

# 基于融合相关性的协同分摊噪声软测量建模 \*

梁楠<sup>1</sup> 高世伟<sup>2</sup> 张伟<sup>1</sup> 田添<sup>3</sup> 薛瑞争<sup>4</sup>

(1. 河南省科学院应用物理研究所有限公司 郑州 450000; 2. 西北师范大学计算机科学与工程学院 兰州 730000;  
3. 河南建筑职业技术学院 郑州 450064; 4. 河南国健医疗设备有限公司 商丘 476002)

**摘要:** 基于数据驱动的软测量建模方法在流程工业中有着广泛的应用。流程工业中,辅助数据常常会受到异构、杂糅的噪声的污染,且工业数据中线性相关与非线性相关共存,而噪声问题和不合理的关系表达均会严重影响软测量模型的预测结果。在协同分摊噪声算法的基础上提出一种基于融合相关性的协同分摊噪声算法进行软测量建模。首先,采用融合了关注线性相关性的 Pearson 系数和关注非线性相关性的 Spearman 系数的融合相关性系数优化协同分摊噪声算法,使协同分摊噪声算法中数据可信度计算更合理,更符合工业数据中线性相关与非线性相关共存的情况。然后,结合卷积神经网络(convolutional neural networks, CNN)搭建软测量模型。在脱丁烷塔数据集上进行多降噪方法、多模型和多回归方法的交叉组合实验,结果表明,该优化后的降噪算法较基础的协同分摊噪声算法、小波变换降噪、降噪自编码器有着较强的降噪能力;所搭建的软测量模型有着较优的预测精度及较小的预测误差,其中决定系数(r-square,  $R^2$ )指标和均方误差(mean squared error, MSE)分别为 0.971 6 和 0.001 1。

**关键词:** 数据驱动建模; 软测量; 融合相关性; 协同分摊噪声; HCAN-CNN

中图分类号: TP274; TN06 文献标识码: A 国家标准学科分类代码: 510.40

## Fused correlation-based collaborative shared noise soft-sensing modeling

Liang Nan<sup>1</sup> Gao Shiwei<sup>2</sup> Zhang Wei<sup>1</sup> Tian Tian<sup>3</sup> Xue Ruizheng<sup>4</sup>

(1. Institute of Applied Physics, Henan Academy of Sciences, Zhengzhou 450000, China; 2. School of Computer Science and Engineering, Northwest Normal University, Lanzhou 730000, China; 3. Henan Technical College of Construction, Zhengzhou 450000, China; 4. Henan Guojian Medical Equipment Co., Ltd., Shangqiu 476002, China)

**Abstract:** Data-driven soft-sensing modeling plays a critical role in process industries, yet faces challenges from heterogeneous noise contamination and the coexistence of linear and nonlinear correlations in industrial datasets. These issues significantly compromise model prediction accuracy. To address this, we propose a fused correlation-based collaborative shared noise algorithm for robust soft-sensing modeling. The algorithm integrates Pearson correlation coefficients (linear relationships) and Spearman rank correlation coefficients (nonlinear relationships) to compute data credibility, thereby optimizing noise allocation under mixed correlation conditions. A convolutional neural network (CNN) is subsequently employed to construct the soft-sensing model. Experiments on a debutanizer column dataset demonstrate the superiority of the proposed method. The FC-CSNA outperforms baseline denoising techniques, including wavelet transform, denoising autoencoders, and the original collaborative shared noise algorithm, in noise suppression. The hybrid model achieves state-of-the-art prediction performance, with an  $R^2$  score of 0.971 6 and mean squared error (MSE) of 0.001 1, validating its effectiveness in handling industrial data complexity.

**Keywords:** data-driven modeling; soft sensor; fused correlation; integrated collaborative allocation of noise; HCAN-CNN

## 0 引言

流程工业是国民经济的重要支柱产业,其生产过程的复杂性以及重要性决定了采用自动控制技术的必要性和紧迫性。由于受到生产环境、设备生产工艺和生产成本等因素的限制,一些关键过程变量难以通过硬件传感器(仪表)进行直接的快速测量。近年来,在控制和检测领域兴起的数据驱动软测量建模技术是解决上述问题的有效方法之一<sup>[1]</sup>。

软测量是采用间接测量的思路,在关键变量和辅助变量之间建立数学模型,以实现对关键变量的快速估计或预测的一种技术。仅使用过程数据,且不考虑其物理意义使该建模技术在工业生产过程中得到了广泛应用<sup>[2]</sup>。目前,有大量的工作借助深度学习进行软测量研究,例如基于自编码器及其改进的<sup>[3-7]</sup>、基于对抗网络的<sup>[8-9]</sup>、基于注意力机制的<sup>[10-13]</sup>等等。这些研究极大的促进了软测量技术的发展,但对于传感器的数据噪声问题研究进展缓慢。

通常情况下,由传感器测量和收集的流程数据会混杂着大量的复杂噪声。数据中混杂的噪声种类繁多,且分布不均<sup>[14]</sup>。目前,在软测量领域,主流方法是使用小波过滤降噪(wavelet denoising, WTD)<sup>[15-16]</sup>和降噪自编码器网络(denoising auto-encoder, DAE)<sup>[17-19]</sup>来对复杂噪声进行数据清理。小波过滤方法利用小波分解得到不同强度分布的细节分量,然后进行小波重构,对细节分量进行阈值化降噪处理。文献[15]采用基于 Haar 母波的 5 级分解对精馏塔数据进行小波过滤,再结合高斯过程回归构建软测量模型,在精馏塔底产物苯浓度估计中具有精度高、建模复杂度低的优点。文献[16]采用斯皮尔曼相关性系数进行辅助变量的选择,利用基于 Sym8 小波 Heursure 准则的 3 级分解作为阈值分解参数,并结合支持向量机回归(support vector regression, SVR)对转子的热变形进行分析和预测,实验结果表明,在对工业数据进行噪声清洗后,该模型的预测性能较基础模型有着较大的提升。去噪自编码器网络是一种基于神经网络强大学习能力的降噪方法。将特定的噪声添加到数据中,然后训练 DAE 网络来捕获该特定噪声的特征,并重构不含该特定噪声的数据原始输入。一旦训练完成,网络就变成了特定噪声的过滤网络。文献[17]采用具有层次感知注意力机制的堆叠去噪自编码器对数据进行噪声清洗,随后采用具有尺度感知注意力机制的多尺度残差网络构建软测量模型,在脱丁烷塔数据集上的实验表明,去噪多尺度残差深度网络(denoising and multiscale residual deep network, DMRDN)较对比网络有着较强的预测性能。文献[19]结合堆叠的降噪自编码器与人工神经网络构建

软测量模型,利用深层架构提取输入数据的高阶特征,与浅层学习方法相比,深层降噪模型有着较优的预测性能和泛化能力。

WTD 降噪方法可看作是对数据添加了一个函数约束,阈值的选择会严重影响降噪效果,且进行多次阈值选择后,会造成后选覆盖先选的现象。另外, WTD 方法的适用性较差,对已知特性的噪声降噪效果较好,对其他噪声降噪效果较差。DAE 网络的降噪能力来源于向数据中添加的特定噪声,用于训练的噪声的选择会严重影响基于 DAE 的模型预测效果。另外, DAE 方法不适用于多种噪声混杂的数据。WTD 和 DAE 方法是从不同角度进行噪声清洗,但这两种方法均是将数据噪声进行分类讨论,但数据中的噪声是异构的、杂糅的,这种对单一类型噪声的清理操作是不合理的。

在 2022 年, Gao 等<sup>[20]</sup>提出一种基于密度峰值聚类的协同分摊噪声(collaborative apportionment noise, CAN)算法,将数据中的各类噪声看作一个整体,统称为复杂噪声,再由数据自身的可信度、偏差和偏置度进行噪声清理。通过在脱丁烷塔过程数据和工业蒸汽量数据上的实验对比验证了 CAN 算法较 WTD 和 DAE 有着较好的泛化能力。但 CAN 算法在设计上存在缺陷,在 CAN 算法中,数据需要清洗的噪声量主要由数据各属性的可信度来加权分配。其中数据的可信度由 Pearson 相关性系数确定,而 Pearson 相关性系数主要反映数据间的线性相关性,但工业生产中收集的流程数据是线性和非线性关系共存的,且非线性关系占主导地位。

综上所述,本文提出融合协同分摊噪声(hybrid collaborative apportionment noise, HCAN)算法,并结合卷积神经网络(convolutional neural networks, CNN)构建了 HCAN-CNN 模型。在 HCAN 算法中,采用 Pearson 和 Spearman 相关性融合的相关性系数来确定数据各属性的可信度,使得可信度计算更为合理。

## 1 相关理论

### 1.1 协同分摊噪声算法

协同分摊噪声算法<sup>[20]</sup>是在两个假设条件下设计开发的。首先,认为获得的数据为单一工况条件下收集的数据。其次,认为样本中心点数据是不理想条件下存在的唯一完全无噪声数据。

CAN 算法是由可信度(credibledegree, CD)对各属性进行加权分摊噪声;由数据与样本中心点之间的偏差(deviation, Dev)来确定噪声削弱方向;由各数据与样本中心点之间的偏置(bias degree, BD)来确定噪声含量;由强度因子(reduction degree, RD)超参数进行调节,控制 CAN 算法的清理过程。若输入的原始辅助数据为  $X$ ,进

行噪声清理后的输出辅助数据为  $X'$ , 则 CAN 算法可表示为:

$$X' = X - (1 - CD) \times BD \times Dev \times RD \quad (1)$$

其中,  $CD$  由各辅助数据与关键变量之间的 Pearson 相关性系数归一化后确定。若辅助数据为  $X$ , 关键变量为  $Y$ , 则  $CD$  的计算过程可表示为:

$$CD = |\rho_{(X,Y)}| / \sum |\rho_{(X,Y)}| \quad (2)$$

$Dev$  由各数据与样本中心的差值确定。若辅助数据为  $X$ , 样本中心数据为  $\mu$ , 则  $Dev$  的计算过程可表示为:

$$Dev = X - \mu \quad (3)$$

$BD$  由各数据与样本中心数据之间的欧氏距离进行归一化后确定。 $BD$  的计算过程可表示为:

$$BD = d_{(X,\mu)} / \sum d_{(X,\mu)} \quad (4)$$

## 1.2 相关性系数

Pearson 相关性系数 (pearson correlation coefficient, PCC) 又称为线性相关性系数, 两个连续的随机变量  $X$  与  $Y$  之间的 PCC 可定义为  $\rho_{XY}$ :

$$\rho_{XY} = \frac{\sum (x - \bar{X})(y - \bar{Y})}{\sqrt{\sum (x - \bar{X})^2} \sqrt{\sum (y - \bar{Y})^2}} \quad (5)$$

式中:  $\bar{X}$  和  $\bar{Y}$  为随机变量  $X$  和  $Y$  的均值。

Spearman 相关性系数 (spearman correlation coefficient, SCC) 是衡量秩序相关性, 被认为是最好的非线性相关性系数, 用于度量独立变量  $X$  与依赖变量  $Y$  之间的相关方向。SCC 被定义为等级变量之间的 PCC, 可定义为  $\gamma_{XY}$ :

$$\gamma_{XY} = \frac{\sum (x - \bar{X})(y - \bar{Y})}{\sqrt{\sum (x - \bar{X})^2} \sqrt{\sum (y - \bar{Y})^2}} \quad (6)$$

式中:  $\bar{X}$  和  $\bar{Y}$  为独立变量  $X$  和依赖变量  $Y$  的均值。

融合相关性系数 (hybrid correlation coefficient, HCC)<sup>[3]</sup> 是融合了 PCC 和 SCC 的相关性系数。为了更好的反映复杂变量之间的相关关系, 采用调谐参数  $\alpha$  来调整融合系数中  $\rho_{XY}$  与  $\gamma_{XY}$  之间的比值, 采用调谐参数来调整融合系数中 PCC 与 SCC 之间的比值, 使线性和非线性系数之比在 0~1 之间变化。融合相关性系数  $\lambda_{XY}$  可表示为:

$$\lambda_{XY} = \alpha |\rho_{XY}| + (1 - \alpha) |\gamma_{XY}| \quad (7)$$

## 2 HCAN-CNN 软测量模型

考虑到工业流程数据是线性与非线性关系耦合, 只考虑数据间的线性关系的不合理性, 本文采用融合相关性系数来进行融合可信度 (hybrid credible degree, HCD) 计算, 优化文献[20]提出的 CAN 算法, 并结合卷积神经网络构建 HCAN-CNN 软测量模型。HCAN-CNN 模型具有两个阶段: 第 1 个阶段是融合协同分摊噪声阶段, 用来对输入数据进行噪声削弱, 并设定强度因子; 第 2 个阶段是卷积回归阶段, 对设定好强度的已清理数据进行深层特征提取, 最后利用全连接神经网络进行回归分析, 并输出关键变量的预测数据  $Y_{pred}$ 。HCAN-CNN 模型如图 1 所示。

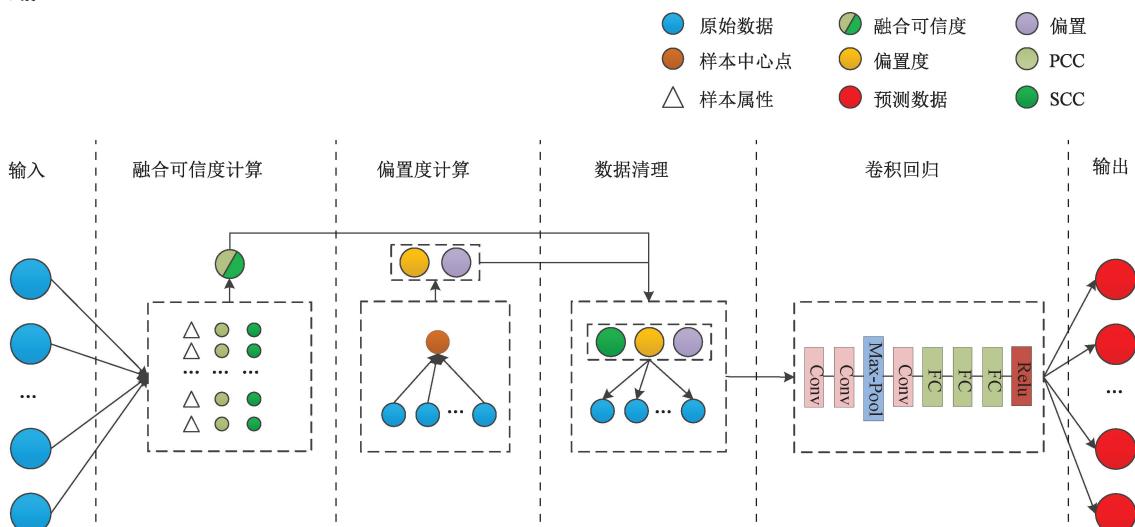


图 1 HCAN-CNN 模型流程  
Fig. 1 HCAN-CNN model flowchart

其中, 为了控制实验变量, 在卷积回归阶段, 采用与文献[20]相同的结构参数, 即设定 3 个卷积层, 在第 2 和

第 3 卷积层之间穿插一个最大池化层, 最后连接一个 3 层的全连接网络。模型总体流程如算法 1 所示。

## 算法 1: HCAN-CNN 软测量模型流程

输入: 原始输入辅助数据  $X$ , 关键变量  $Y$ 输出: 关键变量预测值  $Y_{\text{pred}}$ 

融合协同分摊噪声阶段:

初始化:

设定 HCAN 算法中调谐参数  $\alpha$  和强度因子  $RD$ 

循环计算:

- 根据式(7)计算融合可信度  $HCD$
- 根据式(3)、(4)计算偏差  $Dev$  和偏置  $BD$
- 根据式(1)对原始输入辅助数据  $X$  进行降噪处理

整个数据集遍历完成时结束

中间输出: HCAN 算法降噪处理后辅助变量  $X'$ 

卷积回归阶段:

初始化:

- 设定结构参数
- 设定优化器、卷积核大小、步长、填充方式、激活函数、迭代次数、误差函数

循环计算:

- 进行特征提取、筛选、反向传播
- 参数收敛时结束

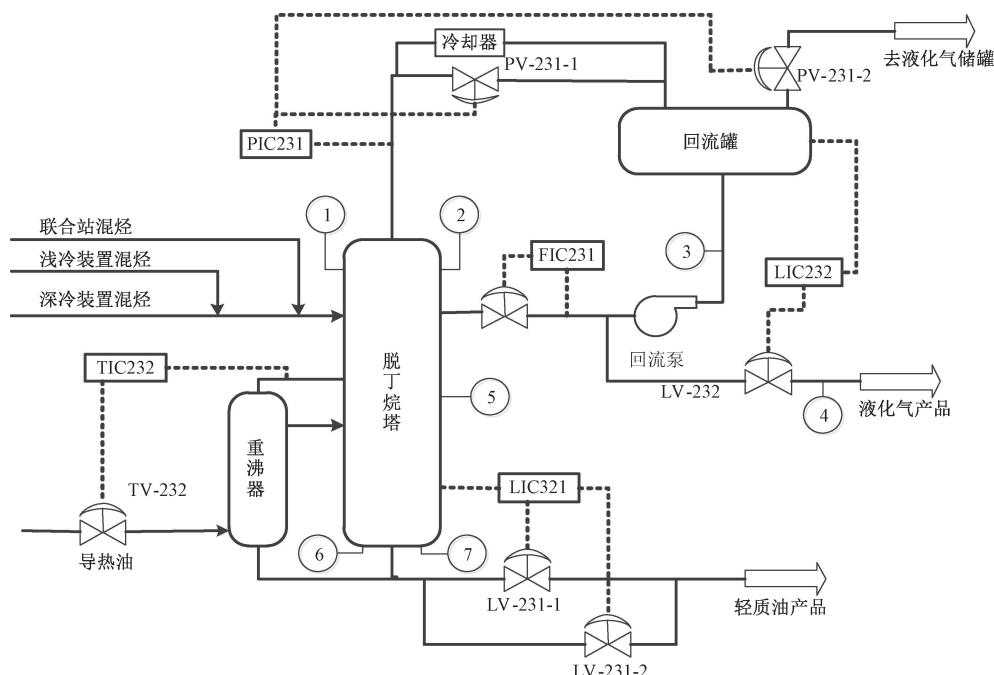
输出: 关键变量预测值  $Y_{\text{pred}}$ 

## 3 实验分析

## 3.1 数据集介绍

烯烃分离装置脱丁烷塔是工业炼油中分离液化石油气和稳定轻烃的重要工业炼化炉,主要用于脱硫和石脑油裂解。脱丁烷塔的主要流程如图 2 所示,该化工过程主要有六种设备,即热交换器、塔顶冷凝器、底部再沸器、回流泵、液化石油气分离器的进料泵和回流蓄能器。为了实现稳定的化工生产且保证气态产物的产品质量,必须使脱丁烷塔底部的丁烷含量最小化,因此对塔内丁烷含量的实时测量是实现精准控制该炼油过程的关键所在。但实际上,能反映丁烷含量的 C4 的浓度并不能直接测量,而是需要借助气相色谱仪对随后的脱异戊烷塔顶部的输出气体进行持续测量分析。

气相色谱仪对脱丁烷塔过程中丁烷含量的分析测量具有严重的滞后性,不能满足该炼化过程的实时精确控制。因此,根据图 2 中 7 个数据采集点进行过程数据采集。该过程中共采集 2 390 条数据,各数据说明如表 1 所示。

图 2 脱丁烷塔主要流程<sup>[9]</sup>Fig. 2 Main process of the debutanizer column<sup>[9]</sup>

## 3.2 模型评价指标

在参考多位研究人员发表的文献后,选择如下 4 种通用模型准确性评价指标: 平均绝对值误差 (mean

absolute error, MAE)、均方误差 (mean squared error, MSE)、均方根误差 (root mean square error, RMSE)、决定系数 (r-square,  $R^2$ )。

表 1 脱丁烷塔数据集说明

Table 1 Debutanizer dataset description

过程变量及关键变量	变量描述
$u_1$	塔顶温度
$u_2$	塔顶压强
$u_3$	回流流量
$u_4$	流向下一流程流量
$u_5$	中间层温度
$u_6$	底部温度 A
$u_7$	底部温度 B
$y$	丁烷含量

$$\begin{bmatrix} u_1(k), u_2(k), u_3(k), u_4(k), u_5(k), u_5(k-1), u_5(k-2), u_5(k-3), \\ (u_6(k) + u_7(k))/2, y(k-1), y(k-2), y(k-3), y(k-4) \end{bmatrix}^T \quad (8)$$

$$\begin{bmatrix} u_1(k), u_2(k), u_3(k), u_4(k), u_5(k), u_5(k-1), u_5(k-2), u_5(k-3), \\ (u_6(k) + u_7(k))/2, y(k-1), y(k-2), y(k-3), y(k-4), 0, 0, 0 \end{bmatrix}^T \quad (9)$$

为了验证 HCAN 方法的可行性,本文选择文献[20]中使用的基于堆栈自编码器神经网络(stacked autoencoder neural network, SAE-NN)和卷积神经网络(CNN)用于基回归模型。在 HCAN 阶段,选择文献<sup>[3]</sup>中的设置,设定调谐参数  $\alpha = 0.5$ 。在回归分析阶段,回归器的参数设置均采用文献[20]中的设置,即将优化器设定为“Adam”,卷积核大小设定为  $2 \times 2$ ,步长设定为 1,填充方式设定为“Same”,激活函数设定为“ReLU”,迭代次数设定为 20,损失函数设定为“MAE”,全连接神经网络的神经元个数设定为 [40, 20, 10, 1],激活函数设定为

### 3.3 实验结果与分析

实验采用文献[3-4, 20-21]中对数据的预处理操作,使用  $u_5$  的历史数据和关键变量  $y$  对数据进行扩展,具体的数据展开如增广矩阵式(8)所示。在将数据传递到回归分析阶段之前,对数据填充 3 列零值,使用增广矩阵式(9)作为 CNN 网络第  $k$  个采样瞬间的原始观测输入变量。总共使用 2 390 个数据样本用于模型的训练和测试,其中 1 000 个样本作为训练数据,其余样本用于测试数据。

“ReLU”,学习率设定为 0.01。

在 HCAN 阶段,采用试错法寻找 HCAN 算法中的  $RD$  参数,这里选择了 6 种模型,分别是采用 PCC 计算数据可信度的 CAN-SAE-NN 模型和 CAN-CNN 模型,采用 SCC 计算数据可信度的 SCC-CAN-SAE-NN 模型和 SCC-CAN-CNN 模型,采用 HCC 计算数据可信度的 HCAN-SAE-NN 模型和 HCAN-CNN 模型,在 [0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0] 范围内不同  $RD$  值的变化趋势,所有试验结果均是多次实验后取得的均值,如图 3 所示。

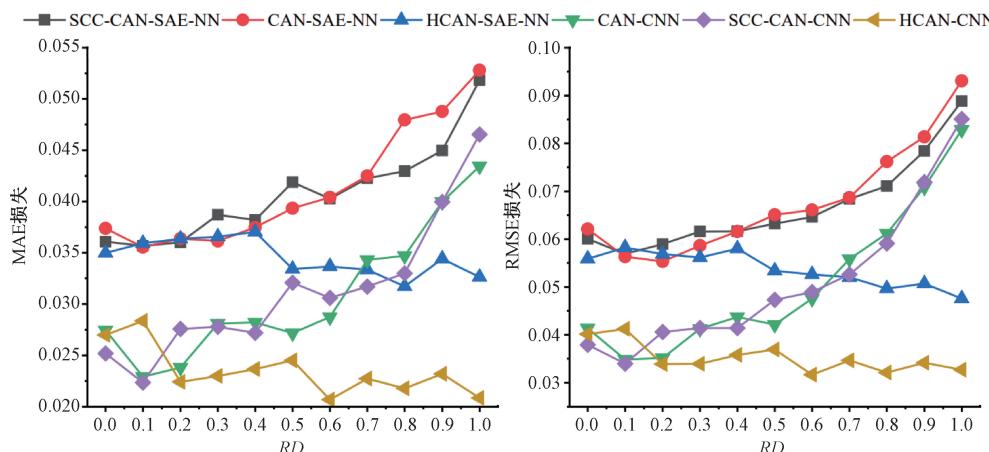


图 3 不同模型在不同 RD 值下的损失变化

Fig. 3 Losses of different models under different RD values

当模型强度因子  $RD = 0.0$  时,表明未对数据进行降噪操作,是一组消融对比实验。当强度因子  $0 < RD \leq 1$  时,表明对数据进行降噪操作,并随着  $RD$  值的增大,降噪操作越强。由图 3 可知,随着强度因子  $RD$  的增加,各模型损失的变化趋势有着较大的不同,为了更清晰的得到我们的结论,在分析时进行分组分析。

其中,基础模型分别是 SAE-NN 和 CNN。在以 SAE-NN 为基模型的组别中一共有 3 个模型,分别是 CAN-

SAE-NN、SCC-CAN-SAE-NN、HCAN-SAE-NN。其中 CAN-SAE-NN 在  $RD$  低时表现较好,但随着  $RD$  增加,MAE 和 RMSE 迅速上升,表现最不稳定;SCC-CAN-SAE-NN 相较于 CAN-SAE-NN 有一定的优化效果,但在  $RD$  较大时损失上升的依旧明显;采用融合相关性的协同分摊噪声算法模型 HCAN-SAE-NN 与前面两者相比更加稳定,尤其是 RMSE 的变化幅度较小,说明 HCAN 机制在 SAE-NN 结构上提升了鲁棒性。在以 CNN 为基模型的组别中也

有3个模型,分别是CAN-CNN、SCC-CAN-CNN、HCAN-CNN。其中CAN-CNN模型在RD较小时MAE和RMSE表现良好,但随着RD的增大,损失上升较快;相较于CAN-CNN模型,SCC-CAN-CNN模型性能有所提升,但在RD增大时仍然出现损失上升趋势;与前两者相比,HCAN-CNN在所有RD值下的MAE和RMSE都是最低的且没有出现明显的上升趋势,说明HCAN机制对CNN结构的提升最为明显。

其中,采用PCC计算可信度的CAN-SAE-NN和CAN-CNN模型与采用SCC计算可信度的SCC-CAN-SAE-NN和SCC-CAN-CNN4种模型MAE损失和RMSE损失的变化趋势表现为先减小后逐步增大,且4种模型的损失最低点均小于消融组。而采用HCC计算可信度的HCAN-SAE-NN和HCAN-CNN模型的MAE损失与RMSE损失在RD值增大时表现出稳步下降的趋势,且损失值均小于消融组。上述实验可表明PCC、SCC和HCC3种方法在计算数据可信度方面有着不同的表现,其中,由HCC优化的CAN方法表现较优,在不同RD值下均能降低模型损失。

综上所述,可以得出结论,采用融合相关性的协同分摊噪声算法HCAN较CAN以及SCC-CAN有着明显的优势,HCAN机制的改进是有效的。可表明PCC、SCC和HCC3种方法在计算数据可信度方面有着不同的表现,其中,由HCC优化的CAN方法表现较优,在不同RD值下均能降低模型损失。

为了证明本文所提出采用HCC优化CAN方法以及提出的HCAN-CNN模型性能,采用多降噪方法、多模型和多回归方法进行交叉对比实验。在所涉及的方法和模型中,降噪方法有刻出来高斯噪声的DAE、基于sym8的小波降噪WTD、基于PCC计算可信度的CAN、基于SCC计算可信度的SCC-CAN和基于HCC计算可信度的HCAN;不同的模型及回归方法有SAE-NN和CNN;对比

模型有基于PCC的VWSAE- $\rho$ 、基于SCC的VWSAE- $\gamma$ 、基于HCC的Hybrid VWSAE和质量驱动模型SQAE-NN以及去噪多尺度残差深度网络DMRDN。其中为了明显观察到模型性能的提升,在交叉对比实验中,将SCC-CAN-SAE-NN和SCC-CAN-CNN模型的RD值设定为0.1,将HCAN-SAE-NN和HCAN-CNN模型的RD值分别设定为1.0和0.6。

采用多种方法及模型进行交叉组合实验,结果如表2所示。从表2可以看出,HCAN-CNN模型较优化前有着明显的提升,具体表现在有着较低的MAE、MSE和RMSE损失,有着较高的R<sup>2</sup>指标。由多次独立实验取均值可以看出,HCAN-CNN模型较对比模型有着较优的指标得分。

为了更直观的比较模型的性能,本文提取了前述6种模型的预测结果,用于对预测误差进行对比分析,如图4~9所示。

表2 脱丁烷塔数据集下多模型多方法交叉实验结果对比

Table 2 Cross-experimental results of multiple models and methods under the Debutanizer column dataset

模型	RD	MAE	MSE	RMSE	R <sup>2</sup>
SCC-CAN-SAE-NN	0.1	0.035 7	0.003 3	0.056 9	0.911 0
CAN-SAE-NN <sup>[9]</sup>	0.2	0.036 4	0.003 1	0.055 3	0.916 1
DAE-CNN <sup>[9]</sup>	—	0.041 3	0.003 1	0.054 9	0.917 4
HCAN-SAE-NN	1	0.032 6	0.002 3	0.047 6	0.937 7
VWSAE- $\rho$ <sup>[10]</sup>	—	—	—	0.038 9	0.943 8
VWSAE-NN <sup>[11]</sup>	—	0.027 7	—	0.037 9	0.944 4
WTD-CNN <sup>[9]</sup>	—	0.025 4	0.001 9	0.039 8	0.949 2
DMRDN <sup>[6]</sup>	—	—	—	0.037 9	0.956 1
VWSAE- $\gamma$ <sup>[10]</sup>	—	—	—	0.033 6	0.958 0
Hybrid VWSAE <sup>[10]</sup>	—	—	—	0.030 8	0.961 5
SQAE-NN <sup>[12]</sup>	—	0.022 0	—	0.030 3	0.964 6
CAN-CNN <sup>[9]</sup>	0.1	0.023 0	0.001 2	0.034 9	0.966 6
SCC-CAN-CNN	0.1	0.022 3	0.001 2	0.034 0	0.968 2
HCAN-CNN	0.6	0.020 7	0.001 1	0.031 7	0.971 6

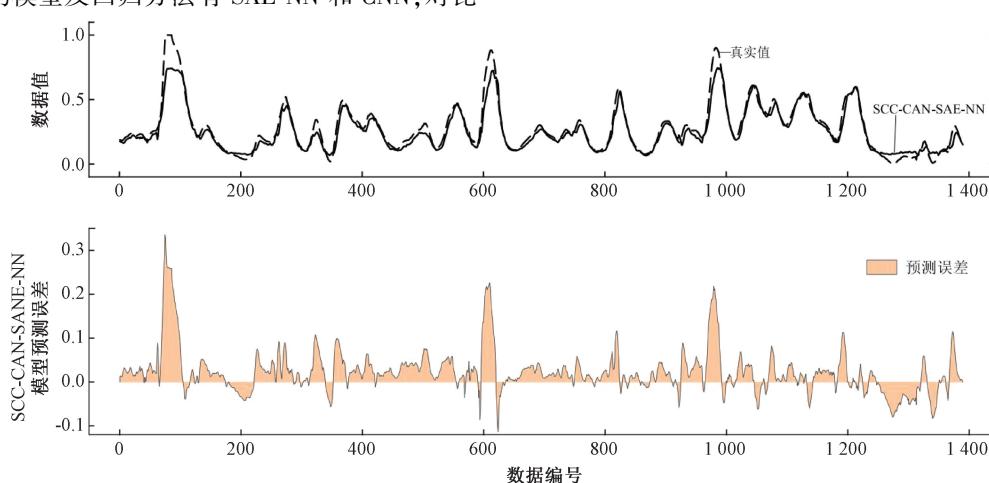


图4 SCC-CAN-SAE-NN模型在脱丁烷塔数据集上的预测结果与预测误差

Fig. 4 The prediction results and prediction errors of the SCC-CAN-SAE-NN model

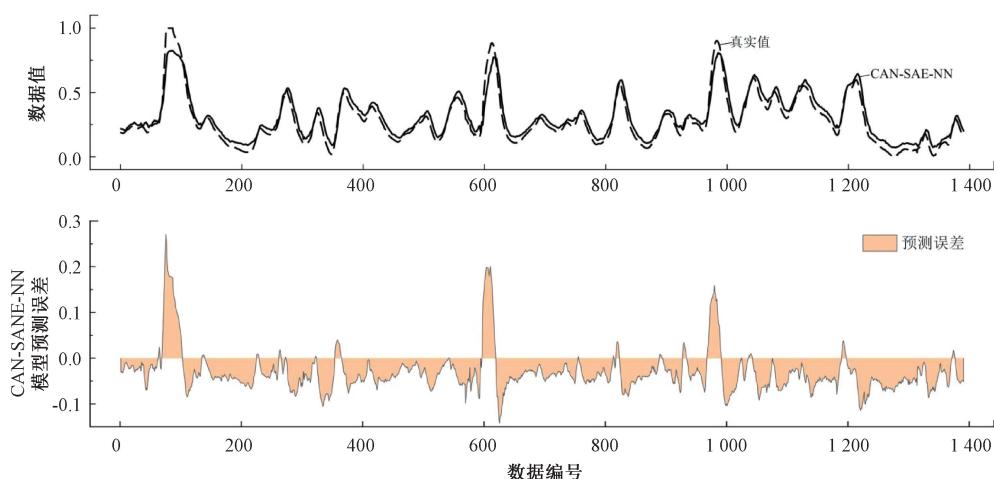


图 5 CAN-SAE-NN 模型在脱丁烷塔数据集上的预测结果与预测误差

Fig. 5 The prediction results and prediction errors of the CAN-SAE-NN model

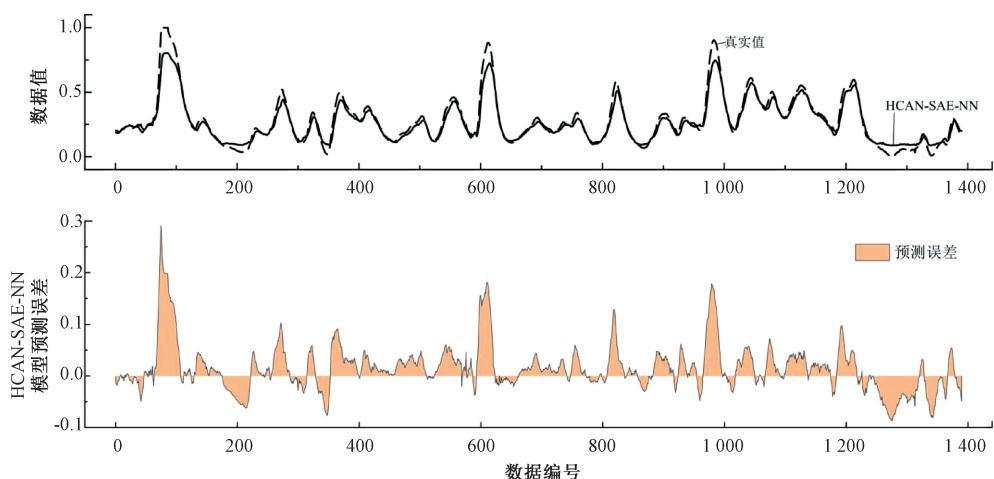


图 6 HCAN-SAE-NN 模型在脱丁烷塔数据集上的预测结果与预测误差

Fig. 6 The prediction results and prediction errors of the HCAN-SAE-NN model

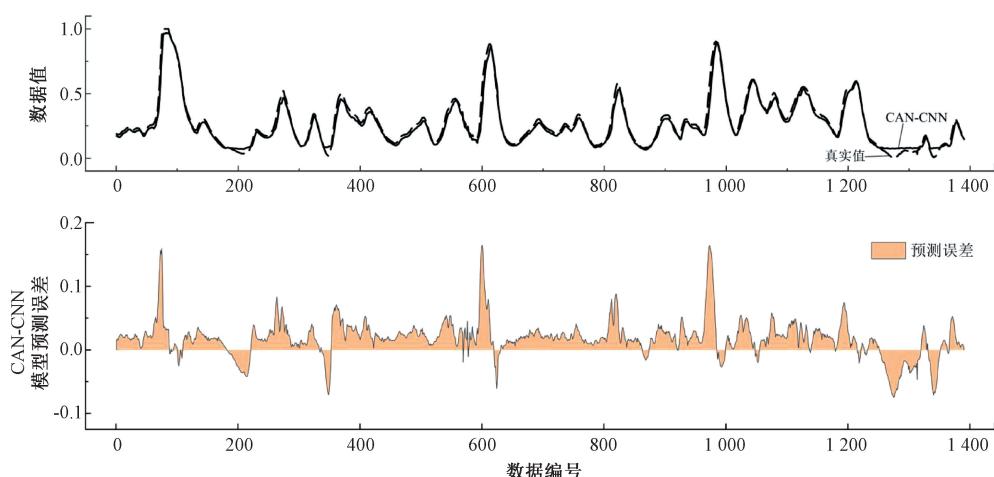


图 7 CAN-CNN 模型在脱丁烷塔数据集上的预测结果与预测误差

Fig. 7 The prediction results and prediction errors of the CAN-CNN model

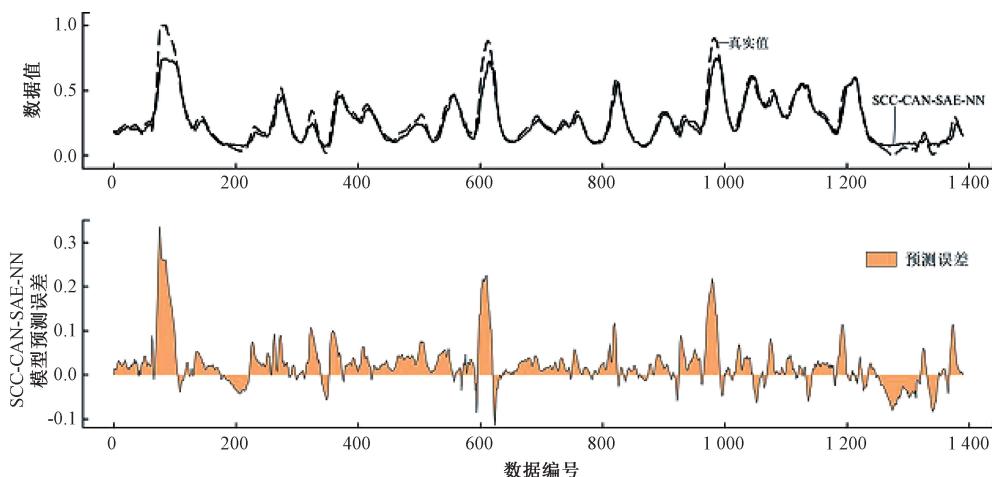


图 8 SCC-CAN-CNN 模型在脱丁烷塔数据集上的预测结果与预测误差

Fig. 8 The prediction results and prediction errors of the SCC-CAN-CNN model

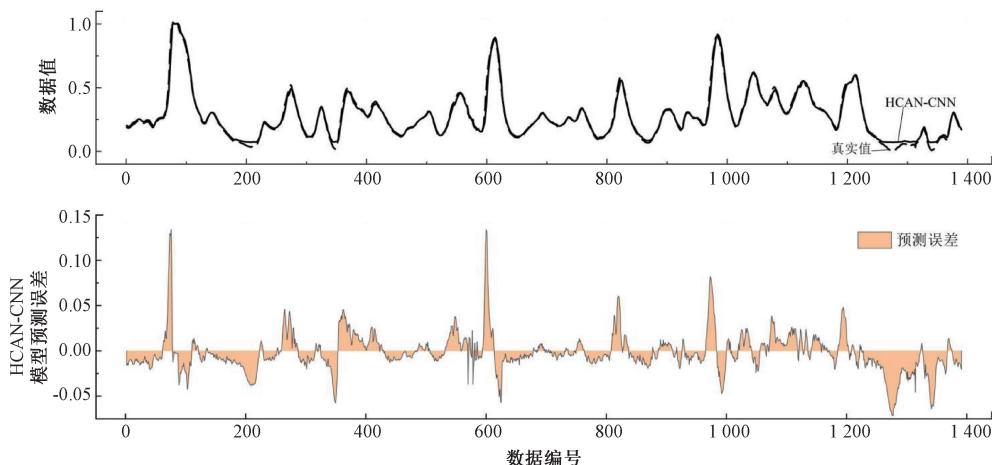


图 9 HCAN-CNN 模型在脱丁烷塔数据集上的预测结果与预测误差

Fig. 9 The prediction results and prediction errors of the HCAN-CNN model

图 4~9 中,以预测结果与真实数据的对比可以看出,HCAN-CNN 模型具有较好的预测效果,预测值较对 比模型更加贴合真实数据。同时,从各模型的预测误差可以看出,HCAN-CNN 模型的误差范围更小。从误差面积比较来看,HCAN-CNN 模型的误差面积更小,表明其预测误差较其他对比模型更小。

## 4 结 论

本文针对工业数据中线性相关与非线性相关共存,且数据中存在复杂噪声问题,提出以融合相关性系数计算数据的可信度,来优化采用 Pearson 相关性系数计算数据可信度的协同分摊噪声算法,并结合卷积神经网络搭建 HCAN-CNN 软测量模型。在本文中采用多降噪方法、多模型和多回归方法的交叉组合进行实验,结果表明优

化后的 HCAN 方法较优化前有着较高的提升,且优化后较优化前可信度计算方法更合理。与其他降噪方法对比表明,HCAN 方法具有更好的噪声抑制效果。与多模型和多回归方法对比表明,本文提出的 HCAN-CNN 模型具有更高的预测精度以及更小的预测误差。

虽然本文提出的 HCAN 较优化前有着较大的提升,但 CAN 算法理论建立在单一工况下,不易直接推广到多工况场景下。因此,下一步的工作是设计多工况条件下对复杂噪声的削弱操作,同时提高多限制条件下的模型预测精度。

## 参考文献

- [1] SUN Q, GE Z. A survey on deep learning for data-driven soft sensors [J]. IEEE Transactions on Industrial Informatics, 2021, 17(9): 5853-5866.
- [2] 沈冰冰. 基于深度概率隐变量模型的软测量建模方

- 法研究[D]. 杭州:浙江大学, 2022.
- SHEN B B. Soft sensing methods based on deep probabilistic latent variable models [D]. Hangzhou: Zhejiang University, 2022.
- [3] YUAN X, OU C, WANG Y, et al. Deep quality-related feature extraction for soft sensing modeling: A deep learning approach with hybrid VW-SAE [J]. Neurocomputing, 2020, 39(6): 375-382.
- [4] YUAN X, HUANG B, WANG Y, et al. Deep learning-based feature representation and its application for soft sensor modeling with variable-wise weighted SAE [J]. IEEE Transactions on Industrial Informatics, 2018, 14(7): 3235-3243.
- [5] HE X, LIU T, ZHANG Y, XIE R, et al. Soft sensor modeling based on vector-quantized weighted-Wasserstein VAE for polyester polymerization process [J]. IEEE Transactions on Industrial Informatics, 2024, 20: 11338-11347.
- [6] TANG X, YAN J, LI Y, et al. Semi-supervised deep conditional variational autoencoder for soft sensor modeling [J]. IEEE Sensors Journal, 2024, 24: 7153-7164.
- [7] WU J, SHEN F, YE L, et al. Data augmentation using time conditional variational autoencoder for soft sensor of industrial processes with limited data [J]. IEEE Transactions on Instrumentation and Measurement, 2024, 73: 1-14.
- [8] LI X, ZHU Q, HE Y, et al. Data mode-related generative adversarial network for industrial soft sensor application [J]. IEEE Transactions on Industrial Informatics, 2024, 20: 4198-4205.
- [9] GUO R, CHEN Q, LIU H, et al. Adversarial robustness enhancement for deep learning-based soft sensors: An adversarial training strategy using historical gradients and domain adaptation [J]. Sensors (Basel, Switzerland), 2024, 24(12): 3909.
- [10] GUO R, LIU H, XIE G, et al. A self-interpretable soft sensor based on deep learning and multiple attention mechanism: from data selection to sensor modeling [J]. IEEE Transactions on Industrial Informatics, 2023, 19: 6859-6871.
- [11] FANG Z, GAO S, DANG X, et al. Transformer enhanced by local perception self-attention for dynamic soft sensor modeling of industrial processes [J]. Measurement Science and Technology, 2024, 35: 055123.
- [12] MA L, ZHAO Y, WANG B, et al. A multistep sequence-to-sequence model with attention LSTM neural networks for industrial soft sensor application [J]. IEEE Sensors Journal, 2023, 23: 10801-10813.
- [13] YANG Y, JIA Y, YU H, et al. Spatiotemporal graph attention network for soft sensor modeling of suspended magnetization roasting process [J]. IEEE Transactions on Instrumentation and Measurement, 2024, 73: 1-15.
- [14] ZOHAIB A K, USMAN Z, OLIVIER D B, et al. Adaptive estimation and reduction of noises affecting a self-mixing interferometric laser sensor [J]. IEEE Sensors Journal, 2020, 20(17): 9806-9815.
- [15] MORTEZA Z, ALI K G, ALI A S. Soft sensor design for distillation columns using wavelets and gaussian process regression [C]. 2019 6th International Conference on Control, Instrumentation and Automation (ICCIA). IEEE, 2019: 1-6.
- [16] ZHAO Y, LIU H, GUO R, et al. Air preheater rotor deformation soft sensor based on wavelet analysis and SVR [C]. 2020 Chinese Automation Congress (CAC). IEEE, 2020: 4490-4495.
- [17] GAO R, ZHU H, WANG G, et al. A denoising and multiscale residual deep network for soft sensor modeling of industrial processes [J]. Measurement Science and Technology, 2022, 33(10): 105117.
- [18] GUO R, LIU H, WANG W, et al. A hybrid-driven soft sensor with complex process data based on DAE and mechanism-introduced GRU [C]. 2021 IEEE 10th Data Driven Control and Learning Systems Conference (DDCLS). IEEE, 2021: 553-558.
- [19] YA W, DI T, LIN Y, et al. A data-driven soft sensor modeling method based on deep learning and its application [J]. IEEE Transactions on Industrial Electronics, 2017, 64(5): 4237-4245.
- [20] GAO S, ZHANG Q, TIAN R, et al. Collaborative apportionment noise-based soft sensor framework [J]. IEEE Transactions on Instrumentation and Measurement, 2022, 71: 1-12.
- [21] YUAN X, ZHOU J, HUANG B, et al. Hierarchical quality-relevant feature representation for soft sensor modeling: A novel deep learning strategy [J]. IEEE Transactions on Industrial Informatics, 2020, 16(6): 3721-3730.

## 作者简介



梁楠, 2003 年于同济大学获得学士学位, 2008 年和 2013 年于西北工业大学获得硕士学位和博士学位, 现为河南省科学院副研究员, 主要研究方向为电子信息技术和人工智能。  
E-mail: 9662062@qq.com

**Liang Nan** received his B. Sc. degree from Tongji University in 2003, M. Sc. degree and Ph. D. degree from the Northwest Polytechnical University in 2008 and 2013, respectively. Now he is an associate researcher in Henan Academy of Science. His main research interests include electronic information technology and artificial intelligence.



**高世伟**,2002 年于青岛科技大学获得学士学位,2005 年于中科院获得硕士学位,2009 年于西北工业大学获得博士学位,现为西北师范大学教授级高工,主要研究方向为工业大数据。

E-mail: gaoshiwei@nwnu.edu.cn

**Gao Shiwei** received his B. Sc. degree from Qingdao University of Science and Technology in 2002, received his M. Sc. degree from Chinese Academy of Sciences in 2005, and received his Ph. D. degree from Northwestern Polytechnical University in 2009. Now he is a professor-level senior engineer at Northwest Normal University. His main research interest includes industrial big data.

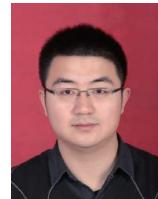


**张伟**(通信作者),2007 年于新疆大学获得学士学位,2012 年于解放军信息工程大学获得硕士学位,现为河南省科学院应用物理研究所有限公司正高级工程师,主要研究方向为电子与通信技术。

E-mail: KG08\_13@hnaas.ac.cn

**Zhang Wei** (Corresponding author) received his B. Sc.

degree from Xinjiang University in 2007, and received his M. Sc. Degree from the PLA Information Engineering University in 2012. Now he is a principal senior engineer in Henan Academy of Science Institute of Applied Physics Co., Ltd. His main research interests include electronic and communication technology.



**田添**,2006 年于武汉理工大学获得学士学位,现为河南建筑职业技术学院副教授,主要研究方向为通信技术与应用、物联网与智能系统。

E-mail: 280317649@qq.com

**Tian Tian** received his B. Sc. degree from Wuhan University of Technology in 2006. Now he is an associate professor at Henan College of Architecture and Technology. His main research interests include communication technology and applications, Internet of Things and Intelligent Systems.



**薛瑞争**,1999 年于郑州大学获得学士学位,现为河南国健医疗设备有限公司高级工程师,主要研究方向为人工智能仿真。

E-mail: 149016898@qq.com

**Xue Ruizheng** received his B. Sc. degree from Zhengzhou University in 1999. Now he is a senior engineer at Henan Guojian Medical Equipment Co., Ltd. His main research interest includes artificial intelligence simulation.