

一种随机配置网络软测量模型的稀疏学习方法

王前进^{1,2}, 代伟^{2,3†}, 陆群¹, 辅小荣¹, 马小平³

(1. 盐城工学院 电气工程学院, 江苏 盐城 224051; 2. 中国矿业大学 信息与控制工程学院, 江苏 徐州 221116;
3. 中国矿业大学 人工智能研究院, 江苏 徐州 221116)

摘要: 随机配置网络 (Stochastic configuration network, SCN) 构建一个不等式约束条件对隐性参数进行随机分配, 同时对其范围进行自适应选择, 具有收敛速度快、建模精度高等优点. 由于随机算法的本质特性, 不可避免产生低值、冗余节点, 为提高 SCN 软测量模型的稀疏性, 本文提出一种简约随机配置网络 (Parsimonious stochastic configuration network, PSCN). PSCN 在网络增量构建目标函数中引入 L_1 范数, 建立一个新的不等式约束条件来保障隐性节点的生成质量. 并进一步, 针对新建目标函数的非凸性和非光滑性, 采用交替方向乘子法 (Alternating direction method of multipliers, ADMM) 对整个 PSCN 网络的输出权重进行更新. 最后, 将本文方法应用于基准数据集和实际工业过程软测量问题中, 证明其可有效简化模型结构, 同时具有较高的泛化性能.

关键词: 随机配置网络; L_1 正则化; 软测量模型; 不等式约束; 交替方向乘子法

中图分类号: TP18

文献标志码: A

DOI: 10.13195/j.kzyjc.2021.1047

开放科学(资源服务)标识码(OSID):



A Sparse Learning Method for SCN Soft Measurement Model

WANG Qian-jin^{1,2}, DAI Wei^{2,3†}, LU Qun¹, FU Xiao-rong¹, MA Xiao-ping³

(1. School of Electrical Engineering, Yancheng Institute of Technology, Yancheng 224051, China; 2. School of Information and Control Engineering, China University of Mining and Technology, Xuzhou 221116, China; 3. Artificial Intelligence Research Institute, China University of Mining and Technology, Xuzhou 221116, China)

Abstract: For stochastic configuration network (SCN), it randomly produces the hidden parameters and adaptively selects their scopes using an inequality constraint. As a result, SCN exhibits superior performance in convergence speed and modeling accuracy. It is inevitable to produce low-value and redundant hidden nodes due to the inherent feature of randomized algorithm. To improve the sparsity of SCN soft sensor model, a parsimonious stochastic configuration network (PSCN) is proposed in this paper. L_1 norm is plugged into the cost function of PSCN, and a new inequality constraint is built to obtain the high-quality hidden nodes. Next, considering the non-smoothness and non-convexity of cost function with L_1 norm, the alternating direction method of multipliers (ADMM) is employed to update the output weights of whole network. Finally, the proposed method is applied to benchmark data sets and soft measurement issue in industrial process, and simulation results show that it can effectively simplify the network structure and possess the higher generalization.

Keywords: Stochastic configuration network; L_1 regularization; soft measurement model; inequality constraint; alternating direction method of multipliers

0 引言

复杂工业过程中, 运行指标的准确测量是实现运行优化闭环控制的基础. 但由于环境恶劣, 常规仪表测量效果不佳, 导致运行指标不能连续在线测量, 不得不通过建模技术来获得^[1-2]. 建模技术的研究主要集中在 2 个方面: 基于机理的建模以及基于数据的建模^[3].

机理模型需要完全掌握工业过程的内部特征, 但由于其高度复杂性, 往往难以获取先验机理知识. 另外, 机理模型一般是基于一些理论假设, 且模型重要参数难以准确获取, 因此, 利用机理模型去估计运行指标会导致一定的偏差^[4]. 基于数据驱动的建模技术, 不需要了解工业过程的复杂变化, 仅利用输入输出数据就可以建立运行指标的估计模型, 因此基

收稿日期: 2021-06-17; 修回日期: 2021-08-10.

基金项目: 国家自然科学基金项目(62003293, 61973306, 62003292); 江苏省自然科学基金项目(BK20191043, BK20200086); 盐城工学院校级科研项目资助(xjr2019018); 东北大学流程工业综合自动化国家重点实验室开放课题(2020-KF-21-10)资助.

†通讯作者. E-mail: daiwei_neu@126.com.

于数据的建模在工业过程中被广泛用于运行指标的估计^[5-6].

目前, 随机权网络 (Random weight neural network, RWNN) 是一种强有力的建模工具, 它具有极快的学习速度、较好的泛化能力等优点^[7-8]. RWNN 结构包含输入层、隐含层和输出层, 是一种单隐层前馈神经网络 (Single-Layer Forward Neural Network, SLFN), 其网络规模由隐性节点的个数所决定. RWNN 的构造分两步进行: 首先在给定区间内随机产生隐性节点, 再通过求解一个线性方程得到输出权重. 正由于其优越的特性, RWNN 已广泛应用于工业过程建模中^[9]. 然而, 存在两个因素制约着 RWNN 的实际应用: 1) 不能合适确定网络结构, 即不能根据训练数据自动确定隐性节点数; 2) 从固定区间中随机选取隐性参数, 难以获取有效的隐性节点, 导致网络收敛较慢, 甚至无法收敛^[10].

RWNN 的一个典型代表是由 Pao 等人在 1994 年提出的随机向量函数链接网络 (Random Vector Function-link Network, RVFLN)^[11]. RVFLN 的特点是在输入层与输出层之间增加了一条直连链, 并在限定范围内随机得到输入节点和增强节点的隐性参数. 文 [12] 针对 RVFLN 的函数近似问题, 基于自适应函数近似的一般方法, 给出了 RVFLN 对有界有限维集上连续函数具有万能逼近能力的理论证明. 至今, 国内外已衍生出诸多性能优良的 RVFLN 变体, 包括变分模态分解 RVFLN [13]、稀疏预训练 RVFLN [14]、特定类的加权 RVFLN [15]、鲁棒多核 RVFLN [16] 等. 最近, 基于 RVFLN 衍生出一种新型的随机学习算法, 即宽度学习系统, 用于高效构建 RWNN 模型^[17-19]. 正由于其简单的训练过程、好的泛化性能以及万能逼近特性, RWNN 在实际工业过程中得到广泛应用.

为解决基于 RWNN 建模中的一些实际问题和常见错误, 文 [20] 对其进行理论分析与仿真实验, 得出一个结论: 增量式随机权网络 (Incremental random weight neural network, IRWNN) 仅在固定区间内随机产生隐性节点, 不具有万能逼近特性. 因此, 若要其具有万能逼近能力则需要额外的附加条件. 文 [21] 通过建立一个不等式约束条件来随机配置隐性节点, 提出了一种新的增量式随机建模技术, 即随机配置网络 (Stochastic configuration network, SCN). 其特征如下: 1) 利用不等式约束条件去产生高质量隐性节点, 并自动调节隐性参数选取范围; 2) 采用局部\全局最小二乘法来计算输出权重. 因此, SCN 不但学习

速度极快, 而且具有万能逼近能力. 此外, SCN 可根据实际需要来自动确定网络结构.

近年, 为了提高 SCN 的性能, 一些学者提出了一些改进策略. 如, 文 [22] 提出了两种不等式约束并对关键参数的选取进行指导, 以提高配置效率. 文 [23] 针对难以采用矩阵伪逆运算的大数据建模问题, 在 SCN 框架下, 提出了一种基于负相关学习的异构随机神经元集成学习方法. 文 [24] 针对 SCN 输出矩阵常为病态的问题, 采用截断奇异值分解对其进行修正, 提出了 TSVD-SCN 方法, 来提升算法性能. 文 [25] 将数据并行和模型并行相结合, 提出了一种混合的随机学习算法, 以提高其大数据建模效率. 文 [26] 提出了基于动量项的 SCN 的运行指标估计方法, 将其误差反馈项添加到输出层, 动态调节网络训练参数, 以获得更紧致的结构和更好的泛化能力. 文 [27] 提出了基于 L_2 范数正则化的 SCN, 进一步改善了算法的泛化性能.

此外, 文 [28] 和 [29] 分别提出了两种基于 SCN 的鲁棒建模方法, 以应对建模问题中存在的异常值和噪声问题. 文 [30] 采用基于核密度估计的鲁棒 SCN 建模方法建立了高炉炼铁过程铁水质量预测模型. 文 [31] 基于上下界估计法和深度 SCN, 提出了一种新的预测区间构造方法, 用于原油物性的预测.

然而, 上述 SCN 算法虽可通过不等式约束来保证隐性节点的质量, 但由于构造学习算法的随机性, 仍会引入对模型性能影响不大的隐性节点. 作用不大的隐性节点较多, 则会增加所建网络的复杂度, 容易产生过拟合, 造成泛化性能下降. 此外, 对于一些存储空间有限的实际工业系统, 其通常以单片机或 PLC 等为硬件平台. 这些系统需要建立一个能够满足精度要求且尺寸尽可能小的模型, 来缓解存储压力. 因此, 针对工业实际, 建立结构最优且具有较高泛化性能的学习算法是亟待解决的关键问题.

神经网络模型通过隐层非线性函数将输入特征空间映射到一个高维特征空间. 由于增量构造算法的随机特性, 高维特征空间中可能存在很多对建模任务无用的隐性节点, 会使网络尺寸变得较大. 目前, L_1 正则化是一种有效的网络结构简化方法, 它能自动剔除无效的隐性节点, 从而控制网络的规模^[32-33]. 本文针对现有 SCN 算法存在的模型规模控制问题, 引入 L_1 正则化技术, 提出一种泛化能力强的简约随机配置网络 (Parsimonious stochastic configuration network, PSCN), 即: SCN 的稀疏学习方法. 首先, 在网络模型增量构建过程中建立带有 L_1 惩罚项的损

失函数, 并相应的构建一个新的不等式约束条件来随机生成隐性节点; 然后, 针对模型目标函数不可微的问题, 采用交替方向乘子法 (Alternating direction method of multipliers, ADMM) 来计算整个网络的输出权重, 从而得到一种 SCN 的稀疏学习方法. 最后, 基于基准数据集和实际工业过程数据对 PSCN 方法进行实验和比较分析, 表明其在简化了模型结构的同时保持较好的泛化能力.

1 相关算法

1.1 IRWNN

给定一个训练样本集合 $(\mathbf{x}_i, \mathbf{t}_i)_{i=1}^N$, 其中, N 为训练样本个数. 其样本输入和相应的输出分别定义为:

$$\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N\}, \mathbf{x}_i = [x_{i,1}, x_{i,2}, \dots, x_{i,d}]^T \in \mathbb{R}^d$$

$$\mathbf{T} = \{\mathbf{t}_1, \mathbf{t}_2, \dots, \mathbf{t}_N\}, \mathbf{t}_i = [t_{i,1}, t_{i,2}, \dots, t_{i,m}]^T \in \mathbb{R}^m$$

含有 $L-1$ 个隐性节点的 SLFN 输出为:

$$\mathbf{f}_{L-1}(\mathbf{X}) = \sum_{j=1}^{L-1} \beta_j \mathbf{g}_j(\mathbf{w}_j^T \mathbf{X} + b_j) (\mathbf{f}_0 = 0) \quad (1)$$

式中, $\mathbf{w}_j$, b_j 和 $\beta_j = [\beta_{j1}, \beta_{j2}, \dots, \beta_{jm}]^T$ 为第 j 个隐性节点的训练参数, m 为输出节点数, $\mathbf{g}_j$ 为第 j 个隐性节点的激活函数.

当前网络残差 $\mathbf{e}_{L-1}$ 可表示为:

$$\mathbf{e}_{L-1} = \mathbf{f} - \mathbf{f}_{L-1} = [\mathbf{e}_{L-1,1}, \mathbf{e}_{L-1,2}, \dots, \mathbf{e}_{L-1,m}] \quad (2)$$

IRWNN 分两步来构造网络. 首先, 对于网络中已添加的 $L-1$ 个隐性节点, 其训练参数在增量学习过程中保持不变, 并随机生成第 L 个隐性节点; 然后, 通过使增加节点后的网络残差最小来获得输出权值 β_L . IRWNN 的输出可写成:

$$\mathbf{f}_L(\mathbf{X}) = \mathbf{f}_{L-1}(\mathbf{X}) + \beta_L \mathbf{g}_L(\mathbf{w}_L \cdot \mathbf{X} + b_L) \quad (3)$$

式中, $\beta_L = \frac{\langle \mathbf{e}_{L-1}, \mathbf{g}_L \rangle}{\|\mathbf{g}_L\|^2}$.

1.2 SCN

针对 IRWNN 存在的学习速度慢、不能收敛等问题, Wang 等人提出了三种不同的 SCN 算法, 即 SC-I、SC-II 和 SC-III^[21]. 其中, SC-III 算法在收敛速度、建模精度等方面均优于其他两种算法. 因此, 在本文余下部分中, 我们用 SCN 来表示 SC-III 算法.

令 $\Gamma := \{\mathbf{g}_1, \mathbf{g}_2, \mathbf{g}_3 \dots\}$ 为一组实值函数, $\text{span}(\Gamma)$ 为 Γ 张成的函数空间.

在每次的网络构造过程中, SCN 先在所建不等式约束下产生一组候选隐性节点, 再从中选择一个质量最好的隐性节点作为新增节点. 不等式约束的

表达式如下:

$$\langle \mathbf{e}_{L-1}, \mathbf{g}_L \rangle^2 \geq b_g^2 \delta_L \quad (4)$$

式中, $\forall \mathbf{g} \in \Gamma$, 对于 $b_g \in \mathbb{R}^+$, 有 $0 < \|\mathbf{g}\| < b_g$; δ_L 有如下形式:

$$\delta_L = (1 - r - u_L) \|\mathbf{e}_{L-1}\|^2 \quad (5)$$

式中, $0 < r < 1$, $0 < u_L \leq (1 - r)$, 且 u_L 满足 $\lim_{L \rightarrow +\infty} u_L = 0$.

在新增节点添加到网络中后, SCN 采用全局最小二乘法对输出权重进行更新:

$$[\beta_1^*, \beta_2^*, \dots, \beta_L^*]^T = \arg \min \|\mathbf{f} - \sum_{j=1}^L \beta_j \mathbf{g}_j\| \quad (6)$$

式 (6) 的简化形式如下:

$$\beta^* = \arg \min_{\beta} \|\mathbf{G}_L \beta - \mathbf{t}\|^2 = \mathbf{G}_L^\dagger \mathbf{t} \quad (7)$$

式中, $\mathbf{G}_L = [\mathbf{g}_1, \mathbf{g}_2, \dots, \mathbf{g}_L]$, $\mathbf{G}_L^\dagger$ 为矩阵 $\mathbf{G}_L$ 的 Moore-Penrose 广义逆^[34].

SCN 通过重复上述步骤来构造 SLFN, 直到满足事先给定的精度要求或者达到预先设定的隐性节点数.

2 PSCN

2.1 稀疏学习算法

从式 (7) 可知, SCN 通过采用伪逆算法进行输出权值确定. 若矩阵 $\mathbf{G}_L$ 含有线性相关的向量, 则求取的输出权值的幅值会很大, 容易得到病态解, 造成系统不稳定^[35]. 为了解决此问题, RSC-II^[27] 将 L_2 范数添加到 SCN 的目标函数中, 则 RSC-II 的目标函数可表示为:

$$\min : J = \frac{1}{2} \|\beta\|_2^2 + \frac{1}{2} \lambda_2 \|\mathbf{t} - \mathbf{G}_L \beta\|_2^2 \quad (8)$$

式中, λ_2 为正则化参数.

通过岭回归广义逆对式 (8) 进行求解, 可得:

$$\beta = (\mathbf{G}_L^T \mathbf{G}_L + \frac{\mathbf{I}}{\lambda_2})^{-1} \mathbf{G}_L^T \mathbf{t} \quad (9)$$

从式 (9) 可知, RSC-II 在 $\mathbf{G}_L^T \mathbf{G}_L$ 的对角线上增加了一个正数项, 避免了病态解的问题, 同时可进一步降低输出权值的幅值. 因此, RSC-II 能够通过减小输出权值的幅值来提高模型的泛化能力, 但它不能得到稀疏性的模型解, 不能通过剔除无用隐性节点来控制模型规模.

为了实现 SCN 模型的稀疏学习, 本文将 L_1 范数惩罚项引入到 SCN 算法中, 则 PSCN 的目标函数

为:

$$\begin{aligned} \min : J &= \lambda_1 \|\beta\|_1 + \frac{1}{2} \sum_{i=1}^N \epsilon_i^2 \\ s.t. : g(\mathbf{x}_i)\beta &= \mathbf{t}_i - \epsilon_i, \forall i \end{aligned} \quad (10)$$

式中, λ_1 为正则化系数, ϵ 为网络残差.

在 PSCN 的构造过程中, 需要通过求解式 (10) 来更新整个网络的输出权重. 优化问题 (10) 可简化为矩阵形式:

$$J(\beta) = \lambda_1 \|\beta\|_1 + \frac{1}{2} \|\mathbf{G}_L \beta - \mathbf{t}\|^2 \quad (11)$$

将优化问题 (11) 进一步写成 ADMM 形式:

$$\begin{aligned} \min f(\mathbf{x}) + g(\beta) \\ s.t. : \mathbf{x} - \beta = 0 \end{aligned} \quad (12)$$

式中, $f(\mathbf{x}) = \frac{1}{2} \|\mathbf{G}_L \mathbf{x} - \mathbf{t}\|^2$, $g(\beta) = \lambda_1 \|\beta\|_1$.

构造式 (12) 的增广拉格朗日函数 $L_\rho(\mathbf{x}, \beta, \mu)$ 为:

$$\begin{aligned} L_\rho(\mathbf{x}, \beta, \mu) &= f(\mathbf{x}) + g(\beta) + \mu^T (\mathbf{x} - \beta) \\ &\quad + \frac{\rho}{2} \|\mathbf{x} - \beta\|_2^2 \end{aligned} \quad (13)$$

式中, $\mu \in \mathbb{R}^{L \times m}$ 为对偶变量, ρ 为惩罚系数. 利用 ADMM 算法对上述优化问题 (13) 进行求解, 其具体求解过程如下:

$$\begin{aligned} \mathbf{x}^{k+1} &= \arg \min_{\mathbf{x}} L_\rho(\mathbf{x}, \beta^k, \mu^k) \\ \beta^{k+1} &= \arg \min_{\beta} L_\rho(\mathbf{x}^{k+1}, \beta, \mu^k) \\ \mu^{k+1} &= \mu^k + \rho(\mathbf{x}^{k+1} - \beta^{k+1}) \end{aligned} \quad (14)$$

式中, k 为迭代次数. 引入尺度对偶变量 $\mu_1 = \frac{\mu}{\rho}$, 将式 (14) 转化为:

$$\begin{aligned} \mathbf{x}^{k+1} &= \arg \min_{\mathbf{x}} (f(\mathbf{x}) + \frac{\rho}{2} \|\mathbf{x} - \beta^k \\ &\quad + \mu_1^k\|_2^2) \end{aligned} \quad (15)$$

$$\begin{aligned} \beta^{k+1} &= \arg \min_{\beta} (g(\beta) + \frac{\rho}{2} \|\mathbf{x}^{k+1} - \beta \\ &\quad + \mu_1^k\|_2^2) \end{aligned} \quad (16)$$

$$\mu_1^{k+1} = \mu_1^k + \mathbf{x}^{k+1} - \beta^{k+1} \quad (17)$$

对式 (15) 求导可得:

$$\nabla = \mathbf{G}_L^T \mathbf{G}_L \mathbf{x} - \mathbf{G}_L^T \mathbf{t} + \rho(\mathbf{x} - \beta^k + \mu_1^k) \quad (18)$$

令 $\nabla = 0$, 有:

$$\mathbf{x}^{k+1} = (\mathbf{G}_L^T \mathbf{G}_L + \rho \mathbf{I})^{-1} (\mathbf{G}_L^T \mathbf{t} + \rho(\beta^k - \mu_1^k)) \quad (19)$$

根据文 [36], 由式 (16) 可得:

$$\beta^{k+1} = S_{\lambda_1/\rho}(\mathbf{x}^{k+1} + \mu_1^k) \quad (20)$$

式中, S 表示软阈值算子:

$$S_{\kappa(a)} = \begin{cases} a - \kappa & a > \kappa \\ 0 & |a| \leq \kappa \\ a + \kappa & a < -\kappa \end{cases} \quad (21)$$

利用 ADMM 迭代得到的解等价于优化问题 (12) 的解^[37].

2.2 不等式约束

SCN 和 RSC-II 采用不等式约束来实现候选节点池的搭建, 并从中得到最佳隐性节点. 但由于所建不等式对隐性参数质量的约束力有限, 便有可能造成生成的隐性节点质量不高. 为此, 本文建立一个新的不等式约束对隐性参数的分配加以约束, 来获得高质量的隐性节点, 这也是本文的主要工作之一.

将优化问题 (10) 转化为:

$$\begin{aligned} J(\beta_L) &= \Upsilon + \frac{1}{2} \|\mathbf{e}_L\|^2 \\ &= \Upsilon + \frac{1}{2} \|\mathbf{e}_{L-1} - \beta_L \mathbf{g}_L\|^2 \\ &= \Upsilon + \frac{\mathbf{g}_L^2}{2} \left\| \frac{\langle \mathbf{e}_{L-1}, \mathbf{g}_L \rangle}{\mathbf{g}_L^2} - \beta_L \right\|^2 \end{aligned} \quad (22)$$

其中, $\Upsilon = \lambda_1(|\beta_1| + |\beta_2| + \dots + |\beta_L|)$.

在式 (22) 中, $\beta_1, \beta_2, \dots, \beta_{L-1}$ 是已知的. 类似于式 (16) 的求解过程, 可得:

$$\beta_L = S_{(\lambda_1/\mathbf{g}_L^2)}\left(\frac{\langle \mathbf{e}_{L-1}, \mathbf{g}_L \rangle}{\mathbf{g}_L^2}\right) \quad (23)$$

由式 (23) 可得:

$$\begin{aligned} &\|\mathbf{e}_{L-1}\|^2 - \|\mathbf{e}_L\|^2 \\ &= \|\mathbf{e}_{L-1}\|^2 - \|\mathbf{e}_{L-1} - \beta_L \mathbf{g}_L\|^2 \\ &= 2\langle \mathbf{e}_{L-1}, \mathbf{g}_L \rangle \beta_L - \beta_L^2 \mathbf{g}_L^2 \\ &= \frac{2\langle \mathbf{e}_{L-1}, \mathbf{g}_L \rangle (\langle \mathbf{e}_{L-1}, \mathbf{g}_L \rangle \pm \lambda_1)}{\mathbf{g}_L^2} \\ &\quad - \frac{(\langle \mathbf{e}_{L-1}, \mathbf{g}_L \rangle \pm \lambda_1)^2}{\mathbf{g}_L^2} \\ &= \frac{(\langle \mathbf{e}_{L-1}, \mathbf{g}_L \rangle \mp \lambda_1)(\langle \mathbf{e}_{L-1}, \mathbf{g}_L \rangle \pm \lambda_1)}{\mathbf{g}_L^2} \\ &= \frac{\langle \mathbf{e}_{L-1}, \mathbf{g}_L \rangle^2 - \lambda_1^2}{\mathbf{g}_L^2} \end{aligned} \quad (24)$$

根据式 (24), 再基于 SCN 的不等式约束 (4), 可建立如下形式的不等式约束:

$$\langle \mathbf{e}_{L-1}, \mathbf{g}_L \rangle^2 - \lambda_1^2 \geq b_g^2(1 - r - u_L) \|\mathbf{e}_{L-1}\|^2 \quad (25)$$

与式 (4) 相比, 本文所建不等式对隐性参数质量的监督作用更强, 从而有利于得到高质量的隐性节点, 并可通过调节 λ_1 来改变监督的约束程度.

2.3 PSCN 的收敛性分析

令

$$\begin{aligned}\beta^* &= [\beta_1^*, \beta_2^*, \dots, \beta_L^*]^T \\ &= \arg \min_{\beta} \frac{1}{2} \|\mathbf{f} - \sum_{j=1}^L \beta_j \mathbf{g}_j\|^2 + \lambda_1 \|\beta\|_1\end{aligned}$$

$$\begin{aligned}\mathbf{e}_L^* &= \mathbf{f} - \sum_{j=1}^L \beta_j^* \mathbf{g}_j \\ &= [\mathbf{e}_{L,1}^*, \mathbf{e}_{L,2}^*, \dots, \mathbf{e}_{L,m}^*]\end{aligned}$$

记

$$\begin{aligned}\tilde{\beta}_L &= S_{(\lambda_1/\mathbf{g}_L^2)}(\frac{\langle \mathbf{e}_{L-1}^*, \mathbf{g}_L \rangle}{\mathbf{g}_L^2}) \\ \tilde{\mathbf{e}}_L &= \mathbf{e}_{L-1}^* - \tilde{\beta}_L \mathbf{g}_L\end{aligned}$$

其中, $\tilde{\beta}_L = [\tilde{\beta}_{L,1}, \tilde{\beta}_{L,2}, \dots, \tilde{\beta}_{L,m}]^T$, $\mathbf{e}_0^* = \mathbf{f}$.

定理 1. 假设 $\text{span}(\mathbf{\Gamma})$ 在 L_2 空间中稀疏, 且 $\forall \mathbf{g} \in \mathbf{\Gamma}$, 对于 $b_g \in \mathbb{R}^+$, 有 $0 < \|\mathbf{g}\| < b_g$. 给定充分小的正实数 λ_1 满足 $\lambda_1^2 < \langle \mathbf{e}_{L-1}^*, \mathbf{g}_L \rangle^2$, $0 < r < 1$, 非负实数序列 u_L 且满足 $u_L \leq (1-r)$ 和 $\lim_{L \rightarrow \infty} u_L = 0$. 存在一个非负实数序列 ε_L 满足

$$\|\mathbf{e}_L^*\|^2 \geq \varepsilon_L \quad (26)$$

对于 $L = 1, 2, \dots$, 做出如下定义:

$$\eta_L = \frac{\|\mathbf{e}_{L-1}^*\|^2 - \varepsilon_L}{\|\mathbf{e}_{L-1}^*\|^2} \quad (27)$$

$$\chi = \lim_{L \rightarrow \infty} \prod_{k=1}^L (1 - \eta_k) \quad (28)$$

$$\gamma = \lim_{L \rightarrow \infty} \prod_{k=1}^L (r + u_k) < 1 \quad (29)$$

若网络的随机基函数 $\mathbf{g}_L$ 由如下不等式约束得到:

$$\langle \mathbf{e}_{L-1}^*, \mathbf{g}_L \rangle^2 - \lambda_1^2 \geq b_g^2 (1 - r - u_L) \|\mathbf{e}_{L-1}^*\|^2 \quad (30)$$

输出权重 β 由式 (19), 式 (20) 和式 (17) 依次迭代获取, 则 $\sqrt{\chi} \|\mathbf{f}\| \leq \lim_{L \rightarrow +\infty} \|\mathbf{f} - \mathbf{f}_L^*\| \leq \sqrt{\gamma} \|\mathbf{f}\|$, 其中, $\mathbf{f}_L^* = \sum_{j=1}^L \beta_j^* \mathbf{g}_j$, $\beta_j^* = [\beta_{j,1}^*, \beta_{j,2}^*, \dots, \beta_{j,m}^*]^T$.

证明. 容易可知, $\|\mathbf{e}_L^*\|^2 \leq \|\tilde{\mathbf{e}}_L\|^2 = \|\mathbf{e}_{L-1}^* - \tilde{\beta}_L \mathbf{g}_L\|^2 \leq \|\mathbf{e}_{L-1}^*\|^2 \leq \|\tilde{\mathbf{e}}_{L-1}\|^2$, 结合 $\lambda_1^2 < \langle \mathbf{e}_{L-1}^*, \mathbf{g}_L \rangle^2$, 可得:

$$\begin{aligned}&\|\mathbf{e}_L^*\|^2 - \|\mathbf{e}_{L-1}^*\|^2 \\ &\leq \|\tilde{\mathbf{e}}_L\|^2 - \|\mathbf{e}_{L-1}^*\|^2 \\ &= \|\mathbf{e}_{L-1}^* - \tilde{\beta}_L \mathbf{g}_L\|^2 - \|\mathbf{e}_{L-1}^*\|^2 \\ &= \frac{-\langle \mathbf{e}_{L-1}^*, \mathbf{g}_L \rangle^2 + \lambda_1^2}{\mathbf{g}_L^2} \\ &< 0\end{aligned} \quad (31)$$

由式 (31) 可知, 网络残差序列 $\{\|\mathbf{e}_L^*\|^2\}$ 是单调递减的.

根据文 [38] 可知, L_1 范数惩罚项的引入会导致有偏估计. 因此, $\|\mathbf{e}_L^*\|^2 \neq 0$, 且存在一个充分小的正实数 ε_L 使得式 (26) 成立.

根据式 (26) 可得:

$$\|\mathbf{e}_{L-1}^*\|^2 - \|\mathbf{e}_L^*\|^2 \leq \|\mathbf{e}_{L-1}^*\|^2 - \varepsilon_L \quad (32)$$

将式 (27) 代入式 (32) 可得:

$$\|\mathbf{e}_{L-1}^*\|^2 - \|\mathbf{e}_L^*\|^2 \leq \eta_L \|\mathbf{e}_{L-1}^*\|^2 \quad (33)$$

上式 (33) 可写成:

$$(1 - \eta_L) \|\mathbf{e}_{L-1}^*\|^2 \leq \|\mathbf{e}_L^*\|^2 \quad (34)$$

由式 (30) 和式 (31) 可得:

$$\begin{aligned}&\|\mathbf{e}_L^*\|^2 - (r + u_L) \|\mathbf{e}_{L-1}^*\|^2 \\ &\leq \|\tilde{\mathbf{e}}_L\|^2 - (r + u_L) \|\mathbf{e}_{L-1}^*\|^2 \\ &= \|\mathbf{e}_{L-1}^* - \tilde{\beta}_L \mathbf{g}_L\|^2 - (r + u_L) \|\mathbf{e}_{L-1}^*\|^2 \\ &= (1 - r - u_L) \|\mathbf{e}_{L-1}^*\|^2 - \frac{\langle \mathbf{e}_{L-1}^*, \mathbf{g}_L \rangle^2 - \lambda_1^2}{\mathbf{g}_L^2} \\ &\leq (1 - r - u_L) \|\mathbf{e}_{L-1}^*\|^2 - \frac{\langle \mathbf{e}_{L-1}^*, \mathbf{g}_L \rangle^2 - \lambda_1^2}{b_g^2} \\ &\leq 0\end{aligned} \quad (35)$$

联立式 (34) 和式 (35) 可得:

$$\chi \|\mathbf{e}_0^*\|^2 \leq \|\mathbf{e}_L^*\|^2 \leq \gamma \|\mathbf{e}_0^*\|^2 \quad (36)$$

故可得 $\sqrt{\chi} \|\mathbf{f}\| \leq \lim_{L \rightarrow \infty} \|\mathbf{e}_L^*\| \leq \sqrt{\gamma} \|\mathbf{f}\|$. 证毕.

注 1. 当 λ_1 充分小时, 可通过设置学习参数 r 来保证式 (30) 始终成立. 另外, 在实际应用中, 式 (30) 也可作为网络模型训练的终止条件. 在网络模型构建的过程中, 若 λ_1^2 的取值始终大于或等于 $\langle \mathbf{e}_{L-1}^*, \mathbf{g}_L \rangle^2$, 则终止 SLFN 的构建.

注 2. 在每一次的构造过程中, 网络的候选节点是在监督机制下得到的, 虽然其可以使模型误差变小, 但由于学习算法的随机性, 候选节点可能均使模型误差下降的非常小, 即使从中选择一个质量最好的作为新增节点, 对建模质量的贡献也不大. 如果后面出现与之相似但对模型贡献更大的节点, 那么其将成为无效节点. 因此, 本文利用 1 范数正则化方法来去掉这些无用隐性节点.

注 3. 从定理 1 可知, PSCN 首先利用不等式约束 (30) 生成高质量的新增节点, 再通过 1 范数正则化方法来产生稀疏的模型解, 使部分隐性节点的输出权值为 0, 进而简化模型结构. 此外, 1 范数正则化方法可有效避免较大输出权值的出现, 使 PSCN 具

有较高的泛化性能.

注 4. 所提 PSCN 算法, 网络中的输出权值是由 ADMM 算法迭代计算得到. 因而, 相比于 SCN 算法, PSCN 的学习速度会有所下降. 但所提出的稀疏学习方法可进一步约减隐性节点, 在保持较高的预测精度的同时, 简化了模型尺寸, 适合于存储有限的工业系统.

注 5. 在传统的学习模型中, 正则化参数 λ_1 可用于调节正则化项与误差项之间的相对重要性, 在 PSCN 中, λ_1 具有同样作用. 此外, 从式 (25) 可知, λ_1 的取值一定程度上影响了不等式约束的程度. 较大的 λ_1 将使得不等式约束过大, 不利于候选节点池的建立; 而当 λ_1 取值较小时, 稀疏化作用不明显. 因此 λ_1 应根据不同的数据集在适当的范围内通过实验选取.

2.4 PSCN 算法描述

为便于候选隐性节点的生成, 引入变量 ξ_L :

$$\xi_L = \langle \mathbf{e}_{L-1}^*, \mathbf{g}_L \rangle^2 - \lambda_1^2 - \mathbf{g}_L^2 (1 - r - u_L) \|\mathbf{e}_{L-1}\|^2 \quad (37)$$

PSCN 的算法流程如下:

给定训练数据 $\{\mathbf{x}, \mathbf{t}\}_{i=1}^N$, 定义如下变量: $L_{\max}$ 为预先设定的隐性节点数, ϵ 为期望精度, $T_{\max}$ 为预配置的隐性节点数, $\Upsilon = \{\lambda_{\min} : \Delta\lambda : \lambda_{\max}\}$ 为隐性参数的选取范围.

步骤 1. 网络初始化: 令 $\mathbf{e}_0 := [\mathbf{t}_1, \mathbf{t}_2, \dots, \mathbf{t}_N]^T$, 给定 $\lambda_1 > 0, 0 < r < 1$, 两个空集: Ω 和 $\mathbf{W}$;

步骤 2. 网络训练阶段: 生成新增节点和更新输出权重.

当 $L \leq L_{\max}$ 和 $\|\mathbf{e}_0\|_F > \epsilon$ 时,

生成新增节点:

- 1) For $\lambda \in \Upsilon$,
- 2) For $k = 1, 2, \dots, T_{\max}$,
- 3) 从区间 $[-\lambda, \lambda]^d$ 和 $[-\lambda, \lambda]$ 内分别随机选取隐性参数 $\mathbf{w}_L$ 和 b_L ;
- 4) 根据式 (37) 得到 ξ_L , 令 $u_L = (1 - r)/(L + 1)$;
- 5) If $\xi_L \geq 0$
- 6) 将 $\mathbf{w}_L$ 和 b_L 存储在 $\mathbf{W}$ 中, ξ_L 存储在 Ω 中;
- 7) Else 返回 2)
- 8) End If
- 9) End For
- 10) If $\mathbf{W}$ 不为空集
- 11) 在 Ω 中找到使 ξ_L 最大的隐性参数: $\mathbf{w}_L^*$

和 b_L^* , 并令 $\mathbf{G}_L = [\mathbf{g}_1^*, \mathbf{g}_2^*, \dots, \mathbf{g}_L^*]$;

12) Break(返回 16));

13) Else 在 $\tau \in (0, 1 - r)$ 中随机选取 τ , 并更新 $r := r + \tau$, 再返回 2);

14) End If

15) End For

更新输出权重:

16) 通过式 (19), 式 (20) 和式 (17) 依次迭代

来求解 β^* , 直到满足预先设定的迭代次数.

17) 计算 $\mathbf{e}_L = \mathbf{e}_0 - \mathbf{G}_L \beta^*$;

18) 令 $\mathbf{e}_0 := \mathbf{e}_L, L := L + 1$;

最后, 算法结束, 并得到最优的 $\beta_1^*, \beta_2^*, \dots, \beta_L^*$ 和 $\mathbf{w}^* = [\mathbf{w}_1^*, \mathbf{w}_2^*, \dots, \mathbf{w}_L^*]$ 以及 $\mathbf{b}^* = [b_1^*, b_2^*, \dots, b_L^*]$.

3 性能评估

为验证算法的有效性, 将本文所提 PSCN 算法与 IRWNN、SCN^[21] 和 RSC-II 3 种算法进行比较分析, 其中, RSC-II 算法从集合 $\{2^{-20}, 2^{-19}, \dots, 2^0, \dots, 2^{19}, 2^{20}\}$ 中选择一个最优的值作为正则化系数 λ_2 . 此外, 选取均方根误差 (Root mean squares error, RMSE)、决定系数 (determination coefficient, R^2) 和可解释方差 (explained variance, E_{VAR}) 三种性能指标, 测量算法的泛化能力.

RMSE 用于描述预测模型输出相对于真实测量值的波动情况, 其值越小越好, 值越小说明预测的精度越高. RMSE 可表示为:

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^N [\hat{y}_i - y_i]^2}{N}} \quad (38)$$

式中, $\hat{y}_i$ 为预测模型的输出值, y_i 为测试样本输出.

对于 R^2 和 E_{VAR} , 其最佳值为 1, 值越小表明模型预测精度越差. R^2 和 E_{VAR} 的定义如式 (39) 和式 (40) 所示:

$$R^2 = 1 - \frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{\sum_{i=1}^N (y_i - \bar{y})^2} \quad (39)$$

$$E_{\text{VAR}} = 1 - \frac{\text{Var}(y_i - \hat{y}_i)}{\text{Var}(y_i)} \quad (40)$$

式中, $\bar{y}$ 为测试样本输出的平均值, $\text{Var}(\cdot)$ 表示方差运算.

针对每一个回归任务, 网络的输入输出数据均归一化到区间 $[0, 1]$ 上. 所有测试实验均独立进行 50 次, 并取其平均值 (Mean) 和标准差 (Standard deviation, Std) 作为最终的实验结果, 以消除偶然因素的影响. 所有测试实验均在 CPU 为 i5, 3.4G Hz, 内存为 8G RAM 的 PC 机上运行, 其运行环境为 Matlab 2016 a. 另外, 最好的仿真结果均用黑体表示.

3.1 公共数据集

实验采用了来自于 KEEL (Knowledge Extraction based on Evolutionary Learning) 数据库⁰ 的 4 个基准数据集来验证 PSCN 算法的有效性. 所有数据集的基本信息如表 1 所示.

表 1 基准数据集

数据集	输入	输出	训练样本	测试样本
Laser	4	1	744	249
Auto MPG6	5	1	313	79
Concrete	8	1	515	515
Compactiv	21	1	6144	2048

3.2 比较实验

在 Auto MPG6 和 Laser 两个例子中, 4 种算法的实验参数设定如表 2 所示. 对于 Concrete 和 Compactiv, 除了 PSCN 算法中, 正则化系数分别为 $\lambda_1 = 0.05$ 和 $\lambda_1 = 0.5$ 外, 其余实验参数的设定和其他 3 种算法实验参数的选取均与表 2 相同.

表 2 Auto MPG6 和 Laser 数据集中 4 种算法的实验参数设置

参数	IRWNN	SCN	RSC-II	PSCN
$T_{\max}$	1	200	200	200
ρ_1	—	—	—	0.9
Υ	{1}	{1 : 0.1 : 5}	{1 : 0.1 : 5}	{1 : 0.1 : 5}
正则化系数	—	—	2^{10}	0.005

在 4 个基准数据集计算得到的 4 种算法的 RMSE 和模型规模如表 3 和表 4 所示. 从表 3 可以看出, 与其他 3 种算法相比, PSCN 算法的泛化性能有一定的提升. 与 IRWNN、SCN 和 RSC-II 算法相比之下, PSCN 算法的测试误差为最小, 其结果也更稳定. 从表 4 可以看出, 4 个回归问题中, 4 种算法的初始化模型规模均为 $L_{\max} = 70$, 经过训练后, IRWNN、SCN 和 RSC-II 3 种模型的最终有效的节点数均为 70, 而 PSCN 算法的最终模型节点数分别为 34.92、49.52、46.42 和 42.36. 可以看出, PSCN 模型的隐性节点数为最少的, 说明其模型最紧致, 稀疏性最好. 由于 1 范数正则化方法能够实现模型解的稀疏性, 使得部分节点的输出权值为 0. 因此, PSCN 算法所包含的有效隐性节点数均少于其他 3 种算法, 这说明了 PSCN 可有效简化模型结构, 同时具有较高的泛化性能.

表 5 和表 6 分别为 4 种算法在 R^2 和 E_{VAR} 两个性能指标上的实验结果. 可以看出, IRWNN 算法的 R^2 和 E_{VAR} 均为最小, 模型预测能力最差. 与 IRWNN 算法相比, SCN 和 RSC-II 算法的 R^2 和 E_{VAR} 得到了较大的提升, 具有较高的预测精度. 其中, RSC-II 算

法在预测精度方面优于 SCN 算法. 对于本文所提 PSCN 算法, 其在 Concrete 和 Compactiv 两个例子中得到的 R^2 和 E_{VAR} 均明显高于其他 3 种算法, 且在其余两个例子中获得的 R^2 和 E_{VAR} 略高于 RSC-II 算法. 需要指出的是, PSCN 算法所需要的隐性节点数为最少, 这也从不同方面验证了本文所提 PSCN 算法在简化模型结构, 同时保持较高泛化能力方面的优越性.

为了检验所提 PSCN 算法的简化模型结构的能力, 进一步研究在相同训练容忍误差 ϵ 下, 计算 4 种算法包含的隐性节点数, 如图 1 所示. 其中 IRWNN 算法需要大量隐性节点才能达到给定的期望精度, 故本文没有给出相关实验结果. 从图 1 可以看出, 在相同的容忍误差 ϵ 下, 相比于其他两种算法, 所提 PSCN 算法所包含的隐性节点数为最少, 这也说明了 PSCN 算法可获得更紧致的模型结构. 表 7 给出了 4 种算法对应模型规模的具体实验结果.

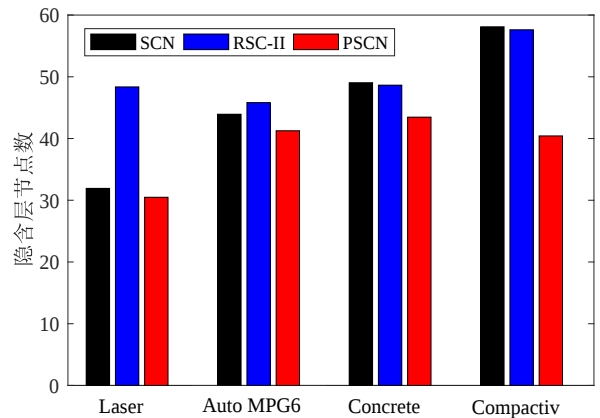


图 1 相同期望精度下 4 种算法在基准数据集上的隐性节点数

图 2 为迭代过程中 Auto MPG6 数据集所对应的目标函数 (8) 的演化路径. 可以看出, 相对较少的迭代次数就能够使得目标函数 (8) 收敛. 另外, 在每次的节点增加过程中, 计算损耗主要集中在输出权值 β 的迭代求解上. 表 8 为所提 PSCN 算法的运算效率以及模型规模和测试集 RMSE 随迭代次数增加时的变化情况. 可以看出, 随着迭代次数的增加, PSCN 算法的模型规模和测试集 RMSE 变化不大, 而运算时间则呈不同程度增长. 因此可以通过设定合适的迭代次数来降低 PSCN 算法的计算复杂度, 以缓解实际工业系统的计算压力. 因此, 本文所建 PSCN 模型适用于对估计精度要求较高, 同时具有较强算能的工业系统.

⁰<https://sci2s.ugr.es/keel/category.php?cat=reg>

为了便于观察 PSCN 算法在泛化性能上的优势, 这里以 Auto MPG6 为例来加以说明. 图 3 为 4 种建模算法下的模型权值分布情况. 从图 3 可以看出, IRWNN 模型的输出权值不为 0 且较小, 但模型预测精度最差. SCN 和 RSC-II 模型得到的输出权值较大且不为 0. 对于本文所提 PSCN 模型, 一部分的输出权值都为 0, 模型最终有效的隐性节点数为 47. 另外, 模型的输出权值均较小. 这也说明了由于 1 范数正则化的作用, PSCN 算法简化了模型规模, 同时对输出权值的幅值具有较强的抑制作用, 具有较高的泛化性能.

表 9 为 PSCN 模型对 λ_1 的灵敏度. 从表 9 可以看出, 随着 λ_1 值的减小, PSCN 模型所包含的隐性节点越来越多, 模型的 RMSE 呈小幅波动. 这说明 λ_1 控制着 PSCN 模型的尺寸, 而且能够保证模型的泛化性能. 经实验研究, 当 λ_1 的取值较大时, PSCN 模型难以建立有效的候选节点池. 根据表 9 可知, 每个数据集的 λ_1 都有一个有效的范围. 因此, 在这一有效范围内, 我们选择表现较好的 λ_1 的值作为实验参数.

PSCN 算法的泛化能力优于其他 3 种算法, 其主要原因在于:

1) IRWNN 算法在固定区间内随机选取隐性参数, 具有一定的盲目性, 导致难以生成高质量的隐性节点, 而 SCN 算法采用不等式约束来获得高质量隐性节点, 同时通过全局最小二乘法更新整个网络的

输出权重. 因此, 与 IRWNN 算法相比, SCN 算法具有较好的预测性能.

2) RSC-II 算法利用 2 范数正则化方法来平衡建模精度与模型复杂度, 可有效抑制过拟合问题, 并通过减小输出权值的幅值来提升泛化性能. 因此, RSC-II 算法与 IRWNN 和 SCN 两种算法相比, 在预测性能方面具有明显的优势;

3) 然而, RSC-II 算法虽然在不等式约束及 2 范数正则化作用下提高了所建模型的精度, 但其无法去除模型中的无用隐性节点来简化模型结构. PSCN 算法利用 1 范数正则化方法进一步剔除无效隐性节点, 来简化模型结构, 同时能有效抑制输出权值出现大幅值, 改善了网络的泛化性能.

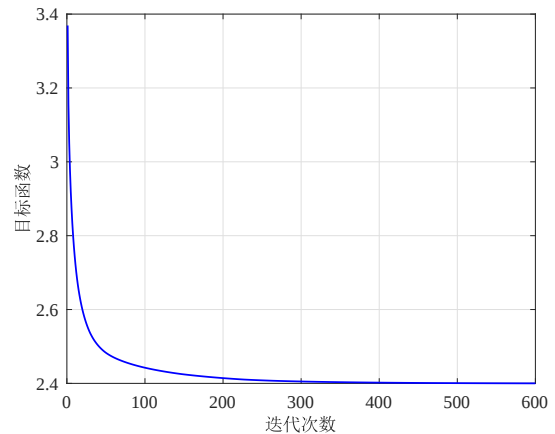


图 2 迭代过程中 Auto MPG6 数据集所对应的目标函数 (8) 的演化

表 3 4 种算法在基准数据集上 RMSE 的比较结果

数据集	IRWNN		SCN		RSC-II		PSCN	
	Mean	Std	Mean	Std	Mean	Std	Mean	Std
Laser	0.2177	0.0145	0.0643	0.0084	0.0543	0.0026	0.0532	0.0019
Auto MPG6	0.2069	0.0099	0.1470	0.0125	0.1283	0.0080	0.1207	0.0015
Concrete	0.3075	0.0098	0.2142	0.0151	0.2033	0.0105	0.1893	0.0024
Compactiv	0.2936	0.0278	0.0982	0.0157	0.0930	0.0092	0.0847	0.0032

表 4 4 种算法在基准数据集上模型规模的比较结果

数据集	IRWNN 模型规模	SCN 模型规模	RSC-II 模型规模	PSCN 模型规模
Laser	70	70	70	70(34.92)
Auto MPG6	70	70	70	70(49.52)
Concrete	70	70	70	70(46.42)
Compactiv	70	70	70	70(42.36)

表 5 4 种算法在基准数据集上 R^2 的比较结果

数据集	IRWNN		SCN		RSC-II		PSCN	
	Mean	Std	Mean	Std	Mean	Std	Mean	Std
Laser	0.6102	0.0591	0.9661	0.0086	0.9750	0.0028	0.9768	0.0019
Auto MPG6	0.7397	0.0261	0.8671	0.0196	0.8996	0.0127	0.9094	0.0015
Concrete	0.4234	0.0329	0.7202	0.0425	0.7451	0.0218	0.7811	0.0024
Compactiv	0.4270	0.1142	0.9221	0.0520	0.9439	0.0131	0.9541	0.0032

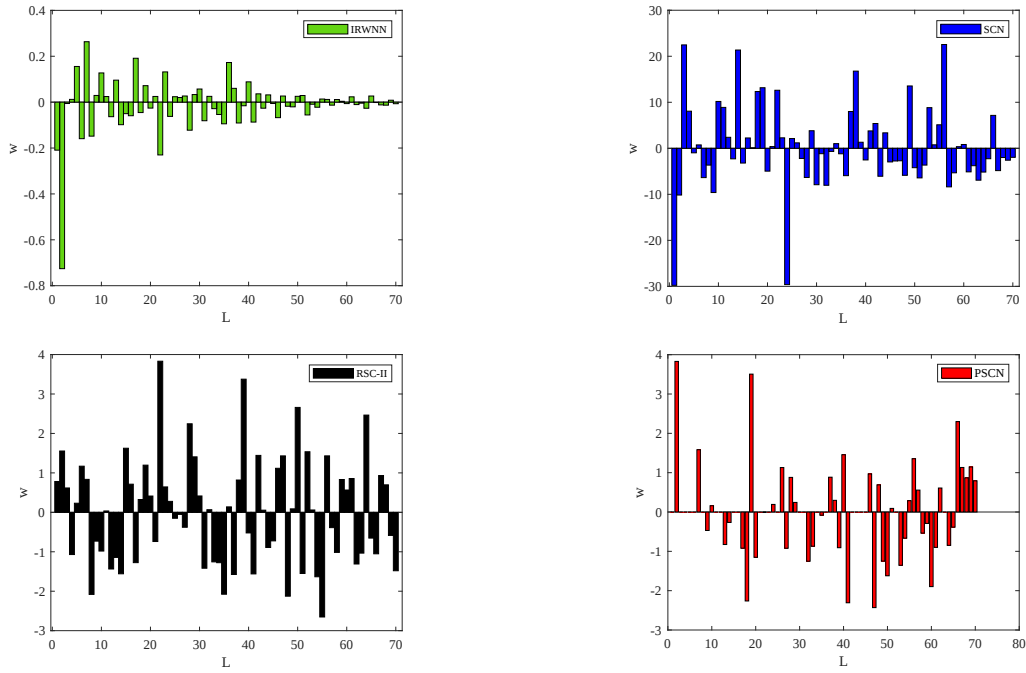


图 3 4 种算法在 Auto MPG6 数据集上各输出权值的分布

表 6 4 种算法在基准数据集上 E_{VAR} 的比较结果

数据集	IRWNN		SCN		RSC-II		PSCN	
	Mean	Std	Mean	Std	Mean	Std	Mean	Std
Laser	0.6138	0.0588	0.9664	0.0085	0.9751	0.0029	0.9770	0.0019
Auto MPG6	0.7402	0.0261	0.8711	0.0186	0.9021	0.0123	0.9114	0.0015
Concrete	0.4248	0.0322	0.7219	0.0419	0.7465	0.0216	0.7818	0.0024
Compactiv	0.4283	0.1141	0.9222	0.0520	0.9440	0.0131	0.9542	0.0032

表 7 相同期望精度下 4 算法在基准数据集上隐性节点数的比较结果

数据集	算法	期望精度	隐性节点数
Laser	IRWNN	$\epsilon = 0.04$	—
	SCN		31.92
	RSC-II		48.36
	PSCN		30.48
Auto MPG6	IRWNN	$\epsilon = 0.12$	—
	SCN		43.93
	RSC-II		45.82
	PSCN		41.26
Concrete	IRWNN	$\epsilon = 0.15$	—
	SCN		49.04
	RSC-II		48.64
	PSCN		43.46
Compactiv	IRWNN	$\epsilon = 0.08$	—
	SCN		58.10
	RSC-II		57.62
	PSCN		40.42

表 8 不同迭代次数下 PSCN 算法对 Auto MPG6 数据集的性能对比

性能	迭代次数							
	100	200	300	400	600	800	1000	1200
RMSE	0.1139	0.1168	0.1172	0.1180	0.1196	0.1193	0.1190	0.1184
模型规模	58.80	55.44	53.80	51.72	51.14	49.86	49.68	50.28
运行时间	1.8035	1.8709	1.9248	1.9396	2.0398	2.0892	2.1489	2.2002

表 9 PSCN 算法在基准数据集上对正则化系数 λ_1 的灵敏度

数据集	灵敏度											
	λ_1											
Laser	1/50		1/100		1/150		1/200		1/250		1/300	
	RMSE	L	RMSE	L	RMSE	L	RMSE	L	RMSE	L	RMSE	L
	0.0554	23.7200	0.0536	28.4000	0.0530	32.0200	0.0537	34.4400	0.0542	37.5200	0.0547	39.5400
Auto MPG6	1/50		1/100		1/150		1/200		1/250		1/300	
	RMSE	L	RMSE	L	RMSE	L	RMSE	L	RMSE	L	RMSE	L
	0.1166	32.9600	0.1155	40.3400	0.1174	45.6400	0.1197	48.8600	0.1222	52.5600	0.1236	53.8600
Concrete	1/20		1/40		1/60		1/80		1/100		1/120	
	RMSE	L	RMSE	L	RMSE	L	RMSE	L	RMSE	L	RMSE	L
	0.1881	46.6600	0.1892	54.4400	0.1931	59.1200	0.1929	60.9800	0.1943	63.5600	0.1975	64.3800
Compactiv	1/2		1/4		1/8		1/16		1/32		1/64	
	RMSE	L	RMSE	L	RMSE	L	RMSE	L	RMSE	L	RMSE	L
	0.0843	42.6000	0.0828	47.9600	0.0831	52.6000	0.0843	57.2600	0.0868	61.3800	0.0897	64.8400

3.3 工程案例

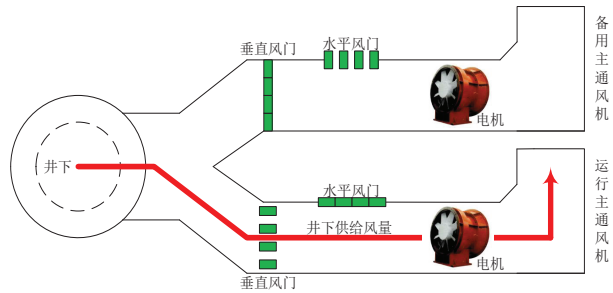


图 4 煤矿主风机切换过程结构图

从图 4 可以看出, 煤矿主通风机切换过程是将一台运行的主通风机切换到另一台备用主通风机运行, 以保证整个矿井的高效、持续生产. 在切换过程中, 需要提供稳定且适量的井下供给风量来维持一个舒适的地下采煤环境. 为实现这一目的, 关键在于精确获取井下供给风量信息, 保障两台主通风机的安全与稳定运行, 同时实现切换过程的稳定性与安全性, 以提供满足地下需求的供给风量. 然而由于井下供给风流成分复杂, 工作环境恶劣, 井下供给风量测量装置的取压孔容易产生堵塞现象, 致使风量仪表测量效果不佳, 造成通风机切换过程长期闭环优化控制无法实现. 因此, 其估计精度成为切换过程闭环优化控制的主要瓶颈问题.

另外, 煤矿主通风机切换过程控制系统一般以 PLC 为控制核心, 来实现两台主通风机的控制与切换, 同时存储其运行状态及供给风量信息. 由于 PLC 的存储空间是有限的, 其迫切需要一个既能够精确估计井下供给风量信息, 又占用空间小的模型. 本文将所提 PSCN 算法应用于煤矿主通风切换过程的井下供给风量估计, 以此来说明所提算法的性能.

根据文 [39] 和 [40], 需要建立以地下矿井风阻 R_0 、4 个风门风阻 R_{1c} , R_{1s} , R_{2c} , R_{2s} 和 2 台主通风机的压头 H_{1d} , H_{2d} 为输入, 井下供给风量 U_0 为输出的软测量模型, 为切换过程的闭环优化控制提供反馈信息. 基于文 [41] 所建的主通风机切换过程分布式半实物仿真平台, 采集 1400 组训练数据, 100 组测试数据. 另外, 这里实验参数 (λ_1 除外) 的选取与表 2 相同, 初始化模型规模为 $L_{\max} = 70$.

表 10 为基于 PSCN 的井下供给风量软测量模型对正则化系数 λ_1 的灵敏度. 从表 10 可以看出, λ_1 的值越大, 模型规模越小, 模型的估计精度呈不同程度的变化. 因此, 在实际应用中, 需要考虑在满足系统硬件存储能力的条件下, 同时也要满足实际工业的估计精度要求, 选择合适的正则化系数 λ_1 .

现取 $\lambda_1 = 0.025$ 来展示 100 个测量样本的实际估计效果. 从图 5 可以看出, 基于 PSCN 估计模型的输出和实际测量值基本吻合, 因此, 可以将其用于井下供给风量的准确估计. 此外, 基于 PSCN 估计模型的 R^2 和 E_{VAR} 分别为 0.9976 和 0.9978, 这也说明了 PSCN 软测量模型具有较高的预测精度. 因此, 本文所提 PSCN 算法不仅可用于实际工业过程运行指标的估计, 而且在缓解工业系统存储压力方面具有较大的优势.

表 10 基于 PSCN 的井下供给风量软测量模型对正则化系数 λ_1 的灵敏度

算法	λ_1	RMSE		模型规模
		Mean	Std	
PSCN	0.085	0.0251	0.0019	28.66
	0.065	0.0241	0.0016	32.48
	0.045	0.0226	0.0018	35.66
	0.025	0.0216	0.0021	42.52
	0.005	0.0208	0.0019	62.08

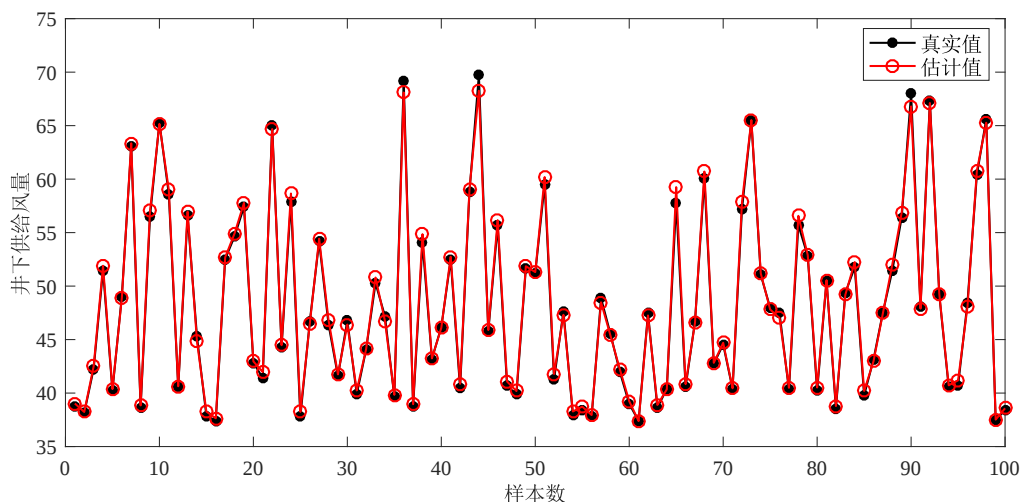


图5 PSCN 井下供给风量估计模型预测结果

4 结论

本文基于 L_1 范数正则化提出了一种简约随机配置网络, 即 PSCN 方法. PSCN 将 L_1 范数添加到网络增量构建目标函数中, 建立了一个不等式约束来产生高质量的新增节点, 并采用 ADMM 算法对非光滑和不可微的网络输出权重目标函数进行优化求解, 从而实现了 SCN 软测量模型的稀疏学习. 最后, 基于基准数据集和实际工业数据的实验仿真表明: 与 IRWNN、SCN 和 RSC-II 方法相比, 本文所提的 PSCN 方法能有效抑制输出权重的幅值, 简化模型的结构, 同时保持了较高的泛化性能. 为进一步提升算法性能, 未来的研究内容是将深度神经网络的分层架构引入到所提 PSCN 算法中; 针对 PSCN 模型的计算损耗, 提出一种高效的输出权值确定方法, 以减少所提模型的计算复杂度.

参考文献(References)

- [1] Brásio A S R, Romanenko A, Fernandes N C P, Santos L O. First principle modeling and predictive control of a continuous biodiesel plant[J]. Journal of Process Control, 2016, 47: 11-21.
- [2] Meyer E J, Craig I K. The development of dynamic models for a dense medium separation circuit in coal beneficiation[J]. Minerals Engineering, 2010, 23(10): 791-805.
- [3] 陈龙, 刘全利, 王霖青, 赵珺, 王伟. 基于数据的流程工业生产过程指标预测方法综述[J]. 自动化学报, 2017, 43(6): 944-954.
(Chen Long, Liu Quan-Li, Wang Lin-Qing, Zhao Jun, Wang Wei. Data-driven prediction on performance indicators in process industry: a survey[J]. Acta Automatica Sinica, 2017, 43(6): 944-954.)
- [4] Xie S W, Xie Y F, Huang T W, Gui W H, Yang C H. Generalized predictive control for industrial processes

based on neuron adaptive splitting and merging RBF neural network[J]. IEEE Transactions on Industrial Electronics, 2019, 66(2): 1192-1202.

- [5] Wang B, Li Z C, Dai Z W, Lawrence N, Yan X F. Data-driven mode identification and unsupervised fault detection for nonlinear multimode processes[J]. IEEE Transactions on Industrial Informatics, 2020, 16(6): 3651-3661.
- [6] 陈龙, 韩中洋, 赵珺, 王伟. 数据驱动的综合能源系统运行优化方法研究综述[J]. 控制与决策, 2021, 36(2): 283-294.
(Chen Long, Han Zhong-Yang, Zhao Jun, Wang Wei. Review of research of data-driven methods on operational optimization of integrated energy systems[J]. Control and Decision, 2021, 36(2): 283-294.)
- [7] Zhou P, Lv Y B, Wang H, Chai T Y. Data-driven robust RVFLNs modeling of a blast furnace iron-making process using cauchy distribution weighted M-estimation[J]. IEEE Transactions on Industrial Electronics, 2017, 64(9): 7141-7151.
- [8] 周平, 刘进进. 基于 Stacking 的高炉铁水质量区间预测[J]. 控制与决策, 2021, 36(2): 335-344.
(Zhou Ping, Liu Jin-Jin. Prediction of blast furnace hot metal quality interval based on stacking[J]. Control and Decision, 2021, 36(2): 335-344.)
- [9] Zhou P, Li W, Wang H, Li M, Chai T. Robust online sequential RVFLNs for data modeling of dynamic time-varying systems with application of an ironmaking blast furnace[J]. IEEE Transactions on Cybernetics, 2020, 50(11): 4783-4795.
- [10] Gorban A N, Tyukin I Y, Prokhorov D V, Sofeikov K I. Approximation with random bases: Pro et contra[J]. Information Sciences, 2016, 364-365: 129-145.
- [11] Pao Y H, Park G H, Sobajic D J. Learning and generalization characteristics of the random vector functional-link net[J]. Neurocomputing, 1994, 6(2): 163-180.

- [12] Igel'nik B, Pao Y H. Stochastic choice of basis functions in adaptive function approximation and the functional-link net[J]. IEEE Transactions on Neural Networks, 1995, 6(6): 1320-1329.
- [13] Bisoi R, Dash P K, Mishra S P. Modes decomposition method in fusion with robust random vector functional link network for crude oil price forecasting[J]. Applied Soft Computing, 2019, 80: 475-493.
- [14] Zhang Y, Wu J, Cai Z, et al. An unsupervised parameter learning model for RVFL neural network[J]. Neural Networks, 2019, 112: 85-97.
- [15] Sahani M, Dash P K. FPGA-based online power quality disturbances monitoring using reduced-sample HHT and class-specific weighted RVFLN[J]. IEEE Transactions on Industrial Informatics, 2019, 15(8): 4614-4623.
- [16] Sahani M, Dash P K. Deep convolutional stack autoencoder of process adaptive VMD data with robust multikernel RVFLN for power quality events recognition[J]. IEEE Transactions on Instrumentation and Measurement, 2021, 70: 1-12.
- [17] Chen C L P, Liu Z. Broad learning system: An effective and efficient incremental learning system without the need for deep architecture[J]. IEEE Transactions on Neural Networks and Learning Systems, 2018, 29(1): 10-24.
- [18] Chen C L P, Liu Z, Feng S. Universal approximation capability of broad learning system and its structural variations[J]. IEEE Transactions on Neural Networks and Learning Systems, 2019, 30(4): 1191-1204.
- [19] Zhao H, Zheng J, Deng W, et al. Semi-supervised broad learning system based on manifold regularization and broad network[J]. IEEE Transactions on Circuits and Systems I: Regular Papers, 2020, 67(3): 983-994.
- [20] Li M, Wang D. Insights into randomized algorithms for neural networks: Practical issues and common pitfalls[J]. Information Sciences, 2017, 382-383: 170-178.
- [21] Wang D, Li M. Stochastic configuration networks: Fundamentals and algorithms[J]. IEEE Transactions on Cybernetics, 2017, 47(10): 3466-3479.
- [22] Zhu X, Feng X, Wang W, Jia X, He R. A further study on the inequality constraints in stochastic configuration networks[J]. Information Sciences, 2019, 487: 77-83.
- [23] Wang D, Cui C. Stochastic configuration networks ensemble with heterogeneous features for large-scale data analytics[J]. Information Sciences, 2017, 417: 55-71.
- [24] Sheng Z Y, Zeng Z Q, Qu H Q, Zhang Y. Optical fiber intrusion signal recognition method based on TSVD-SCN[J]. Optical Fiber Technology, 2019, 48: 270-277.
- [25] 代伟, 李德鹏, 杨春雨, 马小平. 一种随机配置网络的模型与数据混合并行学习方法[J]. 自动化学报, 2019, 45(x): 1-11.
(Dai Wei, Li De-Peng, Yang Chun-Yu, Ma Xiao-Ping. A model and data hybrid parallel learning method for stochastic configuration networks[J]. Acta Automatica Sinica, 2019, 45(x): 1-11.)
- [26] Wang Q, Dai W, Ma X, Shang Z. Driving amount based stochastic configuration network for industrial process modeling[J]. Neurocomputing, 2020, 394: 61-69.
- [27] 王前进, 杨春雨, 马小平, 张春富, 彭思敏. 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 Qian-Jin, Yang Chun-Yu, Ma Xiao-Ping, Zhang Chun-Fu, Peng Si-Min. Underground airflow quantity modeling based on SCN[J]. Acta Automatica Sinica, 2021, 47(8): 1963-1975.)
- [28] Lu J, Ding J. Mixed-distribution based robust stochastic configuration networks for prediction interval construction[J]. IEEE Transactions on Industrial Informatics, 2020, 16(8): 5099-5109.
- [29] Lu J, Ding J, Dai X, Chai T. Ensemble stochastic configuration networks for estimating prediction intervals: a simultaneous robust training algorithm and its application[J]. IEEE Transactions on Neural Networks and Learning Systems, 2020, 31(12): 5426-5440.
- [30] Xie J, Zhou P. Robust stochastic configuration network multi-output modeling of molten iron quality in blast furnace ironmaking[J]. Neurocomputing, 2020, 387: 139-149.
- [31] Lu J, Ding J. Construction of prediction intervals for carbon residual of crude oil based on deep stochastic configuration networks[J]. Information Sciences, 2019, 486: 119-132.
- [32] 田丹, 张国山, 谢英红. 具有融合罚约束的低秩结构化稀疏表示目标跟踪算法[J]. 控制与决策, 2019, 34(11): 2479-2484.
(Tian Dan, Zhang Guo-Shan, Xie Ying-Hong. Object tracking via low-rank and structural sparse representation with fused penalty constraint[J]. Control and Decision, 2019, 34(11): 2479-2484.)
- [33] 朱建勇, 黄鑫, 杨辉, 聂飞平. 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.
(Zhu Jian-Yong, Huang Xin, Yang Hui, Nie Fei-Ping. Selection method for froth image characters based on sparse neural network[J]. Control and Decision, 2021, 36(7): 1627-1636.)
- [34] Rao C R, Mitra S K. Generalized inverse of matrices and its applications[M]. New York: Wiley, 1971.
- [35] 韩敏, 姜涛, 冯守渤. 基于VMD循环随机跳跃状态网络的时间序列长期预测[J]. 控制与决策, 2020, 35(9): 2175-2181.
(Han Min, Jiang Tao, Feng Shou-BO. Long-term prediction of time series based on VMD cyclic reservoir with random jumps network[J]. Control and Decision, 2020, 35(9): 2175-2181.)
- [36] Boyd S, Parikh N, Chu E, et al. Distributed optimization and statistical learning via the alternating direction

- method of multipliers[J]. Foundations & Trends in Machine Learning, 2010, 3(1): 1-122.
- [37] 洪金华, 张荣, 郭立君. 基于 $L_{1/2}$ 正则化的三维人体姿态重构[J]. 自动化学报, 2018, 44(6): 1086-1095. (Hong Jin-Hua, Zhang Rong, Guo Li-Jun. 3D human body pose reconstruction via $L_{1/2}$ regularization[J]. Acta Automatica Sinica, 2018, 44(6): 1086-1095.)
- [38] Fan J, Li R. Variable selection via nonconcave penalized likelihood and its oracle properties[J]. Journal of the American statistical Association, 2001, 96(456): 1348-1360.
- [39] 王前进, 代伟, 杨春雨, 马小平. 煤矿主通风机切换系统建模与分析. 煤炭学报[J], 2018, 43(S2): 606-614. (Wang Qian-Jin, Dai Wei, Yang Chun-Yu, Ma Xiao-Ping. Modeling and analysis of coal mine main fan switchover system[J]. Journal of China Coal Society, 2018, 43(S2): 606-614.)
- [40] Wang Q J, Dai W, Ma X P, et al. Multiple models and neural networks based adaptive PID decoupling control of mine main fan switchover system[J]. IET Control Theory & Applications, 2018, 12(4): 446-455.
- [41] Wang Q J, Ma X P, Yang C Y, et al. Modeling and control of mine main fan switchover system[J]. ISA Transactions, 2019, 85: 189-199.

作者简介

王前进(1987—), 讲师, 博士, 从事数据驱动建模与控制、机器学习算法等研究, E-mail: wangqianjinabc@163.com;

代伟(1984—), 教授, 博士, 从事工业系统智能感知与运行控制、工业大数据分析等研究, E-mail: daiwei_neu@126.com;

陆群(1988—), 讲师, 博士, 从事移动机器人视觉伺服、非线性系统等研究, E-mail: landgod1@126.com;

辅小荣(1968—), 副教授, 从事复杂工业过程建模、运行优化与控制等研究, E-mail: fuxiaorong@126.com;

马小平(1961—), 教授, 博士生导师, 从事过程控制、网络控制系统、故障诊断等研究, E-mail: xpma@cumt.edu.cn.