



基于自注意力机制与卷积ONLSTM网络的软测量算法

李祥宇, 隋璘, 熊伟丽

引用本文:

李祥宇,隋,熊伟丽. 基于自注意力机制与卷积ONLSTM网络的软测量算法[J]. 智能系统学报, 2023, 18(5): 957–965.

LI Xiangyu, SUI Lin, XIONG Weili. Soft sensor algorithm based on self-attention mechanism and convolutional ONLSTM network[J]. *CAAI Transactions on Intelligent Systems*, 2023, 18(5): 957–965.

在线阅读 View online: <https://dx.doi.org/10.11992/tis.202211037>

您可能感兴趣的其他文章

用于关系抽取的注意力图长短时记忆神经网络

Attention graph long short-term memory neural network for relation extraction

智能系统学报. 2021, 16(3): 518–527 <https://dx.doi.org/10.11992/tis.202008036>

面向推荐系统的分期序列自注意力网络

Recommendation system with long-term and short-term sequential self-attention network

智能系统学报. 2021, 16(2): 353–361 <https://dx.doi.org/10.11992/tis.202005028>

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>

基于CNN-BLSTM的化妆品违法违规行为分类模型

Classification model for judging illegal and irregular behavior for cosmetics based on CNN-BLSTM

智能系统学报. 2021, 16(6): 1151–1157 <https://dx.doi.org/10.11992/tis.202104001>

层次化双注意力神经网络模型的情感分析研究

Hierarchical double-attention neural networks for sentiment classification

智能系统学报. 2020, 15(3): 460–467 <https://dx.doi.org/10.11992/tis.201812017>

基于双向消息链路卷积网络的显著性物体检测

Salient object detection based on bidirectional message link convolution neural network

智能系统学报. 2019, 14(6): 1152–1162 <https://dx.doi.org/10.11992/tis.201812003>

DOI: 10.11992/tis.202211037

网络出版地址: <https://kns.cnki.net/kcms2/detail/23.1538.TP.20230427.1801.004.html>

基于自注意力机制与卷积 ONLSTM 网络的软测量算法

李祥宇¹, 隋璘¹, 熊伟丽^{1,2}

(1. 江南大学 物联网工程学院, 江苏 无锡 214122; 2. 江南大学 轻工过程先进控制教育部重点实验室, 江苏 无锡 214122)

摘要: 针对实际工业过程的非线性和动态性特点, 并考虑过程变量中存在的冗余信息, 提出一种带自注意力机制的卷积有序神经元长短时记忆网络 (ordered neurons long short-term memory, ONLSTM) 多层时序预测模型。首先利用卷积神经网络降低局部特征维度, 对输入变量进行局部特征提取, 并通过构建层级重要性指标对长短时记忆网络 (long short-term memory, LSTM) 隐藏层神经元进行特定排序, 以辨识层级结构信息, 提高网络模型的重要信息判断能力; 其次将自注意力机制引入 ONLSTM 网络, 根据各输入变量之间内部相关性, 自适应地为其分配不同的注意力权重, 以提高模型预测性能; 最后将模型应用于青霉素发酵过程的产物浓度预测, 并与其他先进网络模型进行对比, 验证了模型的有效性。

关键词: 自注意力机制; 有序神经元长短时记忆网络; 软测量; 青霉素发酵; 特征提取; 卷积; 冗余信息; 深度学习中图分类号: TP274 文献标志码: A 文章编号: 1673-4785(2023)05-0957-09

中文引用格式: 李祥宇, 隋璘, 熊伟丽. 基于自注意力机制与卷积 ONLSTM 网络的软测量算法[J]. 智能系统学报, 2023, 18(5): 957-965.

英文引用格式: LI Xiangyu, SUI Lin, XIONG Weili. Soft sensor algorithm based on self-attention mechanism and convolutional ONLSTM network[J]. CAAI transactions on intelligent systems, 2023, 18(5): 957-965.

Soft sensor algorithm based on self-attention mechanism and convolutional ONLSTM network

LI Xiangyu¹, SUI Lin¹, XIONG Weili^{1,2}

(1. School of Internet of Things Engineering, Jiangnan University, Wuxi 214122, China; 2. Key Laboratory of Advanced Process Control for Industry (Ministry of Education), Jiangnan University, Wuxi 214122, China)

Abstract: According to the nonlinear and dynamic characteristics of actual industrial processes and considering the redundant information in process variables, this paper presents a multilayer time-series prediction model of convolutional ordered neurons long short-term memory network (ONLSTM) with a self-attention mechanism. First, the convolutional neural network is used to reduce the dimensions of local features, extract the specific local features of the input variables, and rank the neurons in the LSTM hidden layer specifically by constructing the hierarchical importance index to identify the hierarchical structure information and improve the ability of networks to judge important information of the network model. Second, the self-attention mechanism is introduced into the ONLSTM network. This mechanism dynamically assigns different attention weights to the input variables according to their internal correlation to improve the prediction performance of the model. Finally, the model is applied to predict product concentration in the penicillin fermentation process, following which it is compared with other advanced network models to verify the effectiveness of the proposed model.

Keywords: self-attention mechanism; ordered neurons long short-term memory; soft sensor; penicillin fermentation; feature extraction; convolution; redundant information; deep learning

收稿日期: 2022-11-15. 网络出版日期: 2023-04-28.

基金项目: 国家自然科学基金项目 (61773182); 国家重点研发计划子课题 (2018YFC1603705-03).

通信作者: 熊伟丽. E-mail: greenpre@163.com.

现代工业过程中, 关键质量和参数的准确监测, 对确保生产过程安全可靠及产品质量满足生产要求至关重要。然而实际生产过程中存在大量

难以直接或实时测量的过程变量,导致过程控制中难以通过闭环实时反馈信号,造成反馈控制失控^[1-2]。

软测量技术通过挖掘易测辅助变量与难测主导变量间的关系建立数学模型,以实现关键质量变量的监测^[3-4]。软测量建模一般可分为两类:基于机理分析建模与基于数据驱动建模。基于机理分析建模是通过生产过程中工艺机理进行分析以建立数学模型,建模过程复杂且难度较大。基于数据驱动建模则无须对工艺机理有深刻理解,只需要对生产过程中所采集的数据进行分析建模。常用的数据驱动建模方法有主成分回归^[5]、偏最小二乘回归^[6]、高斯过程回归^[7]、支持向量机^[8]及人工神经网络^[9]等机器学习模型。

深度学习作为机器学习的一个重要分支,在面对大数据量、深层特征信息时具有更出色的数据解析能力,其在数据驱动软测量中也拥有高度适应性^[10-11]。Hochreiter 等^[12]提出一种长短时记忆(long short-term memory, LSTM)网络,其作为一种常见的深度学习模型,在处理时间序列信息及提取时序特征问题时有着良好的建模性能。Yuan 等^[13]提出一种有监督 LSTM,并将其应用于青霉素发酵过程,可有效提取过程变量间的非线性动态特征。此外,为提高 LSTM 预测性能,研究者将其与其他算法模型相结合,组成混合预测模型。Zheng 等^[14]将卷积神经网络(convolutional neural networks, CNN)与 LSTM 相结合并应用于青霉素发酵软测量建模中,有效提升了模型对样本间相关特征的发掘能力。常树超等^[15]提出一种时空协同的图卷积 LSTM 网络,兼顾过程变量时序关系及质量变量空间关系,采用多通道独立学习方法对输入变量时空特征关系处理分析。孙凯等^[16]将非负熵杀算法与 LSTM 网络相结合,辨识过程变量的冗余信息,对网络输入权重进行压缩,剔除冗余变量,从而提高模型预测精度。

另一方面,为了更好地处理各种采样数据特征信息,研究者在提升 LSTM 相关模型性能的同时,对 LSTM 的网络结构也进行不断改进。Lui 等^[17]提出一种有监督双向 LSTM 网络,通过双向架构更深层地挖掘过程变量与质量变量间的非线性特征信息。Xie 等^[18]提出一种双流 λ 门控循环单元(gated recurrent unit, GRU)网络模型,通过双流结构,融合时间和动态因果流,更高效地提取学习特征。Shen 等^[19]通过对 LSTM 内部隐藏层神经元进行重要性排序,提出一种有序神经元长短时

记忆网络(ordered neurons long short-term memory, ONLSTM),并将其运用于自然语言处理和图像处理中。在软测量建模中,不同辅助变量对预测结果影响不同,因此保留高影响性关键特征尤为重要,而 ONLSTM 所引入的层级概念能够很好地满足所提要求。Shi 等^[20]将 ONLSTM 网络运用于工业故障检测中,有效解决了高速铣削过程中频繁发生的自激振动问题。

尽管在解决时间序列长期依赖的问题上,相关 LSTM 模型具备良好的应用效果,但是其难以高效专注于不同时间步长下的输入变量。针对这一问题, Bahdanau 等^[21]提出一种基于注意力机制的编码器-解码器网络,在序列建模中有着良好的应用性能,但面对长时间序列时,会造成重要信息丢失,导致模型性能随着时间序列的增长而下降。因此,将编码器-解码器模型替换为 LSTM 网络可以有效解决长期依赖的问题。当将 LSTM 网络与注意力机制相结合时,可以有效处理数据样本之间的动态性和相关性。Yuan 等^[22]提出一种时空注意力长短时记忆网络模型,不仅可以自适应处理输入特征,同时对数据动态特性也有很好的适用性。Li 等^[23]提出一种自注意力(self-attention, SA)机制与 GRU 网络的软测量建模方法,根据输入变量间依赖关系分配注意力权重,减少外部信息的影响,从而提高模型预测精度。

本文设计了一种结合自注意力机制的卷积有序神经元长短时记忆网络(SA-CNN-ONLSTM)混合软测量模型,并将其运用于青霉素发酵过程。所提模型首先利用 CNN 对输入变量进行局部特征提取并降维;然后将所提取特征输入多层 ONLSTM 网络进行时序特征提取,并通过层级判断输入变量的重要程度,过滤特征变量冗余信息;最后结合自注意力机制动态调整特征权重,利用变量间内部依赖关系,对高相关性变量赋予高权重,同时优化全连接层激活函数,以提高模型预测性能。

1 LSTM 网络的基本原理

深度学习神经网络模型中,由于传统递归神经网络(recurrent neural network, RNN)存在梯度消失问题,所以提出一种改进型 RNN,即 LSTM 网络。LSTM 在 RNN 基础上对其隐藏层神经元结构进行改进,通过引入门结构和记忆细胞,使得模型能够更好地处理时序过程中长期依赖的问题,从而有效解决 RNN 中梯度消失或爆炸的缺陷。LSTM 单元结构如图 1 所示。

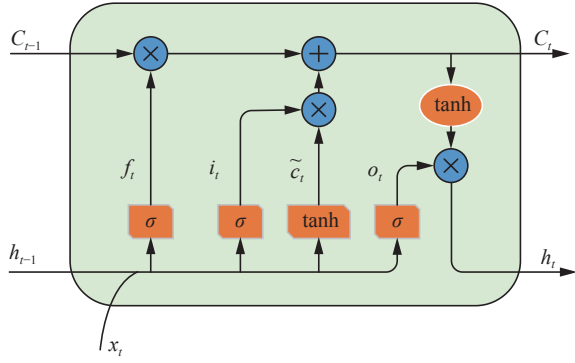


图 1 LSTM 结构图

Fig. 1 Structure of LSTM

在 LSTM 网络结构中神经元包含 3 个门结构和 1 个记忆细胞, 3 个门分别为遗忘门、输入门和输出门, 其在 LSTM 网络中起到判断输入信息重要程度的作用, 通过门的开关实现对时序信息的记忆, 从而防止梯度消失。对于一个 LSTM 神经元, 其外部输入为前一时刻的细胞状态 C_{t-1} 、前一时刻的隐藏状态 h_{t-1} 和当前时刻输入过程变量 x_t 。3 个门控制计算过程为

$$f_t = \sigma(W_{fx}x_t + W_{fh}h_{t-1} + b_f) \quad (1)$$

$$i_t = \sigma(W_{ix}x_t + W_{ih}h_{t-1} + b_i) \quad (2)$$

$$o_t = \sigma(W_{ox}x_t + W_{oh}h_{t-1} + b_o) \quad (3)$$

式中: f_t 、 i_t 、 o_t 分别为遗忘门、输入门和输出门的输出, σ 为激活函数, 本文选用 Sigmoid 函数, W_{f*} 、 W_{i*} 、 W_{o*} 和 b_f 、 b_i 、 b_o 分别为不同门所对应的权重和偏置值。

记忆细胞作为 LSTM 神经元中的重要组成部分, 能够将时序处理过程中相关信息的遗忘与记忆关系进行学习, 将筛选出的重要信息不断传递下去, 从而发挥挖掘样本时序特征的作用。记忆细胞的计算过程为

$$\tilde{C}_t = \tanh(W_{cx}x_t + W_{ch}h_{t-1} + b_c) \quad (4)$$

在 LSTM 中对隐含状态和细胞状态的更新计算过程为

$$C_t = f_t \circ C_{t-1} + i_t \circ \tilde{C}_t \quad (5)$$

$$h_t = o_t \circ \tanh(C_t) \quad (6)$$

式中: C_t 和 h_t 为更新后当前时刻的细胞状态和隐含状态, \circ 为矩阵相应元素乘积运算。

2 核心算法模型

2.1 ONLSTM 网络

ONLSTM 网络在传统 LSTM 网络基础上对隐藏层神经元顺序信息进行运算排序。相较于传统 LSTM 中神经元对所传递信息重要程度的判断不足, ONLSTM 通过引入信息层级概念, 构建信息重要性评价指标, 并按照信息层级的高低, 采

用分区更新的方式对输入信息进行取舍, 从而提高网络模型对信息重要性的判断能力。

ONLSTM 单元结构如图 2 所示。图中橙色、黄色和蓝色部分为对传统 LSTM 网络的改进部分, 分别表示主遗忘门通路、主输入门通路以及信息重叠部分通路, ONLSTM 的输入门、输出门和遗忘门的计算公式同传统 LSTM 网络计算公式相同, 不同之处在于对细胞状态 C_t 的更新函数进行了改进。

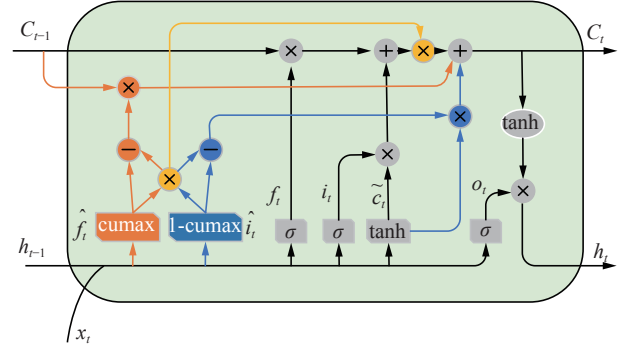


图 2 ONLSTM 结构

Fig. 2 Structure of ONLSTM

为了更准确地提取信息的分层结构, 在神经元中引入了新的激活函数:

$$\hat{g} = \text{cumax}(\cdot) = \text{cumsum}(\text{softmax}(\cdot)) \quad (7)$$

式中: $\text{cumax}()$ 表示累计和函数, 在此基础上在隐藏层神经元中引入“主遗忘门” \hat{f}_t 和“主输入门” \hat{i}_t , 计算公式为

$$\hat{f}_t = \text{cumax}(\text{softmax}(W_{fx}x_t + W_{fh}h_{t-1} + b_f)) \quad (8)$$

$$\hat{i}_t = 1 - \text{cumax}(\text{softmax}(W_{ix}x_t + W_{ih}h_{t-1} + b_i)) \quad (9)$$

式中: $W_{\hat{f}}$ 、 $W_{\hat{i}}$ 和 $b_{\hat{f}}$ 、 $b_{\hat{i}}$ 分别为主遗忘门和主输入门所对应的权重和偏置值。激活函数 \hat{g} 将细胞状态转换为二进制形式。在加入主遗忘门和主输入门后, 改进后的神经元内的细胞状态的更新规则为

$$\omega_t = \hat{f}_t \circ \hat{i}_t \quad (10)$$

$$\hat{f}_t' = f_t \circ \omega_t + (\hat{f}_t - \omega_t) = \hat{f}_t \circ (f_t \circ \hat{i}_t + 1 - \hat{i}_t) \quad (11)$$

$$\hat{i}_t' = i_t \circ \omega_t + (\hat{i}_t - \omega_t) = \hat{i}_t \circ (\hat{i}_t \circ \hat{f}_t + 1 - \hat{f}_t) \quad (12)$$

$$C_t = \hat{f}_t' \circ C_{t-1} + \hat{i}_t' \circ \tilde{C}_t \quad (13)$$

其中 ω_t 为主遗忘门和主输入门重叠部分。

2.2 自注意力机制

自注意力机制在序列建模中具有良好应用性能, 其利用软测量建模中时序信息的特征自身相关性, 直接计算特征内部依赖关系, 自适应确定输入特征权重的大小^[24], 同时与 LSTM 相结合, 对隐藏层输出特征进行优化, 使模型在长期依赖问题上有更好的性能表现。

自注意力机制中确定权重系数的计算过程为

$$u_a = \tanh(W h_t + b) \quad (14)$$

$$\theta = \text{softmax}(u_a^T u_r) = \frac{\exp(u_a^T u_r)}{\sum \exp(u_a^T u_r)} \quad (15)$$

$$s_t = \sum_{i=1}^n \theta_i h_i \quad (16)$$

式中: W 表示权重矩阵, h_t 表示当前时刻前一 LSTM 模型隐藏层输出, b 表示偏置值, u_a 表示输入隐藏层非线性变换层, u_r 表示输入序列向量, θ 表示自注意力特征权重, s_t 表示所赋不同概率注意力特征权重与各隐藏层输出相积之和。

3 基于 SA-CNN-ONLSTM 模型的软测量建模

本文所提 SA-CNN-ONLSTM 预测模型主要包括 4 层, 分别为 CNN 层、ONLSTM 层、自注意

力层和全连接层。考虑发酵过程具有动态性和非线性的特点, 首先选择引入 CNN 层, 通过卷积运算对输入样本数据局部特征提取并降维; 再在 CNN 层后叠入多层 ONLSTM 层, 利用其能够有效解决时序特征长期依赖问题的优点, 进一步提取样本时序特征, 解决发酵过程动态性的问题, 同时舍弃训练过程中冗余信息, 减小其对预测结果的影响; 进一步引入自注意力机制层, 通过自注意力机制, 更新特征权重, 调节 ONLSTM 神经元内隐藏层输出; 最后改进全连接层的激活函数, 减小数据非线性的影响, 提取整个模型的特征输出。所提模型结构如图 3 所示。

②有序神经元长短时记忆网络层 (ONLSTM layer)

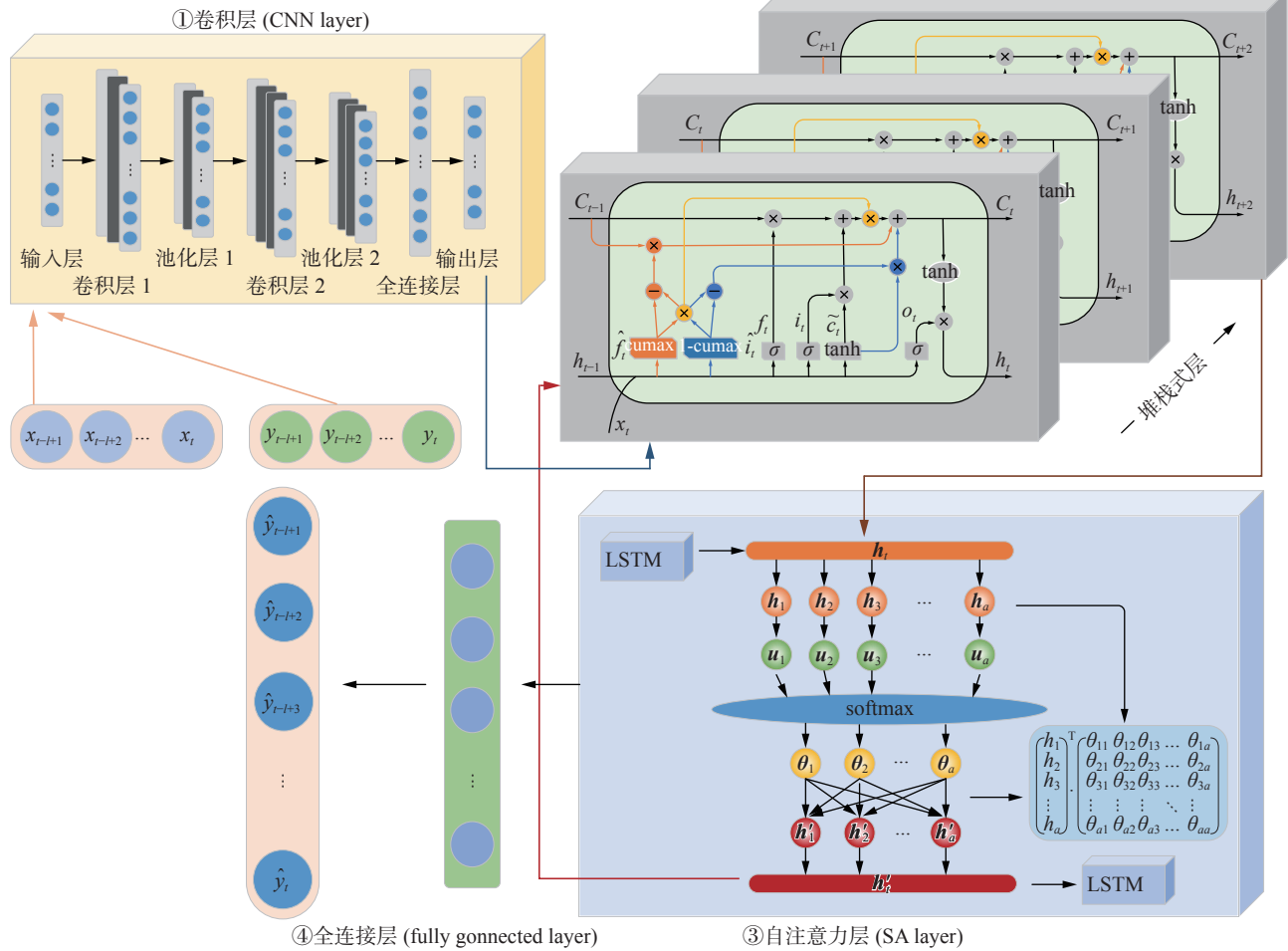


图 3 SA-CNN-ONLSTM 模型结构

Fig. 3 SA-CNN-ONLSTM model structure

3.1 时序特征处理

在处理时序特征问题时, 由于传统 LSTM 在参数更新过程中没有考虑神经元的顺序信息, 所以在判断神经元间信息重要程度关系上准确性不足, 使得后续神经元参数更新时对包含重要信息的神经元保留时间不足, 对包含次要信息的神经元保留时间又过长。

所提混合模型的 ONLSTM 层对 LSTM 隐藏层神经元内部结构进行改进, 引入层级概念来表示输入时序信息的重要程度。当面对输入样本具有数据量大且强非线性特征时, 变量间具有较高的关联性, 同时由于大数据量使得输入样本中存在着冗余变量影响着模型预测精度。而 ONLSTM 模型中引入层级的概念, 能够有效提取输入数据

中的重要信息,从而剔除冗余信息。层级在 ONLSTM 单元内部通过运算得到,通过引入主遗忘门和主输入门的方式,将输入到隐藏层神经元的时序特征信息于两个门内通过运算筛选分类,采用分区更新的方式将不同层级的信息传递到 ONLSTM 模型中的不同通道,从而对不同重要性的信息赋予相应的权值,将包含高重要性信息的神经元保留并传递下去。

所提模型中的 ONLSTM 层将多层 ONLSTM 相叠加,通过堆叠多层 ONLSTM,将前一层 ONLSTM 层的输出传递到后一层 ONLSTM 层,使其作为该 ONLSTM 的输入从而建立堆栈式 ONLSTM 层,利用堆叠多层 ONLSTM 层能够进一步提取输入变量的深层特征,与单层 ONLSTM 模型相比,堆栈式 ONLSTM 模型结构对输入时序特征的非线性特征具有更强的拟合能力,模型具有更强的稳健性。

3.2 相关特征处理

注意力机制的核心是通过输入特征与权重加权求和,判断输入特征对预测目标的重要程度。软测量建模中由于无法像自然语言处理任务中有额外信息,所以标准注意力机制在解决软测量建模的回归预测问题时仍有不足。因此改进标准注意力机制,得到自注意力机制,即所提混合模型中的自注意力机制层。

自注意力机制层在整个模型中起着输出特征优化作用。自注意力机制输入为前一 ONLSTM 层的隐藏层神经元输出,将自注意力机制与 ONLSTM 网络模型相结合,依据输入变量之间的内部依赖关系,对 ONLSTM 层的输出进行动态加权,对关键特征变量赋以高权重,表示该特征对模型预测结果的影响较大。自注意力机制以特征相关性为标准,对特征信息进行筛选,并将模型的预测重点聚焦于关键特征上。

3.3 算法流程

设定输入样本集为 $X=\{x_1, x_2, \dots, x_n\}$, 输出样本集为 $Y=\{y_1, y_2, \dots, y_n\}$, 基于 SA-CNN-ONLSTM 的软测量建模步骤如下:

- 1) 利用 Pensim 仿真平台采集辅助变量与主导变量,并对采集数据进行预处理;
- 2) 将预处理后的数据集划分为训练集 X_{train} 、 Y_{train} 和 X_{test} 、 Y_{test} ;
- 3) 利用训练输入样本 X_{train} 输入到 CNN 层提取样本局部特征 d ,再将 d 代入堆栈式 ONLSTM 层,并得到 ONLSTM 隐藏层输出 h_{train} ;
- 4) 将步骤 3 训练得到的 h_{train} 代入自注意力机制层,提取相关特征信息,并更新训练权重系数;
- 5) Y_{train} 作为训练输出样本集,利用 Adam 算

法对模型参数寻优,训练 SA-CNN-ONLSTM 模型;

- 6) 将测试输入样本集 X_{test} 和测试输出样本集 Y_{test} 代入到已经训练好的 SA-CNN-ONLSTM 模型中,对青霉素浓度进行预测,预测结果为 Y_{predict} ,并根据相应的评价指标评估模型预测性能。

4 仿真实验

青霉素发酵过程是一个复杂的生化反应过程,准确地预测青霉素浓度对整个发酵过程的优化控制具有重要意义^[25]。本文以青霉素发酵过程为案例,验证所提算法应用于软测量建模的优越性。青霉素发酵基本流程如图 4 所示。

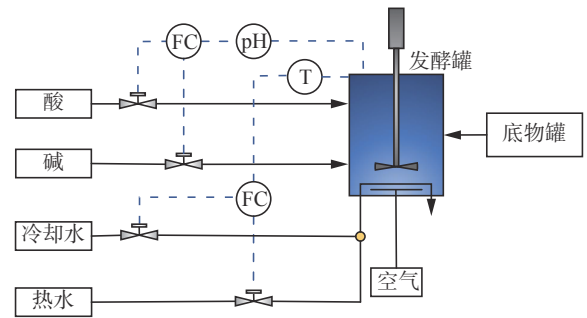


图 4 青霉素发酵过程流程

Fig. 4 Flow of penicillin fermentation process

本文中的数据通过 Pensim 仿真平台得到。该平台由 Cinar 教授针对实际青霉素发酵过程开发的仿真平台。通过该平台采集样本数据前,设定采样时间为 400 h,间隔时间为 0.2 h,即总共采样 20 000 组样本数据,其余设定值均采用平台默认设定值。本文选取曝气速率、 CO_2 浓度、pH 值、发酵罐温度等 11 个变量作为辅助变量,青霉素浓度作为关键质量变量。采样变量及设定值见表 1。

表 1 Pensim 采样变量及设定值

Table 1 Pensim sampling variables and set values

序号	采样变量	设定值
1	曝气速率/(L/h)	8.6
2	搅拌机功率/W	30
3	底物进料速率/(L/h)	0.042
4	底物供给温度/K	296
5	底物质量浓度/(g/L)	15
6	pH 值	5
7	发酵罐温度/K	298
8	CO_2 浓度/(mmole/L)	0.5
9	培养液体积/L	100
10	反应热量/K	—
11	冷水流加速率/(L/h)	—
12	青霉素浓度	—

在青霉素发酵过程仿真实验中,训练集和测试集分别取 5 000 组和 1 000 组数据。本文中采

用 SA-CNN-ONLSTM 模型作为青霉素浓度预测模型,在深度学习神经网络模型中,模型的参数对网络的性能起到至关重要的作用,所以设置合适的网络模型参数能够有效提高模型的预测精度。本文通过试验法不断调整模型的参数,从而选取实验效果最为理想的模型参数用于预测青霉素浓度。

本文中模型预测的评价指标选择均方根误差 $RMSE(E_{RMSE})$, 平均绝对误差 $MAE(E_{MAE})$ 以及相关指数 $R^2(E_{R^2})$ 。其计算公式如下:

$$E_{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - y_{i(pre)})^2} \quad (17)$$

$$E_{MAE} = \frac{1}{n} \sum_{i=1}^n |y_i - y_{i(pre)}| \quad (18)$$

$$E_{R^2} = 1 - \frac{\sum_{i=1}^n (y_i - y_{i(pre)})^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (19)$$

在 SA-CNN-ONLSTM 模型的 ONLSTM 层中,时间步长决定着每个输入滑动子窗口的序列长度,影响着模型预测精度,本文选取时间步长候选集 $\{3,4,5,6,7,8,9,10\}$,不同时间步长下模型预测的 RMSE 如图 5 所示。当时间步长为 6 时,模型预测精度最高。其余参数通过多次实验确定,各层参数设置见表 2,当设定 CNN 部分的基本参数过滤器个数 filters 为 10,层数为 1,ONLSTM 部分每层 ONLSTM 的隐藏层神经元个数为 100 和 50,层数为 2 时,预测青霉素浓度的结果最优。

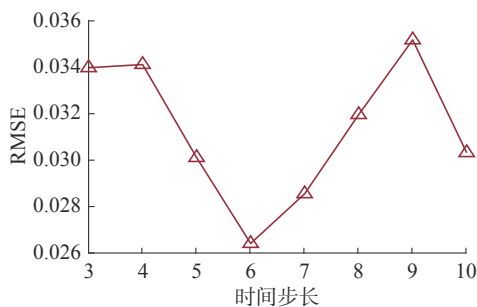


图 5 各时间步长下 SA-CNN-ONLSTM 模型的 RMSE
Fig. 5 RMSE of SA-CNN-ONLSTM model at each time step

由于青霉素发酵过程具有非线性特点,而激活函数将非线性特性引入到神经网络之中,通过对神经元加以非线性特性,可使神经网络模型更好地逼近非线性特征函数。选取合适的全连接层激活函数可有效筛选重要输入神经元信息,以提高模型预测精度。全连接层中较为常见的激活函数为: Sigmoid、ReLU、Elu、Leaky ReLU 和 Selu,将

其分别代入模型全连接层中,选择性能最优的作为全连接层激活函数。表 3 为几种常见激活函数的性能对比,根据评价指标结果,本文选用 Selu 作为全连接层激活函数,模型预测效果最优。

表 2 SA-CNN-ONLSTM 模型参数设定值
Table 2 SA-CNN-ONLSTM model parameter settings

参数名称	设定值
CNN 层数	1
CNN 层 filters	10
ONLSTM 层数	2
ONLSTM 隐藏层神经元个数	100/50
学习率	0.005
时间步长	6
自注意力层 unit	11

表 3 不同激活函数对模型性能的评价指标
Table 3 Evaluation index of model performance by different activation functions

激活函数	RMSE	MAE	R ²
Sigmoid	0.026 85	0.018 84	0.996 01
ReLU	0.018 42	0.015 16	0.997 91
Elu	0.023 90	0.019 32	0.996 84
Leaky ReLU	0.021 74	0.016 52	0.997 39
Selu	0.015 51	0.013 54	0.998 40

自注意力机制作为一种特殊的注意力机制形式,其旨在根据输入变量自身内部特征相关性对输入变量进行加权赋值,高相关度的输入变量将被赋予高权重。在本文 SA-CNN-ONLSTM 模型中以自注意力层的形式对 ONLSTM 层的隐藏层输出加权。在青霉素发酵过程仿真实验中,设置自注意力层神经元数量与辅助变量维度相等。实验中,自注意力层各输入样本自注意力权重如图 6 所示,在不同神经元下的输入变量所赋权重动态调整,通过对权重取平均值,得出不同神经元下输入变量所赋权重占比,结果如图 7 所示。输入变量 5 的自注意力权重占比最低,而输入变量 4、6、7 占比较高,即输入变量 5 对青霉素浓度的预测结果影响最小,而输入变量 4、6、7 影响较大。

为了更好地验证所提 SA-CNN-ONLSTM 算法在青霉素浓度预测中的性能优越性,本文将该算法分别与 LSTM、CNN-LSTM、CNN-ONLSTM、AT-CNN-ONLSTM 和 SA-CNN-ONLSTM 算法模型进行对比,比较 5 种算法模型对青霉素浓度的预测效果,实验结果如图 8、9、10 和表 4 所示。由图 8 可知,SA-CNN-ONLSTM 算法模型随着训

练次数的增加, 损失值下降速度最快, 即模型收敛性最优。

根据实验仿真结果可知, 当将 LSTM 与 CNN 结合时, 利用 CNN 对过程变量的局部依赖关系进行处理, 提取局部特征后可提高模型预测精度。在此基础上通过改变 LSTM 神经元内部结构, 利用层级对隐藏层神经元进行运算排序并分区更新, 构成 ONLSTM 算法, 可有效剔除时序特征中的冗余特征, 并将高重要度的神经元信息长时间保留, 从而进一步提高预测精度。

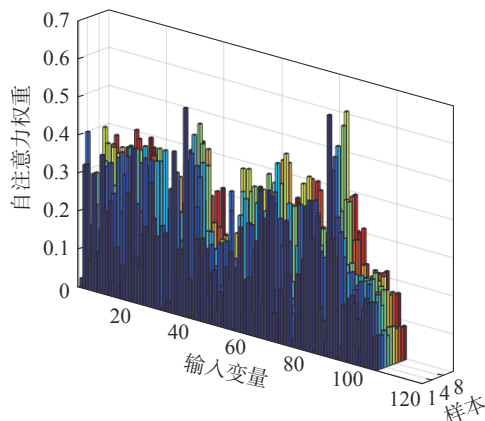


图 6 自注意力层各输入样本自注意力权重

Fig. 6 Self attention weight of each input sample in the self attention layer

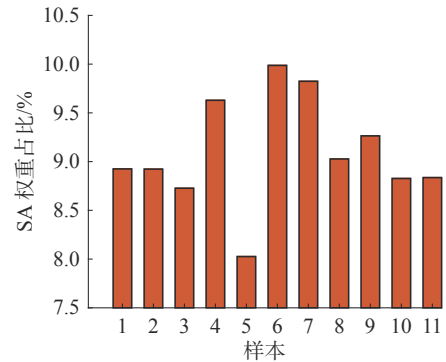


图 7 自注意力层各样本自注意力权重占比

Fig. 7 Proportion of self attention weight of each sample in the self attention layer

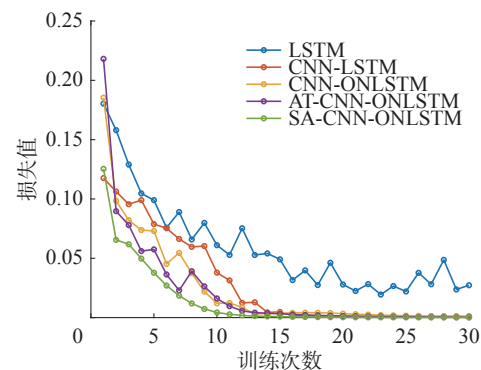


图 8 各网络模型损失值变化曲线

Fig. 8 Change curves of loss value of each network model

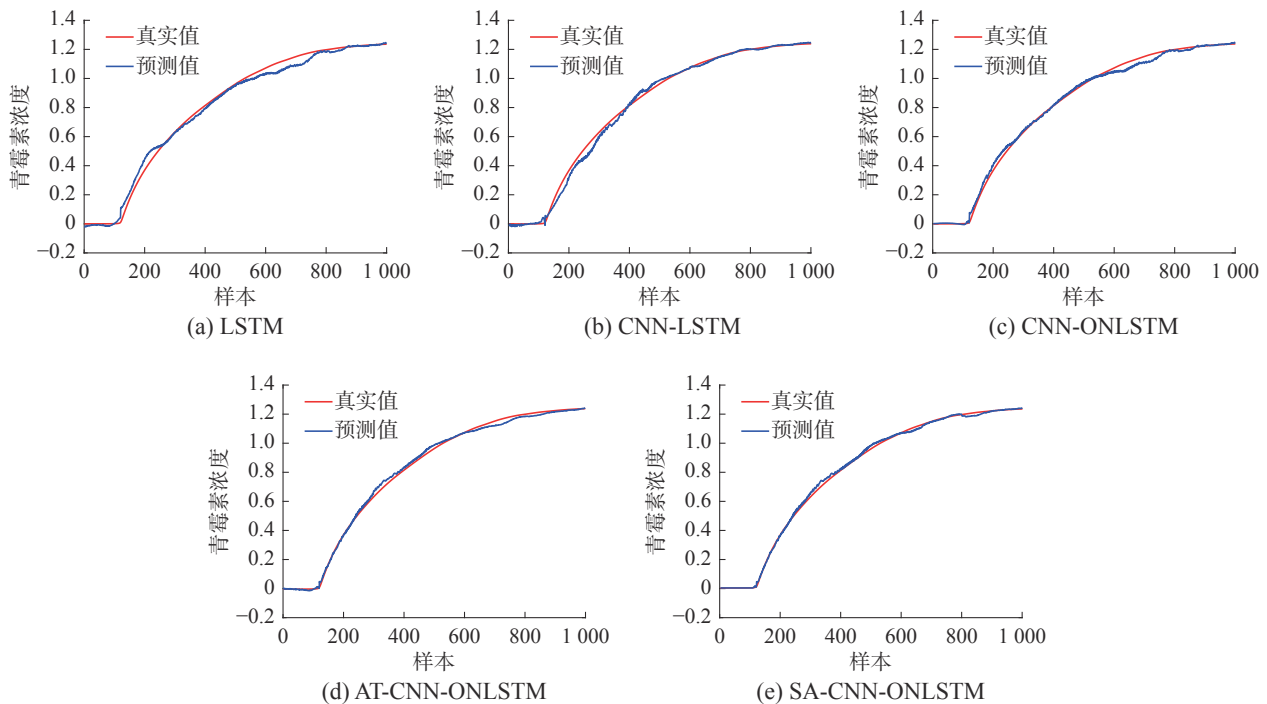


图 9 不同模型下青霉素浓度预测拟合曲线

Fig. 9 Penicillin concentration prediction curve under different models

以上 3 种模型验证了采用卷积 ONLSTM 能够减小冗余变量对模型预测性能的影响, 在此基

础上引入注意力机制。通过该机制能有效提取 CNN-LSTM 所提取信息中的重要特征, 提高该特

征所赋权重,从而增强此特征对预测结果的影响。实验结果显示,采用 AT-CNN-ONLSTM 算法模型的 RMSE 为 0.01974,高于前面 3 种神经网络算法模型。而当应用自注意力机制,对输入特征的相关性进行处理,更加准确地对相关特征加权,实验结果表明,采用 SA-CNN-ONLSTM 模型进行预测时, RMSE 为 0.01551,为 5 种算法模型中最小,预测精度达到最高。

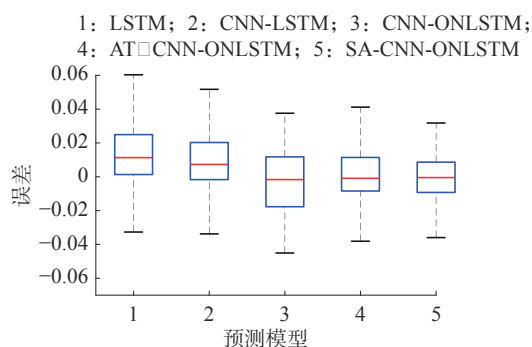


图 10 各网络模型对青霉素浓度预测误差箱线图

Fig. 10 Box chart of prediction error of penicillin concentration by each network model

表 4 各网络模型对青霉素浓度预测结果

Table 4 Prediction results of penicillin concentration by various network models

方法	RMSE	MAE	R^2
LSTM	0.033 09	0.025 73	0.993 36
CNN-LSTM	0.027 50	0.019 95	0.995 82
CNN-ONLSTM	0.023 92	0.018 63	0.996 84
AT-CNN-ONLSTM	0.019 74	0.015 02	0.997 89
SA-CNN-ONLSTM	0.015 51	0.013 54	0.998 40

5 结束语

本文提出一种自注意力机制与卷积 ONLSTM 相结合的混合预测模型,以青霉素发酵过程为案例进行软测量建模,解决实际工业过程中非线性动态建模问题。所提模型通过 CNN 对样本数据的局部特征进行特征提取和降维,并利用 ONLSTM 网络提取样本数据时序特征,根据层级重要性评价指标过滤冗余信息。另一方面,通过引入自注意力机制动态调节权重系数,以提高重要变量特征权重,增强模型预测性能。实验结果表明,与其他深度学习模型相比,所提模型能有效降低输入冗余信息对模型的影响,具有更高的预测精度,可为实际工业过程的优化控制提供有力的技术支撑。

参考文献:

[1] KADLEC P, GABRYS B, STRANDT S. Data-driven soft

sensors in the process industry[J]. *Computers & chemical engineering*, 2009, 33(4): 795–814.

[2] YUAN Xiaofeng, WANG Yalin, YANG Chunhua, et al. Weighted linear dynamic system for feature representation and soft sensor application in nonlinear dynamic industrial processes[J]. *IEEE transactions on industrial electronics*, 2018, 65(2): 1508–1517.

[3] WANG Jie, ZHAO Chunhui. Mode-cloud data analytics based transfer learning for soft sensor of manufacturing industry with incremental learning ability[J]. *Control engineering practice*, 2020, 98: 104392.

[4] 祁成, 史旭东, 熊伟丽. 基于二阶相似度的即时学习软测量建模方法 [J]. *智能系统学报*, 2020, 15(5): 910–918.

QI Cheng, SHI Xudong, XIONG Weili. A just-in-time learning soft sensor modeling method based on the second-order similarity[J]. *CAAI transactions on intelligent systems*, 2020, 15(5): 910–918.

[5] ZHOU Ping. Autoencoder and PCA based RVFLNs modeling for multivariate molten iron quality in blast furnace ironmaking[J]. *Zidonghua Xuebao/acta automatica sinica*, 2018, 44(10): 1799–1811.

[6] WANG Zi xiu, HE Q P, WANG Jin. Comparison of variable selection methods for PLS-based soft sensor modeling[J]. *Journal of process control*, 2015, 26: 56–72.

[7] SHENG Xiaochen, MA Junxia, XIONG Weili. Smart soft sensor design with hierarchical sampling strategy of ensemble Gaussian process regression for fermentation processes[J]. *Sensors*, 2020, 20(7): 1957.

[8] 赵超, 李俊, 戴坤成, 等. 基于自适应加权最小二乘支持向量机的青霉素发酵过程软测量建模 [J]. *南京理工大学学报*, 2017, 41(1): 100–107.

ZHAO Chao, LI Jun, DAI Kuncheng, et al. Soft sensor modeling for penicillin fermentation process based on adaptive weighted least squares support vector machine[J]. *Journal of Nanjing University of Science and Technology*, 2017, 41(1): 100–107.

[9] SUN Kai, HUANG S H, WONG D S H, et al. Design and application of a variable selection method for multilayer perceptron neural network with LASSO[J]. *IEEE transactions on neural networks and learning systems*, 2017, 28(6): 1386–1396.

[10] SUN Qingqiang, GE Zhiqiang. A survey on deep learning for data-driven soft sensors[J]. *IEEE transactions on industrial informatics*, 2021, 17(9): 5853–5866.

[11] SHANG Chao, YANG Fan, HUANG Dexian, et al. Data-driven soft sensor development based on deep learning technique[J]. *Journal of process control*, 2014, 24(3): 223–233.

[12] HOCHREITER S, SCHMIDHUBER J. Long short-term

- memory[J]. *Neural computation*, 1997, 9(8): 1735–1780.
- [13] YUAN Xiaofeng, LI Lin, WANG Yalin. Nonlinear dynamic soft sensor modeling with supervised long short-term memory network[J]. *IEEE transactions on industrial informatics*, 2020, 16(5): 3168–3176.
- [14] ZHENG Jiaqi, MA Lianwei, WU Yi, et al. Nonlinear dynamic soft sensor development with a supervised hybrid CNN-LSTM network for industrial processes[J]. *ACS omega*, 2022, 7(19): 16653–16664.
- [15] 常树超, 赵春晖. 一种时空协同的图卷积长短期记忆网络及其工业软测量应用[J]. *控制与决策*, 2022, 37(1): 77–86.
- CHANG Shuchao, ZHAO Chunhui. A spatio-temporal synergistic graph convolution long short-term memory network and its application for industrial soft sensors[J]. *Control and decision*, 2022, 37(1): 77–86.
- [16] 孙凯, 隋璘, 张芳芳, 等. 基于非负交叉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.
- SUN Kai, SUI Lin, ZHANG Fangfang, et al. Dynamic soft sensor algorithm based on nonnegative garrote and long short-term memory neural network[J]. *Control theory & applications*, 2023, 40(1): 83–93.
- [17] LUI C F, LIU Yiqi, XIE Min. A supervised bidirectional long short-term memory network for data-driven dynamic soft sensor modeling[J]. *IEEE transactions on instrumentation and measurement*, 2022, 71: 1–13.
- [18] XIE Ruimin, HAO Kuangrong, HUANG Biao, et al. Data-driven modeling based on two-stream λ gated recurrent unit network with soft sensor application[J]. *IEEE transactions on industrial electronics*, 2020, 67(8): 7034–7043.
- [19] SHEN YIKANG, TAN S, SORDONI A, et al. Ordered neurons: integrating tree structures into recurrent neural networks[EB/OL]. (2019–05–08)[2022–01–01]. <https://arxiv.org/abs/1810.09536>.
- [20] SHI Fei, CAO Hongrui, WANG Yuke, et al. Chatter detection in high-speed milling processes based on ONLSTM and PBT[J]. *The international journal of advanced manufacturing technology*, 2020, 111(11): 3361–3378.
- [21] BAHDANAU D, CHO K, BENGIO Y. Neural machine translation by jointly learning to align and translate[EB/OL]. (2016–05–19)[2022–01–01]. <https://arxiv.org/abs/1409.0473>.
- [22] YUAN Xiaofeng, LI Lin, SHARDT Y A W, et al. Deep learning with spatiotemporal attention-based LSTM for industrial soft sensor model development[J]. *IEEE transactions on industrial electronics*, 2021, 68(5): 4404–4414.
- [23] LI Lin, WANG Yalin, YUAN Xiaofeng, et al. Quality prediction model for process sequential data of irregular measurements with sampling-interval-attention LSTM [C]//2020 Chinese Automation Congress. Shanghai: IEEE, 2021: 7186–7191.
- [24] ZHU Xiuli, HAO Kuangrong, XIE Ruimin, et al. Soft sensor based on eXtreme gradient boosting and bidirectional converted gates long short-term memory self-attention network[J]. *Neurocomputing*, 2021, 434: 126–136.
- [25] GOPAKUMAR V, TIWARI S, RAHMAN I. A deep learning based data driven soft sensor for bioprocesses[J]. *Biochemical engineering journal*, 2018, 136: 28–39.

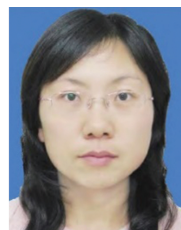
作者简介:



李祥宇, 硕士研究生, 主要研究方向为复杂工业过程建模。



隋璘, 博士研究生, 主要研究方向为机器学习、软测量建模。



熊伟丽, 教授, 博士生导师, 主要研究方向为基于数据挖掘、机器学习和大数据解析的复杂工业过程建模、控制及优化, 智能软测量技术与应用, 以及面向酿造过程、污水处理过程的智能自动化系统设计与开发。主持国家自然科学基金面上项目、国家自然科学基金青年项目、中国博士后基金项目、江苏省产学研前瞻性研究项目等省部级以上项目近 10 项。作为骨干人员参与完成国家重点研发计划课题、国家 863 计划等 7 项。获得江苏省科学技术二等奖 1 项、中国商业联合会科技进步一等奖、中石化自动化应用协会科技进步一等奖共 3 项。已授权发明专利 27 项, 其中国际发明专利 4 项, 以第一/责任作者发表学术论文近百篇。