控制与决策
Control and Decision

基于Hessian正则的自适应损失半监督特征选择

朱建勇， 周振辰， 杨辉， 聂飞平

引用本文:
朱建勇, 周振辰, 杨辉, 等. 基于Hessian正则的自适应损失半监督特征选择[J]. 控制与决策, 2021, 36(8): 1862−1870.

在线阅读 View online: https://doi.org/10.13195/j.kzyjc.2019.1510

基于Hessian正则的自适应损失半监督特征选择

朱建勇[1,2], 周振辰[1,2], 杨 辉[1,2†], 聂飞平[3]

(1. 华东交通大学 电气与自动化工程学院，南昌 330013；2. 江西省先进控制与优化重点实验室，南昌 330013；3. 西北工业大学 光学影像分析与学习中心，西安 710072)

摘 要: 传统的基于拉普拉斯图的半监督特征选择算法处理高维、少标签样本时,缺乏外推能力且对数据异常值的鲁棒性差. 基于此,提出一种基于Hessian正则的自适应损失半监督稀疏特征选择算法. 首先,为提升线性映射能力,利用Hessian正则保留数据的局部流形结构;其次,为增强模型对具有较小或者较大损失数据的鲁棒性,引入自适应损失函数,通过调节自适应损失参数确定最小损失;再次,采用$l_{2,p}$范数稀疏投影矩阵,提升特征的区分度,增加模型适应度;最后,采用递归迭代优化求解目标函数. 仿真实验验证了所提方法的有效性和优越性.

关键词: 半监督；特征选择；自适应损失；稀疏约束；$l_{2,p}$ 范数

中图分类号: TP391 **文献标志码:** A

DOI: 10.13195/j.kzyjc.2019.1510 **开放科学(资源服务)标识码(OSID):**

引用格式: 朱建勇,周振辰,杨辉,等. 基于Hessian正则的自适应损失半监督特征选择[J]. 控制与决策, 2021, 36(8): 1862-1870.

Adaptive loss semi-supervised feature selection based on Hessian regularization

ZHU Jian-yong[1,2], *ZHOU Zhen-chen*[1,2], *YANG Hui*[1,2†], *NIE Fei-ping*[3]

(1. College of Electrical and Automation，East China Jiaotong University，Nanchang 330013，China；2. Key Laboratory of Advanced Control and Optimization of Jiangxi Province，Nanchang 330013，China；3. Center for Optical Image Analysis and Learning，Northwestern Polytechnical University，Xi'an 710072，China)

Abstract: The traditional semi-supervised sparse feature selection based on Laplacian graph has received extensive attention for its higher efficiency. However, due to the lack of extrapolation ability of the Laplacian operator, the limited labeled data is still not well utilized and is too sensitive for outliers. Therefore, an adaptive loss semi-supervised sparse feature selection algorithm based on Hessian regularization is proposed. Firstly, Hessian is used to preserve the local manifold structure of data in order to improve the linear mapping capability. At the same time, an adaptive loss function is exploited to measure the label fitness by adjusting the adaptive loss parameters, which significantly enhances model's robustness to data with a small or substantial loss. Moreover, $l_{2,p}$-norm is leveraged to constrain the prediction matrix, which not only improves the distinguishing degree of features, but also increases the adaptability of the proposed model. Then, a recursive iterative optimization algorithm is proposed to solve the proposed model. Finally, systematic experimental results on real public data sets illustrate the effectiveness and superiority of the proposed approach on related tasks.

Keywords: semi-supervised；feature selection；adaptive loss；sparse constraint；$l_{2,p}$-norm

0 引 言

近年来,随着信息技术的革新,数据维数变得日益复杂,甚至超过了数百万个特性,加重了“维度灾难”. 机器学习中一些实际应用,如人脸识别[1]、图像检索[2-3]、视频语义识别[4-6]、基因诊断[7]等,极易产生高维数据. 例如在基因诊断任务中,实验中测量基因表达水平的基因表达数据通常由数千个基因组成. 在对这些数据进行分类时,所学习的模型容易过拟合,泛化能力较差. 这些应用中的高维数据仅有一小部分特征与类别高度相关,而大多数特征是不相关或者冗余的,直接处理这些高维数据容易导致大量的计算消耗[8]. 因此,特征提取和特征选择成为主要的

收稿日期: 2019-10-29；**修回日期:** 2020-01-18.

基金项目: 国家自然科学基金重点项目(61733005)；国家自然科学基金项目(61563015, 61963015, 61863014)；江西省自然科学基金项目(20171ACB21039, 20192BAB207024)；江西省教育厅科技项目(GJJ150552).

责任编委: 阳春华.

†通讯作者. E-mail: yhshuo@263.net.

降维策略[9]. 特征提取将原始空间的特征映射到一个低维的特征空间,改变了原始特征空间. 特征选择从原始高维特征空间中选择出一个判别性的特征子集,保留原始特征空间,具有一定的可解释性[10]. 根据对标签信息的可用性,特征选择分为有监督、无监督和半监督3类. 有监督特征选择可以根据标签信息与特征集之间的联系评价特征的冗余性,需要大量标签样本选择代表性特征子集[11]. 无监督特征选择完全抛弃了标签信息,仅依靠无标签数据评估特征的相关性[12-13]. 随着数据的快速增长,为这些数据标注标签信息往往需要耗费较多的人力和财力[14]. 由于在机器学习相关任务中获取的数据通常由少量标签数据和大量无标签数据组成,较多的研究人员不断探索能够同时利用标签数据和无标签数据的半监督特征选择算法.

半监督特征选择方法利用标签数据的信息以及标签数据和无标签数据的局部结构进行训练来评价特征相关性,尽而选择判别性的特征,有效地提升有监督学习模型泛化能力和无监督学习模型的精确性[15]. 多数半监督特征选择算法是基于过滤器,单独对每个特征进行评价,通过对特征进行排序,选择高排名的特征并将其应用于预测器[16]. 这类方法忽略了特征与特征之间的相关性,即某些特征本身提供的信息较少,但是与其他特征结合时提供的信息较多[17]. 文献[4]提出基于包裹器的半监督特征选择算法,考虑不同特征之间的相关性,通过分别构建标签信息矩阵、局部样条回归编码数据分布的样条散射输出矩阵,将这两类矩阵相结合,实现对训练集的识别信息和局部几何结构的捕获,促进了判别性特征子集的形成. 然而,该方法涉及迭代特征子集搜索,处理高维数据比较耗时. 嵌入式半监督方法将特征选择作为训练过程的一部分,即在学习器训练过程中自动地进行特征选择. 在此基础上,为了更好地描述数据的局部或者全局的流行结构,文献[18]引入图拉普拉斯正则化描述数据的结构,同时结合了流行学习探索特征空间. 文献[19]通过最大化不同类之间的分类界限并利用生成标记和未标记数据的概率分布的几何特性来选择特征,加强了对判别性特征的选择. 文献[20]基于多特征融合的思想,通过多种角度描述对象特征,构造Hessian与Laplacian图并赋之不同的权重,在学习中利用每个特征的流形结构信息,保持全局标签的一致性,使得分类更加准确.

传统基于图拉普拉斯正则的半监督特征选择算法在标签数据较少的情况下,缺乏对新数据点的推断性能及对数据中异常值较差的鲁棒性. 因此,本文提出基于Hessian正则的自适应损失半监督稀疏特征选择框架. 首先分析了Hessian正则具有更丰富的零空间,能较好地利用数据固有的局部几何特性,有利于学习函数值随测地距离线性变化的函数[21-23]. 此外,多数半监督特征选择算法采用l_2范数作为损失函数来度量预测标签误差,但具有显著损失的异常值将导致模型表现比较敏感,鲁棒性较差. 使用l_1范数作为损失函数,可以在一定程度上缓解对异常值的敏感度,但是又会对小损失比较敏感[24]. 为了克服基于l_1范数和l_2范数损失函数的缺点,本文采用自适应损失来度量预测标签的误差,通过自适应近邻分配策略,得到最优Hessian矩阵,增强特征选择模型的鲁棒性. 此外,使用$l_{2,p}$范数作为隐式正则项约束投影矩阵W,通过设置不同的p值,可以获得更多的稀疏回归系数.

1　相关工作

在介绍基于Hessian正则的自适应损失半监督稀疏特征选择框架(AHFS)之前,给出模型中使用的符号含义如表1所示. 对于矩阵Q, Q^{T}和$\mathrm{Tr}(Q)$分别表示矩阵的转置和矩阵的迹; $X \in R^{n\times d}$表示数据样本矩阵,其中x_i表示第i个数据样本; Y表示训练样本的标签矩阵; F表示预测标签矩阵; D、G表示对称矩阵.

表1　符号的定义

符号	含义	变量	名称
n	样本数	H	Hessian 矩阵
d	特征维度	Y	标签矩阵
c	样本类别	G	对称矩阵
M	局部流行结构	W	投影矩阵
φ	映射函数	F	预测标签矩阵

对于任意矩阵$M \in R^{m\times n}$,向量$m(i,:)$表示矩阵的第i行, $m(:,j)$表示第j 列. $l_{2,p}$范数的计算公式如下:

$$\|M\|_{2,p} = \left(\sum_{i=1}^{m}\left(\sum_{j=1}^{n}|m_{ij}|^2\right)^{\frac{p}{2}}\right)^{\frac{1}{p}} = \left(\sum_{i=1}^{m}\|m_{(i,:)}\|_2^p\right)^{\frac{1}{p}}. \quad (1)$$

1.1　样条回归半监督特征选择

文献[4]通过构造类内图矩阵充分利用标签数据的判别性信息,为利用大量未标记数据潜在的局部几何特性,使用Sobolev空间中开发的样条来插入散在几何设计中的数据. 通过组合类内和样条散射矩阵,保持数据的局部结构和潜在的几何特性. 然后,利用图嵌入的思想,引入$l_{2,1}$范数约束投影矩阵W,计算最优的投影矩阵W,目标函数如下:

$$\arg\min_{W} \mathrm{Tr}(WMW) + \lambda\|W\|_{2,1},$$
$$\text{s.t. } W^{\mathrm{T}}W = I. \tag{2}$$

其中: $M = S_W + \mu\Xi$ 和 $S_W\Xi$ 分别表示类间矩阵和样条散射矩阵, $\Xi = X\zeta X^{\mathrm{T}}, \zeta \in R^{n\times n}$ 表示每对样本的局部相似性. 类间矩阵通过下式计算:

$$S_W = \sum_{j=1}^{c} \frac{1}{N_j} \sum_{X\in\xi_j} (x-m_j)(x-m_j)^{\mathrm{T}}. \tag{3}$$

m_j 表示第 j 类的平均向量, c 表示样本数据的总类别.

1.2 优化迹比准则半监督特征选择

文献[25]分析特征选择中的迹比准则,提出了一种基于噪声不敏感迹比准则的半监督特征选择方法(TRCFS),以解决降维中迹比准则倾向于选择方差很小的特征的问题. 首先,通过优化迹比准则进行缩放预处理,利用离群点检测的标签传播方法获取未标记数据的软标签. 然后,构建类内矩阵、类间矩阵. 最后,通过噪声不敏感跟踪比准则进行特征选择,目标函数为

$$\arg\min_{W} \mathrm{Tr}(W^{\mathrm{T}}S_bW)/\mathrm{Tr}(W^{\mathrm{T}}S_\omega W). \tag{4}$$

根据标签矩阵 F 可以分别计算类矩阵 S_ω、S_b 为

$$S_\omega = \frac{1}{n}X(B - F_cDF)X^{\mathrm{T}}, \tag{5}$$

$$S_b = \frac{1}{n}X\Big(F_cDF_c^{\mathrm{T}} - \frac{1}{n}B\mathbf{1}\mathbf{1}^{\mathrm{T}}B^{\mathrm{T}}\Big)X^{\mathrm{T}}. \tag{6}$$

其中: F_c 表示 F 的前 c 列, B、D 表示对角矩阵.

1.3 图嵌入半监督特征选择

基于图拉普拉斯的半监督特征选择算法分为2个步骤: 1) 根据样本数据,通过调整近邻构造相似图; 2) 根据构造的相似图进行特征选择. $Q = \{V, S\}$ 表示构造的无向权重图, V 表示顶点集合, $S = (w_{ij})_{n\times n} \in R^{n\times n}$ 表示相似矩阵,元素 w_{ij} 表示两个数据点 x_i 与 x_j 之间的相似度. 通常,相似度矩阵 S 在具有高斯函数的原始高维特征空间中是预定的,即

$$w_{ij} = \begin{cases} \exp\Big(-\dfrac{\|x_i - x_j\|_2^2}{2\delta^2}\Big), \\ \quad x_i \in N_k(x_j) \text{ or } x_j \in N_k(x_i); \\ 0, \text{ otherwise.} \end{cases} \tag{7}$$

其中 $N_k(x_i)$ 表示 x_i 在原始的高维空间 k 近邻集合. 基于图论,图拉普拉斯矩阵 $L = Z - S, Z$ 中的元素为 $Z_{ii} = \sum_j S_{ij}$. 此外,假定样本数据点 x_i 的标签为 $y_i = j$,则定义二进制标签矩阵 Y 的元素 $Y_{ij} = 1$, 否则 $Y_{ij} = 0$. 传统的基于图拉普拉斯算子的半监督特征选择方法是为了解决以下优化问题:

$$\arg\min_{W,F} \mathrm{Tr}(F^{\mathrm{T}}LF) + \mathrm{Tr}((F-Y)^{\mathrm{T}}U(F-Y)) + \mu\|X^{\mathrm{T}}W - F\|_F^2 + \lambda Q(W), \tag{8}$$

其中 U 是对称阵. 式(8)的第1项和第2项有利于预测标签矩阵 F 在流行空间尽可能地平滑;第3项则为标签适应度的损失项,度量预测标签的误差;最后一项表示投影矩阵 W 的稀疏性约束,保证不同特征之间的区分性,有利于特征子集的选取.

2 基于Hessian正则的自适应损失半监督特征选择

2.1 目标函数设计

基于图谱的半监督特征选择算法通常依赖Laplacian矩阵的分解获取全局聚类结构信息,需要耗费大量的时间和空间,且Laplacian算子映射能力不足. 因此,本文拟采用Hessian算子探索数据的局部结构[18],首先构造适合半监督特征选择的正则化函数,假定流行结构 M,定义实值函数 $\varphi: M \to R$,定义函数 φ 的能量函数为

$$E(\varphi) = \int \|\nabla_a\nabla_b\varphi\|^2_{T_x^*M\otimes T_x^*M}\mathrm{d}V(X). \tag{9}$$

其中: $\nabla_a\nabla_b\varphi$ 表示函数 φ 的二次协变导数, T_x^*M 表示流行 M 上点 x 处的局部切线空间, $\mathrm{d}V(X)$ 表示流形 M 上的自然体积元. 流行结构 M 的标准坐标促使能量函数 $E(\varphi)$ 容易度量. 由于在点 x 附近的流形与像欧几里德相似,可以得到

$$\|\nabla_a\nabla_b\varphi\|^2_{T_x^*M\otimes T_x^*M} = \sum_{r,s=1}^{l}\Big(\frac{\partial^2\varphi}{\partial x^r\partial x^s}\Big)^2. \tag{10}$$

因此,在给定点 x 处的二次协变导数的范数等价于函数 φ 标准坐标下Hessian正则的Frobenius范数. 同时,评估局部切线子空间,其低维子空间的 l 个主导特征向量与 $T_{x_i}^*M$ 的正交基一致. 其次,确定点 $x_j \in N_k(x_i)$ 的坐标 x^r,得到

$$\frac{\partial^2\varphi}{\partial x^r\partial x^s}\Big|_{x_i} \approx \sum_{t=1}^{k} Z_{rst}^{(i)} f_t, \tag{11}$$

其中 $Z_{rst}^{(i)}$ 是样本 x_i 的局部Hessian算子,且可以用线性最小二乘拟合一个二阶多项式来计算. 再次,将法坐标系中的二阶多项式 $p^{(i)}(x)$ 拟合为 $\{\varphi(x_t)\}_{t=1}^k$,有

$$p^{(i)}(x) = \varphi(x_i) + \sum_{r=1}^{l} H_r x^r + \sum_{r=1}^{l}\sum_{s=r}^{l} N_r x^r x^s, \tag{12}$$

其中零阶项固定在 $\varphi(x_i)$. 在邻域大小趋于零的极限下,将 $p^{(i)}(x)$ 变为 φ 在 x_i 处的二阶泰勒展开式,得到

$$H_r = \frac{\partial\varphi}{\partial x^r}\Big|_{x_i}, \ N_{rs} = \frac{1}{2}\frac{\partial^2\varphi}{\partial x^r\partial x^s}\Big|_{x_i}. \tag{13}$$

便于拟合多项式,使用标准的线性最小二乘,得到

$$\arg\min_{v\in R^P} \sum_{t=1}^{k} \big((\varphi(x_t) - \varphi(x_i)) - (\phi v)_t\big)^2. \tag{14}$$

其中: $\phi \in R^{k\times P}$ 表示设计矩阵,且 $P = m + m(m+1)/2$; ψ 表示相应的基函数,其是 x_i 二阶法向坐标的单项式. 假设 $\varphi(x_\alpha) = f_\alpha$,因此 φ 在点 x_i 处的Hessian的Frobenius范数估计等效为

$$\|\nabla_a\nabla_b\varphi\|^2 \approx \sum_{r,s=1}^{l}\Big(\sum_{\alpha=1}^{k} Z_{rs\alpha}^{(i)} f_\alpha\Big)^2 = \sum_{\alpha\beta=1}^{k} f_\alpha f_\beta H_{\alpha\beta}^{(i)}, \tag{15}$$

其中 $H_{\alpha\beta}^{(i)} = \sum_{r,s=1}^{l} Z_{rs\alpha}^{(i)} Z_{rs\beta}^{(i)}$. 能量函数可以近似为

$$\hat{E}(\varphi) = \sum_{i=1}^{n}\sum_{\alpha\in N_k(x_i)}\sum_{\beta\in N_k(x_i)} f_\alpha f_\beta H_{\alpha\beta}^{(i)} = F^{\mathrm{T}}HF. \tag{16}$$

因此,本文构造AHFS的目标函数为

$$\begin{aligned}&\min_{W,F,b} \mathrm{Tr}(F^{\mathrm{T}}HF) + \mu(\|X^{\mathrm{T}}W + \\ &\qquad \mathbf{1}b^{\mathrm{T}} - F\|_\tau + \lambda\|W\|_{2,p}^p),\\ &\text{s.t. } F_l = Y_l.\end{aligned} \tag{17}$$

其中: W 表示投影矩阵; $F = [F_l; F_u]$ 表示标签矩阵,由已知样本标签 F_l 和未知样本标签 F_u 组成; $\mathbf{1}$ 表示元素为1的列向量; b 表示基础向量; 系数 μ 和 λ 表示平衡不同项的两个参数.

鉴于投影矩阵 W 的行元素与样本 X 的特征属性相对应,因此,通过计算投影矩阵 W 非零行元素之和,按照行元素和的排序进行特征子集的选取.

2.2　目标函数求解

由于AHFS函数(17)含有 $l_{2,p}$ 范数,定义 $\Gamma(W) = \|W\|_{2,p}^p$,对 $\Gamma(W)$ 关于 W 求偏导数,得到

$$\frac{\partial\Gamma(W)}{\partial W} = 2DW. \tag{18}$$

其中: $D \in R^{d\times d}$ 是一个对称矩阵,矩阵中的第 i 个元素为

$$d_{ii} = \frac{p}{2}\left\|w^i\right\|_2^{p-2}. \tag{19}$$

AHFS 函数的第 2 项,引入基于 l_1 范数和 l_2 范数之间的自适应损失函数,其利用基于矩阵范数的损耗测量的优点,对损耗较小或较大的数据具有更强的鲁棒性. 对于自适应损失项的优化,假定矩阵 $Z = [z^1, z^2, \ldots, z^n]^{\mathrm{T}} \in R^{n\times n}$,定义其损失函数为

$$f_{\mathrm{loss}}^\tau = \|Z\|_\tau = \sum_i \frac{(1+\tau)\|z^i\|_2^2}{\|z^i\|_2^2 + \tau}, \tag{20}$$

τ 是损失函数的自适应参数. 由于每个矩阵的行向量可以单独计算且互不影响,令 $q_i(W,F,b) = x_i^{\mathrm{T}}W + b^{\mathrm{T}} - f_i$, 则自适应损失函数项可以表示为

$$\mu\|X^{\mathrm{T}}W + \mathbf{1}b^{\mathrm{T}} - F\|_\tau = \mu\sum_{i=1}^{n}\|q_i(W,F,b)\|_\tau. \tag{21}$$

对式(21)关于 $\{W,F,b\}$ 求偏导,并令偏导数为0,得到

$$2\mu\sum_{i=1}^{n} g_i q_i(W,F,b) q_i'(W,F,b) = 0. \tag{22}$$

其中

$$g_i = (1+\tau)\frac{\|q_i(W,F,b)\|_2 + 2\tau}{2(\|q_i(W,F,b)\|_2 + \tau)^2}.$$

因此,损失函数可以转为求下式:

$$\begin{aligned}&\mu\sum_{i=1}^{n} g_i\|q_i(W,F,b)\|_2^2 = \\ &\mu\mathrm{Tr}((X^{\mathrm{T}}W + \mathbf{1}b^{\mathrm{T}} - F)^{\mathrm{T}}G(X^{\mathrm{T}}W + \mathbf{1}b^{\mathrm{T}} - F)).\end{aligned} \tag{23}$$

其中: G 是对称矩阵,矩阵中的第 i 个元素为

$$g_{ii} = (1+\tau)\frac{\|x_i^{\mathrm{T}}W + b^{\mathrm{T}} - f_i\|_2 + 2\tau}{2(\|x_i^{\mathrm{T}}W + b^{\mathrm{T}} - f_i\|_2 + \tau)^2}. \tag{24}$$

解决目标函数(17)的关键步骤是解决以下问题:

$$\begin{aligned}&\min_{W,F,b} \mu(\mathrm{Tr}((X^{\mathrm{T}}W + 1b^{\mathrm{T}} - F)^{\mathrm{T}}G(X^{\mathrm{T}}W + 1b^{\mathrm{T}} - \\ &\qquad F)) + \mu\lambda\mathrm{Tr}(W^{\mathrm{T}}DW)) + \mathrm{Tr}(F^{\mathrm{T}}HF),\\ &\text{s.t. } F_l = Y_l.\end{aligned} \tag{25}$$

因此,对式(25)关于 b 求偏导,并令偏导数为0,得到

$$b = \frac{1}{\mathbf{1}^{\mathrm{T}}G\mathbf{1}}F^{\mathrm{T}}G\mathbf{1} - \frac{1}{\mathbf{1}^{\mathrm{T}}G\mathbf{1}}W^{\mathrm{T}}XG\mathbf{1}. \tag{26}$$

将式(26)代入(25),关于 W 求偏导,并令偏导数为0,得到

$$W = (XMX^{\mathrm{T}} + 2\mu\lambda D)^{-1}XMF = CF. \tag{27}$$

其中

$$\begin{aligned}M &= G - \frac{1}{\mathbf{1}^{\mathrm{T}}G\mathbf{1}}G\mathbf{1}\mathbf{1}^{\mathrm{T}}G,\\ C &= (XMX^{\mathrm{T}} + 2\mu\lambda D)^{-1}XM.\end{aligned}$$

将式(26)和(27)代入(25),得到

$$\begin{aligned}&\arg\min_{W,F,b} \mathrm{Tr}(F^{\mathrm{T}}(H + \mu M - \mu MX^{\mathrm{T}}C)F),\\ &\text{s.t. } F_l = Y_l.\end{aligned} \tag{28}$$

其中 $F = [F_l; F_u]$. 假定 $Z = H + \mu M - \mu MX^{\mathrm{T}}C$,并将矩阵 Z 按照已知标签样本个数为限定行分为 4×4 的块矩阵,得到

$$\begin{aligned}&\arg\min_{W,F,b} \mathrm{Tr}\Big(\begin{bmatrix}F_l\\F_u\end{bmatrix}^{\mathrm{T}}\begin{bmatrix}Z_{ll} & Z_{lu}\\ Z_{ul} & Z_{uu}\end{bmatrix}\begin{bmatrix}F_l\\F_u\end{bmatrix}\Big),\\ &\text{s.t. } F_l = Y_l.\end{aligned} \tag{29}$$

对式(29)关于 F_u 求偏导数,并令偏导数为0,得到

$$F_u = Z_{uu}^{-1}Z_{ul}F_l. \tag{30}$$

通过上述步骤迭代求解,求得最优矩阵 W、F 和 b. 综上所述,AHFS目标函数的迭代求解过程如下.

step 1: 输入训练集X、标签矩阵Y_l、选择的特征数k、平衡参数μ和λ;

step 2: 计算初始的Hessian正则矩阵H,初始化对称矩阵D、G;

step 3: repeat;

step 4: 通过式(30)计算F_u,即得到软标签矩阵F;

step 5: 根据式(26)计算b,根据式(27)计算投影矩阵W;

step 6: 更新对称矩阵D和G;

step 7: 更新Hessian正则矩阵H;

step 8: until converge;

step 9: 输出投影矩阵W,选择特征子集$\{r_1, r_2, \ldots, r_k\}$.

2.3 算法复杂性

本文提出AHFS算法的计算复杂度,如迭代求解过程所示,主要分为Hessian正则的计算与优化、投影矩阵以及标签矩阵的计算. 其中Hessian正则的计算复杂度为$O(n^3)$. 在固定Hessian正则标签预测步骤中,主要的计算成本是更新W和F时涉及的矩阵逆运算. 更新W,令$Q = XMX^{\mathrm{T}} + 2\mu\lambda D, R = XMF$,其计算消耗分别为$O(nd^2)$和$O(ndc)$,因此$W$可以重新表示为$W = Q^{-1}R$. 矩阵求逆的计算消耗是$O(d^3)$,根据文献[26]可以避免矩阵求逆运算的消耗. 由于更新W可以转化为更新$Q^{-1}R$,通过梯度下降迭代可以求解最优W,这使得矩阵逆的计算消耗降低为$O(Td^2c)$,其中T为迭代次数. 采用同样的策略优化标签矩阵F. 因此,最优W、F的总计算消耗降低为$O(nd^2)+O(Td^2c)$. 综上,本文提出的AHFS总的计算消耗为$O(Tn^3)+O(nd^2)+O(Td^2c)$.

3 实验及分析

本节将AHFS与流行的半监督特征算法相比较,并分析算法选择特征的性能.

3.1 数 据 集

本次实验共选择6个数据集进行实验,包括3个UCI数据集,分别为Heart、Letter以及Vote;1个图片数据集(COIL20),其包含对20个物体处理后的1 440张图片,每隔5度拍摄一张图片,因此每个物体有72张图片;1个小圆蓝色细胞瘤(SRBCT)基因数据集,由83个样本组成,即8个Burkitt淋巴瘤(BL),33个尤文肉瘤(EWS),22个神经母细胞瘤(NB)和20个横纹肌肉瘤(RMS),每个样本由2 308个基因表达;1个手写图片(USPS)数据集,由10个数字的9 298个图像组成,每个数字图像的大小为16×16. 这6个数据集的描述如表2所示.

表2 数据集描述

名称	样本数	特征数	类别
Heart	270	13	2
Letter	2 000	16	26
Vote	435	16	2
COIL20	1 440	1 024	20
SRBCT	83	2 308	4
USPS	9 298	256	10

3.2 分类准确度评价标准

定义f_i和y_i为给定样本x_i的预测标签和样本自带类标签,分类准确度(ACC)公式如下:

$$\mathrm{ACC} = \frac{\sum_{i=1}^{n}\delta(y_i, \mathrm{map}(f_i))}{n}. \tag{31}$$

其中:n表示样本数;$\delta(x, y)$为比较函数,$x = y$时,函数值为1,否则为0.

3.3 算法比较

对于每个数据集,随机选择50%作为训练集,剩余50%为测试集. 在训练集上,各特征选择算法选择出相应的样本特征,然后在测试集中仅保留筛选出的特征,由支持向量机(SVM)对测试样本进行预测,并计算预测样本的正确率,得出相应的实验分类结果. 本次实验选择5种对比算法,分别是1种有监督特征选择算法,高效的鲁棒特征选择算法(RFS)[7];1种无监督特征选择算法,拉普拉斯的得分法(Laplacian Score)[12];3种半监督特征选择算法,局部敏感半监督特征选择(LSDF)[27]、基于相关性和冗余标准的半监督特性选择(RRPC)[28]、重新调节线性回归的半监督特征选择算法(RLSR)[29]. 为保证对比实验的公平性,从训练集中的每个类内随机选择不同比例的样本作为标签样本,其余样本作为无标记样本. 同时,在与有监督特征选择算法RFS比较时,RFS仅利用有标签样本选择特征子集;与无监督特征选择算法Laplacian Score对比时,选择比例相同的无标记样本进行特征选择. 由于随机地选择样本,这可能导致分类准确性不稳定. 因此,每次实验进行20次,以获得具有高可靠性的实验结果,并采用平均值作为对比的结果. 对于RFS、Laplacian Score、LSDF、RLSR中的参数λ,在集合$\{10^{-3}, 10^{-2}, 10^{-1}, 1, 10^2, 10^3\}$搜索确定最优参数值. 在6个数据集上逐渐增加选择的特征个数,按式(31)计算ACC,得到的结果如图1所示.

从图1可以看出,本文提出的半监督特征选择算法优于所比较的特征选择算法. 一方面,随着选择特征数的增加,AHFS的分类精度也会随之提高. 另一方面,AHFS几乎在所选取数据集上的性能优于RFS,

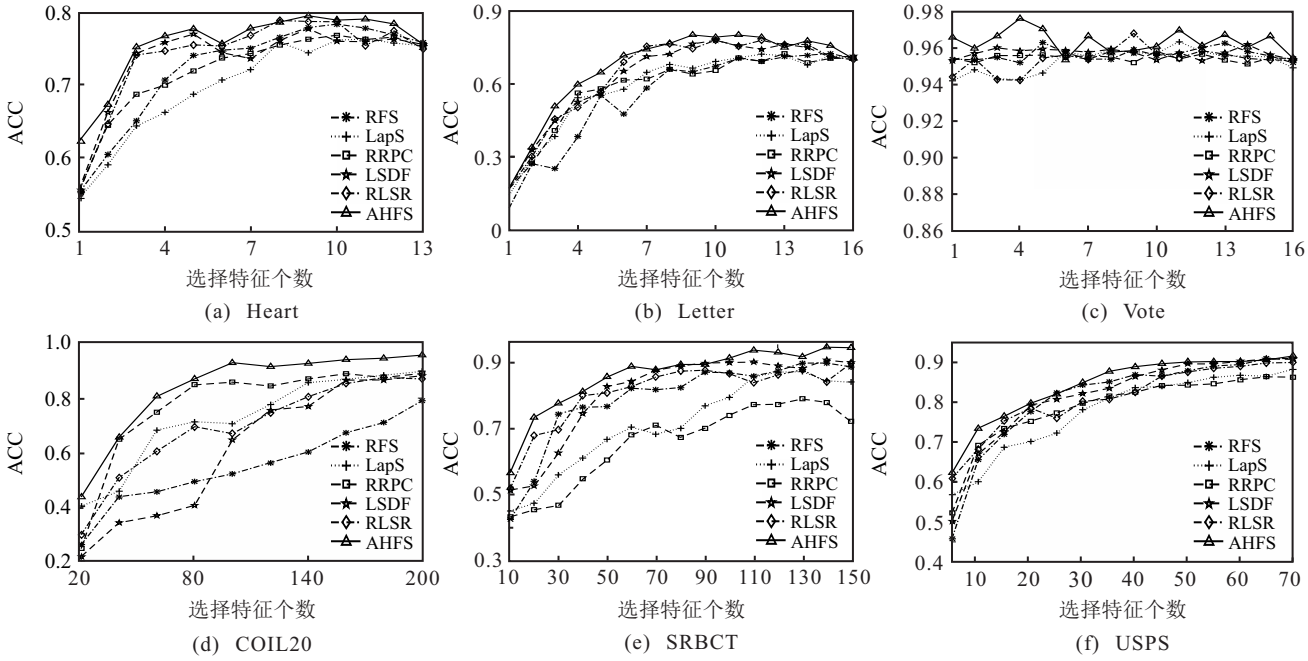

图 1　选择不同特征数的分类精度

这表明半监督方法可以利用未标记的数据固有的结构信息改善有监督方法的性能,验证了半监督特征选择方法的有效性.注意,重复实验多次,发现W行元素和值发生了变化,但各行元素和的排序不会发生变化,所获得的特征子集是一样的. 因此,本文所提方法所获得的特征在同一正则化参数下具有稳定性.

此外,为分析不同比例的标记训练数据对AHFS分类性能的影响,将训练集中标签数据的比率值分别设为{1%, 5%, 10%, 20%, 30%, 50%}. 在数据集Heart,设置对比模型的参数为最优参数,AHFS的平衡参数为1,同时,选择的特征子集数为6,结果如表3所示. 在数据集USPS中,选择的特征维度为25,实验结果如表4所示. 表中的加黑数据为相同标签比例条件下,对比算法中的最优结果.

表3　Heart集不同标记训练数据比率的比较(平均值±标准差)

模型	1%	5%	10%	20%	30%	50%
LSDF	56.66 ± 0.73	57.34 ± 1.16	61.35 ± 2.15	73.20 ± 2.76	76.54 ± 2.44	80.70 ± 2.64
RLSR	54.84 ± 0.64	56.85 ± 1.44	63.20 ± 2.19	71.41 ± 2.08	73.95 ± 2.48	78.39 ± 2.75
RRPC	55.61 ± 0.71	58.54 ± 1.89	65.86 ± 2.86	72.12 ± 2.69	78.67 ± 2.38	81.92 ± 2.14
AHFS	**57.22 ± 0.45**	**61.60 ± 0.92**	**68.85 ± 2.49**	**76.97 ± 1.83**	**79.87 ± 1.93**	**83.24 ± 2.38**

表4　USPS集不同标记训练数据比率的比较(平均值±标准差)

模型	1%	5%	10%	20%	30%	50%
LSDF	43.39 ± 1.16	73.64 ± 1.19	73.85 ± 0.88	77.04 ± 1.14	77.83 ± 1.82	77.85 ± 1.32
RLSR	49.81 ± 1.38	75.67 ± 1.12	70.51 ± 1.13	74.07 ± 1.31	73.79 ± 1.16	76.99 ± 1.05
RRPC	55.53 ± 1.48	67.28 ± 0.92	74.06 ± 1.09	77.18 ± 1.52	76.60 ± 1.25	79.11 ± 1.39
AHFS	**59.97 ± 0.63**	**80.54 ± 1.35**	**84.36 ± 1.00**	**85.37 ± 1.30**	**84.52 ± 1.08**	**85.73 ± 1.32**

对比表3和表4可知,AHFS的平均分类精确率高于所对比模型. 一般而言,拥有的标签数据越多,可获得的准确性越好. 这表明,如果有更多的标记数据可用,则模型能够选择具有更高质量的特征. 如果标签数据较少,则AHFS能够选择出判别性强的特征子集. 例如在USPS数据集,只有20% 的标签数据,模型的ACC结果为84.36,明显优于对比模型.

为说明本文方法应用的广泛性,本文也以深度特征作为输入进行仿真验证. 深度特征是由完全训练好的卷积神经网络处理产生的特征. 严格按照文献[20],卷积神经网络共有5个卷积层和3个权连接层,通过ImageNet集来调节卷积神经网络的参数[30],然后将COIL100数据集(包含100个类,每个类别72张,本文选取前40个类,每个类别选取40张)、

SUN397(108类,754张)[31]作为卷积神经网络的输入,以最后权连接层输出4 096维度的向量作为深层特征,得到深度特征集CNN-COIL、CNN-SUN. 在深度特征集CNN-COIL、CNN-SUN进行验证,结果如表5所示.

表5 基于深度特征的(平均值±标准差)

模型	Baseline	RRPC	LSDF	RLSR	AHFS
CNN-COIL	91.17 ± 0.14	92.23 ± 0.22	93.17 ± 0.36	92.70 ± 0.43	$\mathbf{93.54 \pm 0.13}$
CNN-SUN	42.73 ± 0.24	43.25 ± 0.32	$\mathbf{44.37 \pm 0.15}$	42.41 ± 0.12	44.08 ± 0.20

通过实验可知,对于经过卷积提取的深度特征数据集CNN-COIL, AHFS算法选择出的特征子集将分类性能提升了2.27%;对于深度特征数据集CNN-SUN,分类性能提升了1.35%. 由此可知,AHFS算法在深度空间上依然有着良好的性能.

3.4 平衡参数敏感性分析

本文提出的AHFS模型包含两个平衡参数μ、λ,先将自适应参数设置为$\tau = 0.1, p = 1$. 通过网格搜索策略,分析平衡参数在范围$\{10^{-3}, 10^{-2}, 10^{-1}, 1, 10^{2}, 10^{3}\}$内不同取值对AHFS算法性能的影响,图2、图3和图4分别表示μ、λ在数据集Heart、Letter、USPS对模型AHFS的影响.

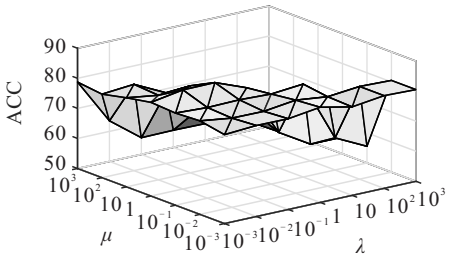

图2 μ、λ在Heart集对AHFS的影响

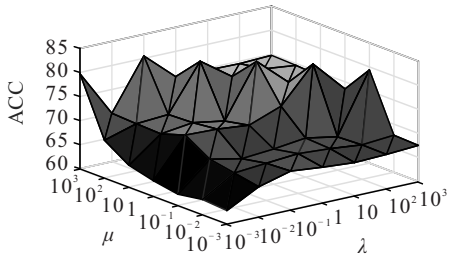

图3 μ、λ在Letter集对AHFS的影响

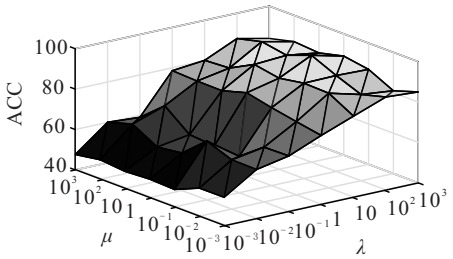

图4 μ、λ在USPS集对AHFS的影响

从3个数据集的实验结果中观察到:一方面,本文提出的方法表现出在较宽范围内受不同平衡参数μ和λ值的影响,参数的不同组合可导致选择不同的特征子集;另一方面,由于数据集相关的属性不同,没有找到类似的规则可为所有应用模型确定最佳参数.

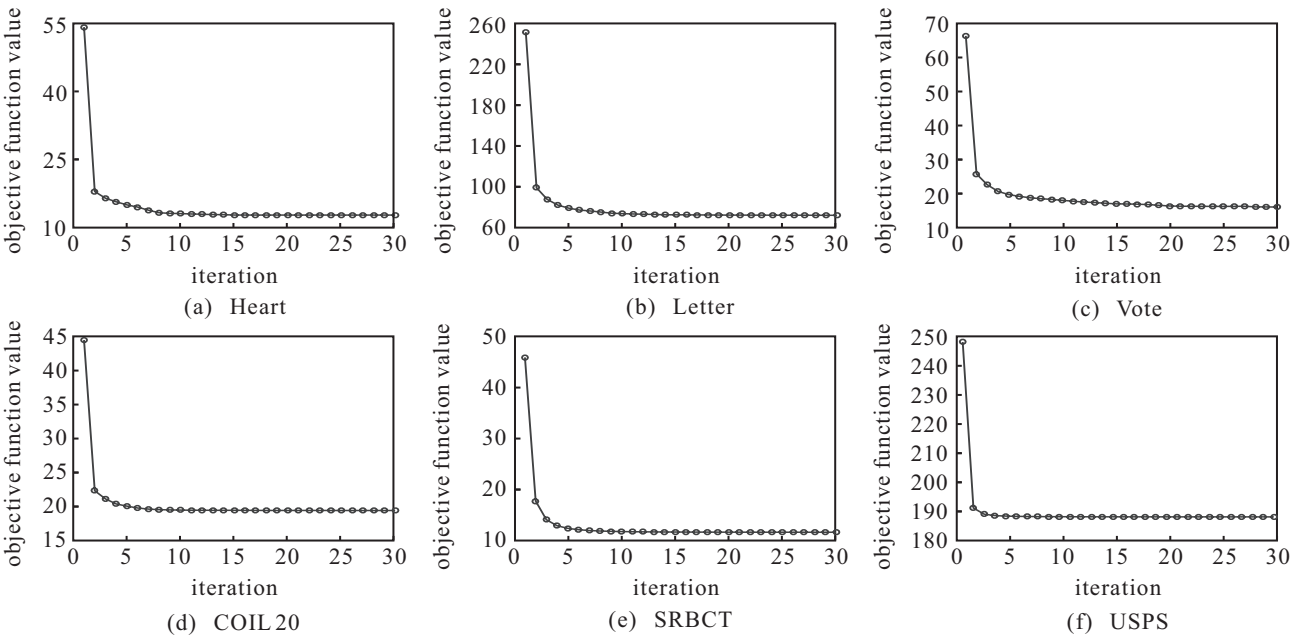

图5 实验中使用的所有数据集上,AHFS的收敛曲线

3.5 复杂性分析

在2.3节分析了AHFS的复杂度,本节以数据集Vote、COIL20、USPS为例,分别从3个数据集的原始特征中选取10、500、1 000个特征,计算各算法的时间消耗,结果如表6所示.

表6　选择对应特征的时间消耗

模型	RRPC	LSDF	RLSR	AHFS
Vote	1.64	0.99	1.39	1.55
COIL20	11.08	9.03	9.78	10.14
USPS	2 630.42	1 460.14	1 840.52	2 183.83

如表6所示,在3个数据集上LSDF的计算消耗都是最低,RRPC的消耗最大. 虽然AHFS涉及了Hessian的更新计算,理论上增加了计算量,但在迭代求解的过程中,该算法收敛速度快,只需几次迭代即可收敛,因此一定程度上减小了算法的计算消耗.

3.6 收敛性分析

本文提出了一种有效递归迭代算法求解目标函数的最优解,通过实验分析AHFS的收敛速度. 在所有数据集上,设置平衡参数μ和λ的值为1,目标值的变化曲线如图5所示. 由图5所得,在6个公开的数据集上实验,迭代30次内达到了收敛. 在数据集Letter、SRBCT、USPS迭代10次内可以达到收敛,证明了该迭代优化算法的快速性和有效性.

3.7 基因学应用

生物信息学的发展为生物和生物医学研究提供了大量的基因组和蛋白质组数据,半监督特征选择的研究对致病原因的分析、诊断发挥着重要的作用. 例如,在基因组学中,DNA微阵列数据可以在实验中测量数千个基因的表达水平,基因表达数据通常包含大量的基因,但样本数量较少. 然而,一种特定的疾病或生物功能通常与几个基因有关. 以急性淋巴细胞白血病(ALL)与急性髓细胞白血病(AML)的患者数据信息来评估每个基因与病症的相关性. 选取72个患者,其中急性淋巴细胞白血病有47人,急性髓细胞白血病有25人,分别选取0.2比率作为已知病人,其余病人信息作为未知案例,每个患者携带7 129个基因表达. 通过半监督特征选择方法获取相关性强的基因子集,依据基因子集可以有效地判断病症的类型. 选择不同基因特征数对应的诊断情况如图6所示.

由图6可以看出,通过AHFS算法选择的基因子集,判断出病症类型的准确率高于对比的选择模型. 因此,在白血病等疾病的基因组学中,本文提出的算法可以较好地利用患者的患病信息获得相关性强的基因子集,可以有效地判断患病的类型.

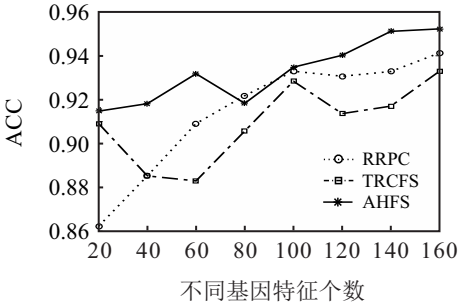

图6　选择不同基因特征子集的诊断率

4 结　论

本文提出了一种基于Hessian正则的自适应损失半监督稀疏特征选择框架(AHFS). 首先,AHFS能够较好地保留数据的局部流形结构,提高对有限标签信息的利用. 其次,AHFS融合了自适应损失,提升了模型对具有较小或者较大损失数据的鲁棒性,增强线性映射能力并通过$l_{2,p}$范数稀疏了投影矩阵,加强了特征之间的区分性,有利于选择判别性强的特征子集. 通过一种高效的替代优化算法解决了提出的挑战性问题,并通过实验验证了算法的收敛性. 在公开数据集上的实验结果表明了所提出的方法优于常用的特征选择算法. 通过在白血病基因学的实例应用,验证了本文算法在实际应用中的有效性. 最后,以卷积神经网络处理的深度特征验证了本文算法在深度空间的性能.

参考文献(References)

[1] Hou C P, Nie F P, Li X L L, et al. Joint embedding learning and sparse regression: A framework for unsupervised feature selection[J]. IEEE Trans Cybernetics, 2013, 44(6): 793-804.

[2] Wu C D, Lu Z W, Yu X S. Image super resolution reconstruction algorithm based on weighted random forest[J]. Control and Decision, 2019, 34(10): 2243-2248.

[3] Fu X, Shen Y T, Li H W, et al. A semi-supervised encoder generative adversarial networks model for image classification [J]. Acta Automation Sinica, 2020, 46(3): 531-539.

[4] Han Y H, Yang Y, Yan Y, et al. Semisupervised feature selection via spline regression for video semantic recognition[J]. IEEE Transactions on Neural Networks and Learning Systems, 2015, 26(2): 252-264.

[5] Jiang Y G, Ye G, Chang S F, et al. Consumer video understanding: A benchmark database and an evaluation of human and machine performance[C]. ICMR′11: Proceedings of the 1st ACM International Conference on Multimedia Retrieval. Trento: ACM, 2011: 1-8

[6] Wang S, Yang Y, Ma Z, et al. Action recognition by exploring data distribution and feature correlation[C]. IEEE Conference on Computer Vision and Pattern Recognition. Rhode Island: IEEE, 2012: 1370-1377.

[7] Nie F, Huang H, Cai X, et al. Efficient and robust feature selection via joint $l_{2,1}$-norms minimization[C]. Proceedings of the 23th International Conference on Neural Information Processing Systems. Vancouver: MIT Press, 2010: 1813-1821.

[8] Liu Y Q, Zhao H W, Wang Y. Video face recognition method based on QPSO and manifold learning[J]. Acta Automation Sinica, 2020, 46(2): 256-263.

[9] Moore B C. Principal component analysis in linear systems: Controllability, observability, and model reduction[J]. IEEE Transactions on Automatic Control, 1981, 26(1): 17-32.

[10] Zhu X F, 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.

[11] Gu Q Q, Li Z H, Han J W. Generalized Fisher score for feature selection[C]. Proceedings of the 27th Conference on Uncertainty in Artificial Intelligence. Barcelona: AUAI Press, 2011: 266-273.

[12] He X F, Cai D, Niyogi P. Laplacian score for feature selection[C]. Advances in Neural Information Processing Systems. Vancouver: MIT Press, 2006: 507-514.

[13] Shi L, Du L, Shen Y D. Robust spectral learning for unsupervised feature selection[C]. IEEE International Conference on Data Mining. Shenzhen: IEEE, 2014: 977-982.

[14] Luo Y, Tao D C, Xu C, et al. Vector-valued multi-view semi-supervsed learning for multi-label image classification[C]. The 27th AAAI Conference on Artificial Intelligence. Bellevue: AAAI, 2013: 647-653.

[15] Liu J W, Liu Y, Luo X L. Semi-supervised learning methods[J]. Chinese Journal of Computers, 2015, 38(8): 1592-1617.

[16] Zhao Z, Liu H. Semi-supervised feature selection via spectral analysis[C]. Proceedings of the SIAM International Conference on Data Mining. Minnesota: SIAM, 2007: 641-646.

[17] Kalakech M, Biela P, Macaire L, et al. Constraint scores for semi-supervised feature selection: A comparative study[J]. Pattern Recognition Letters, 2011, 32(5): 656-665.

[18] Ma Z L, Nie F P, Yang Y, et al. Discriminating joint feature analysis for multimedia data understanding[J]. IEEE Transactions on Multimedia, 2012, 14(6): 1662-1672.

[19] Xu Z, King I, Lyu M R T, et al. Discriminative semi-supervised feature selection via manifold regularization[J]. IEEE Transactions on Neural Networks, 2010, 21(7): 1033-1047.

[20] Zhang L, Zhang D. Visual understanding via multi-feature shared learning with global consistency[J]. Providence: IEEE Transactions on Multimedia, 2016, 18(2): 247-259.

[21] Eells J, Lemaire L. Selected topics in harmonic maps[M]. Providence: American Mathematical Society, 1983: 23-80.

[22] Donoho D L, Grimes C. Hessian eigenmaps: Locally linear embedding techniques for high-dimensional data[J]. Proceedings of the National Academy of Sciences, 2003, 100(10): 5591-5596.

[23] Kim K I, Steinke F, Hein M. Semi-supervised regression using Hessian energy with an application to semi-supervised dimensionality reduction[C]. Advances in Neural Information Processing Systems. Vancouver: MIT Press, 2009: 979-987.

[24] Nie F P, Wang H, Huang H, et al. Adaptive loss minimization for semi-supervised elastic embedding[C]. The 23th International Joint Conference on Artificial Intelligence. Beijing: Morgan Kaufmann, 2013: 1565-1571.

[25] Liu Y, Nie F P, Wu J, et al. Efficient semi-supervised feature selection with noise insensitive trace ratio criterion[J]. Neurocomputing, 2013, 105: 12-18.

[26] Wang D, Nie F P, Huang H. Large-scale adaptive semi-supervised learning via unified inductive and transductive model[C]. Proceedings of the 20th ACM SIGKDD Conference on Knowledge Discovery and Date Mining. New York: ACM, 2014: 482-491.

[27] Zhao J, Lu K, He X. Locality sensitive semi-supervised feature selection[J]. Neurocomputing, 2008, 71(10/11/12): 1842-1849.

[28] Xu J, Tang B, He H, et al. Semisupervised feature selection based on relevance and redundancy criteria[J]. IEEE Transactions on Neural Networks and Learning Systems, 2017, 28(9): 1974-1984.

[29] Chen X, Yuan G, Nie F, et al. Semi-supervised feature selection via rescaled linear Regression[C]. Proceedings of the 27th International Joint Conference on Aritficial Intelligence. Melbourne: Morgen Kaufmann, 2017: 1525-1531.

[30] Saenko K, Kulis B, Fritz M, et al. Adapting visual category models to new domains[C]. European Conference on Computer Vision. Berlin: Springer, 2010: 213-226.

[31] Zhou B, Lapedriza A, Xiao J, et al. Learning deep features for scene recognition using places database[C]. Advances in Neural Information Processing Systems. Montreal: MIT Press, 2014: 487-495.

作者简介

朱建勇 (1977−), 男, 副教授, 博士, 从事复杂工业过程控制与优化、大数据分析等研究, E-mail: zhujyemail@163.com;

周振辰 (1993−), 男, 硕士生, 从事机器学习、数据挖掘的研究, E-mail: zhenchenz@163.com;

杨辉 (1965−), 男, 教授, 博士生导师, 从事复杂系统建模、控制与优化、大数据分析等研究, E-mail: yhshuo@263.net;

聂飞平 (1977−), 男, 教授, 博士生导师, 从事机器学习以及相关应用领域等研究, E-mail: feipingnie@gmail.com.

(责任编辑：齐 霁)