

控制与决策

Control and Decision

基于稀疏化神经网络的浮选泡沫图像特征选择

朱建勇, 黄鑫, 杨辉, 聂飞平

引用本文:

朱建勇, 黄鑫, 杨辉, 等. 基于稀疏化神经网络的浮选泡沫图像特征选择[J]. 控制与决策, 2021, 36(7): 1627–1636.

在线阅读 View online: <https://doi.org/10.13195/j.kzyjc.2019.1788>

您可能感兴趣的其他文章

Articles you may be interested in

基于小波变换与差分变异BSO–BP算法的大坝变形预测

Dam deformation prediction based on wavelet transform and differential mutation BSO–BP algorithm

控制与决策. 2021, 36(7): 1611–1618 <https://doi.org/10.13195/j.kzyjc.2019.1431>

基于神经网络的电力系统暂态稳定分布式自适应控制

Neural network–based distributed adaptive control for power system transient stability

控制与决策. 2021, 36(6): 1407–1414 <https://doi.org/10.13195/j.kzyjc.2019.1168>

基于多维泰勒网的超前d步预测模型

d–step–ahead predictive model based on multi–dimensional Taylor network

控制与决策. 2021, 36(2): 345–354 <https://doi.org/10.13195/j.kzyjc.2019.0722>

复杂背景下全景视频运动小目标检测算法

Panoramic video motion small target detection algorithm in complex background

控制与决策. 2021, 36(1): 249–256 <https://doi.org/10.13195/j.kzyjc.2019.0686>

基于强化学习的倒立摆分数阶梯度下降RBF控制

Reinforcement learning based fractional gradient descent RBF neural network control of inverted pendulum

控制与决策. 2021, 36(1): 125–134 <https://doi.org/10.13195/j.kzyjc.2019.0816>

基于稀疏化神经网络的浮选泡沫图像特征选择

朱建勇^{1,2†}, 黄鑫^{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: 针对泡沫特征复杂繁多不利于建模控制的问题, 提出一种基于稀疏化神经网络的泡沫图像特征选择方法. 相较于大部分稀疏模型以线性回归模型作为损失函数的情况, 选择以更为贴近实际工业过程非线性特点的神经网络模型作为损失函数, 并加入 $L_{2,1}$ 范数约束以起到特征选择的效果; 此方法根据泡沫特征建立解决矿物品位回归问题的特征选择方法, 并采用近点梯度法计算最优解, 通过对第1层权值的综合排序得到特征选择子集; 最后, 利用支持向量机测试输入样本不同特征组合效果, 对比各特征子集得到浮选过程最优特征组合. 工业数据仿真结果表明, 所提出方法可以有效地实现泡沫图像维数约简.

关键词: 泡沫浮选; 稀疏模型; 神经网络; 特征选择; 维度约简

中图分类号: TP183

文献标志码: A

DOI: 10.13195/j.kzyjc.2019.1788

开放科学(资源服务)标识码(OSID):



引用格式: 朱建勇, 黄鑫, 杨辉, 等. 基于稀疏化神经网络的浮选泡沫图像特征选择 [J]. 控制与决策, 2021, 36(7): 1627-1636.

Selection method for froth image characters based on sparse neural network

ZHU Jian-yong^{1,2†}, HUANG Xin^{1,2}, YANG Hui^{1,2}, NIE Fei-ping³

(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: Aiming at the problem that the froth characteristics are complicated and not conducive to the modeling control, this paper proposes a bubble image feature selection method based on sparse neural networks. Compared with most sparse learning methods, the linear regression model is used as the loss function. The neural network model closer to the nonlinear actual industrial process is used as the loss function, and the $L_{2,1}$ -norm constraint condition is added to achieve the effect of feature selection. This method establishes a feature selection method based on the characteristics of the foam to solve the regression problem with the mineral level, and the optimal solution is calculated by the near-point gradient method. The comprehensive ranking of the first layer weights obtains the corresponding feature selection results. Finally, the support vector machine is used to detect the different feature combinations of the input samples, and the optimal feature combination of the flotation process is obtained. The industrial data simulation results show that the proposed method can effectively realize the dimensional reduction of the bubble image.

Keywords: froth flotation; sparse learning; neural network; feature selection; dimensional reduction

0 引言

泡沫浮选是利用矿物颗粒表面物理化学性质差异, 实现不同矿物有效分离的选矿方法. 具体表现为在矿浆中加入浮选药剂, 通过叶轮搅拌并切割空气产生大量气泡, 最后回收含矿泡沫以收集目标矿物. 由于含矿泡沫包含丰富的泡沫品位信息, 浮选工作人员

根据泡沫表面特征判断浮选工况. 近年来随着图像处理技术和计算机技术的不断发展, 机器视觉技术已大量地应用于矿物泡沫浮选过程.

通过矿物浮选泡沫特征的提取及分析, 大致可将泡沫图像特征分为静态特征、动态特征和统计特征等. 静态特征是泡沫浮选工况最显现的特征, 包括泡

收稿日期: 2019-12-22; 修回日期: 2020-02-18.

基金项目: 国家自然科学基金项目(61733005); 国家自然科学基金地区项目(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).

责任编辑: 薛建儒.

†通讯作者. E-mail: zhujyemail@163.com.

沫颜色^[1]和泡沫尺寸^[2]等;动态特征反映了泡沫的整体变化趋势,包括泡沫流速^[3]、稳定性^[4]和承载率^[5]等;统计特征主要指泡沫纹理^[6],反映了泡沫表面附着的矿物颗粒细腻程度,与矿物品位直接相关. 研究人员将这些泡沫特征应用于矿物浮选的软测量^[7]、入矿原料分类^[8]、故障诊断^[9]、液位控制^[10]和药剂量控制^[11]等.

由以上相关文献可知,泡沫图像特征提取方法繁多,同一特征可由多种机器特征描述,而特征之间存在非线性联系,在软测量、故障诊断等模型中易导致信息冗余,在浮选控制中应用太多的泡沫特征不利于实现浮选的有效控制^[12],因此迫切需要维数约简,以降低泡沫特征的复杂度,识别关键泡沫特征. 以往研究侧重根据实践经验判断关键特征,而本文试图在人工智能理论上,通过机器学习工业大数据解决该问题.

维数约简可以分为特征提取和特征选择两类方法. 比较直观的特征选择方法是通过对已有数据进行相关性分析,决定特征组合. 文献[13]提出一种基于模糊特征关系的关键特征提取方法,但该方法的主要特征选择依赖于主观选择,对复杂问题难以保证正确性;文献[14]利用互信息定义特征的条件相关度和冗余度,提出了一种可以解决非线性问题的无监督特征选择方法,但该方法无法适用于连续特征数据. 部分学者将智能算法融入特征选择过程,以提高得到最优解的可能. 文献[15]利用4种Filter方法对特征进行综合评分生成GA初始种群,然后用GA算法进行全局搜索得到最优特征组合,但由于该方法依赖于GA算法的全局搜索能力,存在过早收敛的缺点.

稀疏模型是一种嵌入式的特征选择方法,利用范数的稀疏特性对线性回归模型进行约束,而起到维度约简的目的,且根据约束函数的不同,实现不同的功能. 文献[16]提出的组lasso可以使预设组特征的回归系数更加相近,保留了与重要特征相关的特征;文献[17]提出的融合lasso,对相邻变量的回归系数之差加入惩罚,使相邻回归系数更为相近;文献[18]将线性模型推广为基于核函数的非线性模型得到核lasso;文献[19]提出基于图模型的图lasso;文献[20]利用贝叶斯公式的思想,以回归模型为最大后验估计,正则项为先验分布,损失函数项对应似然函数来求解回归系数;文献[21]提出了一种利用平滑函数替代lasso正则化,将改进后的稀疏函数与前馈神经网络结合的算法,以得到最大程度的稀疏网络;文献[22]提出了RFS(robust feature selection)模型,通过同

时对损失函数和正则项施加 $L_{2,1}$ 范数约束,使稀疏模型对异常值具有更好的鲁棒性.

特征提取方法通过一定的空间变换得到特征投影,消除特征相关性. 较为常用的方法有主成分分析法(PCA)^[23]和线性判别分析法(LDA)^[24],以及针对它们的相关改进方法. 文献[25]将互信息与PCA方法相结合,以互信息来判断特征相关性,但依然不能克服PCA不利于处理非高斯分布数据的缺点;文献[26]针对PCA获得的主成分空间难以解释的问题,将lasso与PCA相结合,提出了一种自适应稀疏PCA方法. 因此,研究维数约简以简化数据结果挖掘数据隐含信息,对浮选泡沫关键特征的获取意义重大.

虽然在人工智能领域特征选择方法众多,但是现有的特征选择方法主要针对分类标签进行图像识别和故障诊断,例如主成分分析、随机森林算法、互信息法、决策树算法,并没有解决工业过程中的连续特征问题. 因此,本文提出一种将稀疏模型和神经网络相结合的泡沫图像特征选择方法. 不同于以往稀疏模型主要用来解决线性分类方法^[27],本文方法建立基于非线性回归的特征选择模型,以期通过维数约简获取关键性泡沫特征. 其基本思路是利用神经网络模型善于求解非线性问题的特点,选择一个合适的神经网络模型,结合稀疏模型思想,使输入层到隐含层的权值更稀疏,并通过权值比重约简泡沫特征.

1 相关工作

1.1 lasso模型

1996年Tibshirani将岭回归估计的 L_2 范数替换为 L_1 范数,得到具有稀疏解的LASSO (least absolute shrinkage and selection operator)^[28]. 假定数据样本表示为 $\{(x_n, y_n)\}_{n=1}^N$,其中 $x_n = (x_{1,n}, \dots, x_{i,n}, \dots, x_{p,n})$ 为 p 维特征向量, y_n 为相应的目标输出,这时的目标输出 y_n 可由特征向量线性组合估计得到,即

$$\phi(x_n) = \beta_0 + \sum_{i=1}^p x_{i,n} \beta_i. \quad (1)$$

其中: $\beta = (\beta_1, \dots, \beta_p)$ 为权值向量, β_0 为偏置项.

lasso可以理解为满足 L_1 范数约束条件下求线性回归问题的最优解 (β_0, β) ,其中 $y = (y_1, \dots, y_N)$ 表示 p 维响应向量, \mathbf{X} 为一个 $N \times p$ 矩阵.

$$\begin{aligned} \min_{\beta_0, \beta} & \left\{ \frac{1}{2N} \|\mathbf{y} - \beta_0 \mathbf{I} - \mathbf{X}\beta\|_2^2 \right\}; \\ \text{s.t.} & \|\beta\|_1 < t. \end{aligned} \quad (2)$$

即保证回归误差尽可能小的前提下,使权值矩阵尽量稀疏. 通过拉格朗日乘子算法将有约束问题去无约束

化,可以构造如下拉格朗日函数:

$$\min_{\beta} \left\{ \frac{1}{2N} \|\mathbf{y} - \mathbf{X}\beta\|_2^2 + \lambda \|\beta\|_1 \right\}. \quad (3)$$

稀疏化模型是一种在稀疏约束条件下求解回归系数的特征选择方法,具有数据要求低、计算简单的特点,可以根据实际问题灵活变化稀疏约束条件。

1.2 BP神经网络

BP神经网络^[29]是一种运用反向传播算法训练多层神经网络的前馈神经网络,也是目前应用最为广泛的神经网络之一。

BP神经网络的目标函数为

$$\min F(x) = \min \|e\|_2^2. \quad (4)$$

其中:误差 $e = y - \hat{y}$, y 表示目标输出, \hat{y} 表示神经网络输出. BP神经网络的基本结构如图1所示。

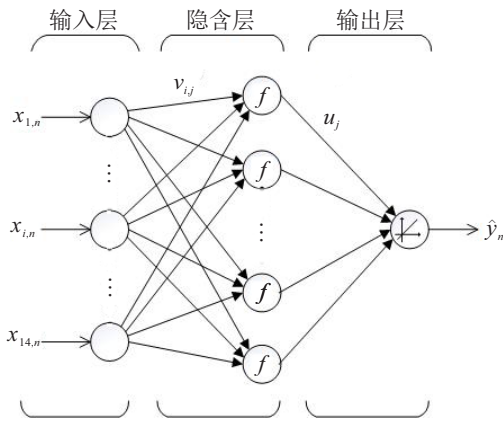


图1 BP神经网络结构

前向传播过程描述为

$$\hat{y}_n = \sum_{j=1}^J u_j f \left(\sum_{i=1}^I v_{i,j} x_{i,n} + b_1 \right) + b_2. \quad (5)$$

其中: i 为输入层神经元个数, j 为隐含层神经元个数, n 为样本个数, f 为隐含层激活函数, $v_{i,j}$ 为输入层神经元权值矩阵, u_j 为隐含层神经元权值矩阵, b_m 为各神经元阈值, m 为神经网络层数. 通过前向传播过程计算得到误差。

误差反向传播过程描述为

$$v_{i,j}(k+1) = v_{i,j}(k) - \eta \frac{\partial F}{\partial v_{i,j}}, \quad (6)$$

$$u_j(k+1) = u_j(k) - \eta \frac{\partial F}{\partial u_j}, \quad (7)$$

$$b_m(k+1) = b_m(k) - \eta \frac{\partial F}{\partial b_m}, \quad (8)$$

其中 η 为学习速度. 通过反向传播过程更新各层权值系数和偏置值, 经过前向传播和反向传播的反复迭代, 最后得到收敛后的最小二乘解。

2 稀疏化神经网络特征选择方法

2.1 稀疏化神经网络设计

虽然已有的稀疏模型众多, 但往往采用线性回归得到的平方误差作为损失函数, 以此方法处理泡沫图像特征选择问题并不符合工业数据非线性的特点. 本文针对浮选过程泡沫图像特征之间互为相关且非线性的特点, 运用lasso模型思想, 将BP神经网络善于解决非线性问题的优点与 $L_{2,1}$ 范数具有稀疏性的特点相结合来解决泡沫浮选过程的图像特征选择问题。

$L_{2,1}$ 范数表示特征矩阵每列的 L_2 范数之和, 有

$$\|\beta\|_{2,1} = \sum_{j=1}^p \|\beta_j\|_2. \quad (9)$$

为了较直观地体现与 $L_{2,1}$ 范数的功能, 可以假设 $\beta = (\beta_1, \beta_2)^T$, 即将特征分为两个特征组, 其中第1组特征为 $\beta_1 = (\beta_{11}, \beta_{12})^T$, 第2组特征为 β_2 , 由此可以得到在三维(特征)空间中的 $L_{2,1}$ 范数等值线, 如图2所示。

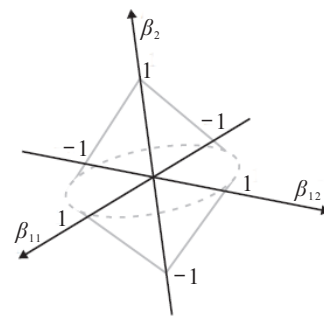


图2 $L_{2,1}$ 范数等值线

在三维空间中, L_1 范数等值线构成四棱锥, 表现为具有较好的稀疏性, 而 L_2 范数等值线构成球体, 可以使特征值更平滑, $L_{2,1}$ 范数兼顾 L_2 范数和 L_1 范数的优点, 使其在具有稀疏性的同时令同组特征值更相近。

本文利用单隐层BP神经网络代替lasso模型作为新的损失函数, 以 $L_{2,1}$ 范数作为约束条件, 构成目标函数如下:

$$\min_{\mathbf{u}, \mathbf{V}, b_m} \left\{ \|\mathbf{y} - f_2(\mathbf{u} f_1(\mathbf{V}\mathbf{X} + b_1) + b_2)\|_2^2 + \lambda \|\mathbf{V}\|_{2,1} \right\}. \quad (10)$$

其中: \mathbf{y} 为目标输出; f_1 为第1层激活函数; f_2 为第2层激活函数; 分别用 \mathbf{V} 、 \mathbf{u} 表示第1层权值矩阵和第2层权值向量, $\mathbf{W} = [\mathbf{V} \ \mathbf{u}]$, $\|\mathbf{V}\|_{2,1}$ 表示第1层(输入层)权值的 $L_{2,1}$ 范数表达; λ 表示稀疏系数, 其为正实数。

在 $L_{2,1}$ 范数约束下, 同一输入层神经元连接到隐

含层的权值会尽量相近,同时第1层的权值矩阵会尽量稀疏,即简化输入层与隐含层之间的连接结构,起到特征选择的作用.

2.2 目标函数求解

梯度下降法是求解最优化问题的常用方法.对于可微凸函数的寻优问题而言,全局最优解 w^* 的充要条件是梯度 $\nabla g(w^*) = 0$,梯度迭代公式为

$$w(k+1) = w(k) - \alpha \nabla g(w(k)). \quad (11)$$

为了能更直观地分析梯度下降法的寻优策略,将迭代过程写为线性化与欧氏距离相加并最小化求解的形式^[30],化简表示为

$$w(k+1) = \arg \min_w \left\{ \frac{1}{2\alpha} \|w - (w(k) - \alpha \nabla g(w(k)))\|_2^2 \right\}, \quad (12)$$

其中已省略与变量无关项.

由于式(10)目标函数中存在不可微点,本文采用近点梯度法求解该目标函数.将目标函数分为可微函数 E 和不可微函数 h 两部分,即

$$\begin{aligned} g &= E + \lambda h, \\ E &= \|\mathbf{y} - f_2(\mathbf{u}f_1(\mathbf{V}\mathbf{X} + \mathbf{b}_1) + \mathbf{b}_2)\|_2^2, \\ h &= \|\mathbf{V}\|_{2,1}. \end{aligned} \quad (13)$$

其中: E 为损失函数, h 为惩罚函数.先求解损失函数 E 的下降梯度,此时固定惩罚函数 h ,得到目标函数 g 的局部近似解

$$\begin{aligned} \mathbf{W}(k+1) &= \arg \min_w \left\{ \|\mathbf{V}(k)\|_{2,1} + \frac{1}{2\alpha} \|\mathbf{W} - (\mathbf{W}(k) - \alpha \nabla g(\mathbf{W}(k)))\|_2^2 \right\}. \end{aligned} \quad (14)$$

再固定损失函数 E (将梯度投影视为常量),对惩罚函数 h 进行求导,得到下一梯度的迭代结果.

根据以上的求解思路,令泡沫图像样本为 $\{(\mathbf{x}_n, y_n)\}_{n=1}^N$, \mathbf{x}_n 表示泡沫图像特征集, y_n 表示矿物品位.选择 logsig 函数作为隐含层激活函数,有

$$f_1(x) = \frac{1}{1 + e^{-x}}; \quad (15)$$

purelin 函数作为输出层激活函数,有

$$f_2(x) = x. \quad (16)$$

将设定参数代入式(10)可以得其目标函数为

$$\min_{\mathbf{V}, \mathbf{u}, \mathbf{b}_m} \left\{ \|\mathbf{y} - \mathbf{u}f_1(\mathbf{V}\mathbf{X} + \mathbf{b}_1) - \mathbf{b}_2\|_2^2 + \lambda \|\mathbf{V}\|_{2,1} \right\}. \quad (17)$$

先固定惩罚函数 h ,以损失函数 E 来近似目标函数 g .用最速梯度下降法求解BP神经网络,其思路

为:先根据前向传播式(5)计算各神经元输出,即

$$\mathbf{o}_1 = \mathbf{V}\mathbf{X} + \mathbf{b}_1, \quad \mathbf{o}_2 = \mathbf{u}f_1(\mathbf{o}_1) + \mathbf{b}_2. \quad (18)$$

然后根据误差反向计算各层偏导数

$$\frac{\partial E}{\partial v_{i,j}} = \frac{\partial E}{\partial \mathbf{o}_1} \times \frac{\partial \mathbf{o}_1}{\partial v_{i,j}}, \quad (19)$$

$$\frac{\partial E}{\partial u_j} = \frac{\partial E}{\partial \mathbf{o}_2} \times \frac{\partial \mathbf{o}_2}{\partial u_j}, \quad (20)$$

$$\frac{\partial E}{\partial b_m} = \frac{\partial E}{\partial \mathbf{o}_m} \times \frac{\partial \mathbf{o}_m}{\partial b_m}. \quad (21)$$

令 \mathbf{s}_m 表示敏感性,有

$$\mathbf{s}_m = \frac{\partial E}{\partial \mathbf{o}_m}. \quad (22)$$

引入导数矩阵

$$\hat{\mathbf{F}}_m(\mathbf{o}_m) = \begin{bmatrix} \dot{f}_m(o_{m,1}) & 0 & \dots & 0 \\ 0 & \dot{f}_m(o_{m,2}) & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \dot{f}_m(o_{m,i}) \end{bmatrix},$$

$$\frac{\partial \mathbf{o}_2}{\partial \mathbf{o}_1} = \mathbf{u}\hat{\mathbf{F}}_1(\mathbf{o}_1). \quad (23)$$

由此可以得到

$$\mathbf{s}_2 = \frac{\partial E}{\partial \mathbf{o}_2} = -2\hat{\mathbf{F}}_2(\mathbf{o}_2)(\mathbf{y} - \hat{\mathbf{y}}) = -2(\mathbf{y} - \mathbf{o}_2), \quad (24)$$

$$\mathbf{s}_1 = \left(\frac{\partial \mathbf{o}_2}{\partial \mathbf{o}_1} \right)^T \frac{\partial E}{\partial \mathbf{o}_2} = \hat{\mathbf{F}}_1(\mathbf{o}_1)(\mathbf{v})^T \mathbf{s}_2. \quad (25)$$

最后更新权值和偏置值为

$$\mathbf{V}(k+1) = \mathbf{V}(k) - \eta \mathbf{s}_1(\mathbf{X})^T, \quad (26)$$

$$\mathbf{u}(k+1) = \mathbf{u}(k) - \eta \mathbf{s}_2(f_1(\mathbf{o}_1))^T, \quad (27)$$

$$b_m(k+1) = b_m(k) - \eta \mathbf{s}_m. \quad (28)$$

可以使 $h(w)$ 得到一个尽可能小且接近不可微点的近似解

$$\mathbf{Z} = \mathbf{V}(k) - \eta \mathbf{s}_1(\mathbf{X})^T. \quad (29)$$

由于每一步梯度下降迭代近似等价于目标函数,而第2层权值矩阵 \mathbf{u} 与约束函数 h 无关,这样在加入约束条件时对第2层权值 \mathbf{u} 更新并无影响,可以将稀疏化神经网络的权值寻优问题转化为一个更简单的近似问题,得到的投影算子

$$\mathbf{V}(k+1) = \arg \min_{\mathbf{V}} \|\mathbf{V} - \mathbf{Z}\|_2^2 + \lambda \|\mathbf{V}\|_{2,1}, \quad (30)$$

即每一步梯度下降的同时考虑 $L_{2,1}$ 范数最小化.

对式(29)进行求导,可以得到

$$\mathbf{V}(k+1) = \left(1 - \frac{\eta \lambda}{\|\mathbf{Z}\|_2} \right) \mathbf{Z}, \quad (31)$$

迭代公式的具体形式为

$$\mathbf{Z} \leftarrow \mathbf{V}(k) - \eta \mathbf{s}_1(\mathbf{X})^T, \quad (32)$$

$$\mathbf{V}(k+1) \leftarrow \left(1 - \frac{\eta\lambda}{\|\mathbf{Z}\|_2}\right) + \mathbf{Z}. \quad (33)$$

2.3 算法步骤

根据以上的目标函数设计及算法求解分析,设计算法程序逻辑,将整个过程分为:数据预处理,特征选择和结果验证3个过程.

预处理是通过归一化处理样本数据,以避免不同量纲对结果的影响.

列向量中心化

$$\frac{1}{N} \sum_{n=1}^N x_{i,n} = 0; \quad (34)$$

方差单位化

$$\frac{1}{N} \sum_{n=1}^N x_{i,n}^2 = 1. \quad (35)$$

特征选择过程是根据2.2节的步骤求解目标函数,得到输入层权值,以权值的绝对值为排序依据,得到针对泡沫图像的特征排序.流程描述见算法1.

算法1 特征选择.

输入: 训练集 S , 稀疏系数 λ , 步长 η , 最大迭代数 T 和停止条件 ε .

过程:

- 1) 随机产生初始值 $\mathbf{V}, \mathbf{u}, b_m$;
- 2) $k = 0$;
- 3) while $k < T$ do
- 4) 根据式(15)和(17)计算目标函数 g ;
- 5) if $g > \varepsilon$ then
- 6) 根据式(19)和(25)计算下降梯度;
- 7) 以式(26)迭代更新第2层权值;
- 8) 以式(27)和(33)迭代更新第1层权值 $\mathbf{V}(k+1)$;
- 9) $k = k + 1$;
- 10) else
- 11) break
- 12) end if
- 13) end while
- 14) 计算第1层权值 \mathbf{V} 的行绝对值之和;
- 15) 根据第14步的结果将对应特征排序.

输出: 特征重要性排序.

结果验证是根据特征选择后的结果,利用SVR^[31]模型进行检验,比较得出固定特征排序下的最优组合结果,结果验证步骤如算法2所示.

算法2 结果验证.

输入: 数据集 D , 特征重要性排序.

过程:

- 1) $p = 0$;
- 2) while $p < 14$ do

3) $p = p + 1$;

4) 根据重要程度提取样本的前 p 个特征,组成新子集 $[D_p]$;

5) 将新子集 $[D_p]$ 输入到SVR模型,得到测试结果;

6) end while

7) 比较不同维度组合的误差.

输出: 最优特征组合.

2.4 收敛性分析

本节主要对SBP算法进行收敛性分析,根据文献[32]中定理1,从单调性和有界性的角度分析,当目标函数满足如下假设时,SBP算法一定可以达到收敛.

假设1 激活函数 f^1 和 f^2 在定义域内可微,在任意有界闭区间内 $f^1, f^2, f^{1'}$ 和 $f^{2'}$ 是连续的.

假设2 存在有界区域 $\Omega \subset R^n$ 使 $\{\mathbf{W}\} \subset \Omega$.

定理1 设 $q(x)$ 是在有界闭区间 $\Omega \subset R^n$ 上定义的函数,且 $q'(x)$ 连续,则令常数 $C > 0$, 有

$$q(x) \leq q(x_0) + q'(x_0)(x - x_0) + \frac{C}{2}(x - x_0)^2, \quad \forall x_0, x \in \Omega. \quad (36)$$

如果以式(10)定义目标函数,以目标函数 g 为李雅普诺夫函数,以式(19)~(33)更新权值 \mathbf{W} ,则在满足假设条件下,由定理1可知以下条件成立:

$$E(\mathbf{W}(k+1)) \leq E(\mathbf{W}(k)) + 2(\mathbf{o}_2(k) - \mathbf{y})\Delta\mathbf{o}_2(k) + \frac{C_1}{2}(\Delta\mathbf{o}_2(k))^2, \quad (37)$$

$$h(\mathbf{V}(k+1)) \leq h(\mathbf{V}(k)) + h'(\mathbf{V}(k))\Delta\mathbf{V}(k) + \frac{C_2}{2}(\Delta\mathbf{V}(k))^2, \quad (38)$$

$$\Delta\mathbf{o}_2(k) = \mathbf{o}_2(k+1) - \mathbf{o}_2(k), \quad (39)$$

$$\Delta\mathbf{V}(k) = \mathbf{V}(k+1) - \mathbf{V}(k). \quad (40)$$

其中:常数 $C_1 > 0, C_2 > 0$.

由式(37)和(38)可以推得

$$g(\mathbf{W}(k+1)) \leq g(\mathbf{W}(k)) + \Delta\mathbf{V}(k)\mathbf{s}_1(\mathbf{V}(k)) + \Delta\mathbf{u}(k)\mathbf{s}_2(\mathbf{u}(k)) + \xi(k) = g(\mathbf{W}(k)) - \frac{1}{\eta}\|\Delta\mathbf{W}(k)\|^2 + \xi(k), \quad (41)$$

$$f_1(k) = f_1(\mathbf{V}(k)\mathbf{X}), \quad (42)$$

$$\Delta\mathbf{u}(k) = \mathbf{u}(k+1) - \mathbf{u}(k). \quad (43)$$

通过推导(完整过程详见附录A)可以将式(41)简化为

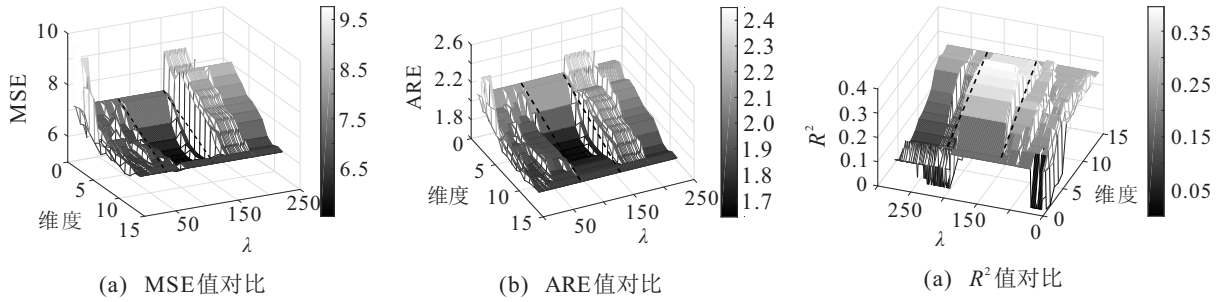


图3 不同参数下的特征选择结果

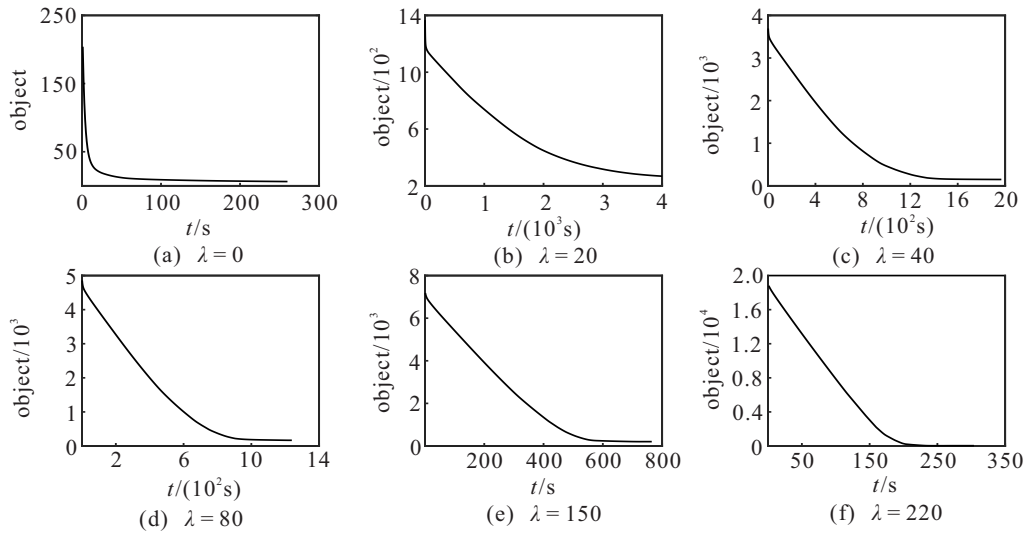


图4 不同参数下的特征选择结果

作为评价标准,得出不同维度特征子集输入下,稀疏系数对算法性能的影响,结果如图3所示.其中图3(a)和图3(b)为MSE、ARE的结果,代表误差,其数值越低表示算法效果越好,图3(c)为 R^2 的结果,代表预测合理性,其数值越高预测结果越合理,图中两条虚线之间为最合适的变量取值范围.由图3可以看出:随着稀疏系数 λ 的变化,整体误差呈先下降后上升趋势,当 λ 很小时,模型泛化性能较低,信息存在冗余,可能出现过拟合现象,到达合适数值区域后,信息冗余降低,泛化性能提高;当 λ 过大时,模型丧失拟合能力,导致误差过大.由此得出,对于铜浮选首槽泡沫图像特征选择问题的最合适稀疏系数区间为 $\lambda \in [70 \ 150]$.

3.3 收敛性分析

通过实际数据仿真检验算法的收敛性,取不同稀疏系数值,观察目标函数收敛情况,其余参数与上文相同,由此得到的目标函数收敛结果如图4所示.不同稀疏系数值代表稀疏约束的强弱的各个阶段,其中 $\lambda = 0$ 表示无稀疏约束状态,此时目标函数等价于神经网络,无特征选择功能.随着稀疏系数 λ 的增大,收敛速度加快,初始目标函数值增大.可以看出,在各种情况下目标函数都可以达到收敛,表明目标函数可以稳定收敛.

3.4 算法比较

通过稀疏化神经网络对训练数据进行学习,得出特征重要性排序,依次选出特征组成新数据集.为了全面地测试稀疏化神经网络(SBP)的各项性能,本文选择序列前向选择(SFS)^[33]、主成分分析法(PCA)^[23]、稀疏化PCA(SPCA)^[26]、稀疏化人工神经网络(SANN)^[21]和RFS^[22]作为对比算法,依照相同的步骤得到新数据集.为了保证对比情况的真实性,采用同一参数的LibSVM^[34]作为特征选择结果的测试模型.用新数据集训练SVR模型,以均方误差、平均相对误差和决定系数作为评价标准,分别得到的对比结果如图5所示.对比图5中不同评价指标下的算法结果,从横向上观察,RFS、SANN、SFS和SBP这些特征选择方法存在低维特征子集测试结果优于高维特征子集的现象,即选择后结果优于全输入(不选择)的结果,说明主观选择的泡沫图像特征中确实存在着数据冗余的现象,需要通过特征选择来提高精度.而传统PCA和SPCA这类降维方法,并没有提高预测精度,不能有效提取出主要信息,也无法解释新生特征的物理意义,并不适用于此类问题;从纵向比较,相较于其他算法,本文的SBP算法在MSE和ARE中得到的最小误差更低,模型精度更高,泛化效果好.随着特征维

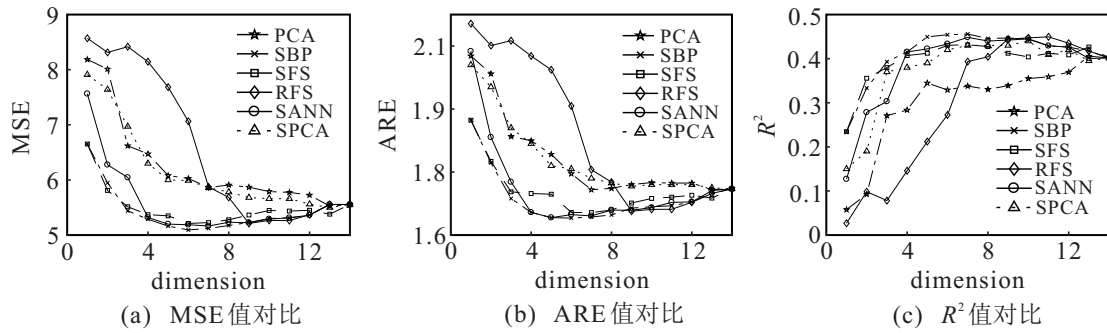


图5 SVR模型测试结果对比

度的变化, SANN、SFS和SBP误差变化呈现一个平稳下降的趋势, 相较于其他方法更为效果稳定. 在低维时, SBP和SFS可以更有效地找到最优组合, 明显优于PCA、SACA和RFS. 但从仿真运行过程时间上看, PCA、SPCA和SFS更加快速, SANN、RFS和SBP精度更高但算法更复杂.

同时, 将245个真实工业数据样本按照等距区间内样本数量比例, 随机选择20%、40%、60%和80%的样本作为训练集, 其余作为测试集, 得到训练比例对算法的影响, 如表3所示.

表3 不同训练比例均方误差

模型	20%	40%	60%	80%
PCA	7.16	6.89	6.54	5.56
SBP	6.88	6.05	5.66	5.12
SFS	6.54	6.04	5.66	5.27
RFS	7.14	6.96	6.56	5.24
SANN	6.93	6.65	6.22	5.36
SPCA	7.36	7.03	6.53	5.43

3.5 实验结论

综合以上的实验图表和仿真结论, 同时考虑到降低工业过程模型复杂度的目的, 对于铜浮选粗选首槽的泡沫图像最佳特征组合是: B值(RGB)、红色分量、黄色分量、稳定度、泡沫平均尺寸和泡沫峰值, 且稀疏化神经网络模型能有效地降低复杂度和预测误差, 达到了预期目标. 这6个特征分别对应了泡沫的颜色、破碎率和尺寸等关键人工特征, 可以代表大部分的泡沫图像信息, 符合真实的人工识别经验. 对于得出的特征中没有泡沫纹理特征问题, 分析认为: 实验对象选择的是粗选首槽泡沫图像, 此时的泡沫纹理影响相对较小, 不是有效特征. 单纯采用灰度值描述泡沫纹理不能有效体现纹理的全部特征信息, 可能误导实验结果. 因此, 该实验结论真实有效, 稀疏化神经网络模型能实现浮选泡沫图像的维数简约.

4 结论

针对泡沫图像特征数量繁多, 泡沫特征与矿物品位之间呈非线性关系, 且特征间存在复杂相关性的问题, 本文提出了一种基于稀疏化神经网络的特征选择方法. 该算法是通过BP神经网络模型对特征变量进行回归估计, 并对第1层网络加入范数约束, 在保留原数据非线性特点的情况下以尽量少的特征拟合出矿物品位. 该方法通过神经网络能有效处理非线性问题的优势, 更好地保留了数据特性, 能够更加有效地处理工业实际过程的特征选择问题, 有利于图像处理和机器视觉在工业实践中的应用.

附录A

李雅普诺夫函数单调性分析详细推导过程(所有符号与上文相同). 将式(37)进一步拆分可以得到

$$\Delta \mathbf{o}_2(k) = \mathbf{u}(k+1)f_1(k+1) - \mathbf{u}(k)f_1(k) = \Delta \mathbf{u}(k)f_1(k) + \mathbf{u}(k)(f_1(k+1) - f_1(k)) + \Delta \mathbf{u}(k)(f_1(k+1) - f_1(k)), \quad (A1)$$

$$\mathbf{u}(k)(f_1(k+1) - f_1(k)) = \mathbf{u}(k)f_1'(k)\Delta \mathbf{V}(k)\mathbf{X} + \mathbf{u}(k)(\Delta \mathbf{V}(k)\mathbf{X})^2 \times \int_0^1 (1-t)f_1''(\mathbf{V}(k) + t\mathbf{V}(k))\mathbf{X}dt. \quad (A2)$$

在式(41)中

$$g(\mathbf{W}(k)) = E(\mathbf{W}(k)) + h(\mathbf{W}(k)), \quad (A3)$$

$$\Delta \mathbf{u}(k)\mathbf{s}_2(\mathbf{u}(k)) = 2(\mathbf{o}_2(k) - \mathbf{y})\Delta \mathbf{u}(k)f_1(k), \quad (A4)$$

$$\Delta \mathbf{V}(k)\mathbf{s}_1(\mathbf{V}(k)) = 2(\mathbf{o}_2(k) - \mathbf{y})\mathbf{u}(k)f_1'(k)\Delta \mathbf{V}(k)\mathbf{X}, \quad (A5)$$

$$\xi_m(k) = \delta_1 + \delta_2 + \delta_3 + \delta_4. \quad (A6)$$

由式(50)~(55)可得

$$\delta_1 = (\mathbf{o}_2(k) - \mathbf{y})\Delta \mathbf{u}(k)(f_1(k+1) - f_1(k)) \leq C_3\|\Delta \mathbf{W}(k)\|^2, \quad (A7)$$

$$\delta_2 = (\mathbf{o}_2(k) - y)\Delta\mathbf{u}(k)(\Delta\mathbf{V}\mathbf{X})^2 \leq C_4\|\Delta\mathbf{W}(k)\|^2, \quad (\text{A8})$$

$$\delta_3 = \frac{1}{2}C_1(-\mathbf{u}(k)f_1(k))^2 \leq C_5\|\Delta\mathbf{W}(k)\|, \quad (\text{A9})$$

$$\delta_4 = \frac{\lambda}{2}C_2\|\Delta\mathbf{V}(k)\|^2 \leq C_6\|\Delta\mathbf{W}(k)\|^2. \quad (\text{A10})$$

因此

$$\begin{aligned} \xi_m &= \delta_1 + \delta_2 + \delta_3 + \delta_4 \leq \\ &(C_3 + C_4 + C_5 + C_6)\|\Delta\mathbf{W}(k)\|^2 \leq \\ &C_7\|\Delta\mathbf{W}(k)\|^2. \end{aligned} \quad (\text{A11})$$

由式(57)~(61)可以将式(53)简化为

$$g(\mathbf{W}(k+1)) \leq g(\mathbf{W}(k)) - \left(\frac{1}{\eta} - C_7\right)\|\Delta\mathbf{W}(k)\|^2. \quad (\text{A12})$$

由此可以看出,当 $\eta \leq 1/C_7$ 时, $g(\mathbf{W}(k+1)) \leq g(\mathbf{W}(k))$ 成立,因此李雅普诺夫函数具有单调性。

参考文献(References)

- [1] 陈宁, 黄璐, 桂卫华, 等. 泡沫图像统计建模与恒常颜色校正算法研究[J]. 控制理论与应用, 2016, 33(5): 613-620.
(Chen N, Huang L, Gui W H, et al. The statistical modeling and constant color correction algorithms for froth image[J]. Control Theory & Applications, 2016, 33(5): 613-620.)
- [2] 周开军, 桂卫华, 朱红求. 矿物浮选泡沫图像分割参数的最优性分析[J]. 华中科技大学学报: 自然科学版, 2013, 41(S1): 432-435.
(Zhou K J, Gui W H, Zhu H Q. Parameter optimality analysis for mineral floatation froth image segmentation[J]. Journal of Huazhong University of Science and Technology: Natural Science Edition, 2013, 41(S1): 432-435.)
- [3] 刘颖, 张平, 赵珺, 等. 基于尺度不变特征变换的浮选泡沫图像动态特性提取方法[J]. 控制理论与应用, 2016, 33(6): 718-726.
(Liu Y, Zhang P, Zhao J, et al. Dynamic characteristic extraction method of floatation froth image based on scale invariant feature transform[J]. Control Theory & Applications, 2016, 33(6): 718-726.)
- [4] 陈良琴, 王卫星. 基于泡沫表面特征变化及支持向量的气泡稳定度测量[J]. 中国矿业大学学报, 2018, 47(3): 652-661.
(Chen L Q, Wang W X. Bubble stability measurement based on froth surface feature variation and support vector machine[J]. Journal of China University of Mining Technology, 2018, 47(3): 652-661.)
- [5] Hyotyniemi H, Ylinen R. Modeling of visual floatation froth data[J]. Control Engineering Practice, 2000, 8(3): 313-318.
- [6] 马爱莲, 徐德刚, 谢永芳, 等. 基于复杂网络时空特性的泡沫图像动态纹理特征分析[J]. 化工学报, 2017, 68(3): 1023-1031.
(Ma A L, Xu D G, Xie Y F, et al. Analysis of dynamic texture features of floatation froth images based on space-time characteristics of complex networks[J]. CIESC Journal, 2017, 68(3): 1023-1031.)
- [7] Nakhaei F, Irannajad M, Mohammadnejad S. Evaluation of column floatation froth behavior by image analysis: effects of operational factors in desulfurization of iron ore concentrate[J]. Energy Sources, Part A: Recovery, Utilization, and Environmental Effects, 2018, 40(19): 2286-2306.
- [8] Fu Y, Aldrich C. Flotation froth image recognition with convolutional neural networks[J]. Minerals Engineering, 2019, 132: 183-190.
- [9] Khushaal Popli, Masih Sekhvat, Artin Afacan, et al. Dynamic modeling and real-time monitoring of froth floatation[J]. Minerals, 2015, 5(3): 570-591.
- [10] 赵洪伟, 谢永芳, 蒋朝辉, 等. 基于泡沫图像特征的浮选槽液位智能优化设定方法[J]. 自动化学报, 2014, 40(6): 1086-1097.
(Zhao H W, Xie Y F, Jiang Z H, et al. An intelligent optimal setting approach based on froth features for level of floatation cells[J]. Acta Automatica Sinica, 2014, 40(6): 1086-1097.)
- [11] 朱建勇, 桂卫华, 阳春华, 等. 基于泡沫尺寸随机分布的铜粗选药剂剂量控制[J]. 自动化学报, 2014, 40(10): 2089-2097.
(Zhu J Y, Gui W H, Yang C H, et al. Reagent dosage control based on bubble size random distribution for copper roughing[J]. Acta Automatic Sinica, 2014, 40(10): 2089-2097.)
- [12] Zhu J Y, Gui W H, Liu J P, et al. Combined fuzzy based feedforward and bubble size distribution based feedback control for reagent dosage in copper roughing process[J]. Journal of Process Control, 2016, 39: 50-63.
- [13] 熊熙, 乔少杰, 韩楠, 等. 一种基于模糊选项关系的关键属性提取方法[J]. 计算机学报, 2019, 42(1): 190-202.
(Xiong X, Qiao S J, Han N, et al. A fuzzy-option based attribute discriminant method[J]. Chinese Journal of Computers, 2019, 42(1): 190-202.)
- [14] 徐峻岭, 周毓明, 陈林, 等. 基于互信息的无监督特征选择[J]. 计算机研究与发展, 2012, 49(2): 372-382.
(Xu J L, Zhou Y M, Chen L, et al. An unsupervised feature selection approach based on mutual information[J]. Journal of Computer Research and Development, 2012, 49(2): 372-382.)

- [15] 高鹏毅, 陈传波, 张葵, 等. 一种使用多Filter初始化GA种群的混合特征选择模型[J]. 小型微型计算机系统, 2012, 33(11): 2379-2384.
(Gao P Y, Chen C B, Zhang K, et al. Hybrid model initializing the genetic population with multiple filters for feature selection[J]. Journal of Chinese Computer Systems, 2012, 33(11): 2379-2384.)
- [16] Ming Y, Yi L. Model selection and estimation in regression with grouped variables[J]. Journal of the Royal Statistical Society, 2006, 68(1): 49-67.
- [17] Rinaldo A. Properties and refinements of the fused lasso[J]. The Annals of Statistics, 2009, 37(5B): 2922-2952.
- [18] Wang G, Yeung D Y, Lochoovsky F H, The kernel path in kernelized LASSO[C]. Proceedings of the 11th International Conference on Artificial Intelligence and Statistics. San Juan, 2007: 580-587.
- [19] Yuan M, Lin Y. Model selection and estimation in the Gaussian graphical model[J]. Biometrika, 2007, 94(1): 19-35.
- [20] Park T, Casella G. The bayesian lasso[J]. Journal of the American Statistical Association. 2008, 103(482): 681-686.
- [21] Basava Naga Girish Koneru, Vinita Vasudevan. Sparse artificial neural networks using a novel smoothed LASSO penalization[J]. IEEE Transactions on Circuits and Systems II: Express Briefs, 2019, 66(5): 848-852.
- [22] Lan G S, Hou C, Nie F, et al. Robust feature selection via simultaneous sapped norm and sparse regularizer minimization[J]. Neurocomputing, 2018, 283: 228-240.
- [23] Jolliffe I T. Pincipal component analysis[J]. Journal of Marketing Research, 2002, 87(4): 513-519.
- [24] Fisher R A, The use of multiple measurements in taxonomic problems[J]. Annals of Eugenics, 2012, 7(2): 179-188.
- [25] 范雪莉, 冯海泓, 原猛. 基于互信息的主成分分析特征选择算法[J]. 控制与决策, 2013, 28(6): 915-919.
(Fan X L, Feng H H, Yuan M. PCA based on mutual information for feature selection[J]. Control and Decision, 2013, 28(6): 915-919.)
- [26] Seghouane Abd-Krim, Shokouhi Navid, Koch Inge. Sparse principal component analysis with preserved sparsity pattern[J]. IEEE Transactions on Image Processing: A Publication of the IEEE Signal Processing Society, 2019, 28(7): 3274-3285.
- [27] 王大荣, 张忠占. 线性回归模型中变量选择方法综述[J]. 数理统计与管理, 2010, 29(4): 615-627.
(Wang D R, Zhang Z Z. Variable selection for linear regression models: A survey[J]. Journal of Applied Statistics and Management, 2010, 29(4): 615-627.)
- [28] Tibshirani R. Regression shrinkage and selection via the lasso[J]. Journal of the Royal Statistical Society: Series B (Methodological), 1996, 58(1): 267-288.
- [29] David E Rumelhart, Geoffrey E Hinton, Ronald J Williams. Learning representations by back propagating errors[J]. Nature, 1986, 323(6088): 533-536.
- [30] Hastie Trevor Tibshirani, Wainwright M R. Statistical learning with sparsity: The lasso and generalizations[M]. California: Taylor Francis Group, 2015: 84-87.
- [31] Du Shu-Xin, Wu Tie-Jun. Support vector machines for regression[J]. Acta Simulata Systematica Sinica, 2003, 15(11): 1546-1580.
- [32] Wang J, Wen Y Q, Ye Z Y, et al. Convergence analysis of BP neural networks via sparse response regularization[J]. Applied Soft Computing, 2017, 61: 354-363.
- [33] Mao K Z. Fast orthogonal forward selection algorithm for feature subset selection[J]. IEEE Transactions on Neural Networks, 2002, 13(5): 1218-1224.
- [34] 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.

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;

黄鑫(1995—), 男, 硕士生, 从事机器学习、工业过程控制的研究, E-mail: 852290325@qq.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, 从事机器学习及相关应用领域(模式识别、数据挖掘、图像处理、信息检索)等研究, E-mail: feipingnie@gmail.com.

(责任编辑: 孙艺红)