

基于级联注意力特征融合的门控 TCN 软测量方法

孙乐¹, 曹余庆², 谢莉¹

(1. 江南大学 物联网工程学院, 江苏 无锡 214122;

2. 无锡爱德旺斯科技有限公司, 江苏 无锡 214161)

摘要: 针对传统 TCN 模型在工业过程动态建模中容易忽略时间序列连续性和局部依赖关系的问题, 研究并提出了一种动态软测量模型 CAFF-GTCN; 通过设计一种新的级联注意力特征融合模块改进 TCN 中的残差连接, 利用自注意力机制和多尺度通道注意力机制对不同感受野提取的特征进行融合, 保证模型不会丢失重要信息; 同时利用门控机制改进扩张因果卷积, 并结合 SELU 函数增强特征提取能力; 实验结果表明, 所提方法显著提升了预测精度: 在青霉素发酵仿真实验中, 相较于传统 TCN 模型, CAFF-GTCN 模型的 RMSE 和 MAE 分别降低了 45.1% 和 49.4%, R^2 从 0.992 3 提升至 0.998 9; 在硫回收过程实验中, CAFF-GTCN 模型的 RMSE 和 MAE 分别降低了 38.2% 和 42.7%, R^2 从 0.750 3 提升至 0.846 4; 实验结果验证了所提方法在动态特征提取和预测精度方面的有效性和优越性。

关键词: 软测量; 深度学习; 时间卷积网络; 特征融合; 门控机制

Gated TCN with Cascaded Attentional Feature Fusion for Soft Sensor Modeling

SUN Le¹, CAO Yuqing², XIE Li¹

(1. School of Internet of Things Engineering, Jiangnan University, Wuxi 214122, China;

2. Wuxi Advanced Technology Co., Ltd., Wuxi 214161, China)

Abstract: In industrial process dynamic modeling, traditional temporal convolutional network (TCN) models make it easy to ignore the continuity of time series and local dependencies. Based on this, a dynamic soft-sensing model, cascaded attention feature fusion cascaded attentional feature fusion (CAFF-GTCN), is proposed. A novel cascaded attention feature fusion module is designed to improve the residual connections of the TCN, which utilizes self-attention mechanisms and multi-scale channel attention mechanisms to fuse the feature extracted from different receptive fields, avoiding the loss of important information. Meanwhile, a gating mechanism is introduced to modify the dilated causal convolution, which enhances its feature extraction ability in combination with the SELU function. Experimental results show that the proposed method significantly improves prediction accuracy. Compared with the traditional TCN model, the CAFF-GTCN model reduces the RMSE and MAE by 45.1% and 49.4%, respectively, while increasing the R^2 from 0.992 3 to 0.998 9 in the penicillin fermentation simulation experiment, and reduces the RMSE and MAE by 38.2% and 42.7%, respectively, while increasing the R^2 from 0.750 3 to 0.846 4 in the sulfur recovery process experiment. Experimental results verify the effectiveness and superiority of the proposed method in dynamic feature extraction and prediction accuracy.

Keywords: soft sensor; deep learning; temporal convolutional network; feature fusion; gating mechanism

收稿日期: 2025-01-13; 修回日期: 2025-03-02。

基金项目: 国家重点研发计划(2022YFC3401302); 中国博士后科学基金(2021M691276)。

作者简介: 孙乐(1999-), 男, 硕士研究生。

通讯作者: 谢莉(1985-), 女, 博士, 副教授。

引用格式: 孙乐, 曹余庆, 谢莉. 基于级联注意力特征融合的门控 TCN 软测量方法[J]. 计算机测量与控制, 2026, 34(2): 23

0 引言

在复杂的工业过程中, 关键质量变量的检测对于确保高效的工业生产至关重要^[1]。然而, 由于相关技术限制、高测量成本和恶劣环境, 通过硬件传感器和实验室分析实现对这些变量的实时监测极具挑战性^[2]。软测量使用易于测量的变量来建立数学模型, 这些模型已成功地作为虚拟传感器应用于石化、冶金、化工等工业领域, 在闭环控制、故障检测等方面发挥着越来越重要的作用^[3-5]。

软测量建模方法主要分为机理建模和数据驱动建模两大类^[6]。对于大部分工业过程而言, 复杂的内部机理使得机理建模变得十分困难^[7]。相比之下, 数据驱动建模方法对先验知识的依赖性较低, 因而被广泛应用于质量变量的在线监测^[8]。传统的数据驱动软测量建模方法主要包括主元成分回归^[9]、支持向量回归 (SVR, support vector regression)^[10]、人工神经网络^[11]等。近年来, 深度学习被引入软测量建模领域中, 常见的神经网络主要有堆叠自动编码器^[12]、递归神经网络^[13]、卷积神经网络 (CNN, convolutional neural networks) 等^[14], 这些网络对高维复杂的数据具有更为强大的特征学习能力, 是目前主流的软测量建模方法^[15]。

由于复杂的物理化学反应和控制作用, 工业过程通常呈现动态特性, 这意味着一定时间段内的过程变量间存在时间相关性^[16]。长短时记忆网络 (LSTM, long short term memory network) 作为递归神经网络的一种变体, 可以有效处理时间序列数据的非线性关系, 因而常被用于动态软测量建模^[17]。然而 LSTM 仅考虑当前时刻的输入和前一时刻的隐藏层特征, 因此难以捕捉到较远采样时刻的变量之间的长期依赖关系^[18]。相比之下, 时间卷积网络 (TCN, temporal convolutional network) 可以提取变量间的非线性关系和长期动态依赖性, TCN 采用扩张卷积策略, 在卷积核中插入零值来跳过某些时刻的样本, 使网络有足够大的感受野来捕捉更远时刻的信息^[19]。文献 [20] 构建了一种自回归时间卷积 (AR-TCN) 预测模型, 在 TCN 中引入质量变量相关信息, 用于脱丁烷塔和加氢裂化过程的关键变量预测。然而, 扩张卷积也使网络忽视变量的连续性, 变量之间的短期时间相关性可能会因为扩张卷积而丢失, 对基于 TCN 的软测量模型而言, 这是一个不可忽视的问题^[21]。

为了获得更好的特征表示并提高软测量模型的预测性能, 特征融合思想被广泛应用于集成不同时间或空间尺度的特征。文献 [22] 提出了一种基于深度残差监督自动编码器的软测量建模方法, 通过残差连接整合先前

隐藏层特征以增强特征提取能力。然而, 该方法未考虑变量的时间相关性, 因此不适用于具有较强动态特性的工业过程。文献 [23] 提出了一种基于历史特征融合注意力机制的半监督 LSTM 软测量方法, 在特征融合中应用注意力机制, 有效增强了 LSTM 的长期记忆能力, 解决了在不规则采样频率下难以挖掘变量时间相关性的问题。尽管如此, 在实际生产过程中, 过程变量之间的动态特性通常是未知且复杂的, 因此需要从不同尺度下提取变量的时空特征, 以更全面地捕捉变量间的短期和长期依赖关系, 从而提高对质量变量的预测精度^[24]。

综上所述, 为了充分挖掘过程的动态特性, 提出了一种基于级联注意力特征融合的门控 TCN 软测量方法。该方法通过引入级联注意力特征融合模块来改进 TCN 的残差连接, 使网络能够自适应地关注重要信息。利用门控扩张因果卷积和 SELU 函数改进 TCN 网络, 进一步增强网络的特征提取能力。通过青霉素发酵仿真过程和硫回收工艺过程的实验结果表明, 所提方法可以提高软测量模型的预测精度。

1 TCN 网络概述

TCN 是一种基于卷积的神经网络, 主要应用于解决时间序列问题。在大多数时间序列任务中, 因果性是不可忽视的, 这意味着模型只能依赖于当前时刻之前的信息来进行预测。为了使所有序列数据具有严格对应的因果关系, 扩张因果卷积策略被应用于 TCN 中。扩张因果卷积采用一维卷积, 其卷积核只在时间维度上滑动, 并按照固定步长跳过某些历史时刻输入, 从而在参数相同的前提下覆盖更远的历史信息范围。对于一个时间序列 $x = (x_1, x_2, \dots, x_T)$, T 时刻扩张因果卷积如式 (1) 所示:

$$F_T = (x * {}_d f)(T) = \sum_{i=0}^{k-1} f(i) \cdot x_{T-d \cdot i} \quad (1)$$

其中: $* {}_d$ 表示扩张卷积, $f = [f_0, f_1, \dots, f_{k-1}]$ 为大小为 k 的卷积核, $x_{T-d \cdot i}$ 表示只有历史输入信息参与扩张卷积运算, d 为用于控制跳跃步长的扩张率。随着网络加深, 扩张率以指数形式增加, 当 $d = 2$ 时, 第 j 层的扩张率为 2^{j-1} , 使用更大的 d 能够使顶层的输出代表更为广泛的输入。利用扩张因果卷积, TCN 可以捕捉与更远历史输入之间的长期时间相关性。

TCN 由多个残差块堆叠构成, 残差块的具体结构如图 1 所示, 包括两层扩张因果卷积, 并使用 ReLU 激活函数。随着网络深度增加, 模型提取的特征信息也会增多, 这对模型的稳定性提出了更高的要求。因此 TCN 引入了残差连接, 将当前残差块的输入 x_T 与经过若干堆叠层后的输出 F_T 进行叠加融合, 以跨层级的方式传递信息, 其数学表达式为:

$$y_T = x_T + F_T \quad (2)$$

式中, y_T 为当前残差块的最终输出, 它会作为输入被传递至下一个残差块。残差连接使网络中层与层之间的信息传递和梯度传播更加顺畅, 从而有效抑制因梯度消失现象导致的性能饱和。此外, TCN 在残差块中还采用了一些有用策略, 其中 1×1 卷积可以保持输入和输出的形状一致, 权值归一化可以加快模型的收敛速度, Dropout 策略可以防止网络过拟合。

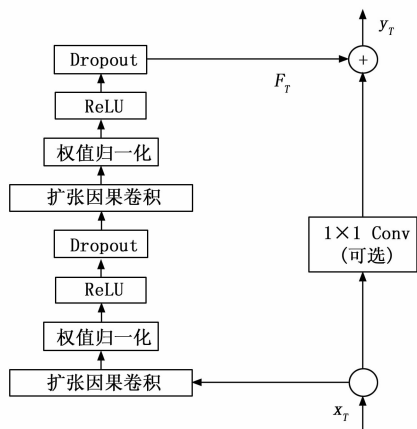


图 1 TCN 网络的残差块结构

2 基于级联注意力特征融合的 GTCN 网络

TCN 可以处理长时间序列中的非线性和动态特性。然而, 当根据具有未知动态的过程数据建立软测量模型时, 选择较大的扩张率可能会使模型忽略局部关键信息。为了解决这个问题, 对 TCN 网络中的残差块进行改进, 提出一种基于级联注意力特征融合的门控 TCN 网络 (CAFF-GTCN, gated TCN with cascaded attentional feature fusion)。

改进的残差块结构如图 2 所示, 首先利用级联注意

力特征融合 (CAFF, cascaded attentional feature fusion) 模块替代残差连接。CAFF 的输入是两种时间尺度下提取的特征: 一种是当前残差块的输入特征, 它包含更多的局部信息; 另一种是通过扩张因果卷积提取的特征, 它包含来自更长历史范围的信息。通过注意力机制实现两种特征的加权融合, 能够减少重要信息的丢失。同时, 采用门控 TCN 网络 (GTCN, gated temporal convolutional network) 结构, 利用门控机制改进残差块的扩张因果卷积, 并将 TCN 中的 ReLU 函数更换为 SELU 函数, 进一步提高模型的动态特征提取能力。

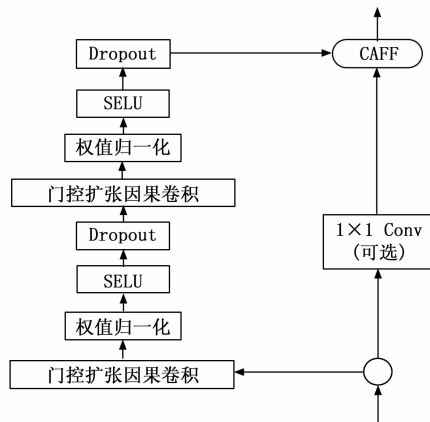


图 2 CAFF-GTCN 网络中改进残差块结构

2.1 级联注意力特征融合模块

CAFF 模块的结构如图 3 所示, 该模块采用自注意力机制 (SAM, self-attention mechanism) 和多尺度通道注意力机制^[25] (MS-CAM, multi-scale channel attention mechanism) 从时间和空间两个角度关注重要信息, 确保重要的信息在网络中向后传输。

在特征的初次融合阶段, SAM 被用于为序列中不

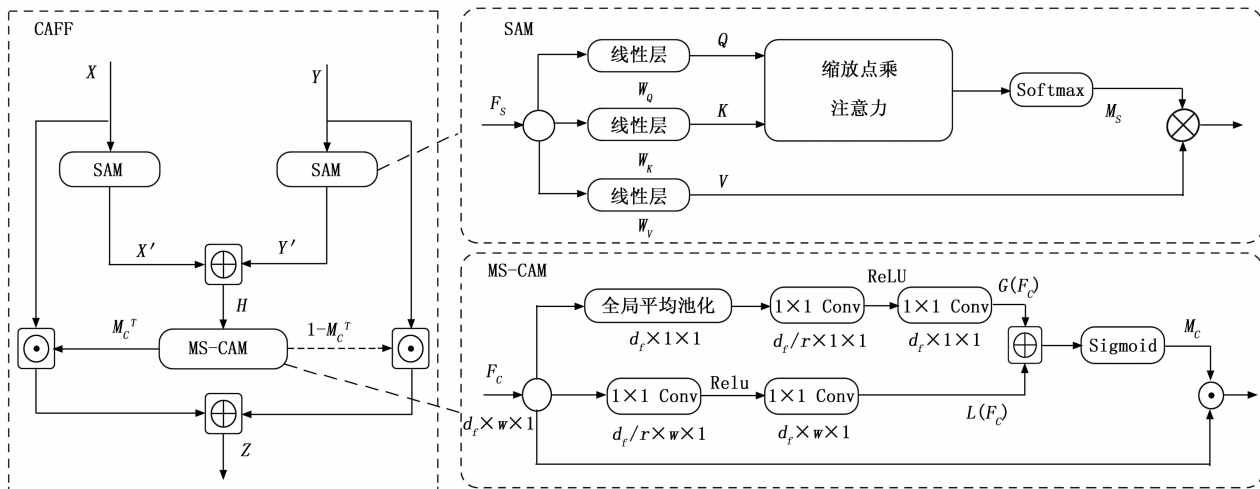


图 3 级联注意力特征融合模块

同位置的信息分配权重。SAM 通过计算向量 *queries* 和向量 *keys* 的相似度,并根据该相似度对向量 *values* 进行加权。对于给定输入特征 $F_S \in \mathbf{R}^{w \times d_f}$,其中 w 为序列长度, d_f 表示特征维度。首先对 F_S 进行线性变换,将其映射到不同的子空间中,并得到查询矩阵 Q 、键矩阵 K 、值矩阵 V :

$$\begin{cases} Q = F_S W_Q \\ K = F_S W_K \\ V = F_S W_V \end{cases} \quad (3)$$

其中: $W_Q \in \mathbf{R}^{d_f \times d_s}$ 、 $W_K \in \mathbf{R}^{d_f \times d_s}$ 、 $W_V \in \mathbf{R}^{d_f \times d_s}$ 为 3 个可训练的参数矩阵, Q 、 K 和 V 分别是由 *queries*、*keys* 和 *values* 构成的矩阵; d_s 表示向量 *queries* 和 *keys* 的维度。然后,通过缩放点乘注意力计算得到自注意力权重矩阵 M_S :

$$M_S = \text{attention}(Q, K, V) = \text{Softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right) \quad (4)$$

其中: d_k 用于防止出现过大的点积运算结果。Softmax 函数将 Q 与 K 之间的相似度归一化得到自注意力权重矩阵 M_S 。将 M_S 与 F_S 相乘得到由 SAM 加权后的特征 F'_S :

$$F'_S = M_S V \quad (5)$$

在 CAFF 模块中,两种特征 X 和 Y 分别经过 SAM 加权,对各自特征中不同位置的信息进行加权,降低对自身不重要信息的关注。在得到各自加权后的特征 X' 和 Y' 后,将两种特征相加得到初次融合后的中间特征 H :

$$H = X' + Y' = M_{S,X} V_X + M_{S,Y} V_Y \quad (6)$$

其中: $M_{S,X}$ 和 $M_{S,Y}$ 分别是特征 X 和 Y 对应的自注意力权重矩阵, V_X 和 V_Y 是相应的值矩阵。

在对 X 和 Y 进行初次特征融合时, SAM 从时间角度确定序列不同位置的注意力权重。为了获得精确的输入权重矩阵,还需要考虑特征的空间关系,将经过 SAM 得到的中间特征 H 进行转置得到 MS-CAM 的输入特征 $F_C \in \mathbf{R}^{d_f \times w \times 1}$,在 MS-CAM 模块中,中间特征 H 中的每一维特征单独作为一个通道进行处理, d_f 表示输入特征的通道数, $w \times 1$ 表示特征图的尺寸。MS-CAM 通过提取输入特征中全局特征 $G(F_C)$ 和局部特征 $L(F_C)$,并使用 Sigmoid 函数计算通道注意力权重矩阵 M_C :

$$g(F_C) = \frac{1}{w} \sum_{i=1}^w F_{C_{:,i,1}} \quad (7)$$

$$G(F_C) = \text{Conv}_2\{\text{ReLU}\{\text{Conv}_1[g(F_C)]\}\} \quad (8)$$

$$L(F_C) = \text{Conv}_2\{\text{ReLU}[\text{Conv}_1(F_C)]\} \quad (9)$$

$$M_C(F_C) = \text{Sigmoid}[G(F_C) \oplus L(F_C)] \quad (10)$$

其中: $g(F_C)$ 为全局平均池化操作, Conv_1 和 Conv_2 分别是卷积核数量为 w/r 和 w 的 1×1 卷积,它们被用

于缩减和增加通道数量, r 为通道的缩减率, \oplus 表示逐元素相加运算。MS-CAM 可以调整每个通道的重要性,即对特征图中不同通道的特征信息进行加权汇总,以增强重要特征通道的响应。通过权重矩阵 M_C 实现 X 和 Y 的融合,得到最终融合后的特征 Z :

$$Z = M_C^T \odot X + (1 - M_C^T) \odot Y \quad (11)$$

其中: \odot 表示点乘操作。CAFF 被用于改进 TCN 中的残差连接,通过反向传播算法对整个网络进行训练,确定每个残差块中 CAFF 模块的参数,保证网络中重要信息的传输。

2.2 门控 TCN 网络

为了增强网络的特征提取能力,采用一种门控 TCN 结构,引入门控机制构造门控扩张因果卷积层。如图 4 所示,门控扩张因果卷积利用 CNN 的并行计算能力,选用两层扩张因果卷积层,两个卷积层各自的卷积核参数权重不共享,一层用于特征的提取,另一层经过 Sigmoid 激活函数得到控制信息流通的门控信号。门控扩张因果卷积表达式为:

$$h_g = \sum_{i=0}^{k-1} f_1(i) \cdot x_{T-d,i} \quad (12)$$

$$h_f = \sum_{i=0}^{k-1} f_2(i) \cdot x_{T-d,i} \quad (13)$$

$$h_i = \text{Sigmoid}(h_g) \cdot (h_f), i = 1, 2, \dots, c \quad (14)$$

其中: h_g 作为门控信号层卷积核的输出, h_f 是特征提取层的卷积核提取结果, h_i 是最终经过门控卷积后得到的结果。卷积操作对应的门控信号越接近于 1,代表记忆下来的信息越多。而越接近于 0,则代表遗忘的信息越多。卷积核按照时间顺序滑动,最终门控卷积得到的输出为:

$$H_g = \begin{bmatrix} h_{1,1} & \dots & h_{w,1} \\ \vdots & \ddots & \vdots \\ h_{1,c} & \dots & h_{w,c} \end{bmatrix} \quad (15)$$

其中: c 表示扩张卷积层中卷积核的数量。

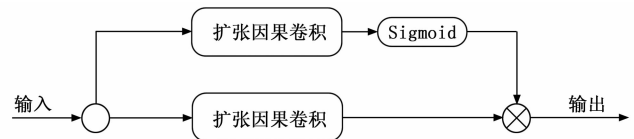


图 4 门控扩张因果卷积

此外,采用扩展型指数线性单元 (SELU)^[26] 替换网络中的 ReLU 激活函数,其数学表达式如下:

$$\text{SELU}(x) = \lambda \begin{cases} x & \text{if } x > 0 \\ \alpha e^x - \alpha & \text{if } x \leq 0 \end{cases} \quad (16)$$

其中: $\lambda \approx 1.05$, $\alpha \approx 1.67$ 。利用 SELU 函数将自归一化属性引入 TCN 网络中,能够在网络训练过程中自动调整各层输出的均值和方差,使其接近零均值和单位

方差。这种特性不仅加快了模型的收敛速度, 还能有效抑制梯度消失问题。此外, 根据式 (16) 可知, SELU 函数在输入为正时表现为线性变换, 而在输入为负时表现为指数变换, 这种非线性特性使得 SELU 函数能够更好地捕捉数据中的复杂模式和动态变化, 从而进一步提升模型的动态特征提取能力。

2.3 基于 CAFF-GTCN 网络的建模步骤

CAFF-GTCN 网络的软测量建模步骤如下:

1) 从工业过程中获取历史样本, 对数据集进行归一化处理, 并将数据集划分为训练集和测试集。

2) 根据先验知识确定输入序列长度, 并将输入样本扩展^[20]为 CAFF-GTCN 网络所需的时间序列数据。

3) 初始化 CAFF-GTCN 网络的结构以及相关超参数, 包括残差块数、卷积核尺寸和数量、扩张率及训练周期等。使用训练集进行模型训练, 利用 Adam 算法优化模型参数。

4) 在测试阶段, 将测试集辅助变量作为输入, 利用训练好的 CAFF-GTCN 模型得到相应预测值并使用均方根误差 (RMSE, root mean squared error)、平均绝对误差 (MAE, mean absolute error) 和决定系数 (R^2) 评估模型的预测性能。

3 仿真实验

此次实验采用青霉素发酵过程仿真数据和工业硫回收过程数据集对提出的 CAFF-GTCN 方法的有效性进行验证, 并与目前流行的 SVR、LSTM 和图卷积网络 (GCN, graph convolutional network)^[27] 的预测性能进行比较。同时, 分别建立了基于 TCN、GTCN 和 CAFF-GTCN 的软测量模型, 以验证提出的 CAFF 模块和 GTCN 网络的有效性。模型的搭建与训练环境为: CPU (12th Gen Intel® Core™ i5-12400 2.50 GHz)、16 GB 内存、Pytorch1.13.1 以及 Python 3.8 仿真环境。

3.1 青霉素发酵仿真

青霉素发酵过程是一种复杂的间歇反应工业过程, 其工艺流程如图 5 所示。本实验利用 Pensim 仿真平台获取发酵过程数据, 该平台以 Birol 模型为内核, 能够准确描述和模拟青霉素发酵过程的动态行为。

本实验以表 1 所示的 11 个变量^[28] 作为模型输入, 建立软测量模型对青霉素浓度进行预测。考虑到实际青霉素发酵过程中, 不允许频繁取样进行产品浓度检测, 以尽量减少对发酵环境的破坏。因此, 本次仿真实验中将采样时间设置为 1 小时, 发酵时长设置为 400 小时, 在不同的初始条件下生成 10 个批次数据。将前 9 个批次的 3 600 个样本数据作为训练集, 剩余 1 个批次的

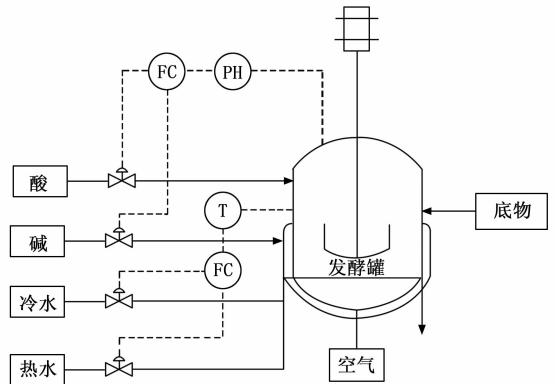


图 5 青霉素发酵过程流程图

400 个样本数据作为测试集。

表 1 青霉素发酵过程变量描述

变量	变量描述	变量	变量描述
x_1	搅拌功率	x_7	pH 值
x_2	通气速率	x_8	二氧化碳浓度
x_3	底物进料温度	x_9	反应罐温度
x_4	底物流加速率	x_{10}	冷却水流量
x_5	溶解氧浓度	x_{11}	碱液流量
x_6	培养体积		

为了让模型有效学习青霉素发酵过程的动态特性, 将 GCN、LSTM、TCN、GTCN 以及 CAFF-GTCN 模型的输入序列长度设置为 5^[28]。根据表 2 所示, 基于 TCN 模型的超参数取值范围, 通过交叉验证的方法确定卷积核尺寸 k 设置为 3, 扩张率 d 为 2, 残差块堆叠数为 3, 卷积核数量为 30。对于其他基于深度网络的软测量模型, 为了进行对比实验, 所有网络的超参数都应保持在与 TCN 相似的水平, LSTM 和 GCN 都有 3 个隐藏层, 每层 30 个神经元。所有方法使用 Adam 优化器进行训练, 损失函数为 MSE, 学习率为 0.001, 批量大小为 30, 训练周期为 50。

表 2 基于 TCN 模型的超参数取值范围

主要参数	取值范围
残差块堆叠数	1,2,3,4
卷积核尺寸	2,3,4
扩张率	1,2,3,4
卷积核数量	10,20,30,40,50
学习率	0.001,0.005,0.01
训练周期	50,100,150,200
批次大小	20,30,40,50

为了对比不同模型的预测性能, 表 3 提供了 6 种方法在测试集上的评估指标。由表可知, SVR 模型的 R^2 是 6 种方法中的最低值 0.975 6, RMSE 和 MAE 为最高值 0.075 5 和 0.065 5, 表明该模型的预测表现最差。通过

对比 R^2 、 $RMSE$ 和 MAE ，其他模型的预测表现均优于 SVR。此外，由于采样周期较长，获得的青霉素仿真数据中变量间的时间相关性较弱，因此 LSTM 的 R^2 高于 TCN 且为 0.993 8，表明 LSTM 具有更好的预测表现。通过改进 TCN 网络的结构，GTCN 增强了动态特征提取能力，其 $RMSE$ 和 MAE 分别为 0.029 3 和 0.023 7，相较于 TCN 降低了 31.1% 和 29.9%，提高了模型的预测精度。CAFF-GTCN 在改进的 GTCN 的基础上引入 CAFF 模块，进一步提高了模型预测精度，其 $RMSE$ 和 MAE 相较于 GTCN 模型分别降低了 45.1% 和 49.4%，并且 R^2 为 6 种方法的最大值 0.998 9，表明提出的 CAFF-GTCN 软测量模型的预测性能最好。

表 3 青霉素发酵过程中不同方法的评估指标

方法	$RMSE$	MAE	R^2
SVR	0.075 5	0.065 5	0.975 6
GCN	0.068 4	0.053 2	0.979 9
LSTM	0.038 0	0.029 6	0.993 8
TCN	0.042 5	0.033 8	0.992 3
GTCN	0.029 3	0.023 7	0.996 3
CAFF-GTCN	0.016 1	0.012 0	0.998 9

此外，图 6 展示了 3 种基于 TCN 的软测量模型的详细预测结果，图 7 呈现了模型的绝对预测误差箱线图。SVR、GCN 的箱线图的箱体宽度也较大，表明预测误差大且较为分散。从 TCN 到 CAFF-GTCN，预测曲线与真实值越加贴近，其箱体宽度逐渐变窄、中位数更接近于 0，代表预测精度逐渐提高，进一步验证了 CAFF 模块和 GTCN 的有效性。从图中可以得出，提出的 CAFF-GTCN 模型的预测表现最佳，因此它更适合描述过程动态，建立高精度的软测量模型。

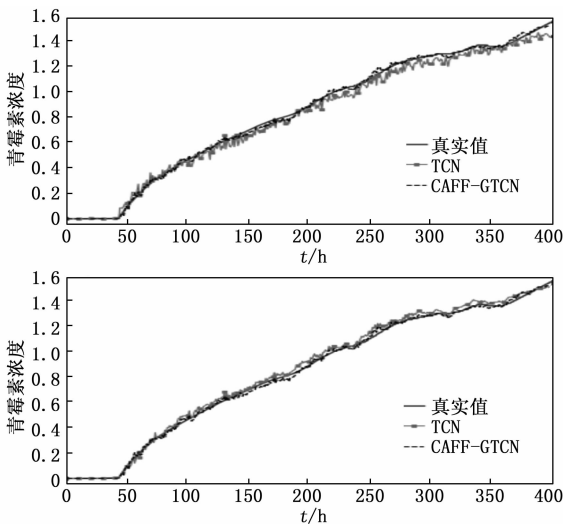


图 6 青霉素发酵过程中基于 TCN 的软测量模型预测表现

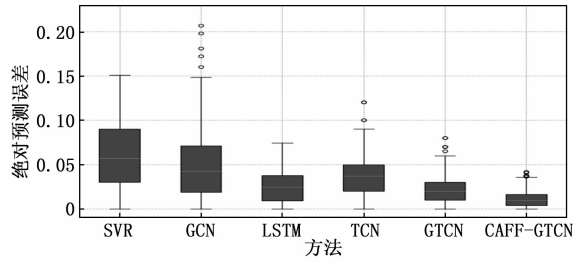


图 7 青霉素发酵过程中不同方法的绝对预测误差箱线图

3.2 硫回收过程实例

为了进一步验证所提方法的有效性，本小节建立了硫回收过程中尾气 SO_2 浓度的软测量模型。硫回收装置 (SRU, sulfur recovery unit) 是炼油厂的重要组成部分，其工艺流程如图 8 所示，相关过程变量如表 4 所示。SRU 的输入包括两种酸性气体，一种是 MEA 气体，其中含有大量的 H_2S 气体；另一种是 SWS 气体，其主要成分是 H_2S 和 NH_3 。MEA 气体和 SWS 气体分别被送往燃烧炉的主炉 B106 和副炉 B103 进行燃烧处理，并在烟气排放到大气之前，将纯硫作为副产品在装置中进行回收。

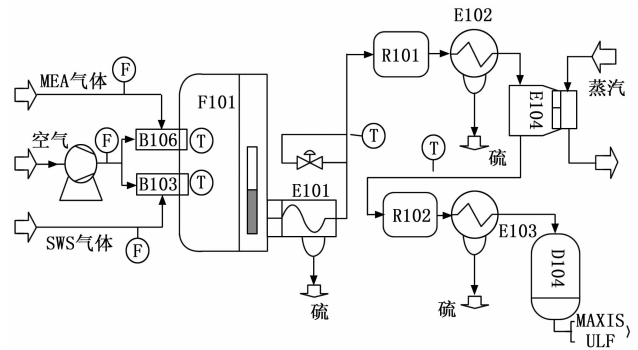


图 8 硫回收过程流程图

表 4 硫回收过程的变量描述

过程变量	相关描述
x_1	MEA 区域 GAS 气流
x_2	MEA 区域 AIR 空气流
x_3	MEA 区域 AIR 二次空气流
x_4	SWS 区域 GAS 气流
x_5	SWS 区域 AIR 空气流
y	SO_2

本次实验使用的数据集来源于文献 [29]，选用前 8 000 个样本构造训练集，剩余的 1 500 个样本作为测试集评估不同模型的预测性能。根据参考文献 [29] 对 SRU 过程中过程变量和质量变量的时间相关性的研究，将 GCN、LSTM、TCN、GTCN、CAFF-GTCN 网络中输入序列长度均设置为 15。由于过程变量与质量变量

间存在较大时滞, 为了保证实验的公平性, 对静态模型 SVR 的输入数据进行动态增广, 将前 14 个历史样本堆叠在当前样本上以形成一维动态向量:

$$\mathbf{x}_i(t) = [x_i(t), x_i(t-1), \dots, x_i(t-14)] \quad (17)$$

其中: $i=1, 2, \dots, 5$ 。与青霉素发酵过程中的软测量建模实验类似, 根据表 2 提供的超参数范围并通过交叉试验法确定 TCN、GTCN、CAFF-GTCN 的最佳网络结构, 将卷积核尺寸设置为 3, 对应卷积核数目为 40, 扩张率 d 为 3, 残差块数量为 3。同理, 作为对比模型的 LSTM 和 GCN 都有 3 个隐藏层, 每层 40 个神经元。所有模型的训练周期为 100, 并且批次浓度大小设置为 30, 设定学习率为 0.001。

表 5 提供了 6 种方法在测试集的评价指标。可以看出, 虽然对静态模型 SVR 的输入进行了动态扩展, 但 R^2 仍为最低值 0.303 7, 且 RMSE 和 MAE 都比其余动态模型高, 表明其动态预测性能较差。LSTM 使用递归结构描述历史信息, 当前时刻特征仅依赖于前一时间信息, 使得 LSTM 不能完全揭示硫回收过程中变量长期依赖关系。相比之下, TCN 使用扩张因果卷积来覆盖更广泛的时间范围以充分学习过程动态, 其 RMSE 和 MAE 低于 LSTM。GTCN 通过门控扩张因果卷积提高特征提取能力, RMSE 和 MAE 相较于 TCN 降低了 16.2% 和 8.5%, R^2 从 0.750 3 提升至 0.825 4。此外, CAFF-GTCN 能够同时关注变量间的长期和短期时间相关性, 其 R^2 最接近 1, RMSE 和 MAE 分别降低了 21.4% 和 17.8%, 表现出最佳的预测性能。

表 5 硫回收过程中不同方法的评估指标

方法	RMSE	MAE	R^2
SVR	0.048 8	0.036 4	0.303 7
GCN	0.031 9	0.026 6	0.691 1
LSTM	0.031 7	0.022 9	0.701 7
TCN	0.029 0	0.021 3	0.750 3
GTCN	0.024 3	0.019 5	0.825 4
CAFF-GTCN	0.022 8	0.017 5	0.846 4

TCN、GTCN 和 CAFF-GTCN 模型在硫回收过程测试集上的预测结果如图 9 所示, TCN 的预测曲线与真实值曲线之间存在较大误差, 表明其预测结果并不理想。对于 GTCN、CAFF-GTCN 模型, 由于其更为强大的动态特征学习能力, 整体上的预测误差明显低于其他方法。此外, 图 10 提供了 6 种方法的绝对预测误差箱形图, 箱形图从 SVR、GCN、LSTM、TCN、GTCN 到 CAFF-GTCN 逐渐变窄且更接近于 0, 表明预测误差愈发集中且更小。TCN 与 LSTM 的箱形图宽度近似, 但是 TCN 具有更低且更集中的绝对预测误差值, 表明 TCN 的整体预测误差比 LSTM 更低。CAFF-GTCN 的

箱线图宽度相较于 TCN 和 GTCN 更窄, 这验证了 CAFF 模块的有效性, CAFF-GTCN 更适合挖掘过程变量间的时间相关性, 建立预测精度更高的动态软测量模型。

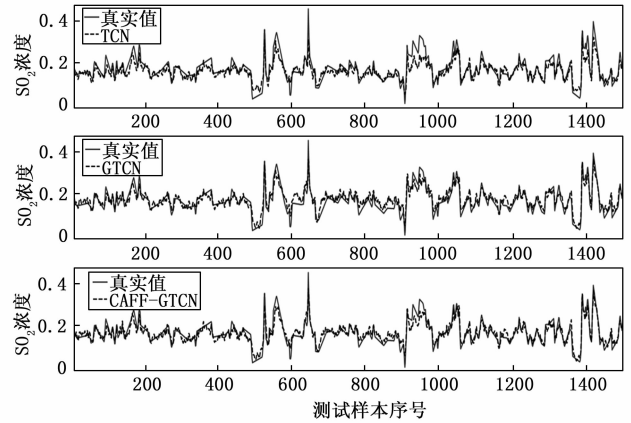


图 9 硫回收过程中基于 TCN 的软测量模型预测表现

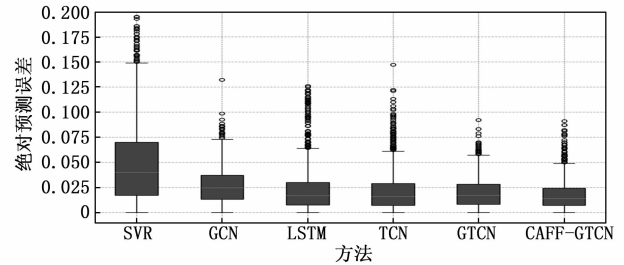


图 10 硫回收过程中不同方法绝对预测误差箱线图

4 结束语

针对复杂且未知的工业过程, 提出一种基于 CAFF-GTCN 的动态软测量方法。该方法通过设计级联注意力特征融合模块, 有效融合不同感受野的特征, 减少了关键局部信息的丢失。此外, 门控 TCN 结构增强了对历史信息的记忆和选择能力。实验结果表明, 在青霉素发酵和硫回收过程的动态建模中, CAFF-GTCN 模型的预测性能显著优于 SVR、GCN、LSTM 等主流软测量模型。在未来的工作中, 将研究扩张率和卷积核的自适应调整策略, 根据过程的动态特性自适应调节感受野大小, 以进一步提升模型的动态特征提取能力和预测性能。

参考文献:

[1] XIBILIA M G, LATINO M, MARINKOVIC Z, et al. Soft sensors based on deep neural networks for applications in security and safety [J]. IEEE Transactions on Instrumentation and Measurement, 2020, 69 (10): 7869 - 7876.

[2] 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.
- [3] WANG G, JIA Q S, ZHOU M C, et al. Artificial neural networks for water quality soft-sensing in wastewater treatment: a review [J]. *Artificial Intelligence Review*, 2022, 55 (1): 565–587.
- [4] 李雅晶, 辛妍丽. 基于 SVR 的燃煤机组 NOX 含量的软测量模型 [J]. *计算机测量与控制*, 2020, 28 (8): 62–66.
- [5] REN L, MENG Z, WANG X, et al. A data-driven approach of product quality prediction for complex production systems [J]. *IEEE Transactions on Industrial Informatics*, 2020, 17 (9): 6457–6465.
- [6] 孙凯, 隋璘, 张芳芳, 等. 基于非负纹杀与长短期记忆神经网络的动态软测量算法 [J]. *控制理论与应用*, 2023, 40 (1): 83–93.
- [7] LIU Y, XIE M. Rebooting data-driven soft-sensors in process industries: A review of kernel methods [J]. *Journal of Process Control*, 2020, 89: 58–73.
- [8] 黄成, 金怀平, 王彬, 等. 基于时空局部学习的集成自适应软测量方法 [J]. *仪器仪表学报*, 2023, 44 (1): 231–241.
- [9] 周阅昇, 熊伟丽. 基于迁移成分分析的发酵过程集成软测量建模 [J]. *系统仿真学报*, 2023, 35 (3): 623–631.
- [10] SHENG X, XIONG W. Soft sensor design based on phase partition ensemble of LSSVR models for nonlinear batch processes [J]. *Mathematical Biosciences and Engineering*, 2020, 17 (2): 1901–1921.
- [11] 单显明, 刘业峰, 那崇正, 等. 基于 ASOS-ELM 的球磨机负荷参数软测量系统设计 [J]. *计算机测量与控制*, 2022, 30 (10): 70–75.
- [12] YUAN X, QI S, WANG Y. Stacked enhanced auto-encoder for data-driven soft sensing of quality variable [J]. *IEEE Transactions on Instrumentation and Measurement*, 2020, 69 (10): 7953–7961.
- [13] 李祥宇, 隋璘, 马君霞, 等. 基于时序迁移与双流加权的 ONLSTM 软测量建模 [J]. *化工学报*, 2023, 74 (11): 4622–4633.
- [14] WANG K, SHANG C, LIU L, et al. Dynamic soft sensor development based on convolutional neural networks [J]. *Industrial & Engineering Chemistry Research*, 2019, 58 (26): 11521–11531.
- [15] JIANG Y, YIN S, DONG J, et al. A review on soft sensors for monitoring, control, and optimization of industrial processes [J]. *IEEE Sensors Journal*, 2020, 21 (11): 12868–12881.
- [16] 何罗苏阳, 熊伟丽. 助训练策略下的多模型软测量建模 [J]. *系统仿真学报*, 2024, 36 (1): 249–259.
- [17] 闫琳琦, 王振雷. 基于 STA-BiLSTM-LightGBM 组合模型的多步预测软测量建模 [J]. *化工学报*, 2023, 74 (8): 3407–3418.
- [18] GENG Z, CHEN Z, MENG Q, et al. Novel transformer based on gated convolutional neural network for dynamic soft sensor modeling of industrial processes [J]. *IEEE Transactions on Industrial Informatics*, 2021, 18 (3): 1521–1529.
- [19] SUN C, ZHANG Y, ZHAO H, et al. A soft sensor model for cement specific surface area based on TCN-ASRU neural network [J]. *IEEE Transactions on Instrumentation and Measurement*, 2023, 72: 1–12.
- [20] YUAN X, QI S, WANG Y, et al. Quality variable prediction for nonlinear dynamic industrial processes based on temporal convolutional networks [J]. *IEEE Sensors Journal*, 2021, 21 (18): 20493–20503.
- [21] SALEHI A, BALASUBRAMANIAN M. DDCNet: Deep dilated convolutional neural network for dense prediction [J]. *Neurocomputing*, 2023, 523: 116–129.
- [22] YANG L, LIU H, CHEN F. Just-in-time updating soft sensor model of endpoint carbon content and temperature in BOF steelmaking based on deep residual supervised autoencoder [J]. *Chemometrics and Intelligent Laboratory Systems*, 2022, 231: 104679.
- [23] TANG Y, WANG Y, LIU C, et al. Semi-supervised LSTM with historical feature fusion attention for temporal sequence dynamic modeling in industrial processes [J]. *Engineering Applications of Artificial Intelligence*, 2023, 117: 105547.
- [24] CHEN J, SONG P, ZHAO C, et al. Spatiotemporal multiscale correlation embedding with process variable re-order for industrial soft sensing [J]. *IEEE Transactions on Instrumentation and Measurement*, 2023, 72: 1–10.
- [25] ZHOU M, LANG S, ZANG T, et al. Attentional feature fusion for end-to-end blind image quality assessment [J]. *IEEE Transactions on Broadcasting*, 2022, 69 (1): 144–152.
- [26] LU Y, GOULD S, AJANTHAN T. Bidirectionally self-normalizing neural networks [J]. *Neural Networks*, 2023, 167: 283–291.
- [27] JIA M, XU D, YANG T, et al. Graph convolutional network soft sensor for process quality prediction [J]. *Journal of Process Control*, 2023, 123: 12–25.
- [28] 刘聪, 谢莉, 杨慧中. 基于改进 DPC 的青霉素发酵过程多模型软测量建模 [J]. *化工学报*, 2021, 72 (3): 1606–1615.
- [29] FORTUNA L, RIZZO A, SINATRA M, et al. Soft analyzers for a sulfur recovery unit [J]. *Control Engineering Practice*, 2003, 11 (12): 1491–1500.