



基于混合分布加权M估计和自适应正则化的随机配置网络

严爱军, 郝晨

引用本文:

严爱军, 郝晨. 基于混合分布加权M估计和自适应正则化的随机配置网络[J]. *智能系统学报*, 2025, 20(6): 1392–1403.

YAN Aijun, HAO Chen. Stochastic configuration networks based on mixed distribution weighted M-estimation and adaptive regularization[J]. *CAAI Transactions on Intelligent Systems*, 2025, 20(6): 1392–1403.

在线阅读 View online: <https://dx.doi.org/10.11992/tis.202501023>

您可能感兴趣的其他文章

一种卷积神经网络集成的多样性度量方法

Diversity measuring method of a convolutional neural network ensemble

智能系统学报. 2021, 16(6): 1030–1038 <https://dx.doi.org/10.11992/tis.202011023>

结合卷积特征提取和路径语义的知识推理

Knowledge-based inference on convolutional feature extraction and path semantics

智能系统学报. 2021, 16(4): 729–738 <https://dx.doi.org/10.11992/tis.202008007>

一类分数阶神经网络的自适应 ∞ 同步

Adaptive ∞ synchronization of a class of fractional-order neural networks

智能系统学报. 2019, 14(2): 239–245 <https://dx.doi.org/10.11992/tis.201709045>

基于GABP-KF的WSN数据漂移盲校准算法

GABP-KF-based blind calibration algorithm of data drift in wireless sensor networks

智能系统学报. 2019, 14(2): 254–262 <https://dx.doi.org/10.11992/tis.201712003>

融合朴素贝叶斯方法的复杂网络链路预测

Link prediction in complex networks with syncretic naive Bayes methods

智能系统学报. 2019, 14(1): 99–107 <https://dx.doi.org/10.11992/tis.201810025>

基于快速密度聚类的RBF神经网络设计

Construction of RBF neural networks via fast density clustering

智能系统学报. 2018, 13(3): 331–338 <https://dx.doi.org/10.11992/tis.201702014>

DOI: 10.11992/tis.202501023

网络出版地址: <https://link.cnki.net/urlid/23.1538.tp.20251016.1819.002>

基于混合分布加权 M 估计和自适应正则化的随机配置网络

严爱军^{1,2,3}, 郝晨^{1,2}

(1. 北京工业大学信息科学技术学院, 北京 100124; 2. 数字社区教育部工程研究中心, 北京 100124; 3. 城市轨道交通北京实验室, 北京 100124)

摘要: 为提升随机配置网络 (stochastic configuration networks, SCNs) 的鲁棒性和泛化性, 提出了一种基于混合分布加权 M 估计和自适应正则化的 SCN 建模方法。采用高斯和柯西混合分布加权 M 估计获得训练样本的惩罚权重, 根据训练数据对模型的贡献度评估其输出权重, 以增强模型鲁棒性; 根据建模残差变化情况和隐节点数分配合适的 L_2 正则化参数, 以保证模型具有较好的泛化性。通过 4 个标准数据集和城市固废焚烧过程的历史数据对该方法的性能进行实验测试。实验结果表明, 基于本文所提方法构建的参数模型在鲁棒性和泛化性上相对于其他对比方法具有优势, 从而拓宽了 SCN 的应用范围。

关键词: 神经网络; 随机配置网络; 参数预测; M 估计; 混合分布; 正则化; 模型鲁棒性; 模型泛化性; 城市固废焚烧

中图分类号: TP183 文献标志码: A 文章编号: 1673-4785(2025)06-1392-12

中文引用格式: 严爱军, 郝晨. 基于混合分布加权 M 估计和自适应正则化的随机配置网络 [J]. 智能系统学报, 2025, 20(6): 1392-1403.

英文引用格式: YAN Aijun, HAO Chen. Stochastic configuration networks based on mixed distribution weighted M-estimation and adaptive regularization[J]. CAAI transactions on intelligent systems, 2025, 20(6): 1392-1403.

Stochastic configuration networks based on mixed distribution weighted M-estimation and adaptive regularization

YAN Aijun^{1,2,3}, HAO Chen^{1,2}

(1. School of Information Science and Technology, Beijing University of Technology, Beijing 100124, China; 2. Engineering Research Center of Digital Community, Ministry of Education, Beijing 100124, China; 3. Beijing Laboratory for Urban Mass Transit, Beijing 100124, China)

Abstract: To enhance the robustness and generalization capability of stochastic configuration networks (SCNs), this paper proposes a novel SCN modeling method based on mixed-distribution weighted M-estimation and adaptive regularization. First, a Gaussian-Cauchy mixed-distribution weighted M-estimation is employed to determine the penalty weights of training samples, and the output weights of the model are evaluated according to the contribution of each training instance, thereby improving model robustness. Second, an adaptive L_2 regularization parameter is assigned based on the variation of modeling residuals and the number of hidden nodes, ensuring that the model maintains good generalization. Finally, the effectiveness of the proposed method is empirically validated using four benchmark datasets and historical data from municipal solid waste incineration processes. Experimental results demonstrate that the parameter models developed using the proposed method outperform comparative methods in both robustness and generalization, thereby broadening the application scope of SCNs.

Keywords: neural network; stochastic configuration networks; parameter prediction; M-estimation; mixture distribution; L_2 regularization; model robustness; model generalization; municipal solid waste incineration

收稿日期: 2025-01-15. 网络出版日期: 2025-10-17.

基金项目: 国家自然科学基金项目 (62373017, 62073006); 北京市自然科学基金项目 (4212032).

通信作者: 严爱军. E-mail: yanaijun@bjut.edu.cn.

随着人工智能技术的兴起, 以人工神经网络 (artificial neural network, ANN) 为代表的机器学习算法受到了学术界的广泛关注, 在图像识别^[1]、工

业过程预测建模^[2]等领域得到了广泛的应用。结合随机学习技术^[3]和随机神经网络^[4], Pao 等^[5]提出了一种随机向量函数链 (random vector functional link, RVFL) 网络, 其对输入层至隐含层之间的权重和偏差采用随机化方式分配, 极大地简化了网络训练过程。然而, 文献 [6] 指出, RVFL 随机分配输入权重和偏置使得其不具备万能逼近性。为了解决上述问题, Wang 等^[7]提出了依赖训练数据集的随机参数选择监督机制, 以隐含层神经元递增的方式自适应快速构建具有万能逼近性质的随机配置网络 (stochastic configuration networks, SCNs) 模型, 由于其学习效率高, 逼近能力强等优点, 已被广泛应用于研磨粒度估计^[8]、污水处理^[9]等工业过程参数建模领域。

然而, 在 SCN 的建模过程中, 存在两方面的因素影响模型的鲁棒性和泛化性: 1) 训练样本的惩罚权重估计。在工业过程的数据采集过程中, 传感器易受设备故障、人为干扰等因素影响, 采集数据存在不同程度的噪声和异常值^[10], 从而降低模型的准确性。针对该问题, M 估计是增强模型鲁棒性常用的方法。文献 [11-12] 将高斯核函数分别引入 RVFL 和 SCN, 采用 M 估计和核密度估计相结合的方法为每个训练样本设置输出权值, 以减少噪声对模型的负面影响; 文献 [13] 将高斯加权 M 估计引入 RVFL, 取得了较好的效果; 文献 [14] 采用最大相关熵准则评估训练样本的惩罚权重, 削弱了训练过程中异常值噪声数据产生的影响; 文献 [15] 采用基于 Huber 和基于四分位距的 M 估计设置惩罚权值, 并分别给出了鲁棒建模方法; 文献 [16] 采用柯西分布加权 M 估计为训练样本设置输出权值, 使得模型鲁棒性能有所提升。然而, 上述文献采用的权函数形式较为单一, 在面对复杂数据或异常值占比较高的数据时, 可能不具备较好的鲁棒性; 2) 正则化策略。随着隐节点数量的增加, 模型会变得更复杂, 可能导致模型在有限的数据中学习噪声特征, 从而影响模型的泛化性能。对此, 采用一定的正则化策略, 有助于提高模型的泛化性能。常用的正则化方法有 L_2 、弹性网^[17]、Dropout^[18] 正则化等。文献 [19] 提出了一种简约 SCN, 将 L_1 正则化引入到 SCN, 并建立了一个新的不等式约束条件, 有效简化了模型结构且增强了模型的泛化性; 文献 [20] 提出了一种基于弹性网络的分布式正则化 SCN, 并通过实验表明了该方法在准确性方面的相对优势; 文献 [21] 结合 L_2 正则化技术提出了一种新的 SCN 算法, 降低了模型结构风险, 提升了

模型精度和泛化能力。然而, 上述方法采用的正则化参数均为固定值, 虽然可以保证训练过程的稳定, 但固定的正则化参数无法适应训练过程中模型性能的变化, 限制了模型的泛化性能。

基于上述研究, 针对训练样本的惩罚权重估计和正则化策略问题, 本文提出一种基于混合分布加权 M 估计和自适应正则化 (mixed distribution weighted M estimation and adaptive regularization, MM-AR) 的随机配置网络建模方法, 主要工作包括: 1) 根据混合分布能够更加灵活适应复杂数据的特性, 采用混合分布加权 M 估计评估训练样本的惩罚权重; 2) 针对 SCN 构建隐含层神经元过程中模型性能的未知性和动态变化, 构造了一种参数自适应的 L_2 正则化项到 SCN 模型中, 可以根据模型在训练过程中的表现和模型的当前节点数适当分配正则化参数, 增强灵活性的同时提高了模型的泛化性。最后通过实验证明了该方法的有效性。

1 SCN 算法及问题分析

1.1 SCN 基本算法

SCN 是一种增量式快速建模方法, 其网络结构由一个隐含层节点开始, 采用监督机制随机配置隐节点的输入权重和偏置, 逐渐增加隐节点个数, 直到达到预设条件, 完成网络配置^[7]。SCN 的构建过程可以表示如下:

对于目标函数 $f: \mathbf{R}^d \rightarrow \mathbf{R}^m$, 给定数据集 $D = (\mathbf{X}, \mathbf{Y}) = \{(\mathbf{x}_i, \mathbf{y}_i) \in \mathbf{R}^d \times \mathbf{R}^m\}$, 假设已经构造了 $L-1$ 个隐节点, 则此时网络输出可以表示为

$$f_{L-1}(\mathbf{X}; \boldsymbol{\beta}) = \sum_{j=1}^{L-1} \beta_j g_j(\mathbf{w}_j^T \mathbf{X} + b_j)$$

式中: β_j 、 $g_j(\cdot)$ 、 w_j 和 b_j 分别表示隐含层第 j 个节点的输出权重、激活函数、输入权重和偏置; f_{L-1} 为当前模型输出。

计算当前网络残差向量:

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

如果 $\|\mathbf{e}_{L-1}\|^2$ 未达到预设误差 ε 或模型未达到最大节点数 L_{\max} , 则增加第 L 个隐节点。

$$\mathbf{h}_L = [g_L(\mathbf{w}_L^T \mathbf{x}_1 + b_L), g_L(\mathbf{w}_L^T \mathbf{x}_2 + b_L), \dots, g_L(\mathbf{w}_L^T \mathbf{x}_N + b_L)]^T$$

$$\mathbf{h}_L \in \mathbf{R}^N \quad (1)$$

$$\xi_{L,q} = \frac{(\mathbf{e}_{L-1,q}^T \cdot \mathbf{h}_L)^2}{\mathbf{h}_L^T \mathbf{h}_L} - (1-r-\mu_L) \|\mathbf{e}_{L-1,q}\|^2 \quad (2)$$

根据式 (2) 监督机制确定其输入权重和偏置。式中: $q=1, 2, \dots, m$; \mathbf{h}_L 表示隐节点 L 的输出; \mathbf{w}_L 和 b_L 分别表示节点 L 的候选参数, $r \in (0, 1)$;

$\{\mu_L\}$ 表示非负实数序列, 其中 $\mu_L \leq 1-r, \lim_{L \rightarrow +\infty} \mu_L = 0$;
满足 $\xi_L = \sum_{q=1}^{\infty} \xi_{L,q} \geq 0$ 最大值的候选节点参数作为
第 L 个节点参数。

计算隐含层输出权重:

$$\beta = \arg \min_{\beta} \|\mathbf{H}\beta - \mathbf{Y}\|^2 = \mathbf{H}^{\dagger} \mathbf{Y}$$

式中: \mathbf{H}^{\dagger} 表示 \mathbf{H} 的摩尔-彭若斯 (Moore-Penrose) 广义逆, $\mathbf{H} = [\mathbf{h}_1 \ \mathbf{h}_2 \ \cdots \ \mathbf{h}_L]$

最后计算网络输出结果 \mathbf{f} :

$$\mathbf{f} = \mathbf{H}\beta$$

1.2 问题分析

SCN 在面对复杂数据或具有异常值的数据样本时的鲁棒性和泛化性可能会受到影响, 尽管传统的 M 估计、正则化等算法能够改善模型性能, 但仍然具有一定局限性:

1) 传统 M 估计的权函数对异常值敏感度较高。在处理大规模数据时, Huber、Tukey、Gaussian 等常用的权函数由于形式较为单一, 缺乏足够的灵活性适应数据变化, 限制了模型捕捉复杂数据关系的能力。在数据中包含异常值时, 可能无法较好拟合数据分布情况, 模型的鲁棒性会受到较大影响。因此, 选择合适的权函数以减少异常值对模型性能的影响仍需进一步研究。

2) 固定的正则化系数无法适应所有训练阶段的需求。传统的 L_2 正则化通过在损失函数中增加惩罚项, 对模型的泛化性具有一定改善。然而, 固定正则化系数不根据模型在训练过程中的表现进行调整, 无法适应训练不同阶段的变化, 限制了正则化对模型泛化性能的改善效果。因此, 如何在训练过程中根据模型性能分配合适的正则化参数仍需进一步研究。

2 MM-AR-SCN 建模方法

2.1 M 估计基本算法

M 估计^[22]是一种常用的抗差估计方法, 其核心思想是通过最小化一个特定的目标函数来估计模型参数, 能够通过训练样本对模型的贡献评估训练样本的惩罚权重, 从而提高模型的鲁棒性。

在标准 SCN 中, 输出权重 β 和实际输出 \mathbf{Y} 可以表示为

$$\beta = [\beta_1 \ \beta_2 \ \cdots \ \beta_L]^T, \mathbf{Y} = [y_1 \ y_2 \ \cdots \ y_N]^T$$

由第 1 节可得, β 通过最小化获得:

$$\beta = \arg \min_{\beta} \sum_{i=1}^N (\mathbf{H}_i \beta - y_i)^2 \quad (3)$$

式中: 损失函数 $J(r_i)$ 和残差 r_i 分别表示为

$$\begin{aligned} J(r_i) &= r_i^2 \\ r_i(\beta) &= \mathbf{H}_i \beta - y_i \end{aligned} \quad (4)$$

在 M 估计中, 引入新的残差目标损失函数 $\rho(r_i)$, 此时式 (3) 变为

$$\hat{\beta} = \arg \min_{\beta} \sum_{i=1}^N \rho(\mathbf{H}_i \beta - y_i) = \arg \min_{\beta} \sum_{i=1}^N \rho(r_i(\beta))$$

为减弱数据样本中异常值的影响, 引入稳健尺度估计量 $\hat{\sigma}$ 将残差标准化, 此时 $\hat{\beta}$ 可表示为

$$\hat{\beta} = \arg \min_{\beta} \sum_{i=1}^N \rho\left(\frac{r_i(\beta)}{\hat{\sigma}}\right) \quad (5)$$

式 (5) 中引入的稳健尺度估计量提供了一个对异常值不敏感的尺度度量, 通过重新缩放残差, 使异常值的影响被限制在一定范围内。具体地, $\hat{\sigma}$ 使用中位数绝对偏差 (median absolute deviation, MAD) 定义, 即

$$\hat{\sigma} = \frac{r_{\text{MAD}}}{0.6745}, r_{\text{MAD}} = \text{median}(|r_i - \text{median}(r_i)|)$$

MAD 是统计离散度的度量, 是一种稳健的度量方法, 适用于包含异常值或极端值的数据集^[23]。MAD 还具有非参数性, 这使得它适用于各种类型的数据分布。

$$\partial \sum_{i=1}^N \rho\left(\frac{\mathbf{H}_i \beta - y_i}{\hat{\sigma}}\right) / \partial \beta = 0, \text{ 可以得到:}$$

$$\sum_{i=1}^N \rho'\left(\frac{\mathbf{H}_i \beta - y_i}{\hat{\sigma}}\right) \mathbf{H}_i^T = 0 \quad (6)$$

为实现模型对数据样本的差异化处理, 使得每个数据样本的权重根据其残差大小, 即对模型的贡献度进行分配, 定义权重因子:

$$p(r_i) = \rho'\left(\frac{r_i}{\hat{\sigma}}\right) / r_i \quad (7)$$

则式 (6) 可以重写为

$$\sum_{i=1}^N p\left(\frac{\mathbf{H}_i \beta - y_i}{\hat{\sigma}}\right) \times (\mathbf{H}_i \beta - y_i) \mathbf{H}_i^T = 0 \quad (8)$$

整理成矩阵形式:

$$\mathbf{H}^T \mathbf{P} \mathbf{H} \beta = \mathbf{H}^T \mathbf{P} \mathbf{Y} \quad (9)$$

式中: \mathbf{P} 是惩罚权重矩阵, $\mathbf{P} = \text{diag}\{p_i\}, i=1, 2, \dots, N; N$ 为数据样本数量, $p_i(\cdot)$ 为权函数, 表示第 i 个样本的惩罚权重。

2.2 混合分布加权 M 估计算法

式 (9) 中的惩罚权重矩阵 \mathbf{P} 由每个样本点的惩罚权重 $p_i(\cdot)$ 组成, 其中, $p_i(\cdot)$ 为权函数, 其选择是 M 估计的重要环节。一个合适的权函数不仅影响模型的鲁棒性, 而且会影响模型的计算效率。常用的权函数包括 Huber 权函数、Hampel 权函数和 Tukey 双权法权函数^[24] 等。

考虑到高斯分布的对称性和有限的方差, 使其非常适合描述在均值附近集中的数据。在数据中异常值比例较低的情况下, 高斯分布更加稳定可靠。而柯西分布具有无限的方差和重尾特性, 这使得柯西分布在面对异常值比例较高的数据样本时更加稳健。因此, 采用混合分布加权函数作为 M 估计的目标函数, 根据异常值的比例适当地调整权重系数, 可以使得模型对于异常值比例较高还是较低的数据样本, 都具有较好的鲁棒性。

基于以上分析, 本文拟采用高斯函数和柯西分布概率密度函数的加权函数作为 M 估计的权重函数。

$$p(a) = \tau N(a; \mu_1, \sigma_1) + (1 - \tau) C(a; \mu_2, \sigma_2)$$

式中: τ 为权重, 其值受异常值比例的影响; $N(a; \mu_1, \sigma_1)$ 为均值 μ_1 、标准差 σ_1 的高斯函数; $C(a; \mu_2, \sigma_2)$ 为位置参数 μ_2 、尺度参数 σ_2 的柯西分布的概率密度函数。高斯分布的概率密度函数的表达式为

$$N(a, \mu_1, \sigma_1) = \frac{1}{\sqrt{2\pi}\sigma_1} \exp\left(-\frac{(a - \mu_1)^2}{2\sigma_1^2}\right)$$

柯西分布的概率密度函数的表达式为

$$C(a; \mu_2, \sigma_2) = \frac{1}{\pi\sigma_2 \left[1 + \left(\frac{a - \mu_2}{\sigma_2}\right)^2\right]}$$

在高斯分布和柯西分布中, μ_1 和 μ_2 分别为对应分布的中心, 也是定义分布峰值的参数; σ_1 和 σ_2 分别决定了对应分布的分散程度, 其值越大, 数据点越分散, 分布越宽。为简化表述, 将高斯分布的均值 μ_1 和柯西分布的位置参数 μ_2 统一定义为 μ , 将高斯分布的标准差 σ_1 和柯西分布的尺度参数 σ_2 统一定义为 σ 。

基于上述分析得知, 如何根据残差的分布确定参数是构造加权函数的重要内容:

1) μ 的确定: 由于加权的目的是协调异常值带来的影响, 因此 μ 应取标准化残差 $r_i/\hat{\sigma}$ 的中间值, 即

$$\mu = \text{median}(r_i/\hat{\sigma})$$

2) σ 的确定: σ 决定了分布图像的形状特性, 它的值应根据建模误差分布的统计特性来确定。若建模误差分布较窄, 则 σ 的值应较大; 否则, σ 应较小。因此, σ 应取标准化残差的标准差的倒数。即

$$\sigma = 1 / \sqrt{\sum_{i=1}^N \left\| r_i/\hat{\sigma} - \sum_{i=1}^N (r_i/\hat{\sigma})/N \right\|^2 / N}$$

2.3 自适应正则化算法

正则化是防止模型过拟合, 增强模型泛化能

力的常用方法, 其中正则化参数的选择是提升模型性能的关键。根据文献 [25] 中提出的“早停策略”, 本节提出了一种参数自适应正则化的算法, 其基本思想是根据模型训练过程中残差的变化情况以及模型当前节点数分配合适的正则化参数, 算法如下。

在式 (4) 所示的损失函数中引入 L_2 正则化项, 输出权重的计算公式变为

$$\hat{\beta} = \arg \min_{\beta} (\|H\beta - Y\|_2^2 + \alpha \|\beta\|_2^2) \quad (10)$$

根据 2.1 节, 在式 (10) 中引入 M 估计后转化为

$$\hat{\beta} = \arg \min_{\beta} \left(\sum_{i=1}^N \rho(H_i\beta - y_i) + \alpha \|\beta\|_2^2 \right)$$

令 $\frac{\partial \left(\sum_{i=1}^N \rho(H_i\beta - y_i) + \alpha \|\beta\|_2^2 \right)}{\partial \beta} = 0$, 根据式 (6)~

(9), 可得:

$$\beta = (H^T P H + \alpha I)^{-1} (H^T P Y)$$

式中 α 为正则化参数。在 SCN 中, 模型的性能受模型复杂度等多方面的影响。令正则化参数 $\alpha_L = cf(e_L, e_{L-1}, L)$, 其中 c 为常数, L 为当前节点数, e_L 为增加第 L 个节点时的残差, $f(e_L, e_{L-1}, L)$ 表示和 e_L 、 e_{L-1} 、 L 有关的函数。

定义一个中间量 δ_L :

$$\delta_L = \frac{\|e_L\|_2^2}{\|e_{L-1}\|_2^2}$$

当 δ_L 很小时, 说明模型残差变化率较大, 此时不需要过多的正则化干预; 当 δ_L 保持过大时, 说明模型性能在一定训练周期内没有显著提升, 继续训练有可能会使模型泛化性降低, 此时需要适当增大正则化参数。此外, 随着隐含层节点 L 的增加, 模型会更复杂, 进而更容易面临过拟合的风险^[26], 因此正则化参数应随着 L 的增加而适当增大。

基于以上分析, 并根据测试, 本文拟采用函数 $f(e_L, e_{L-1}, L) = f(\delta_L, L) = \frac{1}{1 - \delta_L} + L$ 作为正则化参数的自适应函数。为方便调整参数, 引入两个非负常数项 c_1 、 c_2 , 将正则化参数重新定义为

$$\alpha_L = \frac{c_1}{1 - \delta_L} + c_2 L$$

在训练后期, 残差变化率会越来越小, 这会导致正则化参数 α_L 过大, 从而影响模型的整体性能, 经过测试, 采用对数函数将 α_L 控制在较小范围内。

$$\alpha_L = c_1 \ln \left(\frac{1}{1 - \delta_L} + 1 \right) + c_2 L$$

最后,采用交替优化策略^[7]交替更新输出权重 w 和惩罚权重 β ,交替优化的迭代表达式:

$$w_j^{(k+1)} = \tau \frac{1}{\sqrt{2\pi}\sigma} \exp\left(\frac{(e_j^{(k)} - \mu)^2}{2\sigma^2}\right) + (1-\tau) \frac{1}{\pi\sigma \left[1 + \left(\frac{e_j^{(k)} - \mu}{\sigma}\right)^2\right]} \quad (11)$$

$$\beta^{(k+1)} = (H^T P^{(k+1)} H + \alpha I)^{\dagger} H^T P^{(k+1)} Y \quad (12)$$

式中: k 表示迭代次数;当惩罚权重 β 的变化率满足如下不等式时:

$$\max_j \left| \frac{\beta_j^{(k)} - \beta_j^{(k+1)}}{\beta_j^{(k+1)}} \right| < \varepsilon \quad (13)$$

停止迭代同时模型训练结束。在式 (13) 中, $j=1,2,\dots$; ε 表示趋于 0 的正数,此处取 10^{-3} 。

2.4 复杂度分析

2.4.1 时间复杂度分析

在上述混合分布加权 M 估计算法以及自适应正则化算法中,对每个训练样本分别计算残差和权重以及根据模型的残差变化调整正则化参数,该部分时间复杂度为

$$O(N \times T) + O(L \times T) \quad (14)$$

式中: N 是训练样本数量, T 是迭代次数, L 是隐含层节点数量。

在网络训练中,模型训练的时间复杂度为

$$O(N \times L \times T) \quad (15)$$

交替优化更新输出权值和惩罚权重的时间复杂度为

$$O(N \times L \times A) \quad (16)$$

式中 A 是交替优化的迭代次数。

结合式 (14)~(16),可得 MM-AR-SCN 的总时间复杂度为

$$O(N \times T) + O(L \times T) + O(N \times L \times T) + O(N \times L \times A)$$

式中:由于 T 和 A 通常为常数,可以将它们合并为一个常数因子并简化,因此总时间复杂度可以简化为

$$O(N \times L)$$

综上所述,MM-AR-SCN 算法尽管引入了额外的计算步骤,但这些步骤的复杂度与训练过程中的迭代次数 T 和交替优化次数 A 在一定程度上成正比,因此本文提出的算法的时间复杂度仍然保持在与传统 SCN 一致的量级,即 $O(N \times L)$ 。

2.4.2 空间复杂度分析

在上述混合分布加权 M 估计以及自适应正则化算法中,模型存储每个训练样本的残差和权重以及存储正则化参数的空间复杂度为

$$O(N) + O(L) \quad (17)$$

由于每个隐节点有 $d+1$ 个参数(输入权重和偏置),总共有 L 个隐节点,因此该部分空间复杂度为

$$O(L \times (d+1)) \quad (18)$$

隐含层输出矩阵的大小为 $N \times L$,因此该部分空间复杂度为

$$O(N \times L) \quad (19)$$

输出权重的大小为 $L \times m$,因此该部分空间复杂度为

$$O(L \times m) \quad (20)$$

式中 m 为输出层的维度。

结合式 (17)~(20),可得 MM-AR-SCN 的空间复杂度为

$$O(N) + O(L) + O(L \times (d+1)) + O(N \times L) + O(L \times m)$$

与传统 SCN 相比,MM-AR-SCN 的空间复杂度仅增加了存储残差权重以及正则化参数的空间需求,但这些额外的空间需求较小,对空间复杂度的增加没有显著影响。

2.5 算法伪代码

基于上述关于算法实现的介绍,MM-AR-SCN 算法的伪代码为:

1) 输入数据集 $D = \{X, Y\}$, 其中 $X = [x_1 \ x_2 \ \dots \ x_N] \in \mathbf{R}^{N \times d}$, $x_i = [x_{i1} \ x_{i2} \ \dots \ x_{id}]^T \in \mathbf{R}^d$, $Y = [y_1 \ y_2 \ \dots \ y_N] \in \mathbf{R}^{N \times 1}$; 隐含层最大节点数 L_{\max} , 最大配置次数 T_{\max} ; 容忍误差 ε ; 隐含层参数搜索范围; 最大迭代次数 A_{\max} ; $\mathcal{Y} = [\lambda_{\min} : \Delta\lambda : \lambda_{\max}]$

2) 参数初始化: 令 $e_0 = [y_1 \ y_2 \ \dots \ y_N]^T$; 正则化参数 $\alpha=0$; $\Omega = []$; $W = []$; 随机参数 γ ;

3) While $A \leq A_{\max}$ and $e_0 > \varepsilon$ Do

4) While $L \leq L_{\max}$ and $\|e_0\|_F > \varepsilon$ Do

5) For $\lambda \in \mathcal{Y}$ Do

6) For $k = 1, 2, \dots, T_{\max}$ Do

7) 从区间 $[-\lambda, \lambda]$ 中随机分配 w_j 和 b_j ; 根据式 (1)~(2) 计算 $h_L^*, \xi_{L,q}$, 并设置 $\mu_L = (1-r)/(L+1)$;

8) if $\min(\xi_{L,1}, \xi_{L,2}, \dots, \xi_{L,m}) \geq 0$

9) 将 w_j 和 b_j 存入 W 中, 将 $\xi_L = \sum_{q=1}^m \xi_{L,q}$ 存入 Ω 中;

10) Else 返回 6)

11) End if (步骤 8)

12) End for (步骤 6)

13) if W 非空 Do

14) 在 Ω 中找到使 ξ_L 最大的 w_j^*, b_j^* , 设置 $H_L = [h_1^* \ h_2^* \ \dots \ h_L^*]$; 跳转到 18);

15) End if (步骤 13)

16) End for (步骤 5)

- 17) 根据式 (10) 计算 β , $e_L = H_L \beta - Y$;
- 18) 更新 $e_0 = e_L$, $L=L+1$;
- 19) End While (步骤 4)
- 20) 根据式 (11) 和 (12) 交替优化惩罚权重矩阵 W 和输出权值 β ; $A = A+1$;
- 21) End While (步骤 3)
- 22) 输出 MM-AR-SCN 模型。

3 实验结果与分析

为方便表述, 将基于核密度估计 (kernel density estimation, KDE) 的 SCN 记作 KDE-SCN^[12], 将基于最大相关熵准则 (maximum correntropy criterion, MCC) 的 SCN 记作 MCC-SCN^[14], 将基于 L_1 正则化的简约 (parsimonious)SCN 记作 PSCN^[19], 将基于弹性网 (elastic net, EN) 的 SCN 记作 EN-SCN^[20], 并设计了对比实验: 将本文所提方法与 SCN^[7] 以及上述改进 SCN 算法分别进行鲁棒性和泛化性对比; 具体数据信息见表 1。

表 1 数据集信息
Table 1 Data set information

编号	数据集名称	样本总数	输入/输出维度
1	stock	950	9/1
2	treasury	1049	15/1
3	Concrete	1030	8/1
4	Compactiv	1000	21/1
5	MSWI历史数据	1000	61/1

在鲁棒性对比实验中, 随机选择 60% 的样本作为训练集, 20% 的样本作为测试集, 20% 的样本作为验证集。在泛化性对比实验中, 随机选择

表 2 参数设置
Table 2 Parameter setting

数据集	MM- AR-SCN				EN-SCN			PSCN
	τ_0	$\tau_{outlier}$	c_1	c_2	α	λ_{EN}	ρ	λ_P
stock	0.85	0.74, 0.68, 0.27, 0.20	0.001	0.005	0.005	0.001	0.001	0.05
treasury	0.90	0.85, 0.50, 0.15, 0.05	0.001	0.005	0.005	0.001	0.001	0.05
Concrete	0.68	0.60, 0.55, 0.30, 0.11	0.001	0.005	0.005	0.001	0.001	0.05
Compactiv	0.72	0.82, 0.65, 0.30, 0.15	0.001	0.005	0.005	0.001	0.001	0.05
MSWI历史数据	0.50	0.45, 0.30, 0.25, 0.12	0.005	0.010	0.010	0.005	0.005	0.10

3.2 标准数据集实验

3.2.1 鲁棒性对比

采用 4 个 KEEL(kedige of evlovtionary learning) 标准数据集对上述方法的鲁棒性能进行验证, 在训练样本中随机选取比例 $\zeta = \{5\%, 10\%, 20\%, 30\%\}$ 的样本, 并将这些样本的输出进行异常化处理, 具体处理方式

40% 的样本作为训练集, 40% 的样本作为测试集, 20% 的样本作为验证集。所有的数据均归一化至 $[0,1]$ 区间。所有的仿真实验所用的计算机 CPU 型号为 Intel(R) Core (TM) i5-8300H, 内存为 4 GB。

3.1 实验准备

采用均方根误差 e_{RMSE} (root mean square error, RMSE) 和平均绝对误差 e_{MAE} (mean absolute error, MAE) 作为评价指标:

$$e_{RMSE} = \sqrt{\sum_{i=1}^N (y_i - \hat{y}_i)^2 / N} \quad (21)$$

$$e_{MAE} = \left(\sum_{i=1}^N |y_i - \hat{y}_i| \right) / N \quad (22)$$

式中: y_i 表示样本输出的实际值, \hat{y}_i 表示模型的预测值, N 表示样本数量。为了避免随机性对实验结果的影响, 所有实验重复执行 50 次, 在鲁棒性对比实验中, 给出均值 (mean) 和标准差 (std) 以验证模型的鲁棒性能和稳定性; 在泛化性对比实验中, 给出模型在训练集和测试集上的误差对比以验证模型的泛化性能。

所有模型的激活函数均为 Sigmoid 函数; 参数设置为: 所有方法最大隐节点数量 $L_{max}=100$, 最大配置次数 $T_{max}=100$, 训练预期误差 $\epsilon=10^{-3}$, 交替方向乘法 (alternating direction method of multipliers, ADMM) 最大迭代次数 $T_{ADMM}=20$, 交替优化方法中最大循环次数 $A_{max}=10$, 隐含层参数搜索范围 $[-\lambda, +\lambda]=[-5, +5]$; 其余参数见表 2 所示, 其中, α 、 λ_{EN} 、 λ_P 均为正则化参数, ρ 为拉格朗日系数; 所有超参数均由验证集得出。

$$y_{i,outlier} = y_i + \text{sign} \times \text{rand}(0, 1)$$

式中: y_i 和 $y_{i,outlier}$ 分别表示输出数据的真实值和处理后的异常值, $\text{rand}(0,1)$ 表示 $(0,1)$ 之间的随机数, 为使数据分布更加不平衡, 正偏差异常值 ($\text{sign}=1$) 和负偏差异常值 ($\text{sign}=-1$) 的比例设置为 2:1, 观察不同方法在处理包含不同比例异常样本数据的效果。

实验结果如表 3 和表 4 所示, 通过表 3 和表 4 得到, 随着样本中异常值比例的提升, 本文所提

算法的误差相较于其他鲁棒算法均处于较低水平, 这表明了本文所提方法的有效性。

表 3 标准数据集下各种方法的 RMSE 比较
Table 3 RMSE comparison of various methods under standard datasets

编号	$\zeta/\%$	mean, std				
		SCN	KDE-SCN	MCC-SCN	MM-SCN	MM-AR-SCN
Data1	5	0.055 1, 0.002 1	0.042 2, 0.002 8	0.041 9, 0.002 6	0.035 3, 0.001 5	0.035 1 , 0.001 2
	10	0.082 7, 0.004 7	0.045 3, 0.006 5	0.044 8, 0.003 3	0.037 9, 0.004 3	0.037 5 , 0.003 2
	20	0.091 5, 0.005 1	0.045 5, 0.001 5	0.043 2, 0.001 9	0.036 5, 0.001 8	0.035 1 , 0.001 8
	30	0.109 7, 0.004 8	0.052 7, 0.003 8	0.051 2, 0.003 6	0.041 1 , 0.002 7	0.042 2, 0.002 5
Data2	5	0.036 2, 0.004 6	0.012 4, 0.000 8	0.012 1, 0.001 1	0.010 9, 0.006 4	0.010 6 , 0.007 1
	10	0.040 3, 0.007 1	0.014 6, 0.000 7	0.014 1, 0.001 1	0.012 9, 0.000 9	0.012 7 , 0.001 0
	20	0.037 1, 0.003 1	0.014 3, 0.000 6	0.014 9, 0.001 3	0.012 2, 0.001 1	0.012 1 , 0.001 2
	30	0.048 8, 0.006 1	0.021 1, 0.000 13	0.020 7, 0.001 2	0.015 4, 0.003 1	0.015 0 , 0.002 6
Data3	5	0.095 2, 0.003 3	0.073 6, 0.004 5	0.072 7, 0.004 2	0.061 5, 0.003 7	0.060 7 , 0.002 8
	10	0.102 4, 0.003 8	0.075 4, 0.004 1	0.079 1, 0.005 5	0.063 6, 0.004 3	0.062 0 , 0.004 2
	20	0.117 4, 0.003 8	0.072 1, 0.003 3	0.072 0, 0.003 6	0.063 1, 0.003 6	0.061 1 , 0.002 8
	30	0.132 4, 0.132 4	0.087 9, 0.107 9	0.086 6, 0.106 6	0.074 4, 0.104 4	0.070 3 , 0.106 3
Data4	5	0.086 7, 0.009 8	0.069 9, 0.007 2	0.068 3, 0.013 5	0.059 9, 0.008 2	0.057 3 , 0.007 5
	10	0.092 4, 0.010 1	0.079 3, 0.008 5	0.095 6, 0.014 4	0.067 9, 0.014 5	0.066 9 , 0.010 3
	20	0.119 1, 0.014 9	0.081 1, 0.010 1	0.079 4, 0.010 7	0.065 9, 0.012 9	0.063 5 , 0.009 9
	30	0.139 6, 0.015 6	0.093 4, 0.025 5	0.092 6, 0.033 1	0.084 7, 0.026 9	0.083 3 , 0.010 2

注: 加黑数字为最优结果。

表 4 标准数据集下各种方法的 MAE 比较
Table 4 MAE comparison of various methods under standard datasets

编号	$\zeta/\%$	mean, std				
		SCN	KDE-SCN	MCC-SCN	MM-SCN	MM-AR-SCN
Data1	5	0.043 9, 0.001 6	0.033 8, 0.001 9	0.032 5, 0.001 3	0.026 8, 0.001 1	0.026 6 , 0.000 9
	10	0.059 9, 0.002 8	0.034 9, 0.002 2	0.036 5, 0.001 6	0.027 3, 0.001 7	0.026 9 , 0.001 7
	20	0.069 2, 0.003 5	0.034 7, 0.001 3	0.032 9, 0.001 4	0.027 1, 0.001 4	0.025 9 , 0.001 5
	30	0.086 6, 0.003 6	0.038 5, 0.002 5	0.037 7, 0.002 3	0.030 1, 0.001 6	0.029 9 , 0.001 4
Data2	5	0.021 4, 0.001 6	0.007 8, 0.000 3	0.007 5, 0.000 4	0.005 4, 0.001 5	0.005 1 , 0.001 3
	10	0.031 1, 0.003 1	0.008 4, 0.000 5	0.007 8, 0.000 4	0.006 0, 0.000 5	0.005 6 , 0.000 5
	20	0.030 8, 0.002 3	0.008 2, 0.000 4	0.007 5, 0.000 4	0.006 9, 0.000 5	0.006 7 , 0.000 6
	30	0.047 1, 0.004 2	0.016 3, 0.007	0.016 1, 0.000 8	0.010 9, 0.001 1	0.010 6 , 0.001 0
Data3	5	0.082 5, 0.002 3	0.069 7, 0.002 9	0.069 0, 0.002 9	0.064 0, 0.002 5	0.062 9 , 0.002 4
	10	0.084 5, 0.002 7	0.069 1, 0.002 3	0.065 4, 0.002 7	0.062 3, 0.002 9	0.061 6 , 0.002 4
	20	0.091 9, 0.002 6	0.063 9, 0.002 1	0.067 8, 0.002 4	0.054 9, 0.002 5	0.054 5 , 0.002 1
	30	0.101 7, 0.003 4	0.085 4, 0.003 2	0.073 7, 0.003 4	0.065 0, 0.003 8	0.064 5 , 0.004 2
Data4	5	0.046 2, 0.003 2	0.037 3, 0.002 7	0.036 9, 0.003 1	0.036 4 , 0.002 7	0.039 5, 0.002 9
	10	0.052 8, 0.003 4	0.039 2, 0.003 2	0.034 0, 0.003 1	0.029 3, 0.003 4	0.029 0 , 0.003 6
	20	0.074 4, 0.005 9	0.051 8, 0.007 2	0.049 8, 0.003 7	0.040 8, 0.003 6	0.039 5 , 0.005 1
	30	0.097 2, 0.006 0	0.054 8, 0.005 8	0.055 1, 0.004 3	0.041 6, 0.005 1	0.038 7 , 0.003 1

注: 加黑数字为最优结果。

3.2.2 泛化性对比

采用 4 个 KEEL 标准数据集对上述方法的泛化性能进行对比, 观察不同方法在训练集和测试

集上的效果, 结果如表 5 所示。通过表 5 可知, 本文所提方法在训练集和测试集上均具有较低的误差水平, 验证了该方法良好的泛化性。

表 5 不同方法评价指标对比
Table 5 Comparison of evaluation indexes of different methods

数据集	评价指标	SCN	PSCN	EN-SCN	AR-SCN	MM-AR-SCN	
Data1	训练集	RMSE	0.0295	0.0290	0.0249	0.0257	0.0259
		MAE	0.0257	0.0233	0.0204	0.0179	0.0173
	测试集	RMSE	0.0454	0.0428	0.0395	0.0325	0.0342
		MAE	0.0359	0.0328	0.0268	0.0259	0.0271
Data2	训练集	RMSE	0.0121	0.0138	0.0081	0.0069	0.0081
		MAE	0.0045	0.0086	0.0056	0.0045	0.0042
	测试集	RMSE	0.0189	0.0149	0.0121	0.0129	0.0132
		MAE	0.0078	0.0092	0.0075	0.0078	0.0076
Data3	训练集	RMSE	0.0726	0.0697	0.0647	0.0606	0.0588
		MAE	0.0571	0.0459	0.0502	0.0472	0.0453
	测试集	RMSE	0.0864	0.0832	0.0853	0.0872	0.0715
		MAE	0.0662	0.0625	0.0632	0.0621	0.0602
Data4	训练集	RMSE	0.0343	0.0387	0.0259	0.0244	0.0276
		MAE	0.0183	0.0182	0.0195	0.0184	0.0177
	测试集	RMSE	0.0624	0.0583	0.0469	0.0493	0.0479
		MAE	0.0402	0.0379	0.0239	0.0298	0.0224

注: 加黑数字为最优结果。

图 1 为 MM-AR-SCN 算法训练过程中正则化参数随节点增加的变化情况与不同节点下测试集的误差变化曲线图, 最大节点数分别为 100、100、200、200。由图 1 可以得出, 在训练后期, 模型在训练集上的误差变化趋于稳定且模型节点数逐渐增多的时候, 为避免模型出现过拟合现象, 正则化参数也在适当增加, 此时模型在测试集上的误差仍然处于较低水平, 这说明了本文所改进正则化算法的有效性。

3.3 城市固废焚烧烟气含氧量预测实验

随着经济的快速发展和城市化的持续发展, 全球城市固体废物 (municipal solid waste, MSW) 的产量一直在逐步增长^[27], 城市固废焚烧 (municipal solid waste incineration, MSWI) 是一种复杂的、非线性的物化反应过程^[28]。在 MSWI 过程中, 烟气含氧量是过氧空气系数的表征量, 能够在一定程度上表征燃烧状态^[29], 是与焚烧效率和污染物排放密切相关的重要参数。如果烟气含氧量过低, 则表明不完全燃烧热损失增加, 燃烧效率下降的同时还会产生大量有毒有害气体; 如果

烟气含氧量过高, 会增加燃料型 NO_x 等污染物的排放^[30]。因此, 建立准确的烟气含氧量预测模型对于提高燃烧效率、减少污染物排放、实现该过程的平稳运行具有重要意义。

实验数据来自 2021 年北京市某固废焚烧厂的 1000 条历史数据, 数据信息如表 1 所示, 实验参数设置如表 2 所示。采用式 (21) 所示的均方根误差和式 (22) 所示的平均绝对误差作为评价指标。

3.3.1 鲁棒性对比

本节为不同方法在城市固废焚烧过程烟气含氧量预测的鲁棒性对比, 异常值引入方法与 3.2 节相同。观察不同方法在不同异常值比例下的误差情况。

通过表 6 和表 7 可以得出, 当异常值比例 $\zeta=20\%$ 时, MM-AR-SCN 的 RMSE 和 MAE 相比 SCN 分别降低了 26.04% 和 32.95%, 这表明将 M 估计引入 SCN 算法能够在一定程度上改善模型的鲁棒性; 此外, 本文所提方法在不同异常值比例下均保持较好的鲁棒性, 表明了该算法的有效性。

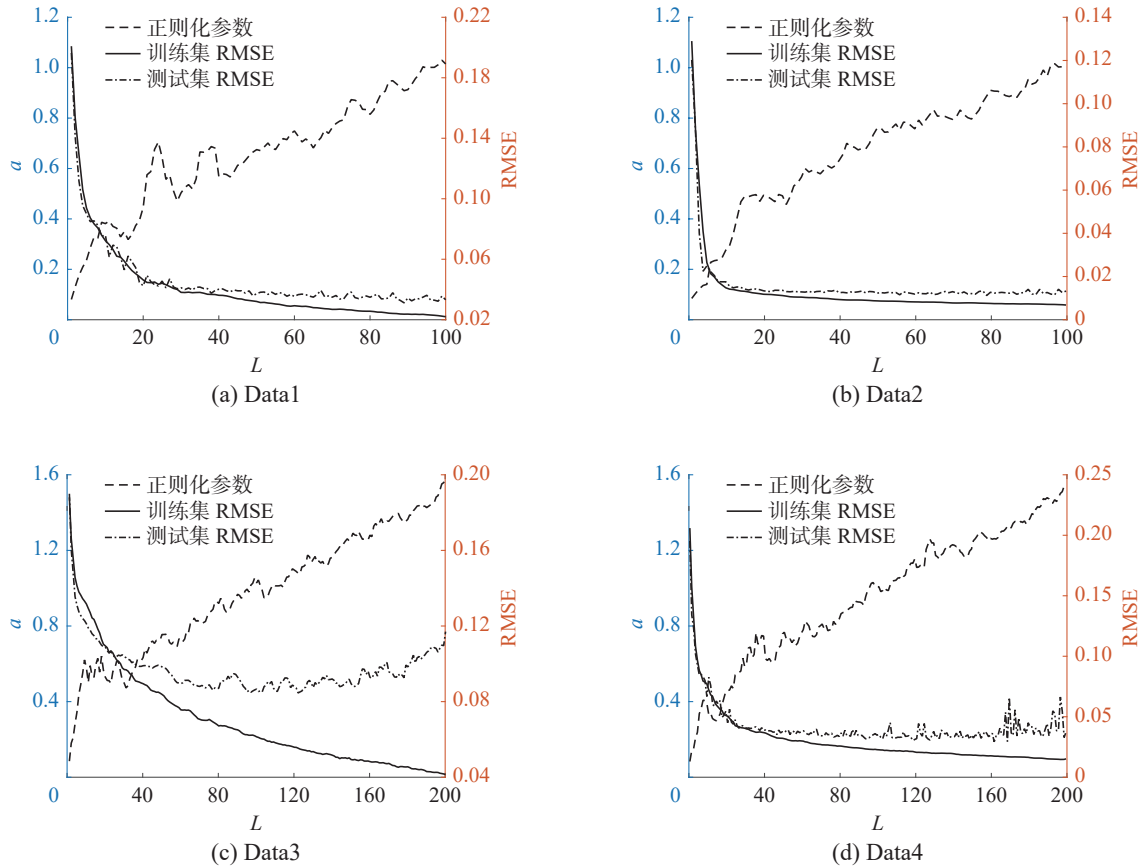


图 1 不同数据集上正则化参数 α 和 RMSE 随节点数 L 增加的变化曲线

Fig. 1 Variation curves of regularization parameters α and RMSE on different datasets as the number of nodes increases L

表 6 MSWI 历史数据下各种方法的 RMSE 比较
Table 6 RMSE comparison of various methods under MSWI historical data

$\zeta/\%$	mean, std				
	SCN	KDE-SCN	MCC-SCN	MM-SCN	MM-AR-SCN
5	1.2213, 0.0238	1.1308, 0.0211	1.1189, 0.0248	1.0475, 0.0180	1.0322 , 0.0271
10	1.4306, 0.0237	1.2273, 0.0256	1.2344, 0.0221	1.0931, 0.0348	1.0680 , 0.0432
20	1.7408, 0.0231	1.4294, 0.0185	1.4068, 0.0253	1.2958, 0.0324	1.2874 , 0.0274
30	1.7849, 0.0240	1.4225, 0.0363	1.4275, 0.0364	1.3171, 0.0317	1.3043 , 0.0373

注: 加黑数字为最优结果。

表 7 MSWI 历史数据下各种方法的 MAE 比较
Table 7 MAE comparison of various methods under MSWI historical data

$\zeta/\%$	mean, std				
	SCN	KDE-SCN	MCC-SCN	MM-SCN	MM-AR-SCN
5	0.9695, 0.0107	0.7674, 0.0143	0.7651, 0.0102	0.7031, 0.0156	0.6989 , 0.0136
10	1.1799, 0.0139	0.7767, 0.0175	0.7805, 0.0137	0.6971 , 0.0206	0.7137, 0.0260
20	1.4371, 0.0159	1.1460, 0.0142	1.1329, 0.0208	0.9794, 0.0231	0.9636 , 0.0209
30	1.6874, 0.0173	1.3240, 0.0188	1.3206, 0.0176	1.1845 , 0.0252	1.1868, 0.0201

注: 加黑数字为最优结果。

3.3.2 泛化性对比

本节为不同方法在城市固废焚烧过程烟气含氧量预测上的泛化性对比, 观察不同方法在训练集和测试集的性能。

表 8 给出了不同正则化方法在训练集和测试集上的 RMSE 和 MAE 对比; 图 2 为 MSWI 历史

数据正则化参数随训练过程变化的曲线图。

通过表 8 可以看出, 所有方法在训练集上的效果相似, 然而本文提出的方法在测试集上效果更好, 这说明将自适应正则化引入 SCN 在一定程度上改善了模型的泛化性, 提升了模型在未知数据上的预测精度。

表 8 不同方法评价指标对比
Table 8 Comparison of evaluation indexes of different methods

数据集	评价指标	SCN	PSCN	EN-SCN	AR-SCN	MM-AR-SCN	
MSWI 历史数据	训练集	RMSE	0.5872	0.6333	0.5647	0.4134	0.4331
		MAE	0.3927	0.4811	0.4295	0.3736	0.4059
	测试集	RMSE	0.8424	0.7026	0.6757	0.7272	0.6654
		MAE	0.5711	0.5256	0.4957	0.4753	0.4689

注: 加黑数字为最优结果。

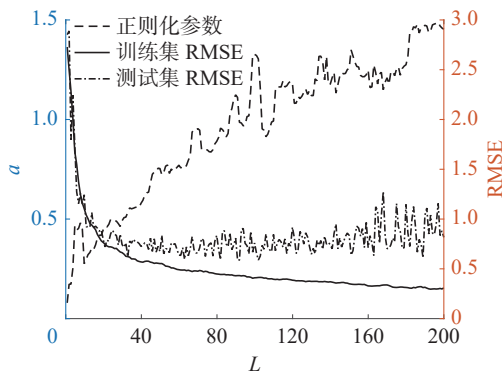


图 2 MSWI 历史数据上正则化参数 a 和 RMSE 随节点数 L 增加的变化曲线

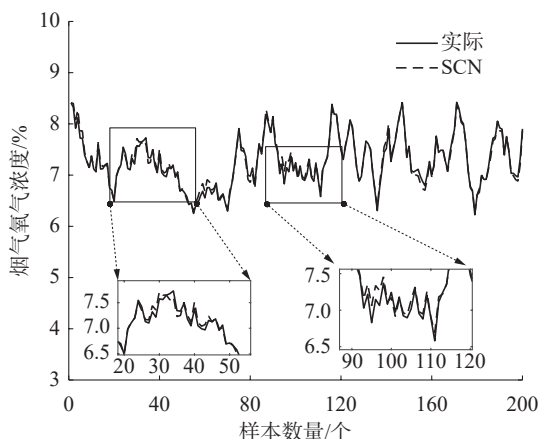
Fig. 2 Variation curves of regularization parameters a and RMSE on the historical data of MSWI as the number of nodes increases L

表 9 不同模型运行时间对比
Table 9 Comparison of running time of different models

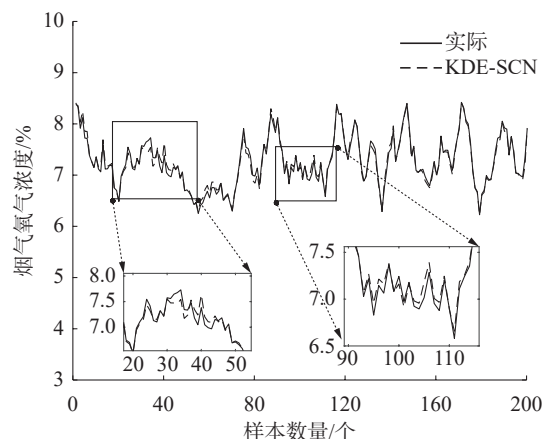
方法	SCN	MM-SCN	AR-SCN	MM-AR-SCN
时间/s	0.2057	0.2132	0.2112	0.2150

表 9 给出了不同模型在处理烟气含氧量数据时运行时间对比结果。结合 2.4 节对模型的复杂度分析可得, MM-AR-SCN 由于加入了混合分布加权 M 估计与自适应正则化算法, 其额外计算量的增加使得 MM-AR-SCN 模型相对于 SCN 而言, 在建模过程中所需的时间略微增加。

为了更直观地显示 MM-AR-SCN 的拟合效果, 给出如图 3 所示部分典型建模方法的烟气含氧量预测拟合曲线图。通过图 3 可以看出, 相较于其他方法, 本文提出的方法拟合效果更好, 通过烟气含氧量的变化, 准确观察焚烧效率, 有助于城市固废焚烧过程运行优化控制的实现。



(a) SCN 烟气含氧量拟合曲线



(b) KDE-SCN 烟气含氧量拟合曲线

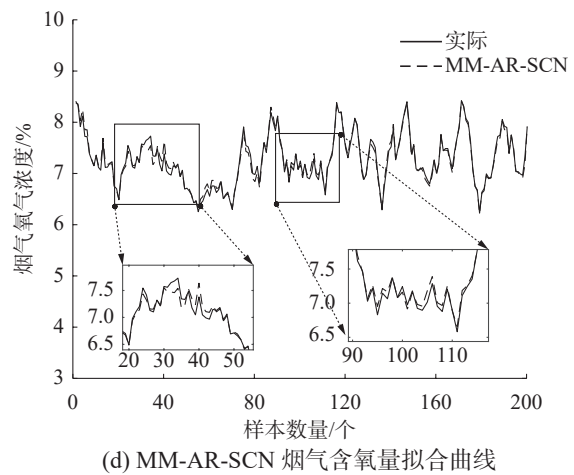
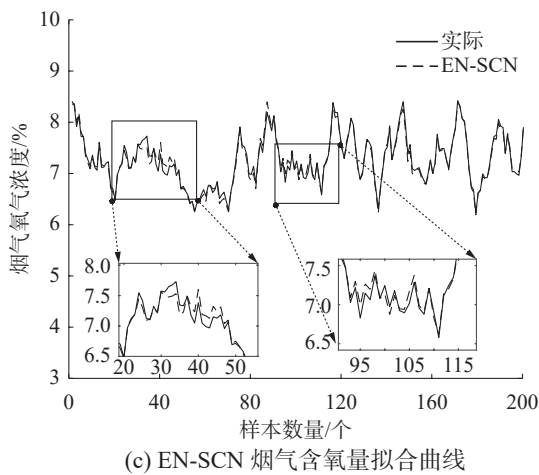


图 3 不同方法的烟气含氧量拟合曲线

Fig. 3 Fitting curves of flue gas oxygen content by different methods

4 结束语

为提升 SCN 的鲁棒性和泛化性, 本文提出了一种采用混合分布加权 M 估计和自适应正则化的 SCN 建模方法 (MM-AR-SCN), 并通过实验说明了该方法的有效性。

1) 采用混合分布加权 M 估计评估输出权重。针对传统 SCN 中标准最小二乘法对异常值较为敏感的问题, 采用混合分布加权 M 估计, 根据每个样本对模型的贡献度为它们分别分配不同的输出权重, 实现样本数据惩罚权重的合理分配, 有效提高了 SCN 在包含异常数据的参数预测任务中的鲁棒性;

2) 提出了一种改进的 L_2 正则化算法。该算法使正则化参数根据模型训练过程中残差的变化情况和节点数自适应分配, 弥补了 L_2 正则化缺乏灵活性, 难以适应训练各阶段需求的不足, 有效缓解了训练过程中可能出现的过拟合问题, 在一定程度上改善了模型的泛化能力。

实验结果表明, 本文所提出的 MM-AR-SCN 方法相较于其他方法, 在 4 个标准数据集和城市固废焚烧过程烟气含氧量预测实验上均表现出了较好的性能, 表明了该算法具有更良好的鲁棒性和泛化性。

然而, 在实际工业过程中, 建模数据中异常值的比例难以得知, 从而导致权重参数很难确定。此外, 所提建模方法中的一些超参数仍需手动调整。因此, 对于数据样本异常值的分析处理和模型参数的自动优化是未来研究的重点。

参考文献:

[1] MAO Chengsheng, YAO Liang, LUO Yuan. ImageGCN: multi-relational image graph convolutional networks for

disease identification with chest X-rays[J]. *IEEE transactions on medical imaging*, 2022, 41(8): 1990–2003.

[2] QIAO Junfei, SUN Zijian, MENG Xi. A comprehensively improved interval type-2 fuzzy neural network for NO_x emissions prediction in MSWI process[J]. *IEEE transactions on industrial informatics*, 2023, 19(11): 11286–11297.

[3] SCARDAPANE S, WANG Dianhui. Randomness in neural networks: an overview[J]. *Wiley interdisciplinary reviews: data mining and knowledge discovery*, 2017, 7(2): e1200.

[4] 乔俊飞, 李凡军, 杨翠丽. 随机神经网络研究现状与展望[J]. *智能系统学报*, 2016, 11(6): 758–767.

QIAO Junfei, LI Fanjun, YANG Cuili. Review and prospect on neural networks with random weights[J]. *CAAI transactions on intelligent systems*, 2016, 11(6): 758–767.

[5] PAO Y H, TAKEFUJI Y. Functional-link net computing: theory, system architecture, and functionalities[J]. *Computer*, 1992, 25(5): 76–79.

[6] LI Ming, WANG Dianhui. Insights into randomized algorithms for neural networks: practical issues and common pitfalls[J]. *Information sciences*, 2017, 382: 170–178.

[7] WANG Dianhui, LI Ming. Stochastic configuration networks: fundamentals and algorithms[J]. *IEEE transactions on cybernetics*, 2017, 47(10): 3466–3479.

[8] DAI Wei, ZHOU Xinyu, LI Depeng, et al. Hybrid parallel stochastic configuration networks for industrial data analytics[J]. *IEEE transactions on industrial informatics*, 2021, 18(4): 2331–2341.

[9] LI Kang, YANG Cuili, WANG Wei, et al. An improved stochastic configuration network for concentration prediction in wastewater treatment process[J]. *Information sciences*, 2023, 622: 148–160.

[10] EL-MELEGY M T. Model-wise and point-wise random sample consensus for robust regression and outlier detection[J]. *Neural networks*, 2014, 59: 23–35.

[11] DAI Wei, CHEN Qixin, CHU Fei, et al. Robust regularized random vector functional link network and its indus-

- trial application[J]. *IEEE access*, 2017, 5: 16162–16172.
- [12] WANG Dianhui, LI Ming. Robust stochastic configuration networks with kernel density estimation for uncertain data regression[J]. *Information sciences*, 2017, 412: 210–222.
- [13] 李温鹏, 周平. 高炉铁水质量鲁棒正则化随机神经网络建模[J]. *自动化学报*, 2020, 46(4): 721–733.
LI Wenpeng, ZHOU Ping. Robust regularized RVFLNs modeling of molten iron quality in blast furnace ironmaking[J]. *Acta automatica sinica*, 2020, 46(4): 721–733.
- [14] LI Ming, HUANG Changqin, WANG Dianhui. Robust stochastic configuration networks with maximum correntropy criterion for uncertain data regression[J]. *Information sciences*, 2019, 473: 73–86.
- [15] DAI Wei, LI Depeng, CHEN Qixin, et al. Data driven particle size estimation of hematite grinding process using stochastic configuration network with robust technique[J]. *Journal of Central South University*, 2019, 26(1): 43–62.
- [16] ZHOU Ping, LYU Youbin, WANG Hong, et al. 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.
- [17] ZOU Hui, 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.
- [18] SRIVASTAVA N, HINTON G, KRIZHEVSKY A, et al. Dropout: a simple way to prevent neural networks from overfitting[J]. *Journal of machine learning research*, 2014, 15: 1929–1958.
- [19] 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 Qianjin, DAI Wei, LU Qun, et al. A sparse learning method for SCN soft measurement model[J]. *Control and decision*, 2022, 37(12): 3171–3182.
- [20] ZHAO Lijie, ZOU Shida, HUANG Mingzhong, et al. Distributed regularized stochastic configuration networks via the elastic net[J]. *Neural computing and applications*, 2021, 33(8): 3281–3297.
- [21] 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 Lijie, ZOU Shida, GUO Shuo, et al. Ball mill load condition recognition model based on regularized stochastic configuration networks[J]. *Control engineering of China*, 2020, 27(1): 1–7.
- [22] FAN Jun, YAN Ailing, XIU Naihua. Asymptotic properties for M-estimators in linear models with dependent random errors[J]. *Journal of statistical planning and inference*, 2014, 148: 49–66.
- [23] ROUSSEEUW P J, CROUX C. Alternatives to the Median absolute deviation[J]. *Journal of the American statistical association*, 1993, 88(424): 1273–1283.
- [24] LÓPEZ-RUBIO E, PALOMO E J, DOMÍNGUEZ E. Robust self-organization with M-estimators[J]. *Neurocomputing*, 2015, 151: 408–423.
- [25] AGLIARI E, ALEMANNI F, AQUARO M, et al. Regularization, early-stopping and dreaming: a hopfield-like setup to address generalization and overfitting[J]. *Neural networks*, 2024, 177: 106389.
- [26] 张成龙, 丁世飞, 郭丽丽, 等. 随机配置网络研究进展[J]. *软件学报*, 2024, 35(5): 2379–2399.
ZHANG Chenglong, DING Shifei, GUO Lili, et al. Research progress on stochastic configuration network[J]. *Journal of software*, 2024, 35(5): 2379–2399.
- [27] WANG Tianzheng, TANG Jian, XIA Heng, et al. Data-driven multi-objective intelligent optimal control of municipal solid waste incineration process[J]. *Engineering applications of artificial intelligence*, 2024, 137: 109157.
- [28] MENG Xi, TANG Jian, QIAO Junfei. NOx emissions prediction with a brain-inspired modular neural network in municipal solid waste incineration processes[J]. *IEEE transactions on industrial informatics*, 2022, 18(7): 4622–4631.
- [29] 汤健, 夏恒, 余文, 等. 城市固废焚烧过程智能优化控制研究现状与展望[J]. *自动化学报*, 2023, 49(10): 2019–2059.
TANG Jian, XIA Heng, YU Wen, et al. Research status and prospects of intelligent optimization control for municipal solid waste incineration process[J]. *Acta automatica sinica*, 2023, 49(10): 2019–2059.
- [30] 孙剑, 蒙西, 乔俊飞. 数据驱动的城市固废焚烧过程烟气含氧量预测控制[J]. *控制理论与应用*, 2024, 41(3): 484–495.
SUN Jian, MENG Xi, QIAO Junfei. Data-driven predictive control of oxygen content in flue gas for municipal solid waste incineration process[J]. *Control theory & applications*, 2024, 41(3): 484–495.

作者简介:



严爱军, 博士生导师, 教授, 主要研究方向为复杂过程建模与智能优化控制方法。发表学术论文 100 余篇。
E-mail: yanaijun@bjut.edu.cn。



郝晨, 硕士研究生, 主要研究方向为复杂过程建模与智能优化控制方法。E-mail: haochen@emails.bjut.edu.cn。