ± 2.02	85.77 ± 1.71	80.53 ± 2.32	96.53 ± 0.22
ORL	78.75 ± 5.59	84.50 ± 5.27	87.25 ± 2.49	85.50 ± 2.40	96.75 ± 2.02
平均	78.51	82.00	83.98	82.01	91.06

通过图 2~7 所示可以清楚地看到 ,BHSLR_FS 算法在六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 ,如 isolets 和 ORL 数据集 ,可以通过控制低秩的数量来得到不同的分类结果。如图 8、9 所示 ,具有低秩结构的 BHSLR_FS 算法比满秩的效果更佳 ,分类准确率更高。

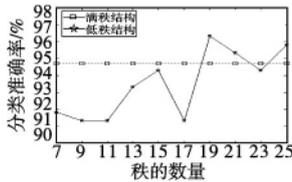


图 8 数据集 isolets

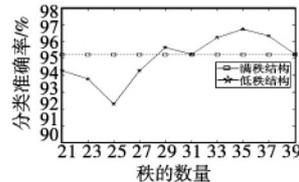


图 9 数据集 ORL

通过分析表 2 可以看出 ,BHSLR_FS 与其他四种对比算法比较 ,均取得了最高的分类准确率 ,具体地 ,与 NFS 算法比较平均提高了 12.55% ;比 RSR 算法比较平均提高了 9.05% ;比 LDA 算法比较平均提高了 9.06% ,证明了本文提出的算法比单一的子空间学习算法性能要好。由于 LungCancer 和 Soybean 数据集的维度较小 ,五种算法所得准确率比较接近 ,在 ORL 数据集上 ,BHSLR_FS 算法的效果最好 ,不但比 NFS 高出了 18% ,而且与 LR 算法相比也提高了 9.5%。这是因为 BHSLR_FS 算法不仅通过 $l_{2,1}$ 范数去除噪声和无关的属性 ,并且还同时考虑了数据的全局结构和局部结构 ,还进一步考虑了属性之间更深层的关系 ,所以能够显著地提高模型的性能。

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 ,实验结果表明 ,本文提出的 BHSLR_FS 算法在每个数据集上都要对比算法好些 ,对于对比算法来说 ,BHSLR_FS 算法具有最高的鲁棒性 ,同时也说明 BHSLR_FS 算法输出了更具辨别力的属性集。

5 结束语

本文基于属性自表达提出了一种新的属性选择算法——BHSLR_FS 算法 ,即通过使用数据本身的性质来构造自表达系数矩阵 ,并且通过稀疏表达来发现自表达系数矩阵中的稀疏结构 ,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的不足 ,然后借助 LDA 算法来进一步调整属性选择的结果 ,最后通过低秩假设来提高算法的运行效率 ,因此可以达到较好的效果。该算法融合了低秩和子空间学习的优点 ,在一定程度上弥补低秩属性选择在保持数据几何结构方面的不足。经实验表明 ,本文算法能够在分类准确率和稳定性上取得显著提高。在今后的工作中 ,将尝试在半监督属性选择方面拓展验证本文提出的算法 ,并尝试使用更先进的技术来改进算法。

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 ,Huang Zi ,Shen Hengtao *et al.* Dimensionality reduction by mixed kernel canonical correlation analysis [J]. *Pattern Recognition* 2012 45(8) : 3003-3016.

[3] 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.

[4] Zhao Yanchang ,Zhang Shichao. Generalized dimension-reduction framework for recent-biased time series analysis [J]. *IEEE Trans on Knowledge & Data Engineering* 2006 18(2) : 231-244.

[5] Zhu Xiaofeng ,Suk H I ,Shen Dinggang. Matrix-similarity based loss function and feature selection for Alzheimer's disease diagnosis [C] // *Proc of IEEE Conference on Computer Vision and Pattern Recognition*. 2014: 3089-3096.

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

[7] Pyatykh S ,Hesser J ,Lei Zheng. Image noise level estimation by principal component analysis [J]. *IEEE Trans on Image Processing* , 2013 22(2) : 687-699.

[8] 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.

[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] Qin Yongsong ,Zhang Shichao ,Zhu Xiaofeng *et al.* Semi-parametric optimization for missing data imputation [J]. *Applied Intelligence* , 2007 27(1) : 79-88.

[11] Zhang Shichao ,Zhang Chengqi ,Yang Qiang. Data preparation for data mining [M]. [S.l.]: Academic Press ,1999.

[12] 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.

[13] 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.

[14] Zhu Xiaofeng ,Zhang Shichao ,Zhang Jilian *et al.* Cost-sensitive imputing missing values with ordering [C] // *Proc of AAAI Conference on Artificial Intelligence*. 2007: 1922-1923.

[15] Wu Xindong ,Zhang Chengqi ,Zhang Shichao. Efficient mining of both positive and negative association rules [J]. *ACM Trans on Information Systems* 2004 22(3) : 381-405.

[16] Zhou Dengyong ,Huang Jiayuan ,Bernhard S. Learning with hyper-graphs: clustering ,classification ,and embedding [C] // *Advances in Neural Information Processing Systems*. 2006: 1601-1608.

[17] Zhang Chengqi ,Qin Yongsong ,Zhu Xiaofeng *et al.* Clustering-based missing value imputation for data preprocessing [C] // *Proc of IEEE International Conference on Industrial Informatics*. 2006: 1081-1086.

[18] Ling C X ,Yang Qiang ,Wang Jianing *et al.* Decision trees with minimal costs [C] // *Proc of International Conference on Machine Learning*. 2004: 4-8.

[19] Chen Y M ,Chiang J H. 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.

[20] UCI repository of machine learning datasets [EB/OL]. [2016-05-27]. <http://archive.ics.uci.edu/ml/>.

[21] Feature selection datasets [EB/OL]. [2016-05-27]. <http://featureselection.asu.edu/datasets.Php>.

[22] LIBSVM: a library for support vector machines [EB/OL]. [2016-05-27]. <http://www.csie.ntu.edu.tw/~cjlin/libsvm>.

[23] Zhu Pengfei ,Zuo Wangmeng ,Zhang Lei *et al.* Unsupervised feature selection by regularized self-representation [J]. *Pattern Recognition* 2015 48(2) : 438-446.