

## 基于lasso和elastic net的宽度学习系统网络结构稀疏方法

褚菲<sup>1,2,3†</sup>, 苏嘉铭<sup>2</sup>, 梁涛<sup>2</sup>, 陈俊龙<sup>4,5</sup>, 王雪松<sup>1,2</sup>, 马小平<sup>1,2</sup>

(1. 中国矿业大学 地下空间智能控制教育部工程研究中心, 江苏 徐州 221116;

2. 中国矿业大学 信息与控制工程学院, 江苏 徐州 221116;

3. 北京矿冶科技集团有限公司 矿冶过程自动控制技术国家重点实验室/矿冶过程自动控制技术北京市重点实验室, 北京 100160;

4. 华南理工大学 计算机科学与工程学院, 广东 广州 510006; 5. 澳门大学 科技学院, 澳门 999078)

**摘要:** 本文提出了一种基于lasso和elastic net的宽度学习系统(BLS)网络结构稀疏方法, 将标准BLS目标函数中的 $L_2$ 范数分别替换为lasso和elastic net, 利用这两种正则化技术来约束网络输出权重, 衡量每个网络节点输出权重对预测的影响程度, 将多余的节点进行剔除, 提高了网络结构的稀疏性. 通过对一些回归数据集进行实验, 可以看到本文提出的方法在不损失预测精度的前提下, 同时简化了网络结构.

**关键词:** 宽度学习系统; 网络结构; lasso; elastic net

**引用格式:** 褚菲, 苏嘉铭, 梁涛, 等. 基于lasso和elastic net的宽度学习系统网络结构稀疏方法. 控制理论与应用, 2020, 37(12): 2543 – 2550

DOI: 10.7641/CTA.2020.00178

## Sparsity method for network structure of broad learning system based on lasso and elastic net

CHU Fei<sup>1,2,3†</sup>, SU Jia-ming<sup>2</sup>, LIANG Tao<sup>2</sup>, CHEN Jun-long<sup>4,5</sup>,  
WANG Xue-song<sup>1,2</sup>, MA Xiao-ping<sup>1,2</sup>

(1. Research Center of Underground Space Intelligent Control Engineering of the Ministry of Education, China University of Mining and Technology, Xuzhou Jiangsu 221116, China;

2. School of Information and Control Engineering, China University of Mining and Technology, Xuzhou Jiangsu 221116, China;

3. State Key Laboratory of Process Automation in Mining & Metallurgy/Beijing Key Laboratory of Process Automation in Mining & Metallurgy, Beijing General Research Institute of Mining & Metallurgy, Beijing 100160, China;

4. School of Computer Science and Engineering, South China University of Technology, Guangzhou Guangdong 510006, China;

5. Faculty of Science and Technology, University of Macau, Macau 999078, China)

**Abstract:** This paper proposes a sparsity method for network structure of broad learning system (BLS) based on lasso and elastic net. The  $L_2$ -norm in the standard BLS objective function is replaced by the lasso and the elastic net respectively. These two regularization techniques are used to constrain the output weight of each network node, so as to measure the impact of each node's output weight on the prediction. In this way, the redundant nodes are eliminated and the sparsity of network structure is improved. Through the experiments on some regression datasets, it can be seen that the proposed method can simplify the network structure without losing the prediction accuracy.

**Key words:** broad learning system; network structure; lasso; elastic net

**Citation:** CHU Fei, SU Jiaming, LIANG Tao, et al. Sparsity method for network structure of broad learning system based on lasso and elastic net. *Control Theory & Applications*, 2020, 37(12): 2543 – 2550

收稿日期: 2020-03-31; 录用日期: 2020-09-01.

†通信作者. E-mail: chufeizhufei@sina.com; Tel.: +86 18796244325.

本文责任编辑: 赵冬斌.

国家自然科学基金项目(61973304, 61503384, 61702195, 61751202), 国家重点研发项目(2019YFA0706200, 2019YFB1703600), 江苏省六大人才高峰项目(DZXX-045), 江苏省科技计划项目(BK20191339), 广州市科技重大专项项目(202007030006), 徐州市科技创新计划项目(KC19055), 矿冶过程自动控制技术国家重点实验室开放课题项目(BGRIMM-KZSKL-2019-10)资助.

Supported by the National Natural Science Foundation of China (61973304, 61503384, 61702195, 61751202), the National Key Research and Development Program of MOST, China (2019YFA0706200, 2019YFB1703600), the Six Talent Peak Project of Jiangsu Province (DZXX-045), the Science and Technology Project of Jiangsu Province (BK20191339), the Science and Technology Major Project of Guangzhou (202007030006), the Science and Technology Innovation Project of Xuzhou (KC19055) and the Open Subject of State Key Laboratory of Process Automation in Mining & Metallurgy (BGRIMM-KZSKL-2019-10).

## 1 引言

随着人工智能的迅速发展,神经网络在人脸识别、语音识别等一系列具有挑战性的任务中得到了广泛的应用<sup>[1-6]</sup>,其中深度学习表现得尤为出色.虽然深度学习具有很强的学习能力,但由于结构复杂,需要对多个参数的调整,这就导致了长期的训练过程.另外,为了提高学习能力,深度学习需要使用大量性能更强的计算机资源<sup>[7]</sup>.

宽度学习系统是一个无需深度架构、有效且高效的学习系统,它是基于随机向量函数链接神经网络的思想设计的<sup>[8-10]</sup>.与深度学习相比,宽度学习系统(broad learning system, BLS)只包含特征层、增强层和输出层,具有网络结构简单、参数少的优点. BLS利用岭回归算法得到网络的权值,避免了复杂的迭代计算过程,在较短的计算时间内达到令人满意的精度.此外, BLS有多种增量学习算法,根据实际需求添加特征节点、增强节点,无需完整的再训练过程即可快速重新建模<sup>[11]</sup>.文献[12]证明了BLS的普遍逼近性质,并提出了BLS的几种变型结构.文献[13]提出了循环宽度学习系统,在增强节点中加入反馈连接,使网络具有记忆历史信息的能力,用于时间序列预测.文献[14]基于流形正则化的框架,对BLS进行了扩展,提出了半监督宽度学习系统,该算法可以利用大量未标记样本和少量标记样本实现半监督分类,计算速度快,泛化能力强.文献[15]将由特征节点组成的BLS的左侧部分替换为Takagi-Sugeno模糊子系统,从而建立了一个新的神经模糊模型——模糊BLS.文献[16]提出了小波BLS自适应滤波器,用于预测和消除遥控操作中的生理震颤.文献[17]提出了无监督BLS,用于高光谱图像聚类.尽管取得了这些进展,但少有文章讨论宽度学习系统网络结构稀疏的问题.

一个合适的网络结构对BLS的训练尤为重要.一方面,网络节点过少,可能导致网络无法充分挖掘数据信息,影响模型的性能.另一方面,网络节点过多,会产生一个复杂的模型,可能会造成过拟合的问题.目前,宽度学习系统利用岭回归算法得到网络的输出权重,岭回归倾向于保持原有的网络节点,不具有特征选择的作用.当样本数据具有高维性时,网络结构变得更为复杂,大大提高了训练成本.

为了解决选择合适的网络结构的问题,本文提出了一种基于lasso和elastic net的BLS网络结构稀疏方法. Lasso<sup>[18]</sup>具有特征选择的作用,将权值直接缩减至0,能够产生更加稀疏的模型,但这一过程可能存在过度简化网络的风险,导致模型的性能下降. Elastic net<sup>[19]</sup>是将 $L_1$ 范数和 $L_2$ 范数组合在一起,在 $L_1$ 范数与 $L_2$ 范数之间进行权衡.本文提出的BLS网络结构稀疏方法如下:首先选择较大的初始网络结构,即节点个数比解决问题所需的节点数量多;然后随机生成特征层和增强层的权重,分别利用lasso和elastic net衡量每个网络节点对输出预测的影响程度,将影响程度低的节点进行剔除,最后得到一个更为稀疏的网络结构.

本文的其余部分如下:第2节对宽度学习系统进行简要的回顾;第3节详细描述本文提出的基于lasso和elastic net的BLS网络结构稀疏方法;第4节对几种常用的回归数据集进行仿真实验,并对实验结果分析与讨论;第5节总结全文工作.

## 2 宽度学习系统

宽度学习系统是由C.L.Phipil Chen等人提出的一种新型神经网络,为分类和回归问题提供了一个有效的学习框架,如图1所示<sup>[11]</sup>. BLS在某些领域可替代深度学习,简化了训练过程.本节将简要回顾宽度学习系统.

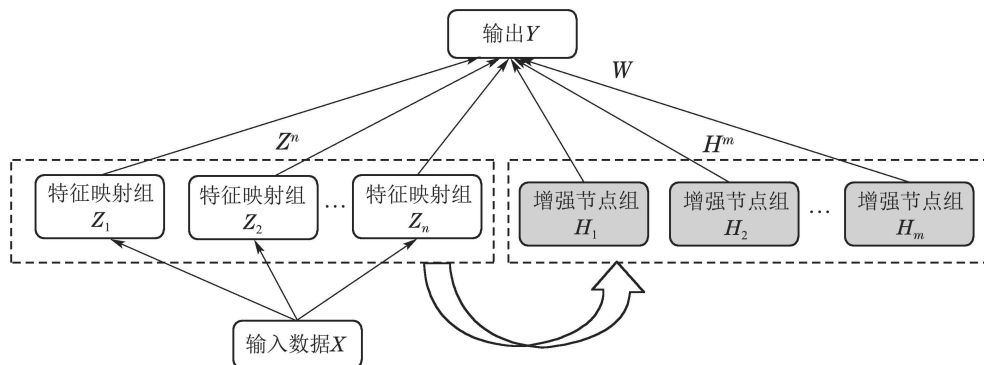


图1 宽度学习系统结构图

Fig. 1 Structure of broad learning system

假设训练数据集 $\{(x_i, y_i) | x_i \in \mathbb{R}^M, y_i \in \mathbb{R}^C, i = 1, \dots, N\}$ ,其中: $N$ 是训练样本的数量, $M$ 和 $C$ 分别表示输入和输出数据的维度.在网络中,假设共有 $n$ 组

特征节点,每组包含 $k$ 个节点.因此,第 $i$ 组特征节点可以表示为

$$Z_i = \phi(XW_{e_i} + \beta_{e_i}), i = 1, \dots, n, \quad (1)$$

其中:  $W_{e_i}, \beta_{e_i}$  是由网络随机生成,  $\phi(\cdot)$  是一个映射函数. 所有的特征节点可以表示为  $Z^n \equiv [Z_1 \cdots Z_n]$ . 然后, 假设有  $m$  组增强节点, 每组包含  $q$  个节点. 那么, 第  $j$  组增强节点可以表示为

$$H_j = \xi(Z^n W_{h_j} + \beta_{h_j}), j = 1, \cdots, m, \quad (2)$$

其中:  $W_{h_j}, \beta_{h_j}$  是由网络随机生成,  $\xi(\cdot)$  是一个激活函数. 那么, 所有的增强节点可以表示为  $H^m \equiv [H_1 \cdots H_m]$ .

因此, 最终的宽度学习网络输出可以表示为

$$Y = AW, \quad (3)$$

其中:  $A = [Z^n | H^m]$ ,  $W$  是将特征节点和增强节点连接到输出层的输出权重.

最后, 输出权重  $W$  由以下公式进行优化求解:

$$\arg \min_W \|Y - AW\|_2^2 + \lambda \|W\|_2^2. \quad (4)$$

针对式(4), 将对  $W$  的导数置为 0, 得到输出权重的解为

$$W = (\lambda I + A^T A)^{-1} A^T Y, \quad (5)$$

其中:  $\lambda$  是正则化参数,  $I$  是单位矩阵.

### 3 BLS网络结构稀疏方法

本文提出的一种基于lasso和elastic net的BLS网络结构稀疏方法的主要思想是, 利用lasso和elastic net约束输出权重, 衡量每个节点对预测的相关程度, 将不相关或低相关性的网络节点剔除, 得到一个更为紧凑的网络结构, 并且在一定程度上不会降低网络的泛化能力.

在所提出的方法的第1步中, 设置一个足够多节点的初始网络结构. 本文确定初始节点个数采取了经验法: 初始节点数应足够大, 远大于实际所需要的节点数<sup>[20]</sup>. 根据标准的BLS, 随机生成特征层和增强层的权重, 得到式(1)和式(2). 然后, 利用增广拉格朗日乘子法<sup>[21-22]</sup>得到了连接特征节点、增强节点到输出层的输出权重.

#### 3.1 正则化技术

对于一个回归预测模型, 一方面, 模型的预测精度很重要; 而另一方面, 模型的大小更为重要. 当两个模型预测精度不相上下, 其中一个模型结构更为紧凑, 那么这个模型的训练时间会大大减少, 从而降低了训练成本. 传统的最小二乘法通常不能在满足一定程度的预测精度下同时保证模型结构的紧凑<sup>[23]</sup>, 所以为了解决这个问题, 在过去的几十年中, 有学者提出了一些正则化技术, 这其中就包括lasso, elastic net以及岭回归. 而BLS正是利用了岭回归这一方法, 本文不再作详细介绍.

Lasso方法具有特征筛选的作用, 将不相关的权值缩减至0, 消除这些节点在网络中的作用, 这有助于实

现最小规模的网络结构<sup>[18]</sup>. 但当有多个高相关的特征时, lasso倾向于只选择其中的一个最高相关的特征, 而忽略其他同样对预测输出有重要影响的特征, 这可能导致预测精度下降. Elastic net在lasso方法与岭回归之间进行权衡, 融合二者的优点<sup>[19]</sup>. 在处理多个高相关的特征时, elastic net倾向于一起选择, 这与lasso大为不同. 所以在实际应用时, 要根据问题选择最合适的方法.

本文将上述两种正则化技术应用于BLS, 针对式(4), 下面分别给出基于lasso和elastic net的BLS优化目标函数:

$$\arg \min_W \|Y - AW\|_2^2 + \lambda \|W\|_1, \quad (6)$$

$$\arg \min_W \|Y - AW\|_2^2 + \lambda_1 \|W\|_1 + \lambda_2 \|W\|_2^2. \quad (7)$$

本文将这两种算法称为lasso宽度学习系统(lasso broad learning system, L1BLS)、弹性网络宽度学习系统(elastic net broad learning system, ENBLS).

#### 3.2 优化求解

在本节中, 讨论L1BLS和ENBLS的优化求解过程. 本文利用增广拉格朗日乘子法来对式(6)和式(7)求解. 根据增广拉格朗日乘子法, 式(6)改写为以下形式:

$$\begin{aligned} \arg \min_{Q, W} \|Y - AW\|_2^2 + \lambda \|Q\|_1, \\ \text{s.t. } Q = W. \end{aligned} \quad (8)$$

然后, 将式(8)转化为拉格朗日函数形式:

$$\begin{aligned} \mathcal{L}(Q, W, C_1) = \\ \|Y - AW\|_2^2 + \lambda \|Q\|_1 + \\ C_1^T (Q - W) + \frac{\mu}{2} \|Q - W\|_2^2, \end{aligned} \quad (9)$$

其中:  $C_1$  表示拉格朗日乘子,  $\mu$  是惩罚因子.

然后, 利用块坐标下降法<sup>[24]</sup>对式(9)的  $Q, W, C_1$  变量优化求解, 下面是具体过程:

1) 求解  $W$ : 保证变量  $Q, C_1$  不变并移除与  $W$  的无关项, 那么  $W$  可由以下公式求解:

$$\arg \min_W \frac{\mu}{2} \|F - W\|_2^2 + \|Y - AW\|_2^2, \quad (10)$$

其中  $F = Q + \frac{C_1}{\mu}$ . 然后通过以下公式得到  $W$ :

$$W = (\mu I + 2A^T A)^{-1} (\mu F + 2A^T Y), \quad (11)$$

其中  $I$  表示单位矩阵.

2) 求解  $Q$ : 保证变量  $W, C_1$  不变并移除与  $Q$  的无关项, 那么  $Q$  可由以下公式求解:

$$\arg \min_Q \lambda \|Q\|_1 + \frac{\mu}{2} \|M - Q\|_2^2, \quad (12)$$

其中  $M = W - \frac{C_1}{\mu}$ . 然后通过以下公式得到  $Q$ :

$$Q = \text{shrink}(M, \frac{\lambda}{\mu}) \triangleq \text{sgn } M \times \max\{|M| - \frac{\lambda}{\mu}, 0\}. \quad (13)$$

$C_1$ 在每次迭代中通过以下公式进行更新:

$$C'_1 = C_1 + \mu(Q - W). \quad (14)$$

针对式(7)进行优化求解, 式(7)改写为以下形式:

$$\arg \min_{E, W} \|Y - AW\|_2^2 + \lambda_1 \|E\|_1 + \lambda_2 \|W\|_2^2, \quad (15)$$

s.t.  $E = W$ .

然后, 将式(15)转化为拉格朗日函数形式:

$$\begin{aligned} \mathcal{L}(E, W, C_2) = & \|Y - AW\|_2^2 + \lambda_1 \|E\|_1 + \lambda_2 \|W\|_2^2 + \\ & C_2^T (E - W) + \frac{\mu}{2} \|E - W\|_2^2, \end{aligned} \quad (16)$$

其中:  $C_2$ 表示拉格朗日乘子,  $\mu$ 是惩罚因子.

然后, 利用块坐标下降法对式(16)的 $E, W, C_2$ 变量优化求解, 下面是具体过程:

1) 求解 $W$ : 保证变量 $E, C_2$ 不变并移除与 $W$ 的无关项, 那么 $W$ 可由以下公式求解:

$$\arg \min_W \frac{\mu}{2} \|G - W\|_2^2 + \|Y - AW\|_2^2 + \lambda_2 \|W\|_2^2, \quad (17)$$

其中 $G = E + \frac{C_2}{\mu}$ . 然后通过以下公式得到 $W$ :

$$W = (\mu I + 2A^T A + 2\lambda_2 I)^{-1} (\mu G + 2A^T Y), \quad (18)$$

其中 $I$ 表示单位矩阵.

2) 求解 $E$ : 保证变量 $W, C_2$ 不变并移除与 $E$ 的无关项, 那么 $E$ 可由以下公式求解:

$$\arg \min_E \lambda_1 \|E\|_1 + \frac{\mu}{2} \|N - E\|_2^2, \quad (19)$$

其中 $N = W - \frac{C_2}{\mu}$ . 然后通过以下公式得到 $E$ :

$$E = \text{shrink}(N, \frac{\lambda_1}{\mu}) \triangleq \text{sgn } N \times \max\{|N| - \frac{\lambda_1}{\mu}, 0\}. \quad (20)$$

$C_2$ 在每次迭代中通过以下公式进行更新:

$$C'_2 = C_2 + \mu(E - W). \quad (21)$$

综上所述, 本文提出的基于lasso和elastic net的BLS网络结构稀疏方法步骤如下所示:

- 步骤 1** 输入训练样本 $X, Y$ ;
- 步骤 2** 随机产生 $W_{e_i}, \beta_{e_i}$ ;
- 步骤 3** 计算 $Z_i = \phi(XW_{e_i} + \beta_{e_i})$ ;
- 步骤 4** 设置特征节点组 $Z^n \equiv [Z_1 \cdots Z_n]$ ;

**步骤 5** 随机产生 $W_{h_j}, \beta_{h_j}$ ;

**步骤 6** 计算 $H_j = \xi(Z^n W_{h_j} + \beta_{h_j})$ ;

**步骤 7** 设置增强节点组 $H^m \equiv [H_1 \cdots H_m]$ ;

**步骤 8** 设置 $A = [Z^n | H^m]$ ;

**步骤 9** 优化求解目标函数(6)或者(7);

**步骤 10** 得到输出权值 $W$ .

## 4 实验与讨论

在本节中, 将提出的基于lasso和elastic net的BLS网络结构稀疏方法应用于回归问题, 验证其可行性. 另外, 由于各种不可控制的变化, 例如仪器误差, 实际获取的样本数据可能会受到某些噪声或异常值的污染<sup>[25]</sup>, 影响训练过程, 导致模型的预测精度下降. 针对该问题, JunWei Jin和C.L.Philip Chen提出了鲁棒宽度学习系统(robust broad learning system, RBLS)<sup>[26]</sup>, 旨在提高模型的鲁棒性. 本节引用RBLS这一方法, 其中包括 $L_1$ 鲁棒宽度学习系统( $L_1$ -robust broad learning system, L1RBLS)、弹性鲁棒宽度学习系统(elastic net robust broad learning system, ENRBLS), 分别在无离群值和有离群值条件下, 将5种算法(BLS, L1BLS, ENBLS, L1RBLS, ENRBLS)进行对比.

笔者从加州大学欧文分校(University of California, Irvine, UCI)数据库中选择了4个回归数据集<sup>[27]</sup>, 数据集的详细信息见表1.

表1 回归数据集的信息

Table 1 Information of regression data sets

数据集	输入变量	样本数	
		训练集	测试集
Basketball	4	64	32
Cleveland	13	202	101
Quake	3	1452	726
Abalone	8	2784	1393

根据第3节所述, 选择足够大的初始网络结构, 设置特征节点组 $N_1 = 20$ , 每组包含20个特征节点, 增强节点 $N = 300$ . 所以, 初始网络一共有700个节点. 正则化参数在 $\{2^{-5}, 2^{-4}, \dots, 2^5\}$ 范围内由网格搜索法得到. 以下实验均在该初始网络结构下进行, 以便统一比较网络结构简化的效果. 本文采用均方根误差(root mean square error, RMSE)来衡量模型预测性能, 下面为其表达式:

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{N}}, \quad (22)$$

其中:  $y_i$ 是实际值,  $\hat{y}_i$ 是模型预测值,  $N$ 是样本个数.

表2-5给出的具体信息如下:

1) 计算了在不同数据集上训练后的测试集的

RMSE以及测试时间, 根据第3.2节的方法, 选取一个最佳的模型, 该模型的具体参数以及简化后的网络节点个数在表格中罗列了出来(参数包括: 初始网络节点、正则化参数 $2^k$ 的指数 $k$ ).

2) 针对这4个数据集, 向训练样本中随机加入了

离群值, 离群值等级设为 $p = 30\%$ . 同样计算了5种算法在被离群值污染后的数据集上训练后的测试集的RMSE, 其他信息同1.

3) 在每个数据集中的所有算法中, 关于RMSE的最好的两个结果用粗体突出显示.

表 2 Basketball数据集中不同算法的模型性能  
Table 2 Model performance of different algorithms on Basketball dataset

算法	$p = 0\%$				$p = 30\%$			
	参数	t/s	RMSE	节点个数	参数	t/s	RMSE	节点个数
BLS	700, 4	0.0011	0.0780	700	700, 5	0.0015	0.1269	700
L1BLS	700, -1	0.0011	<b>0.0775</b>	9.2	700, 0	0.0014	0.1237	5.66
ENBLS	700, -2, 0	0.0010	<b>0.0766</b>	48.36	700, 1, 5	0.0009	<b>0.0925</b>	14.04
L1RBLS	700, -1	0.0010	0.0803	10.94	700, 0	0.0021	0.0930	6.22
ENRBLS	700, -3, 1	0.0009	0.0779	130.86	700, 0, 4	0.0009	<b>0.0848</b>	40.5

表 3 Cleveland数据集中不同算法的模型性能  
Table 3 Model performance of different algorithms on Cleveland dataset

算法	$p = 0\%$				$p = 30\%$			
	参数	t/s	RMSE	节点个数	参数	t/s	RMSE	节点个数
BLS	700, 5	0.0023	0.1320	700	700, 5	0.0019	0.1737	700
L1BLS	700, 0	0.0015	0.1317	42.46	700, 1	0.0015	0.1952	34.92
ENBLS	700, -1, 5	0.0015	<b>0.1258</b>	202.82	700, -1, 5	0.0013	<b>0.1714</b>	266.32
L1RBLS	700, 1	0.0016	0.1517	46.6	700, 1	0.0014	0.1808	53.98
ENRBLS	700, 0, 5	0.0014	<b>0.1315</b>	167.08	700, -1, 5	0.0013	<b>0.1478</b>	308.08

表 4 Quake数据集中不同算法的模型性能  
Table 4 Model performance of different algorithms on Quake dataset

算法	$p = 0\%$				$p = 30\%$			
	参数	t/s	RMSE	节点个数	参数	t/s	RMSE	节点个数
BLS	700, 3	0.0105	0.1733	700	700, 5	0.0111	0.2401	700
L1BLS	700, -1	0.0102	<b>0.1729</b>	13.98	700, -1	0.0105	0.2407	14.78
ENBLS	700, -1, -1	0.0103	<b>0.1729</b>	24.28	700, -2, 3	0.0101	0.2410	140.24
L1RBLS	700, -3	0.0100	0.1787	43.22	700, -3	0.0101	<b>0.1885</b>	52.02
ENRBLS	700, -1, 5	0.0109	0.1736	158.24	700, -3, 5	0.0106	<b>0.1785</b>	350.22

表 5 Abalone数据集中不同算法的模型性能  
Table 5 Model performance of different algorithms on Abalone dataset

算法	$p = 0\%$				$p = 30\%$			
	参数	t/s	RMSE	节点个数	参数	t/s	RMSE	节点个数
BLS	700, 0	0.0202	<b>0.0760</b>	700	700, 1	0.0214	0.1017	700
L1BLS	700, -3	0.0192	<b>0.0759</b>	80.7	700, -1	0.0204	0.1013	36.94
ENBLS	700, -2, -5	0.0196	0.0762	54.4	700, -1, -5	0.0182	0.1016	38.2
L1RBLS	700, -3	0.0189	0.0780	130.24	700, -2	0.0196	<b>0.0808</b>	82
ENRBLS	700, -1, -1	0.0190	0.0774	83.88	700, 0, 2	0.0189	<b>0.0786</b>	114.28

表2-5展示了5种算法在不同数据集上的性能。Basketball样本数量少, Quake样本数量多, 但维度都较低, Cleveland样本数量中等, 但维度较高, Abalone样本数量多, 维度中等。在无离群值的情况下, 针对Basketball和Quake数据集, L1BLS和ENBLS表现最为出色, 与BLS相比, 不仅预测精度有所提升, 而且节点个数大幅减少, 简化了网络结构。针对Cleveland数据集, ENBLS, ENRBLS预测精度最高, 相比BLS, 这两种算法一定程度上简化了网络结构, 而L1BLS, L1RBLS显然网络结构简化程度更大, 但预测精度略微下降。针对Abalone数据集, BLS预测性能较好, 而L1BLS在提升预测性能的同时, 产生了一个更为紧凑的网络结构。

在有离群值( $p = 30%$ )情况下, 可以看到, JunWei Jin和C.L.Philip Chen提出的鲁棒宽度学习系统受到的影响是最小的, 其中ENRBLS在各个数据集上的预测性能均表现优异, 但相比于L1BLS, ENBLS, L1RBLS, 网络结构的简化程度略低。针对处理有离群值的数据集, L1BLS, ENBLS预测性能相比于无离群值时有一定程度的下降, 但同样能够大幅简化网络结构。在测试时间方面, 由于L1BLS, ENBLS大幅简化了网络结构, 所以在无离群值和有离群值的情况下, L1BLS和ENBLS相较于BLS均减少了测试时间。

为了更加直观地展现不同算法的性能, 图2-5给出了L1BLS, ENBLS, L1RBLS, ENRBLS相比于BLS的RMSE的变化率以及网络结构的压缩率。

图2和图3展示了在无离群值情况下, 每种算法与BLS相比预测精度的提升程度以及网络结构的压缩率。该提升程度或者压缩率由以下公式得到:

$$\text{Percentage} = \frac{\text{RMSE}/\text{Nodes}_{\text{BLS}} - \text{RMSE}/\text{Nodes}_{\text{ImprovedBLS}}}{\text{RMSE}/\text{Nodes}_{\text{BLS}}} \times 100, \quad (23)$$

其中:  $\text{RMSE}/\text{Nodes}_{\text{BLS}}$ 表示原始BLS的RMSE以及节点数,  $\text{RMSE}/\text{Nodes}_{\text{ImprovedBLS}}$ 表示改进后的BLS的RMSE以及节点数。

在无离群值情况下, L1RBLS相较于BLS预测精度下降较多。针对Abalone数据集, ENBLS, L1RBLS, ENRBLS与BLS相比预测精度均有所下降。而在另外3个数据集上, 各个算法预测精度略微有提升, 但总体相近。在网络结构方面, 各个算法相比于BLS, 能够实现更为紧凑的网络结构, 并且在所有情况下, 压缩率都高于60%, 尤其在Basketball, Quake, Abalone数据集上L1BLS和ENBLS压缩率达到了80%以上, 网络结构得到了明显的简化。

图4和图5展示了在有离群值情况下, 每种算法与BLS相比预测精度的提升程度以及网络结构的压缩率。该提升程度或者压缩率由式(23)得到。

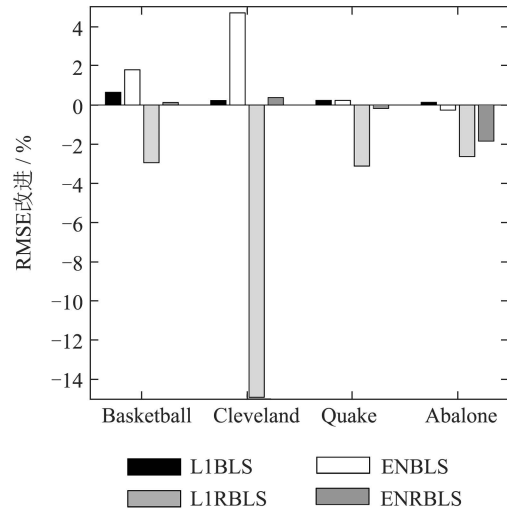


图2 正则化BLS的RMSE改进(无离群值)  
Fig. 2 RMSE improvement of regularized BLS (no outliers)

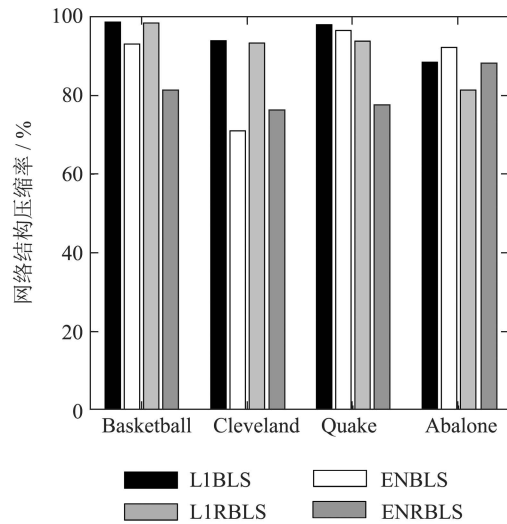


图3 正则化BLS的网络结构压缩率(无离群值)  
Fig. 3 Network structure compression ratio of regularized BLS (no outliers)

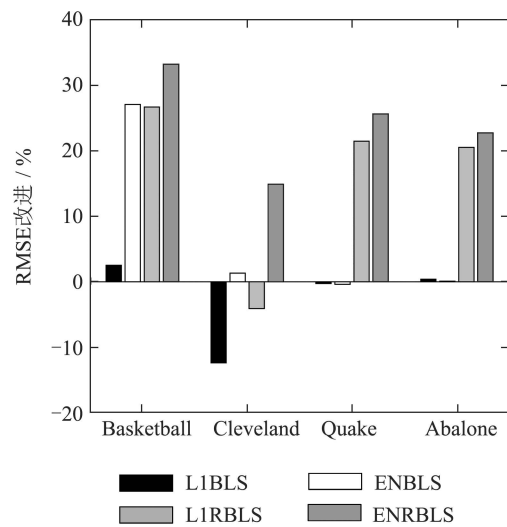


图4 正则化BLS的RMSE改进(离群值 $p = 30%$ )  
Fig. 4 RMSE improvement of regularized BLS (outliers  $p = 30%$ )

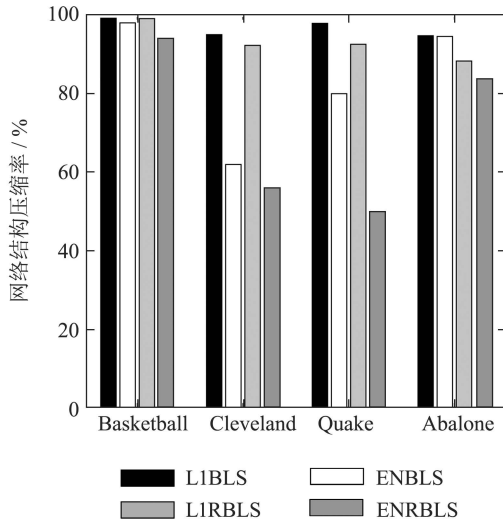


图 5 正则化BLS的网络结构压缩率(离群值 $p = 30\%$ )

Fig. 5 Network structure compression ratio of regularized BLS (outliers  $p = 30\%$ )

在有离群值情况下,在Basketball, Quake, Abalone数据集上, L1RBLs, ENRBLs相较于BLS, 预测精度都有提升,体现了良好的鲁棒性。在网络结构方面, L1BLS, ENBLS, L1RBLs, ENRBLs这4种算法不同程度地简化了网络结构,压缩率均高于50%,特别地, L1BLS在每个数据集上的网络结构压缩率均达到了90%以上。

这些结果表明,在无离群值的情况下,基于lasso和elastic net的BLS网络结构稀疏方法能够在不损失预测精度的前提下,产生一个更为紧凑的网络结构。在有离群值的情况下,本文提出的L1BLS和ENBLS精度略有下降,但同样能够大幅简化网络结构。

## 5 总结

本文提出了一种基于lasso和elastic net的BLS网络结构稀疏方法,其主要思想是利用Lasso和Elastic net代替BLS目标函数中的 $L_2$ 范数,将对输出预测影响程度低的节点进行剔除,权衡模型预测精度与网络结构的关系,最终得到一个稀疏的网络结构。本文选取了几个常用的UCI回归数据集,将该方法与BLS进行了性能比较。实验表明,该方法在保持模型预测精度的前提下,同时实现了一个更为紧凑的网络结构。

## 参考文献:

- GONG M, ZHAO J, LIU J, et al. Change detection in synthetic aperture radar images based on deep neural networks. *IEEE Transactions on Neural Networks and Learning Systems*, 2015, 27(1): 125 – 138.
- HOU W, GAO X, TAO D, et al. Blind image quality assessment via deep learning. *IEEE Transactions on Neural Networks and Learning Systems*, 2014, 26(6): 1275 – 1286.
- JIANG Zhaohui, YIN Juping, GUI Weihua, et al. Prediction for blast furnace silicon content in hot metal based on composite differential evolution algorithm and extreme learning machine. *Control Theory & Applications*, 2016, 33(8): 1089 – 1095.
- XIE S, GIRSHICK R, DOLLAR P, et al. Aggregated residual transformations for deep neural networks. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Hawaii, USA: IEEE, 2017: 1492 – 1500.
- LIU W, WEN Y, YU Z, et al. Spheraface: deep hypersphere embedding for face recognition. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Hawaii, USA: IEEE, 2017: 212 – 220.
- GE Rui, WANG Zhaohui, XU Xin, et al. Action recognition with hierarchical convolutional neural networks features and bi-directional long short-term memory model. *Control Theory & Applications*, 2017, 34(6): 790 – 796.
- (葛瑞, 王朝晖, 徐鑫, 等. 基于多层卷积神经网络特征和双向长短期记忆单元的行为识别. *控制理论与应用*, 2017, 34(6): 790 – 796.)
- CHEN C L P, LIU Z. Broad learning system: a new learning paradigm and system without going deep. *The 32nd Youth Academic Annual Conference of Chinese Association of Automation*. Hefei, China: IEEE, 2017: 1271 – 1276.
- PAO Y H, PARK G H, SOBAJIC D J, et al. Learning and generalization characteristics of the random vector functional-link net. *Neurocomputing*, 1994, 6(2): 163 – 180.
- TYUKIN I Y, PROKHOROV D V. Feasibility of random basis function approximators for modeling and control. *IEEE Control Applications & Intelligent Control*. St. Petersburg, Russia: IEEE, 2009: 1391 – 1396.
- PAO Y H, TAKEFUJI Y. Functional-link net computing: theory, system architecture, and functionalities. *IEEE Computer*, 1992, 25(5): 76 – 79.
- CHEN C L P, LIU Z. Broad learning system: an effective and efficient incremental learning system without the need for deep architecture. *IEEE Transactions on Neural Networks and Learning Systems*, 2018, 29(1): 10 – 24.
- CHEN C L P, LIU Z, FENG S. Universal approximation capability of broad learning system and its structural variations. *IEEE Transactions on Neural Networks and Learning Systems*, 2019, 30(4): 1191 – 1204.
- XU M, HAN M, CHEN C L P, et al. Recurrent broad learning systems for time series prediction. *IEEE Transactions on Cybernetics*, 2020, 50(4): 1405 – 1417.
- ZHAO H, ZHENG J, DENG W, et al. Semi-supervised broad learning system based on manifold regularization and broad network. *IEEE Transactions on Circuits and Systems-I: Regular Papers*, 2020, 67(3): 983 – 994.
- FENG S, CHEN C L P. Fuzzy broad learning system: a novel neuro-fuzzy model for regression and classification. *IEEE Transactions on Cybernetics*, 2020, 50(2): 414 – 424.
- LIN J, LIU Z, CHEN C L P, et al. A wavelet broad learning adaptive filter for forecasting and cancelling the physiological tremor in teleoperation. *Neurocomputing*, 2019, 356: 170 – 183.
- KONG Y, CHENG Y, CHEN C L P, et al. Hyperspectral image clustering based on unsupervised broad Learning. *IEEE Geoscience and Remote Sensing Letters*, 2019, 16(11): 1741 – 1745.
- TIBSHIRANI R. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society Series B-Methodological*, 1996, 58(1): 267 – 288.
- ZOU H, HASTIE T. Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society Series B-Statistical Methodology*, 2005, 67(2): 301 – 320.

- [20] MARTINEZ-MARTINEZ J M, ESCANDELL-MONTERO P, SORIA-OLIVAS E, et al. Regularized extreme learning machine for regression problems. *Neurocomputing*, 2011, 74(17): 3716 – 3721. University of California, 1998, URL: <http://www.archive.ics.uci.edu/ml>.
- [21] KANNAN B K, KRAMER S N. An augmented lagrange multiplier based method for mixed integer discrete continuous optimization and its applications to mechanical design. *Journal of Mechanical Design*, 1994, 116(2): 405 – 411.
- [22] ROCKAFELLAR R T. Augmented lagrange multiplier functions and duality in nonconvex programming. *Siam Journal on Control*, 1974, 12(2): 268 – 285.
- [23] GELADI P, KOWALSKI B R. Partial least-squares regression: a tutorial. *Analytica Chimica Acta*, 1986, 185(1): 1 – 17.
- [24] TSENG P. Convergence of a block coordinate descent method for nondifferentiable minimization. *Journal of Optimization Theory and Applications*, 2001, 109(3): 475 – 494.
- [25] RUPPERT D. Robust statistics: the approach based on influence functions. *Technometrics*, 1987, 150(2): 240 – 241.
- [26] JIN J W, CHEN C L P. Regularized robust broad learning system for uncertain data modeling. *Neurocomputing*, 2018, 322: 58 – 69.
- [27] BLAKE C, MERZ C. *Uci repository of machine learning databases*. Irvine, California: Department of Information and Computer Science, 1998, URL: <http://www.archive.ics.uci.edu/ml>.

#### 作者简介:

**褚菲** 副教授, 博士生导师, 目前研究方向为复杂工业过程建模、控制与优化、机器学习、运行状态评价, E-mail: [chufeizhufei@sina.com](mailto:chufeizhufei@sina.com);

**苏嘉铭** 硕士研究生, 目前研究方向为复杂工业过程建模、控制与优化、机器学习, E-mail: [sujiaming@cumt.edu.cn](mailto:sujiaming@cumt.edu.cn);

**梁涛** 硕士研究生, 目前研究方向为复杂工业过程建模、控制与优化、机器学习, E-mail: [ts17060134p3@cumt.edu.cn](mailto:ts17060134p3@cumt.edu.cn);

**陈俊龙** 教授, 博士生导师, 目前研究方向为智能系统与控制、计算智能、混合智能、无人系统、数据科学, E-mail: [philipchen@scut.edu.cn](mailto:philipchen@scut.edu.cn);

**王雪松** 教授, 博士生导师, 目前研究方向为机器学习、生物信息学、人工智能, E-mail: [wangxuesongcumt@163.com](mailto:wangxuesongcumt@163.com);

**马小平** 教授, 博士生导师, 目前研究方向为控制理论及应用、网络控制、计算机技术及应用, E-mail: [xpma@cumt.edu.cn](mailto:xpma@cumt.edu.cn).