

# 控制与决策

Control and Decision

## 基于光滑化 $L_1$ 正则项的随机配置网络

刘晶晶, 刘业峰, 马祎航, 富月

引用本文:

刘晶晶, 刘业峰, 马航, 富月. 基于光滑化 $L_1$ 正则项的随机配置网络[J]. 控制与决策, 2024, 39(3): 813–818.

在线阅读 View online: <https://doi.org/10.13195/j.kzyjc.2022.1612>

---

## 您可能感兴趣的其他文章

### Articles you may be interested in

#### 基于双边分解与 $L_1$ 暗通道的战场图像增强算法

A battlefield image enhancement algorithm based on bilateral decomposition and  $L_1$  dark channel prior  
控制与决策. 2021, 36(5): 1165–1172 <https://doi.org/10.13195/j.kzyjc.2019.1213>

#### 基于Hessian正则的自适应损失半监督特征选择

Adaptive loss semi-supervised feature selection based on Hessian regularization  
控制与决策. 2021, 36(8): 1862–1870 <https://doi.org/10.13195/j.kzyjc.2019.1510>

#### 非局部低秩正则化视频压缩感知重构

Compressive video sensing reconstruction via nonlocal low-rank regularization  
控制与决策. 2021, 36(11): 2743–2750 <https://doi.org/10.13195/j.kzyjc.2020.0310>

#### 双重驱动的果蝇优化算法及其在PID控制器中的应用

Double drive fruit fly optimization algorithm and its application in PID controller  
控制与决策. 2021, 36(9): 2225–2233 <https://doi.org/10.13195/j.kzyjc.2020.0046>

#### 具有未建模动态和输出约束的耦合系统的分散自适应控制

Decentralized adaptive control for interconnected systems with unmodeled dynamics and output constraints  
控制与决策. 2018, 33(12): 2113–2121 <https://doi.org/10.13195/j.kzyjc.2017.0933>

# 基于光滑化 $L_1$ 正则项的随机配置网络

刘晶晶<sup>1,2</sup>, 刘业峰<sup>2†</sup>, 马祎航<sup>3</sup>, 富月<sup>4</sup>

(1. 沈阳工学院 基础课部, 辽宁 沈抚示范区 113122; 2. 沈阳工学院 辽宁省数控机床信息物理融合与智能制造重点实验室, 辽宁 沈抚示范区 113122; 3. 沈阳工学院 机械工程与自动化学院, 辽宁 沈抚示范区 113122; 4. 东北大学 流程工业综合自动化国家重点实验室, 沈阳 110819)

**摘要:** 为了提高随机配置网络 (stochastic configuration networks, SCN) 的泛化能力, 提出一种适用于 SCN 的光滑化  $L_1$  正则化方法. 针对  $L_1$  正则化算子局部不可微的缺陷, 在曲线不光滑点的邻域内进行光滑处理, 并在此基础上构建 SCN 的光滑误差函数, 提出增量计算权值的算法, 进而以交替方向乘子法为基础给出权值的全局优化算法, 并且在理论上分析算法的收敛性. 与  $L_1$  正则化的稀疏性和  $L_2$  正则化均匀减小参数的特点相比, 所提出方法按重要程度保留数据的全部特征, 使参数既保持在较小的范围内又具有层次分明的分布, 从而使网络具有更好的泛化能力. 最后, 通过数值仿真实验验证了所提出方法的可行性和有效性.

**关键词:** 光滑正则化; 随机配置网络; 泛化能力; 交替方向乘子法; 收敛性分析; 数据特征

中图分类号: TP183

文献标志码: A

DOI: 10.13195/j.kzyjc.2022.1612

引用格式: 刘晶晶, 刘业峰, 马祎航, 等. 基于光滑化  $L_1$  正则项的随机配置网络 [J]. 控制与决策, 2024, 39(3): 813-818.

## Smoothing $L_1$ regularization for stochastic configuration networks

LIU Jing-jing<sup>1,2</sup>, LIU Ye-feng<sup>2†</sup>, MA Yi-hang<sup>3</sup>, FU Yue<sup>4</sup>

(1. Department of Basic Courses, Shenyang Institute of Technology, Shenyang New District 113122, China; 2. Liaoning Key Laboratory of Information Physics Fusion and Intelligent Manufacturing for CNC Machine, Shenyang Institute of Technology, Shenyang New District 113122, China; 3. School of Mechanical Engineering and Automation, Shenyang Institute of Technology, Shenyang New District 113122, China; 4. State Key Laboratory of Synthetical Automation for Process Industries, Northeastern University, Shenyang 110819, China)

**Abstract:** In order to improve the generalization capability of stochastic configuration networks (SCNs), a smooth  $L_1$  regularization method for SCNs is proposed. Aiming at the defect of local non-differentiability of the  $L_1$  regularization operator, smoothing is carried out in the neighborhood of non-smooth points of the curve. The convex error function of the SCN is constructed on this basis, and an algorithm for incremental calculation of the weights of the SCN is proposed. Furthermore, the global optimization algorithm is proposed based on the alternating direction multiplier method, and the convergence of the algorithm is analyzed theoretically. Compared with the sparsity of  $L_1$  regularization and the uniform reduction of parameters by  $L_2$  regularization, the proposed method retains all features of the data according to the degree of importance, the parameters are not only kept in a small range, but also have hierarchical distribution, so that the network has better generalization ability. Finally, the feasibility and effectiveness of the proposed method are verified by some numerical simulations.

**Keywords:** smoothing regularization; stochastic configuration networks; generalization capability; alternating direction multiplier method; convergence analysis; data feature

## 0 引言

随着大数据技术的发展, 以数据驱动的智能控制技术受到了各界的广泛关注, 在复杂的工业系统

中可以借助这类数学模型或算法逼近非线性关系, 实现设备的自主感知和识别等功能. 设备作为一个智能体, 为了能在被动感知或主动检测到异常情况时运

收稿日期: 2022-09-11; 录用日期: 2022-12-01.

基金项目: 国家自然科学基金项目 (62073226); 辽宁省自然科学基金重点领域联合开放基金项目 (2022-KF-11-01, 2020-KF-11-09, 2021-KF-11-05); 沈抚示范区本级科技计划项目 (2020JH13, 2021JH07); 沈阳工学院青年骨干教师科研基金项目 (QN202210).

责任编辑: 夏元清.

†通讯作者. E-mail: liuyefeng@situ.edu.cn.

用自己的思维做出决策,需要拥有机器学习和自主智能决策能力,进而与其他多智能体协同实现整个工业系统的智能化.在机器学习中,过拟合是一个常见的问题,研究抑制过拟合的方法十分重要.

对权值参数进行惩罚是一种抑制过拟合现象的常用方法,即正则化方法.正则化方法也为贝叶斯正则化<sup>[1-2]</sup>,在神经网络模型的处理中,常用的方法是在误差函数后面加上权值范数,如 $L_2$ 范数、 $L_1$ 范数、 $L_{1/2}$ 范数和 $L_0$ 范数. $L_2$ 正则化可实现权值衰减; $L_1$ 正则化可稀疏网络结构<sup>[3-8]</sup>; $L_{1/2}$ 正则化进一步稀疏网络结构<sup>[9-10]</sup>; $L_0$ 正则化会导致NP难问题<sup>[11]</sup>,不能直接利用优化算法求解<sup>[12]</sup>.针对不可微问题,光滑化的 $L_1$ 正则化方法<sup>[13-16]</sup>、 $L_{1/2}$ 正则化方法<sup>[17-18]</sup>、 $L_0$ 正则化方法<sup>[19]</sup>也随即被提出.

随着神经网络类型的不断增多,正则化技术逐渐应用于更多网络中.随机配置网络(stochastic configuration networks, SCN)于2017年被提出<sup>[20]</sup>,是一种增量生成式神经网络,在大规模的数据分析中具有成功应用<sup>[21]</sup>.SCN使用最小二乘法求解输出层与隐层之间的权值,偶尔会导致输出权值过大,影响网络的泛化性能.为了解决这一问题, $L_2$ 正则化的SCN<sup>[22-23]</sup>、 $L_1$ 正则化的SCN<sup>[24]</sup>和流形正则化SCN<sup>[25]</sup>被陆续提出,提高了SCN的泛化能力.

带有 $L_1$ 正则项的SCN使用ADMM算法迭代求解最优的输出权值,最终使用软阈值函数表示最优解,将很多权值置零,稀疏了网络结构,并提高了泛化能力.但是,这一方法具有两面性,因为修剪掉一些网络连接的同时,也意味着舍弃了数据的某些次要特征.本文针对这一问题,以提高网络泛化能力为目标,保留网络的全部连接,在权值保持在较小范围的基础上按照重要程度赋值,从而更深入地挖掘数据特征.研究 $L_1$ 正则项在不可微点的光滑处理方法,将 $L_1$ 正则项转化为光滑函数,从而方便使用更常规的微分计算方法求出最优解;通过改进不等式监督机制分析网络的收敛性;光滑化的 $L_1$ 正则化方法使参数特征主次分明,使重要的网络连接发挥主要作用,同时保留了次要特征,相比 $L_2$ 正则化使所有参数均匀变小以及 $L_1$ 正则化的稀疏处理,光滑化 $L_1$ 正则化方法折中处理两种方案,使网络具有更好的性能;光滑 $L_1$ 正则化中的可变参数可调节网络对于 $L_1$ 和 $L_2$ 正则化的倾向程度,当参数足够小时,实现对 $L_1$ 正则项的光滑逼近,模型更具灵活性.

符号说明: $\|\cdot\|$ 表示2-范数,即 $\|\cdot\|_2$ .

## 1 带有光滑化 $L_1$ 正则项的随机配置网络

### 1.1 随机配置网络

SCN是一种在不等式监督机制下的增量生成式网络,具有通用逼近性质<sup>[20]</sup>.设 $X = (x^1, x^2, \dots, x^J)$ 为训练样本的输入数据, $x^i \in R^D$ , $O = (o^1, o^2, \dots, o^J)$ 为对应的目标输出数据, $O^i \in R^M$ .则网络的隐层输出为

$$z_l^i = g(w_l^T x^i + b_l), \quad l = 1, 2, \dots, L, \quad i = 1, 2, \dots, J. \quad (1)$$

其中: $w_l$ 和 $b_l$ 为连接第 $l$ 个隐层节点与输入节点之间的权值和偏置, $g$ 为有界激活函数.已经配置好 $L$ 个隐层节点的网络实际输出为

$$y^{(L)} = (y_1, y_2, \dots, y_M) = \left( \sum_{l=1}^L \beta_{l1} z_l, \sum_{l=1}^L \beta_{l2} z_l, \dots, \sum_{l=1}^L \beta_{lM} z_l \right). \quad (2)$$

此时对于第 $i$ 个样本,网络的误差为

$$e_L = O^i - y^{(L)}. \quad (3)$$

其中: $\beta_{lm}$ ( $m = 1, 2, \dots, M$ )为第 $l$ 个隐层节点与第 $m$ 个输出间的连接权值, $z_l$ 为第 $l$ 个隐层节点的输出值.

SCN的配置算法为:随机选取满足如下不等式的参数:

$$\sum_{q=1}^m (e_{L-1,q}, z_L)^2 \geq b_g^2 (1 - r - \mu_L) \|e_{L-1}\|^2, \quad (4)$$

配置输入层与隐层之间的连接权值 $w_l$ 和 $b_l$ .其中: $e_{L-1,q}$ 为已配置好 $L-1$ 个隐层节点后网络的第 $q$ 维输出误差, $b_g$ 为激活函数 $g$ 的上界.输出层与隐层之间权值的计算公式为

$$\beta_L = (\beta_{L,1}, \beta_{L,2}, \dots, \beta_{L,M}) = \arg \min_{\beta} \|e_L\|^2. \quad (5)$$

### 1.2 $L_1$ 正则项的光滑处理和输出侧权值配置算法

为了提高SCN的泛化能力,将正则项加入误差函数中. $L_1$ 正则项含有绝对值函数,具有在原点不可微的缺陷,因此需要磨光 $L_1$ 正则项中绝对值函数的尖点,使其成为光滑函数,并将其应用于随机配置网络的误差函数.下文将所提出的网络简称为 $SL_1$ SCN(smooth  $L_1$ -SCN).

使用如下函数代替 $L_1$ 正则项:

$$f(x) = \begin{cases} x, & x \geq a; \\ \frac{x^2}{2a} + \frac{a}{2}, & |x| < a; \\ -x, & x \leq -a. \end{cases} \quad (6)$$

式(6)在 $x = a$ 和 $x = -a$ 处连续、可导且光滑<sup>[15]</sup>.与 $L_1$ 正则项相比,该函数在原点的邻域内用二次函数

代替线性函数,得到了整个定义域上光滑的函数. 建立误差函数为

$$E = \frac{1}{2} \|e_L\|^2 + \lambda \sum_{i=1}^L f(\beta_i) = \sum_{q=1}^m \left( \frac{1}{2} \|e_{L-1,q} - \beta_{L,q} z_L\|^2 + \lambda \sum_{i=1}^L f(\beta_{i,q}) \right). \quad (7)$$

对 $\beta_{L,q}$ 求导( $\beta_{1,q}, \beta_{2,q}, \dots, \beta_{L-1,q}$ 已知),得到

$$\nabla \beta_{L,q} = (-z_L^T)(e_{L-1,q} - \beta_{L,q} z_L) + \lambda f'(\beta_{L,q}) = \beta_{L,q} \|z_L\|^2 - \langle e_{L-1,q}, z_L \rangle + \lambda \gamma. \quad (8)$$

$$\gamma = \begin{cases} 1, & \beta_{L,q} \geq a; \\ \frac{\beta_{L,q}}{a}, & |\beta_{L,q}| < a; \\ -1, & \beta_{L,q} \leq -a. \end{cases} \quad (9)$$

令式(8)等于0,有

$$\beta_{L,q} = \begin{cases} \frac{\langle e_{L-1,q}, z_L \rangle - \lambda}{\|z_L\|^2}, & \Theta_1 \geq \Theta_2; \\ \frac{\langle e_{L-1,q}, z_L \rangle + \lambda}{\|z_L\|^2}, & \Theta_1 \leq -\Theta_2; \\ \frac{\langle e_{L-1,q}, z_L \rangle}{\lambda/a + \|z_L\|^2}, & |\Theta_1| < \Theta_2. \end{cases} \quad (10)$$

其中: $\Theta_1 = \langle e_{L-1,q}, z_L \rangle, \Theta_2 = \lambda + a \|z_L\|^2$ . 按式(10)计算输出权重即可得到最优解.

### 1.3 输出侧权值的全局更新算法

使用式(10)计算 $\beta_L$ 会减缓模型的收敛速度, $f(\beta)$ 是分段函数,使用矩阵方式计算全局更新并不适用,但因其光滑性可使用ADMM算法整体更新输出侧权值 $\beta = (\beta_1, \beta_2, \dots, \beta_L)$ . 构造优化问题

$$\min f(x) + g(\beta_q) = \frac{1}{2} \|Hx - O_q\|^2 + \lambda f(\beta_q); \quad (11)$$

$$\text{s.t. } x - \beta_q = 0. \quad (12)$$

其中 $O_q$ 为目标输出的第 $q$ 维数据. 令 $\mu_1 = \mu/\rho$ ,求解得到

$$x^{k+1} = (H^T H + \rho I)^{-1} (H^T O + \rho(\beta_q^k - \mu_1^k)). \quad (13)$$

$$\beta_q = \begin{cases} x^{k+1} + \mu_1^k - \lambda/\rho, & x^{k+1} + \mu_1^k \geq a + \lambda/\rho; \\ \frac{\rho(x^{k+1} + \mu_1^k)}{\lambda/a + \rho}, & |x^{k+1} + \mu_1^k| < a + \lambda/\rho; \\ x^{k+1} + \mu_1^k + \lambda/\rho, & x^{k+1} + \mu_1^k \leq -a - \lambda/\rho. \end{cases} \quad (14)$$

$$\mu_1^{k+1} = \mu_1^k + x^{k+1} - \beta_q^{k+1}. \quad (15)$$

按式(13)~(15)的顺序迭代更新,得到目标函数最优解<sup>[24]</sup>. 式(14)中 $a$ 取值越小, $\beta_q$ 取值越接近于0,方法便越接近于 $L_1$ 正则化;反之,方法接近于 $L_2$ 正则化.

## 2 不等式监督机制的改进及收敛性分析

改进SCN的不等式监督机制为

$$\sum_{q=1}^M \langle e_{L-1,q} z_L \rangle^2 \geq \max \left\{ b_g^2 \delta_L + M \lambda^2, \frac{(\lambda/a + b_g^2)^2}{2\lambda/a + b_g^2} \delta_L \right\}, \quad (16)$$

$$\delta_L = (1 - r - \mu_L) \|e_{L-1}\|^2. \quad (17)$$

**定理1** 设 $\Gamma = g_1, g_2, \dots$ 为实值函数集,  $\text{Span}(\Gamma)$ 为 $L_2$ 空间上的紧集,  $\forall g \in \Gamma, 0 < \|g\| \leq b_g (b_g > 0)$ . 对于给定的 $0 < r < 1$ 和实数列 $\mu_L$ , 满足 $\lim_{L \rightarrow +\infty} \mu_L = 0$ 且 $\mu_L + r \leq 1$ . 若满足式(16)且隐层与输出层之间的权值按式(10)计算, 则 $\lim_{L \rightarrow +\infty} \|f -$

$$f_L\|^2 = 0, \text{ 其中 } f_L = \sum_{j=1}^L \beta_{j,q} z_j.$$

**证明** 有

$$\begin{aligned} \|e_L\|^2 - \|e_{L-1}\|^2 &= \sum_{q=1}^M (\|e_{L-1,q} - \beta_{L,q} z_L\|^2 - \|e_{L-1,q}\|^2) = \sum_{q=1}^M (-2 \langle e_{L-1,q}, g_L \rangle \beta_L + \beta_L^2 \|z_L\|^2) = \begin{cases} \sum_{q=1}^M \left( \frac{\lambda^2 - \langle e_{L-1,q}, z_L \rangle^2}{\|z_L\|^2} \right), & \Psi \geq 0; \\ \sum_{q=1}^M \left( -\frac{2\lambda/a + z_L^2}{\lambda/a + z_L^2} \langle e_{L-1,q}, z_L \rangle^2 \right), & \Psi < 0. \end{cases} \end{aligned}$$

其中 $\Psi = |\langle e_{L-1,q}, z_L \rangle| - (\lambda + a z_L^2)$ . 由式(16), 当 $\Psi \geq 0$ 时, 有

$$\|e_L\|^2 - \|e_{L-1}\|^2 \leq -\frac{b_g^2(1 - r - \mu_L) \|e_{L-1}\|^2}{\|z_L\|^2} \leq -(1 - r - \mu_L) \|e_{L-1}\|^2.$$

当 $\Psi < 0$ 时, 有

$$\begin{aligned} \|e_L\|^2 - \|e_{L-1}\|^2 &\leq -\frac{2\lambda/a + z_L^2}{\lambda/a + z_L^2} \left( \frac{\lambda/a + z_L^2}{2\lambda/a + z_L^2} \right) (1 - r - \mu_L) \|e_{L-1}\|^2 = -(1 - r - \mu_L) \|e_{L-1}\|^2. \end{aligned}$$

综上,  $\|e_L\|^2 \leq (r + \mu_L) \|e_{L-1}\|^2, \{\|e_L\|^2\}$ 单调递减, 且 $\|e_L\|^2 > 0$ , 故 $\lim_{L \rightarrow +\infty} \|e_L\|^2$ 存在. 由文献[24]方法可证明 $\lim_{L \rightarrow +\infty} \|e_L\|^2 = 0$ .  $\square$

令

$$\beta^* = [\beta_1^*, \beta_2^*, \dots, \beta_L^*] =$$

$$\arg \min_{\beta} \frac{1}{2} \left\| f - \sum_{j=1}^L \beta_j z_j \right\|^2 + \lambda f(\beta),$$

$$e_L^* = f - \sum_{j=1}^L \beta_j^* z_j = [e_{L,1}^*, e_{L,2}^*, \dots, e_{L,M}^*].$$

定义

$$\tilde{\beta}_{L,q} = \begin{cases} \frac{\langle e_{L-1,q}^*, z_L \rangle - \lambda}{\|z_L\|^2}, & \Phi_1 \geq \Phi_2; \\ \frac{\langle e_{L-1,q}^*, z_L \rangle + \lambda}{\|z_L\|^2}, & \Phi_1 \leq -\Phi_2; \\ \frac{\langle e_{L-1,q}^*, z_L \rangle}{\lambda/a + \|z_L\|^2}, & |\Phi_1| < \Phi_2. \end{cases} \quad (18)$$

$$\Phi_1 = \langle e_{L-1,q}^*, z_L \rangle, \Phi_2 = \lambda + a\|z_L\|^2.$$

$$\tilde{e}_L = f - \tilde{f}_L = f - \tilde{\beta}_L z_L = e_{L-1}^* - \tilde{\beta}_L z_L. \quad (19)$$

**定理2** 设  $\Gamma = g_1, g_2, \dots$  为实值函数集,  $\text{Span}(\Gamma)$  为  $L_2$  空间上的紧集,  $\forall g \in \Gamma, 0 < |g| \leq b_g$  ( $b_g > 0$ ). 对于给定的  $0 < r < 1$  和实数列  $\mu_L$ , 满足  $\lim_{L \rightarrow +\infty} \mu_L = 0$  且  $\mu_L + r \leq 1$ . 若满足式(20)且隐层与输出层之间的权值按式(13)~(15)的顺序迭代计算, 则  $\lim_{L \rightarrow +\infty} \|e_L^*\|^2 = 0$ . 有

$$\sum_{q=1}^M \langle e_{L-1,q}^*, z_L \rangle^2 \geq \max \left\{ b_g^2 \delta_L + M\lambda^2, \frac{(\lambda/a + b_g^2)^2}{2\lambda/a + b_g^2} \delta_L \right\}. \quad (20)$$

**证明**  $\|e_L^*\|^2 - \|e_{L-1}^*\|^2 \leq \|\tilde{e}_L\|^2 - \|e_{L-1}^*\|^2 = \|e_{L-1}^* - \tilde{\beta}_L z_L\|^2 - \|e_{L-1}^*\|^2$ , 由不等式约束条件(20), 易得  $\|e_L^*\|^2 \leq (r + \mu_L)\|e_{L-1}^*\|^2$ ,  $\{\|e_L^*\|^2\}$  单调递减. 又  $\|e_L^*\|^2 > 0$ , 故  $\lim_{L \rightarrow +\infty} \|e_L^*\|^2$  存在. 由文献[24]方法可证明  $\lim_{L \rightarrow +\infty} \|e_L^*\|^2 = 0$ .  $\square$

设  $\gamma = \{\mu_{\min} : \Delta\mu : \mu_{\max}\}$ , 最大隐节点个数为  $L_{\max}$ . 令

$$\xi = \sum_{q=1}^M \langle e_{L-1,q}^*, z_L \rangle^2 - \max \left\{ b_g^2 \delta_L + M\lambda^2, \frac{(\lambda/a + b_g^2)^2}{2\lambda/a + b_g^2} \delta_L \right\},$$

则  $SL_1SCN$  算法描述如算法1所示.

**算法1**  $SL_1SCN$  algorithm.

- 1: 初始化输出误差  $e_0 = [o_1, o_2, \dots, o_M]^T, 0 < r < 1, W = \emptyset, \Omega = \emptyset$ ;
- 2: while do  $L \leq L_{\max}$
- 3: 配置隐层节点, 确定输入层与隐层之间的权值和偏置;
- 4: for do  $\lambda \in T$
- 5: for do  $k = 1 : T_{\max}$
- 6: 随机生成  $[-\mu, \mu]$  区间上的一组权值和偏置  $w_L$  和  $b_L$ ;
- 7: 计算  $\xi$  的值;

- 8: if  $\xi \geq 0$  then
- 9: 保存  $w_L$  和  $b_L$  至  $W$ , 保存  $\xi$  至  $\emptyset$ ;
- 10: else
- 11: 返回6;
- 12: end if
- 13: end for
- 14: if  $W \neq \emptyset$  then
- 15: 搜索使  $\xi$  最大的  $w_L^*$  和  $b_L^*$ ;
- 16: break (转至22);
- 17: else
- 18: 随机选取  $\tau \in (0, 1 - r)$ , 更新  $r = r + \tau$ , 返回6;
- 19: end if
- 20: end for
- 21: 计算输出层与隐层之间的权值;
- 22: 通过式(13)~(15)迭代计算权值  $\beta^*$
- 23: 计算  $e_L = e_{L-1} - \beta_L h_L^*, h_L^* = [g_1^*, g_2^*, \dots, g_L^*]$ ;
- 24:  $L = L + 1$ ;
- 25: end while
- 26: return  $\beta = [\beta_1, \beta_2, \dots, \beta_L], w^* = [w_1^*, w_2^*, \dots, w_L^*], b^* = [b_1^*, b_2^*, \dots, b_L^*]$ .

### 3 仿真实验

为了验证算法的有效性, 基于4个回归数据集和3个分类数据集, 将  $SL_1SCN$  与  $PSCN$ 、 $RSC-2$  和  $SCN$  算法进行对比分析. 数据集分别来自  $KEEL$  (knowledge extraction based on evolutionary learning) 数据库的  $AUTO$   $MPG6$ 、 $Laser$ 、 $Dee$ 、 $Friedman$  数据集和来自  $UCI$  数据库 ( $UCI$  Irvine machine learning repository) 的  $Iris$ 、 $Wine$ 、 $Raisin$  数据集, 各数据集属性如表1所示. 为了验证算法对网络泛化能力的影响, 使用均方误差函数 ( $MSE$ ) 作为评判标准, 有

$$MSE = \frac{1}{J_{\text{test}}} \sum_{i=1}^{J_{\text{test}}} (o^i - y^i)^2. \quad (21)$$

其中:  $o^i$  为第  $i$  个样本的目标输出,  $y^i$  为第  $i$  个样本的网络输出,  $J_{\text{test}}$  为测试样本个数.

所有程序均使用  $Matlab$  2019a 编写, 在 8 G 内存、1.6 GHz 的 CPU 上独立重复执行 20 次, 取平均值作为实验结果, 网络参数如表2所示. “迭代次数”为  $ADMM$  算法的迭代计算次数,  $C$  为  $L_2$  正则化系数, 取值参照  $RSC-2$  算法. 为了体现  $SL_1SCN$  的泛化能力, 重点比较各类模型在测试集上的预测效果, 实验结果如表3所示. 图1汇总了前4个数据集在  $SL_1SCN$  模型下的平均训练误差收敛曲线.

表1 数据集信息

数据集	输入属性	输出属性	样本总数	训练样本	测试样本
AUTO MPG6	5	1	392	313	79
Laser	4	1	993	744	249
Dee	6	1	365	300	65
Friedman	5	1	1200	1000	200
Iris	4	3	150	120	30
Raisin	7	2	900	600	300
Wine	13	3	178	148	30

表2 网络参数

数据集	$a$	$\lambda$	$C$	迭代次数	隐层节点
AUTO MPG6	0.000 01	0.005	$2^{10}$	1 000	70
Laser	0.000 001	0.005	$2^{10}$	10 000	70
Dee	0.000 01	0.005	$2^{10}$	2 000	70
Friedman	0.000 01	0.005	$2^{10}$	5 000	70
Iris	0.000 01	0.005	$2^5$	2 000	40
Raisin	0.000 01	0.000 5	$2^5$	2 000	70
Wine	0.000 01	0.000 5	$2^{10}$	1 000	5

表3 20次实验平均测试误差或准确率

数据集	$SL_1SCN$	PSCN	RSC-2	SCN
AUTO MPG6	<b>0.057 2</b>	0.065 7	0.062 3	0.068 1
Laser	<b>0.026 3</b>	0.038 4	0.038 6	0.030 4
Dee	<b>0.088 9</b>	0.089 2	0.089 2	0.089 0
Friedman	<b>0.042 9</b>	0.044 0	0.043 7	0.044 1
Iris / %	<b>95.67</b>	94.67	94.50	94.00
Raisin / %	<b>86.20</b>	86.13	85.95	86.12
Wine / %	<b>97.67</b>	96.67	97.50	96.50

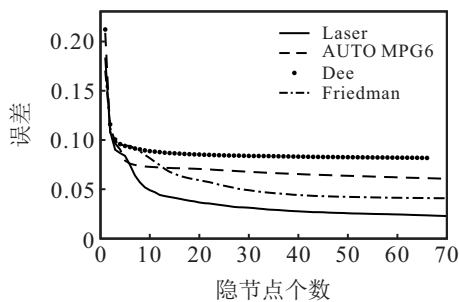


图1  $SL_1SCN$ 的平均训练误差收敛曲线汇总

由表3可见,  $SL_1SCN$ 的预测能力更强, 具有更好的泛化性能. 图2以Laser数据集为例给出了输出层-隐层的权值分布, 显示了4种模型( $SL_1SCN$ 、PSCN、RSC-2-SCN和SCN)输出侧权值 $\beta$ 的分布情况. 可以看出: PSCN有一部分权值为零, 符合其稀疏化的特征; RSC-2算法计算的权值几乎都是非零元, 分布曲线波动的幅度更为均衡; SCN算法在整体上呈现出权值取值范围较大的现象, 最大值接近1 000; 所提出的 $SL_1SCN$ 权值大小区分明显, 具体表现为曲线图视觉上接近于零的部分大多为 $1e-04$ 数量级大小或更小, 其他点远离零点, 但取值也会控制在较小范围内,

体现出权值是按网络连接的重要程度赋值, 更充分地体现了数据特征.

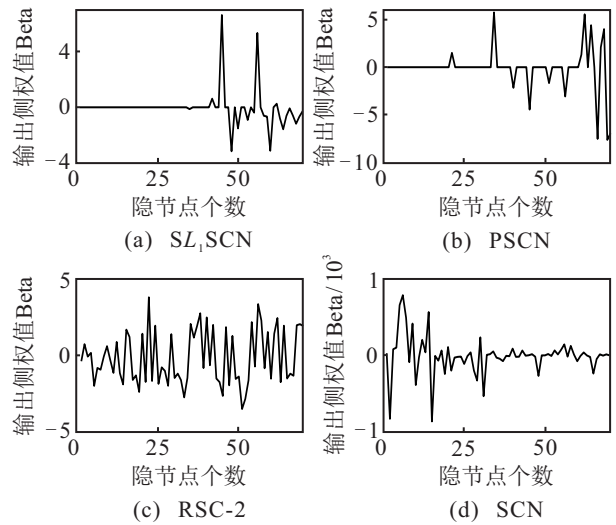


图2 输出层-隐层的权值分布(Laser)

$SL_1SCN$ 的优势需要从几种正则化方法的机制进行解释.  $L_2$ 正则化近乎均匀地缩小了权值参数, 没有舍弃数据的任一特征, 也没有突显任何特征, 处理方式相对缓和, 模型相对稳定;  $L_1$ 正则化将一些权值参数直接赋零, 使网络具有稀疏性, 鲁棒性也相对较强, 但这种处理也舍弃了一些数据特征;  $SL_1SCN$ 区别于PSCN和RSC-2的地方在于它按照重要性保留了样本数据的全部特征, 对于发挥主要作用的特征, 权值参数赋值较大, 而对于次要地位的特征则赋值较小, 甚至接近于零.

$SL_1SCN$ 对正则项的光滑处理, 相当于在 $L_1$ 正则项之中嵌入一部分 $L_2$ 正则项, 因此参数 $a$ 的取值对网络性能有着一定的影响.  $a$ 越小算法越接近于 $L_1$ 正则化方法,  $a$ 越大算法越接近于 $L_2$ 正则化方法, 网络呈现一种从 $L_1$ 向 $L_2$ 过渡的效果. 因此, 在实际应用中应根据数据分布的特点选择合适的 $a$ 值.

## 4 结论

本文提出了一种基于光滑化 $L_1$ 正则项的随机配置网络 $SL_1SCN$ , 将 $L_1$ 正则项的不可微点进行光滑处理, 为随机配置网络构造一个光滑误差函数, 提出了网络权值的优化配置算法, 并在理论上分析了收敛性. 一方面,  $SL_1SCN$ 模型的输出侧权值既能够保持在较小水平又主次分明, 体现了对 $L_1$ 正则化和 $L_2$ 正则化的均衡处理, 网络性能也因此得到提升; 另一方面, 构造带有光滑正则项的误差函数能够使网络权值的优化算法得以扩充, 为进一步基于梯度方法的优化研究提供了可能.  $SL_1SCN$ 既具有良好的学习能力又具有较好的泛化性能, 这是数据驱动智能系统中的两个重要能力. 因此, 在当今的多智能体协同控制和机

器学习背景下,可进一步将 $SL_1$ SCN模型应用于生产线物联网中,以提高智能体设备的自主学习能力。

### 参考文献(References)

- [1] Burden F, Winkler D. Bayesian regularization of neural networks[J]. *Methods in Molecular Biology*, 2008, 458: 25-44.
- [2] Ticknor J L. A Bayesian regularized artificial neural network for stock market forecasting[J]. *Expert Systems with Applications*, 2013, 40(14): 5501-5506.
- [3] Han M, Li D. An norm 1 regularization term ELM algorithm based on surrogate function and Bayesian framework[J]. *Acta Auto Sin*, 2011, 37(11): 1344-1350.
- [4] Cen L, Yu Z L, Dong M H. Speech emotion recognition system based on  $L_1$  regularized linear regression and decision fusion[C]. *Affective Computing and Intelligent Interaction*. Berlin, 2011: 332-340.
- [5] Zou H, Hastie T. Regularization and variable selection via the elastic net[J]. *Journal of the Royal Statistical Society—Series B: Statistical Methodology*, 2005, 67(2): 301-320.
- [6] Zou H. The adaptive lasso and its oracle properties[J]. *Journal of the American Statistical Association*, 2006, 101(476): 1418-1429.
- [7] Zhao P, Yu B. Stagewise lasso[J]. *Journal of Machine Learning Research*, 2007, 8(12): 2701-2726.
- [8] Zhang Z, Xu Y, Yang J, et al. A survey of sparse representation: Algorithms and applications[J]. *IEEE Access*, 2015, 3: 490-530.
- [9] Xu Z B, Zhang H, Wang Y, et al.  $L_{1/2}$  regularizer[J]. *China Information Science*, 2010, 53(6): 1159-1169.
- [10] Xu Z B, Chang X Y, Xu F M, et al.  $L_{1/2}$  regularization: A thresholding representation theory and a fast solver[J]. *IEEE Transactions on Neural Networks and Learning Systems*, 2012, 23(7): 1013-1027.
- [11] Candes E J, Tao T. Decoding by linear programming[J]. *IEEE Transactions on Information Theory*, 2005, 51(12): 4203-4215.
- [12] Wang Y F, Liu P, Li Z H, et al. Data regularization using Gaussian beams decomposition and sparse norms[J]. *Journal of Inverse and Ill-Posed Problems*, 2013, 21(1): 1-23.
- [13] 王若鹏, 徐红敏. 非线性 $l_1$ 问题的光滑近似算法[J]. *纯粹数学与应用数学*, 2013, 29(1): 25-32.  
(Wang R P, Xu H M. Smooth approximation method for nonlinear  $l_1$  problem[J]. *Pure and Applied Mathematics*, 2013, 29(1): 25-32.)
- [14] Bian W, Chen X J. A smoothing proximal gradient algorithm for nonsmooth convex regression with cardinality penalty[J]. *SIAM Journal on Numerical Analysis*, 2020, 58(1): 858-883.
- [15] Bian W, Chen X J. Neural network for nonsmooth, nonconvex constrained minimization via smooth approximation[J]. *IEEE Transactions on Neural Networks and Learning Systems*, 2014, 25(3): 545-556.
- [16] Chen X J. Smoothing methods for nonsmooth, nonconvex minimization[J]. *Mathematical Programming*, 2012, 134(1): 71-99.
- [17] 吴微, 杨洁. 用于神经网络权值稀疏化的 $L_{1/2}$ 正则化方法[J]. *中国科学: 数学*, 2015, 45(9): 1487-1504.  
(Wu W, Yang J.  $L_{1/2}$  regularization methods for weights sparsification of neural networks[J]. *Scientia Sinica: Mathematica*, 2015, 45(9): 1487-1504.)
- [18] Liu Y, Yang D K. Convergence analysis of the batch gradient-based neuro-fuzzy learning algorithm with smoothing  $L_{1/2}$  regularization for the first-order Takagi-Sugeno system[J]. *Fuzzy Sets and Systems*, 2017, 319: 28-49.
- [19] Fan Q W, Liu T. Smoothing  $L_0$  regularization for extreme learning machine[J]. *Mathematical Problems in Engineering*, 2020, 2020: 1-10.
- [20] Wang D H, Li M. Stochastic configuration networks: Fundamentals and algorithms[J]. *IEEE Transactions on Cybernetics*, 2017, 47(10): 3466-3479.
- [21] Wang D H, Cui C H. Stochastic configuration networks ensemble with heterogeneous features for large-scale data analytics[J]. *Information Sciences*, 2017, 417: 55-71.
- [22] 王前进, 杨春雨, 马小平, 等. 基于随机配置网络的井下供给风量建模[J]. *自动化学报*, 2021, 47(8): 1963-1975.  
(Wang Q J, Yang C Y, Ma X P, et al. Underground airflow quantity modeling based on SCN[J]. *Acta Automatica Sinica*, 2021, 47(8): 1963-1975.)
- [23] 赵立杰, 邹世达, 郭烁, 等. 基于正则化随机配置网络的球磨机工况识别[J]. *控制工程*, 2020, 27(1): 1-7.  
(Zhao L J, Zou S D, Guo S, et al. Ball mill load condition recognition model based on regularized stochastic configuration networks[J]. *Control Engineering of China*, 2020, 27(1): 1-7.)
- [24] 王前进, 代伟, 陆群, 等. 一种随机配置网络软测量模型的稀疏学习方法[J]. *控制与决策*, 2022, 37(12): 3171-3182.  
(Wang Q J, Dai W, Lu Q, et al. A sparse learning method for SCN soft measurement model[J]. *Control and Decision*, 2022, 37(12): 3171-3182.)
- [25] 南静, 宁传峰, 建中华, 等. 基于随机配置网络的轻量级人体行为识别模型[J]. *控制与决策*, 2023, 38(6): 1541-1550.  
(Nan J, Ning C F, Jian Z H, et al. A lightweight model for human activity recognition using stochastic configuration networks[J]. *Control and Decision*, 2023, 38(6): 1541-1550.)

### 作者简介

刘晶晶(1988—), 女, 副教授, 硕士, 从事数据驱动的智能控制、智能计算等研究, E-mail: liujingjing@situ.edu.cn;  
 刘业峰(1981—), 男, 教授, 博士, 从事复杂系统建模及优化、智能制造等研究, E-mail: liuyefeng@situ.edu.cn;  
 马祎航(1987—), 男, 讲师, 硕士, 从事生产调度及优化算法等研究, E-mail: mayihang@situ.edu.cn;  
 富月(1978—), 女, 教授, 博士, 从事复杂工业过程建模、控制与优化等研究, E-mail: fuyue@mail.neu.edu.cn.