

DOI: 10.13382/j.jemi.B2407570

基于 CNN-STA-DLSTM 模型的间歇过程质量预测*

惠永永^{1,2} 孙凯文^{1,2} 脱奔奔^{1,2} 陈鹏^{2,3} 赵小强^{1,2}

(1. 兰州理工大学电气工程与信息工程学院 兰州 730050; 2. 兰州理工大学国家级电气与控制工程
实验教学中心 兰州 730050; 3. 兰州石化职业技术大学电子电气工程学院 兰州 730060)

摘要:对于间歇过程变量深层特征提取困难,以及变量的时序性、非线性、动态特性所导致质量预测精度不高的问题,提出了一种基于卷积-时空注意力的双层长短期神经网络(convolutional neural networks spatial and temporal attention with double long short term memory networks, CNN-STA-DLSTM)的间歇过程质量预测模型。首先,对间歇过程数据沿着变量的方向展开成二维矩阵,对二维数据采用 Max-Min 法归一化,接着,使用 PLS 对原始数据降维,保留与质量变量相关性较强的变量,使用 CNN 挖掘过程数据的潜在特征,提高质量相关特征信息的关注;其次,引入时间注意力机制和空间注意力机制构建双层 LSTM 的编码器-解码器结构网络,利用注意力机制自适应地学习时间步长的相关历史信息,以提高模型的长期记忆能力,并加强过程变量与质量变量之间的时空相关性;然后,采用随机-网格搜索法寻找预测模型合适的超参数,并构建了预测模型;最后,使用青霉素发酵仿真平台和带钢热连轧生产过程数据进行实验验证,结果表明所提模型具有更精准的预测效果。

关键词: 间歇过程; 质量预测; 卷积神经网络; 双层长短期神经网络; 随机-网格搜索法

中图分类号: TP277; TN06 **文献标识码:** A **国家标准学科分类代码:** 510.80

Batch process quality prediction based on CNN-STA-DLSTM model

Hui Yongyong^{1,2} Sun Wenkai^{1,2} Tuo Benben^{1,2} Chen Peng^{2,3} Zhao Xiaoqiang^{1,2}

(1. School of Electrical Engineering and Information Engineering, Lanzhou University of Technology, Lanzhou 730050, China;
2. National Experimental Teaching Center of Electrical and Control Engineering, Lanzhou University of Technology,
Lanzhou 730050, China; 3. School of Electronic and Electrical Engineering, Lanzhou Petrochemical
University of Vocational Technology, Lanzhou 730060, China)

Abstract: For the difficulty in extracting deep features of batch process variables, as well as low quality prediction accuracy caused by the temporal, nonlinear, and dynamic characteristics of variables, this article proposes a quality prediction model for batch processes based on convolutional neural networks spatial and temporal attention with double long short term memory networks (CNN-STA-DLSTM). Firstly, the three-dimensional data of the batch process are expanded into a two-dimensional matrix along the direction of the variables, and the two-dimensional data are normalized by the Max-Min method. Then, the partial least squares (PLS) method is used to reduce the dimension of the original data, and the variables with strong correlation with the quality variables are retained. The convolutional neural network (CNN) is used to mine the potential features of the process data and improve the attention of the quality-related feature information. Secondly, the temporal attention mechanism and the spatial attention mechanism are introduced to construct the encoder-decoder structure network of the double-layer LSTM, and the attention mechanism is used to adaptively learn the relevant historical information of the time step, so as to improve the long-term memory ability of the model and strengthen the spatio-temporal correlation between the process variables and the quality variables. Then, the random-grid search method is used to optimize the hyperparameters of the prediction model, and the prediction model is constructed. Finally, the penicillin fermentation simulation platform and the hot strip rolling production process data are used for experimental verification. The results show that the proposed model has more

收稿日期: 2024-06-03 Received Date: 2024-06-03

* 基金项目: 国家自然科学基金(62263021)、兰州市青年科技人才创新项目(2023-QN-36)、甘肃省高校青年博士支持项目(2024QB-037)资助

accurate prediction effect.

Keywords: batch process; quality prediction; convolutional neural network; double-layer LSTM; random-grid search method

0 引言

随着现代社会生活水平的提高,人们对产品的种类、规格及产品的需求越来越高^[1],由于间歇过程具有小批量生产、多种类、批次重复运行以及生产灵活性强等优势,已被广泛的应用于食品、医药、化工和半导体的生产等众多领域^[2]。为了确保产品质量和生产过程的安全性及稳定性,对间歇过程的质量进行在线预测显得尤为重要^[3]。但是,在一些生产过程中,产品质量难以通过实时在线检测来实现,工作人员需要通过实验室化验进行检测,这不仅给操作人员带来了额外的工作量,而且传统检测方法具有滞后性不能及时准确的指导生产,使生产成本增大,不利于企业的发展^[4-5]。因此,为提升产品质量和生产效益,实时在线的质量预测已成为专家学者研究的焦点。

随着工业 4.0 时代的到来,分布式控制系统 (distributed control system, DCS) 和智能化仪表被广泛应用于工业生产过程中,促使大量的过程数据被检测和记录^[6],因此,许多基于数据驱动的方法被广泛的应用于间歇过程的建模、监测及质量预测中^[7-11]。偏最小二乘法^[12] (partial least squares, PLS) 是质量预测模型中的典型代表,然而,传统的 PLS 方法仅处理线性数据,无法有效捕捉系统的非线性特征,为了解决上面的问题,文献[13]提出一种核函数方法对 PLS 进行改进的核偏最小二乘算法 (kernel partial least squares, KPLS),以用于处理质量预测过程中数据的非线性特性;此外,文献[14]提出一种新的基于特征空间 (feature space, FS) 的 KPLS,用于对间歇过程的质量数据进行在线软测量。文献[15]提出了一种基于功能数据描述 (functional data description, FDD) 的新型批处理过程监控框架,该框架将批处理过程中每个变量随时间变化的轨迹视为一个典型的函数基准。但这些方法难以有效利用现代工业生产中所产生的海量数据,无法有效提取间歇过程数据中隐含的深层时序特征^[16],导致在实时质量预测时预测精度不高。因此,研究者们将目光聚集到了深度学习相关的质量预测方法。

深度学习具有强大的特征提取能力,出色的模式识别和自适应能力,能够从海量的数据中提取关键信息。而且相较于传统的方法,深度学习可以处理更加复杂的非线性数据,并且在处理大量数据时具有更高的效率和准确性,因此,在众多领域中的重要变量预测方面被较多的使用。文献[17]提出了一种改进的反向传播神经网络 (back propagation, BP) 质量预测方法,通过降噪自编码

器提取数据的有效特征,降低了其存在的噪声,保证了数据的真实性。文献[18]利用深度网络 (deep belief networks, DBN) 的方法,通过深度置信网络无监督的学习对风速序列提取特征,进而有效的对短期电力功率预测。文献[19]采用分层极限学习机 (hierarchical extreme learning machine, HELM) 进行半监督的深度学习方法进行化工过程的质量预测。然而,上述方法均假设系统的运行状态是处于稳定的,且变量之间不存在时间序列的相关性,进而建立静态模型。但是在实际的工业过程中,观测数据常常是非静态的,其具有动态特性,因此,上述的假设难以满足。近年来,循环神经网络 (recurrent neural network, RNN) 将时序的概念引入到网络结构设计中,在时序数据的分析中表现出较强的适应性,并且可以有效地处理数据的非线性和动态特性,但当遇见较长的时间步长时,在时间反向传播中,网络的误差梯度常常会出现梯度消失和梯度爆炸的问题,难以捕捉到较远时间点之间的信息,进而影响预测精度。为了解决这一问题,一种基于长短期记忆神经网络 (long short-term memory, LSTM) 的质量预测方法被应用到预测中^[20],该方法从每个阶段的长时间序列中提取与质量相关的综合隐藏特征,从而显著降低预测因子维度。文献[21]提出了一种结合注意力机制的 LSTM 非线性动态模型辨识算法,该算法基于 LSTM 模型,考虑了输入变量对于目标变量的重要性,通过为输入序列中影响输出结果的关键特征分配了更多的时间注意力,提高了 LSTM 的泛化能力。注意力机制的加入可以很大程度的减少高维输入带来的计算负担,并且有助于模型在有限的信息中快速的寻找并学习关键的特征,进而提升模型的性能^[22]。然而,上述方法仅针对整个时间序列进行预测,将每个时刻进行单独的分析,并未考虑间歇过程数据在不同时刻之间的关联性,因此对整个过程的时间序列的特征的提取不够灵敏;而且,上述方法中仅仅关注了二维有效信息,即应用了单一的时间注意力机制或者空间注意力机制,不能够完全的聚焦重要信息,容易丢失有效的信息,导致预测精度不高。

针对以上问题,提出了一种基于卷积-时空注意力的双层长短期神经网络 (convolution neural networks spatial and temporal attention with double long short-term memory networks, CNN-STA-DLSTM) 的间歇过程质量预测模型,该模型首先沿着变量的方向按照批次增加的顺序对间歇过程的三维数据展开,保留了过程变量之间的时序相关性,以更好的训练模型并进行预测,通过卷积神经网络 (convolutional neural networks, CNN) 模型提取数据中的

特征变量,充分挖掘过程数据的潜在特征,同时自动学习时间序列数据中的复杂模式和规律,提高计算效率,减小预测模型的计算负担和延迟,确保预测过程的实时性;接着,构建了双层 LSTM 网络的编码器-解码器结构,并引入时间和空间注意力机制,使得模型自适应学习时间步长相关的历史信息,提升长期记忆能力并加强变量之间的时空相关性,从而提高模型的表达能力和预测性能;接着,采用随机-网格搜索法对模型的超参数进行寻优,确保了搜索效率;最终通过实验验证模型在对间歇过程质量预测时具有可靠的预测效果。

1 基本方法

1.1 卷积神经网络

CNN 是多层感知机 (multilayer perceptron, MLP) 的变种,由纽约大学的 Yann LeCun 于 1998 年提出^[23]。CNN 是在深度学习中最常见的一种网络模型,可以有效的提取数据的局部特征。被应用于图像识别,物体识别等计算机视觉,自然语言处理,物理学和遥感科学等领域。CNN 是一种包含卷积运算的前馈神经网络 (feedforward neural networks, FNN),由输入层、卷积层、池化层、全连接层和输出层组成。卷积层和池化层交替连接形成深度结构,其具有共享权重、平移不变等特点,能够实现表征学习,其结构图如图 1 所示。

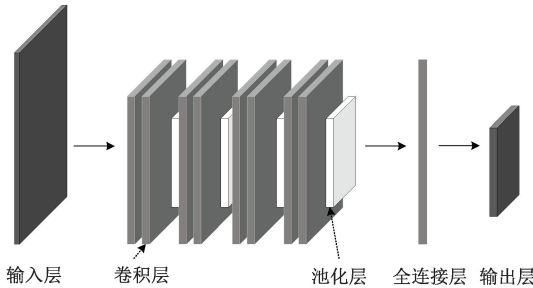


图 1 卷积神经网络 (CNN) 的结构图

Fig. 1 Convolutional neural network (CNN) structure diagram

卷积层是 CNN 的核心,主要实现局部感知和权值共享,其主要负责提取数据特征,卷积层内部由多个卷积核组成,每个卷积核的元素都对应一个权重系数和偏差量。池化层的主要作用是减小了卷积核的尺寸,可以有效的缩小参数矩阵的尺寸,从而减小全连接层中的参数数量。池化层夹在连续的卷积层中间,可以加快计算速度并防止过拟合的现象,且可以对数据进行降维处理。最常见的池化操作有平均池化和最大池化的方式。全连接层在整个卷积神经网络中起到“分类器”的作用,将通过卷积和池化后的数据映射到 CNN 的隐层特征空间中,再进行维度的转换,输出为一维的数据。激活函数为卷积神经

网络提供了非线性输出的保障,主要的激活函数有 Sigmoid、tanh 和 ReLU 函数等。

1.2 长短期神经网络

LSTM 是 RNN 的一种变体,传统的 RNN 随着时间序列间隔的增大,容易出现梯度消失或梯度爆炸等问题,导致模型在训练时会出现不稳定或无法有效进行特征学习,为此,LSTM 的每个单元结构 (LSTM cell) 增加了更多的单元结构,通过引入门控机制来控制信息的流动,解决了长期依赖引发的梯度消失和梯度爆炸问题,此外,利用记忆单元记录长期信息,避免了传统 RNN 因信息重复使用而丢失信息的问题。如图 2 所示为 LSTM 的整体单元结构图,其中包括 3 个输入为上一时刻的单元的状态 C_{t-1} 、上一时刻单元的隐藏状态 h_{t-1} 和当前时刻的输入变量 x_t ; 两个输出为当前时刻的单元状态 C_t 和当前时刻隐含的状态 h_t 。

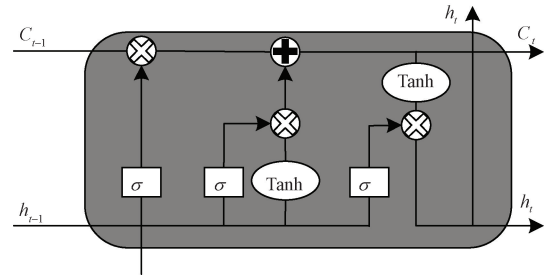


图 2 LSTM 的整体单元结构图

Fig. 2 Overall unit structure diagram of LSTM

LSTM 单元的門控结构由遗忘门 (f_t), 输入门 (i_t) 和输出门 (o_t) 三部分组成,这些門控机制用于控制 LSTM 单元对输入的信息进行选择性地存储,更新和输出。遗忘门通过上一时刻隐藏状态和当前时刻的输入来控制内部信息的遗忘;输入门利用 Sigmoid 函数和 Tanh 函数调节当前时刻的输入与上一时刻隐藏状态的输入信息,进而控制记忆单元的更新信息;输出门的功能是读取刚刚更新过的神经网络状态,对记忆单元进行输出,而具体哪些信息可以输出受输出门的控制,从而通过 3 个控制門实现动态记忆。LSTM 网络的状态计算过程如式 (1) ~ (6) 所示。

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \quad (1)$$

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (2)$$

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \quad (3)$$

$$\tilde{C}_t = \text{Tanh}(w_c \cdot [h_{t-1}, x_t] + b_c) \quad (4)$$

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

$$h_t = o_t \odot \text{Tanh}(C_t) \quad (6)$$

式中: x_t 为当前时刻的输入, h_{t-1} 为上一时刻的隐藏状态, $\tilde{C}_t$ 为网络的中间状态, w, b 分别为各个門控结构所

对应的权重矩阵和偏置向量, $[h_{t-1}, x_t]$ 表示将两个矩阵拼接在一起, σ 和 Tanh 分别表示 Sigmoid 激活函数和 Tanh 激活函数, 其表达式分别如(7)~(8)式所示。

$$\sigma(x) = \frac{1}{1 + e^{-x}} \quad (7)$$

$$\text{Tanh}(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \quad (8)$$

LSTM 通过引入记忆的存储单元和门控结构, 实现了其对历史信息的长时记忆能力和动态处理的能力, 从而能够有效捕捉数据内部以及数据之间的时序关系, 使得在处理数据和长期生产过程数据的预测方面表现出优异的性能。

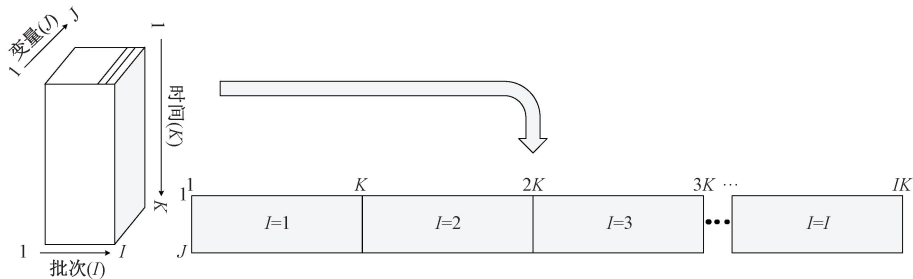


图 3 三维数据的展开方法

Fig. 3 Expansion method of 3D data

由于间歇过程数据中变量的数据量纲差异较大, 而且存在的异常值对后续模型的学习会产生较大的影响, 因此需要对数据进行归一化处理, 使在训练的过程中减少所用的时间以及可以防止异常值对模型的干扰。最大-最小法 (Max-Min) 归一化是对间歇过程数据进行预处理的更有效的方法之一, 其处理方式如式(9)所示。

$$x' = \frac{x - \min(x)}{\max(x) - \min(x)} \quad (9)$$

式中: x 为原始数据, $\min(x)$ 是样本的最小值, $\max(x)$ 是样本的最大值, x' 是归一化后的数据。通过上面的 Max-Min 方法归一化后的数据, 便可以直接进行质量预测。

2.2 建立 CNN-STA-DLSTM 预测模型

所提出的 CNN-STA-DLSTM 预测模型是以编码器-解码器 (encoder-decoder) 作为基础的模型框架, 采用 LSTM 网络作为基本单元构建了双层 LSTM 网络预测模型, 首先将输入的数据通过 CNN 进行滑动窗口的卷积操作, 对输入变量的特征信息进行提取, 并将这些提取到的特征信息输入到空间注意力模块中, 空间注意力模块主要实现空间维度上的注意力机制, 使模型更加关注输入序列中的重要区域, 具体来说, 空间注意力模块提取输入序列的空间特征信息, 在每个时间步中对质量预测相关的输入信息进行加权, 生成新的数据。并将加权后的数据输入到编码层 LSTM 网络的输入以捕获数据中的隐藏状

2 基于 CNN-STA-DLSTM 的间歇过程质量预测

2.1 间歇过程的数据预处理

间歇过程和连续过程数据相比, 多了一个批次的维度, 即 $X(I \times J \times K)$, 其中 I 为批次的次数, J 为变量的个数, K 是采样次数, 因此, 为保证完整的变量信息, 需要将三维样本数据沿着变量的方向展开成二维矩阵 $X(IK \times J)$, 按照批次的顺序排列, 如图 3 所示。

态。接着, 将这些隐藏状态信息输入到时间注意力模块, 通过时间注意力模块, 在时间步中识别并加权与质量预测相关的隐藏状态信息, 生成时间序列向量, 为最终预测建立基于时间的动态模型。最终, 时间序列向量作为解码层 LSTM 网络的输入, 预测输出结果。建立的预测模型的结构图如图 4 所示。

输入层: 对原始数据进行预处理, 因其变量之间的数据量纲差异比较大, 而且在后续模型的训练过程中需要对数据进行迭代运算, 如梯度下降法 (gradient descent method, GD), 需要较长的时间才能找到最优的解决方法, 甚至可能陷入局部最优解, 故需要进行数据归一化, 减小预测的时间, 使其更快的收敛到最优解, 提高模型的训练速度和性能, 更好的实现实时预测的效果。

编码层: 由于间歇过程数据量较大, 所以使用卷积神经网络提取数据 x_t^α 的特征, 提取出了 β 个主要特征, α 表示变量的个数, 因此去掉了 $\alpha - \beta$ 个与质量相关性较低的信息, 不但抑制了干扰信息, 而且通过降低特征维度提高了计算的速度, 减小了预测的时间, 实现了实时预测的效果。

CNN 的构成主要涵盖卷积层、池化层和全连接层, 其中最重要的是卷积层部分。CNN 可以自动的创建滤波器并建立稠密、完整的特征向量, 已被研究证实是一种非常可靠的深层特征提取方法, 相比于相同层数和神经

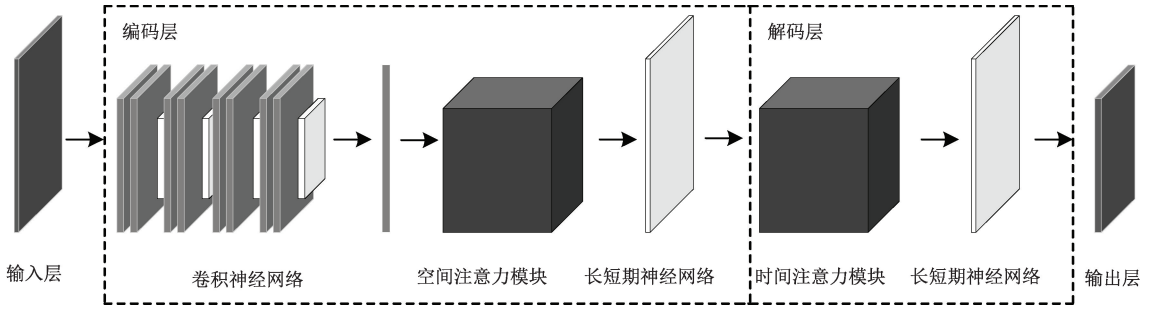


图 4 基于 CNN-STA-DLSTM 的预测模型结构

Fig. 4 Prediction model structure based on CNN-STA-DLSTM

元个数的多层感知器, CNN 具有更高的效率。卷积层的表达式如式 (10) 所示。

$$h_{ij}^k = f((\mathbf{w}^k \otimes x)_{ij}) + b_k \quad (10)$$

式中: x 为卷积层的输入, h_{ij}^k 为卷积层在第 i 行、第 j 列的输出向量, f 为激活函数, $\mathbf{w}^k$ 为第 k 个特征的卷积核权重矩阵, b_k 为该特征的偏置向量, $\otimes$ 表示卷积运算。本文使用 ReLU 作为 CNN 的激活函数, 其表达式如式 (11) 所示。

$$h_{ij}^k = \max(0, h_{ij}^{k'}) \quad (11)$$

池化层是 CNN 的另一特殊的层, 其主要是对压缩卷积层后的非线性特征进行激活, 以至于缩小特征尺寸, 降低数据维度, 提高计算速度, 并提取数据的主要特征, 同时减少下一层的参数和计算量, 防止过拟合。所提模型采用最大池化的方式进行池化, 通过选择邻域内的最大值进行池化, 从而对卷积层提取的特征进行降维, 减少后续模型的数据输入个数, 有效的降低计算量。全连接层将多个卷积层和池化层提取后的特征连接起来, 其过程是将权重矩阵与输入向量相乘, 加上偏置值作为输出值, 其输出值的计算公式如式 (12) 所示, 其中 z 表示全连接层的输出值, x 表示全连接层的输入值, $\mathbf{w}$ 表示权重, b 表示偏置。

$$z = \mathbf{w}^T x + b \quad (12)$$

编码层 LSTM 网络的隐含状态包含与质量预测相关的状态信息, 可以通过前一时刻解码层的隐藏状态 h_{t-1} 和输入 x_t 之间的相关性来描述输入与质量变量的关系。故在 LSTM 网络前应用空间注意力机制, 为了关注每个变量的重要程度, 对其赋予不同的空间注意力权重, 计算流程如图 5 所示, 通过不断的更新空间注意力值以提高模型的预测结果, 其更新过程如式 (13) ~ (14) 所示。

$$s_t^n = V_1^n \tanh(U_1^n x_t + w_1^n h_{t-1} + b_1^n) \quad (13)$$

$$\beta_t^n = \frac{|s_t^n|}{\sum_{k=1}^N |s_t^k|}, 1 \leq n \leq N \quad (14)$$

式中: s_t^n 为 t 时刻的第 n 个变量所对应的空间注意力值, $\tanh$ 表示双曲正切激活函数, 将变量的输入映射到 $[-1,$

$1]$ 之间, N 为每个样本的变量个数, $V_1^n, U_1^n, W_1^n, b_1^n$ 是需要通过学习而不断更新的空间注意力参数, β_t^n 表示空间注意力权重, 通过式 (14) 可以使得每个采样时刻的注意力值的和为 1。通过采用 GD 对注意力参数进行优化, 以获得更好的注意力值。通过空间注意力值加权之后的输入的变量如式 (15) 所示。

$$\tilde{x}_t = (\beta_t^1 x_t^1, \beta_t^2 x_t^2, \dots, \beta_t^N x_t^N) \quad (15)$$

$$\sum_{i=1}^N \beta_t^i = 1 \quad (16)$$

空间注意力加权后的样本 $\tilde{x}_t$ 作为编码层的 LSTM

网络的输入, 在前向传播的过程中从 $\tilde{x}_t$ 中学习其中的隐藏状态, 编码层 LSTM 网络内部门控单元和存储单元的状态更新状态如式 (17) ~ (22) 所示。

$$\tilde{f}_t = \sigma([w_{hf} h_{t-1} + w_{xf} \tilde{x}_t] + b'_{f'}) \quad (17)$$

$$\tilde{i}_t = \sigma([w_{hi} h_{t-1} + w_{xi} \tilde{x}_t] + b'_{i'}) \quad (18)$$

$$\tilde{o}_t = \sigma([w_{ho} h_{t-1} + w_{xo} \tilde{x}_t] + b'_{o'}) \quad (19)$$

$$\tilde{C}_t'' = \tanh([w_{hc} h_{t-1} + w_{xc} \tilde{x}_t] + b'_{c'}) \quad (20)$$

$$\tilde{C}_t' = \tilde{i}_t \odot \tilde{C}_t'' + \tilde{f}_t \odot \tilde{C}_{t-1}' \quad (21)$$

$$\tilde{h}_t = \tilde{o}_t \odot \tanh(\tilde{C}_t') \quad (22)$$

式中: $\odot$ 表示两个向量逐点相乘, $\tilde{C}_t'$ 为过程中的相应存储单元, w_h 和 b' 为更新过程中需要不断学习的权重和偏置参数。LSTM 网络的参数通过 Adam 算法进行优化, 可以更快并且更稳定的收敛且正确的学习, 同时调整内部的参数, 最大程度的最小化损失函数, 提升了预测的性能。

解码层: 时间注意力层, 将编码层的输出的隐含状态作为解码层的输入, 通过参考解码层前一时刻的隐含状态 s_{t-1} 计算编码层 LSTM 网络输出的隐含状态, 然后计算每个解码层隐含状态所对应的时间注意力值, 其计算流程如图 5 所示。在 t 时刻所对应的隐含状态的时间注意

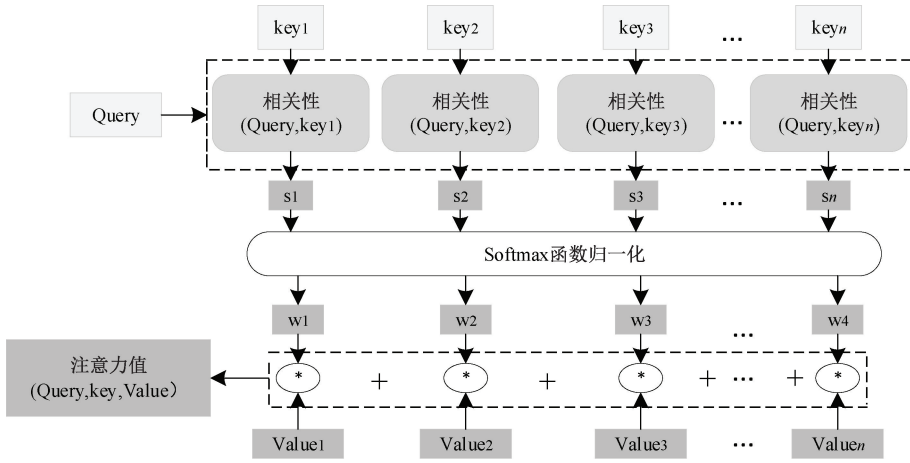


图 5 注意力机制计算流程

Fig. 5 Calculation flow of attention mechanism

力值的计算过程如式 (23) ~ (24) 所示。

$$a_t^k = V_2^k \tanh(W_2^k s_{t-1} + U_2^k h_{t-T+k} + b_2^k) \quad (23)$$

$$\delta_t^k = \frac{|a_t^k|}{\sum_{m=1}^T a_t^m}, 1 \leq k \leq T \quad (24)$$

式中: a_t^k 为 t 时间步长中第 k 个编码层隐藏状态的注意力值, $\tanh$ 表示双曲正切激活函数, 将变量的输入映射到 $[-1, 1]$ 之间, $V_2^k, W_2^k, U_2^k, b_2^k$ 为第 k 个样本需要学习的参数, s_{t-1} 为解码层上一个时刻的隐藏状态, T 表示时间步长的大小, δ_t^k 表示时间注意力值, 通过式 (24) 可以使得每个采样的时刻的时间注意力值的和为 1。对编码层的隐藏状态进行时间注意力机制的加权求和, 获得自适应过程向量, 计算过程如式 (25) 所示。

$$c_t = \sum_{k=1}^T h(t - T + k) \delta_t^k \quad (25)$$

将所得到的自适应过程向量作为解码层 LSTM 网络的输入, 更新解码层的隐藏状态, 结合自适应过程向量及给定的目标序列 $y_t = [y_1, y_2, \dots, y_{t-1}]$, 更新过程如式 (26) ~ (27) 所示。

$$\tilde{s}_{t-1} = W_3 y_{t-1} + V_3 s_{t-1} + b_3 \quad (26)$$

$$s_t = f_t(c_t, \tilde{s}_{t-1}) \quad (27)$$

式中: W_3 和 V_3 为更新过程中的权重矩阵, b_3 表示偏置向量, $f_t(\cdot)$ 表示一个 LSTM 单元。最后对质量变量 $\tilde{y}_T$ 进行预测, 如式 (28) 所示。

$$\tilde{y}_T = F(y_1, y_2, \dots, y_T, x_1, x_2, \dots, x_T) = V_4(f_t([\tilde{s}_{T-1}, C_T])) + b_4 \quad (28)$$

式中: V_4 和 b_4 分别表示权重矩阵和偏置向量, $[\tilde{s}_{T-1}, C_T]$ 表示 $T-1$ 时刻的输出层隐藏状态和 T 时刻过程向量

的全连接处理。解码层使用 GD 对模型训练, 用 Adam 算法实现梯度的优化, 提升模型的预测性能。

2.3 超参数优化

所提模型采用随机-网格搜索法对模型进行优化。传统的网格搜索是一种穷举搜索方法, 它通过遍历超参数的所有可能组合来寻找最优的超参数, 首先需要对每个超参数设定一组候选值, 然后生成这些超参数的笛卡尔积, 形成超参数的组合网格, 然后对每个超参数进行模型训练和评估, 从而找到性能最佳的超参数, 但网格搜索的计算量较大, 当超参数和候选值较多时, 搜索空间会增大, 导致搜索效率较低。随机搜索是一种随机化的方法, 可以通过随机采样超参数的组合来寻找最优超参数, 随机搜索可以在超参数空间中随机抽取一定数量的组合进行评估, 因此使用随机-网格搜索法, 在超参数的空间中随机选取一部分组合进行搜索, 从而减少搜索的空间规模以及减少搜索的时间, 提高搜索的效率, 提升模型预测的实时性。

2.4 算法流程

基于 CNN-STA-DLSTM 模型的质量预测流程图如图 6 所示, 其主要分为两个阶段: 分别为离线建模和在线预测, 离线建模时使用历史数据对模型进行训练, 得到较好的网络参数, 和训练之后的模型, 然后使用该模型进行在线实时的预测。

1) 离线建模

(1) 对间歇过程的三维数据 $X(I \times J \times K)$ 沿变量方向展开成二维矩阵 $X(IK \times J)$;

(2) 采用 Max-Min 方法对矩阵 $X(IK \times J)$ 进行归一化, 使数据标准化;

(3) 对标准化之后的数据使用 CNN 进行特征提取, 得到数据的特征矩阵;

(4) 构建双层 LSTM 模型,将空间注意力模块引入第 1 层 LSTM 之前,以便第 1 层 LSTM 网络关注每个变量的重要程度并对其赋予不同的空间注意力权重;

(5) 将时间注意力模块引入第 2 层 LSTM 之前,以便第 2 层 LSTM 网络关注每个变量在时间维度的重要程度并对其赋予不同的时间注意力权重;

(6) 训练预测模型,通过随机-网格搜索法对模型的超参数进行优化得到最优超参数;

(7) 通过损失函数微调整个网络参数,并保存训练好的 CNN-STA-DLSTM 模型。

2) 在线预测

(1) 通过采样得到测试数据,对测试数据利用离线建模中的数据展开方式以及归一化方法对其进行预处理;

(2) 将处理好的测试数据输入到通过离线建模进行超参数优化后的网络模型中,进行质量预测和性能评估。

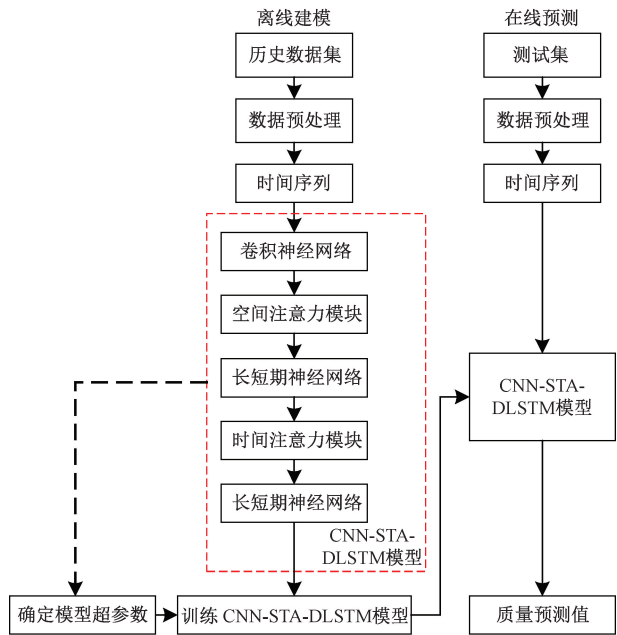


图 6 质量预测流程图

Fig. 6 Quality prediction flow chart

2.5 预测模型评价指标

一般使用损失函数对模型的性能进行评估,损失函数将用来评估模型的预测值与真实值的偏离程度,以及反映模型的性能的好坏,它是一个非负的实值函数。如果模型的损失函数越小,则该模型的鲁棒性和预测性能越好。神经网络的训练是通过前向传播的方式计算损失函数值,并通过反向传播更新网络的参数,从而使得损失函数值不断减小至最小值。预测模型所采用的损失函数为均方根误差 (root mean square error, RMSE)、均方误差 (mean square error, MSE)、平均绝对误差 (mean absolute

error, MAE) 对模型的性能进行评估。其计算公式如式 (29) ~ (31) 所示。

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \tilde{y}_i)^2} \quad (29)$$

$$MSE = \frac{1}{N} \sum_{i=1}^N (y_i - \tilde{y}_i)^2 \quad (30)$$

$$MAE = \frac{1}{N} \sum_{i=1}^N |y_i - \tilde{y}_i| \quad (31)$$

式中: N 为样本数, y_i 为真实值, $\tilde{y}_i$ 为预测值, RMSE、MAE 的值越小,说明模型描述实验数据具有更好的精确度; MSE 值越小,说明模型训练效果和稳定性越好。

3 实验验证与分析

3.1 实验平台

实验所采用的实验平台的软硬件的参数: 处理器为 Core(TM) i9-13900HX, CPU 频率为 2.20 GHz, 显卡为 NVIDIA RTX 4050, 操作系统为 Windows 11。所提的基于 CNN-STA-DLSTM 模型的深度学习框架由 Pytorch 2.0.0, Python 3.9.13, Anaconda 22.9.0 运行。

3.2 仿真实验

1) 青霉素发酵过程

青霉素是一种重要的抗生素,生产过程是典型的间歇过程,其具有非线性、强耦合和时滞性等特性。在青霉素的发酵过程中,温度、通风速率、PH 值、底物浓度等多种因素都影响其发酵过程,任何过程变量的异常或偏差都可能会影响产品的质量。因此,建立精确的模型对过程的控制具有十分重要的作用,但其数据之间及数据内部存在多变量的相互耦合关系,且关键参数难以测量,因此建立精确的预测模型至关重要。青霉素发酵生产流程如图 7 所示。

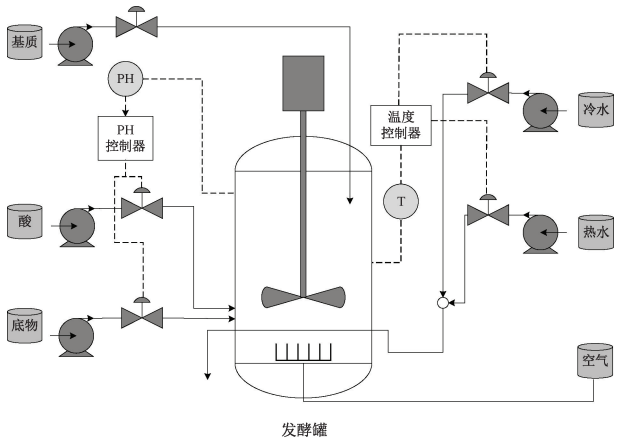


图 7 青霉素发酵流程图

Fig. 7 Penicillin fermentation process chart

为验证所提出的基于 CNN-STA-DLSTM 模型在工业过程中的准确性,以青霉素的发酵过程作为研究对象,采用 Birol^[24] 等扩展了 Bajpai 和 Reuss 的机理模型所开发的青霉素发酵过程仿真建模软件 Pensim2.0^[25],所产生的数据进行实验验证,每批次的发酵时间设置为 400 h,

表 1 青霉素发酵过程的主要变量

Table 1 Main variables of penicillin fermentation process

变量序号	变量名称	变量单位	变量序号	变量名称	变量单位
x_1	通风速率	L/h	x_6	CO2 浓度	mmol/L
x_2	搅拌速率	r/min	x_7	PH	-
x_3	底物流加速率	L/h	x_8	发酵罐温度	K
x_4	补料温度	K	x_9	反应罐体积	L
x_5	溶解氧浓度	mmol/L	x_{10}	产生热量	kcal

为验证 CNN-STA-DLSTM 模型的预测效果,选取青霉素发酵过程的质量变量为青霉素浓度。通过对青霉素发酵过程的研究和数据的预处理,保留了 6 个相关性系数较大的变量作为模型的输入^[26]。

为验证所提模型的优越性,将青霉素浓度作为预测目标,与 LSTM、GSTAE、SAE、STA-LSTM、STCGAT 的质量预测模型通过实验进行比较,6 种预测模型的结构参数如表 2 所示。其中 LSTM^[27] 是长短期神经记忆网络的非线性动态软传感器模型;GSTAE^[28] 网络是加入门控单元的目标相关堆叠自编码器;SAE^[29] 是一种质量堆叠自编

码器;STA-LSTM 是一种时空注意力机制的长短时神经网络模型,能够控制时间序列历史信息的有效性;STCGAT^[30] 是一种采用时空相关组件的模型,其自适应地对局部和全局时空相关性进行学习。将以上模型作为对比模型,合理说明所提模型的精确性。因为神经网络的训练含有随机性,故不能用某一次的预测结果进行对比,这样会使得其不具有客观性,因此使用相同的数据集及模型参数对每个模型进行 10 次的仿真实验,取平均值作为模型的评价指标进行对比,结果如表 3 所示。

表 2 实验对比模型的参数设置

Table 2 Parameter Settings of experimental comparison models

参数名称	LSTM	GSTM	SAE	STA-LSTM	STCGAT	CNN-STA-DLSTM
输入维度	6	6	6	6	6	6
网络隐含层数	1	3	2	2	2	2
隐含层节点数	256	256	256	(20,20)	256	(40,40)
学习率	0.001	0.001	0.001	0.001	0.001	0.001
最大迭代次数	200	200	200	300	200	200
输出维度	1	1	1	1	1	1

分别采用 LSTM、GSTAE、SAE、STA-LSTM、STCGAT 以及 CNN-STA-DLSTM 模型预测青霉素发酵过程中的青霉素浓度,如图 8 所示,展示了各模型微调过程的损失曲线,但需要注意的是,各模型微调过程损失曲线的初始值也是预训练过程的结束点^[31]。如图 8 所示,LSTM 和 GSTAE 在微调过程中损失值分别为 5.324 和 0.084 5,可以发现,GSTAE 的初始损失值较小,这是因为 GSTAE 将质量预测这一目标的关注,引入到了模型中。与其他的模型相比,CNN-STA-DLSTM 模型的损失值显著降低,其初始值最小,为 0.003 6,这是因为该模型不仅考虑通过 CNN 提取数据中的特征,还引入了时空注意力机制捕获了数据中潜在的时空相关性。此外,可以看出,在经过 200 次的训练后,CNN-STA-DLSTM 相比其他 5 种模型,损失值最小,这进一步说明了该模型能更好的为质量预测任务提取质量相关性特征。

码器;STA-LSTM 是一种时空注意力机制的长短时神经网络模型,能够控制时间序列历史信息的有效性;STCGAT^[30] 是一种采用时空相关组件的模型,其自适应地对局部和全局时空相关性进行学习。将以上模型作为对比模型,合理说明所提模型的精确性。因为神经网络的训练含有随机性,故不能用某一次的预测结果进行对比,这样会使得其不具有客观性,因此使用相同的数据集及模型参数对每个模型进行 10 次的仿真实验,取平均值作为模型的评价指标进行对比,结果如表 3 所示。

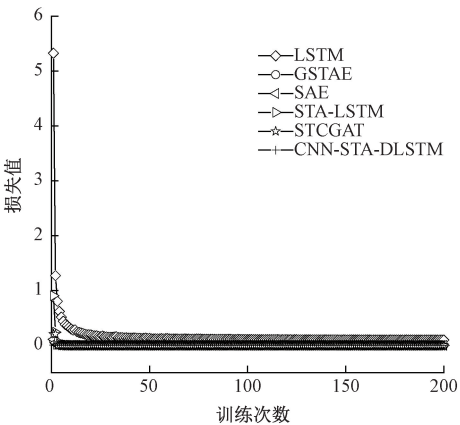


图 8 各网络微调过程损失曲线

Fig. 8 Loss curves of each network fine-tuning process

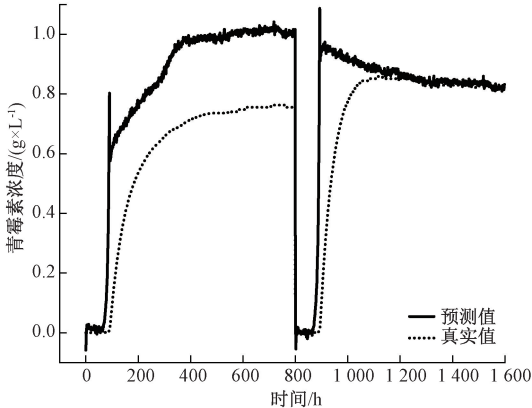
如表 3 所示,当 6 种预测模型的结构参数设置如表 2 所示, GSTAE、SAE、STA-LSTM、STCGAT、CNN-STA-DLSTM 相较于 LSTM 的 RMSE、MAE 以及 MSE 都有明显的减小,且 STA-LSTM 的拟合误差有阶跃性的减小,其中所提 CNN-STA-DLSTM 的拟合误差最小,即真实值与预测值偏离程度最小,说明该模型相较于其他 5 种模型预测精度较高。

表 3 青霉素浓度各网络预测结果

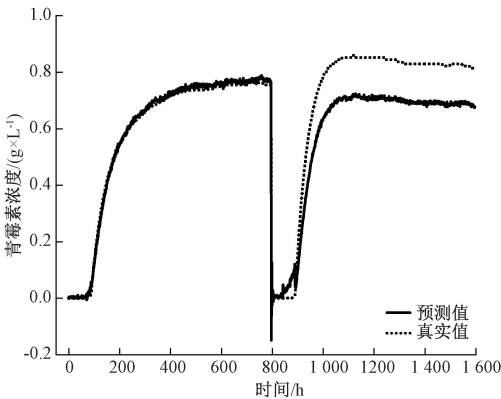
Table 3 Prediction results of penicillin concentration by network

网络模型	RMSE	MAE	MSE
LSTM	0.228 0	0.161 9	0.052 0
GSTAE	0.093 3	0.067 9	0.008 7
SAE	0.080 5	0.056 6	0.006 4
STA-LSTM	0.077 6	0.053 0	0.000 7
STCGAT	0.047 3	0.028 5	0.000 6
CNN-STA-DLSTM	0.020 4	0.009 8	0.000 4

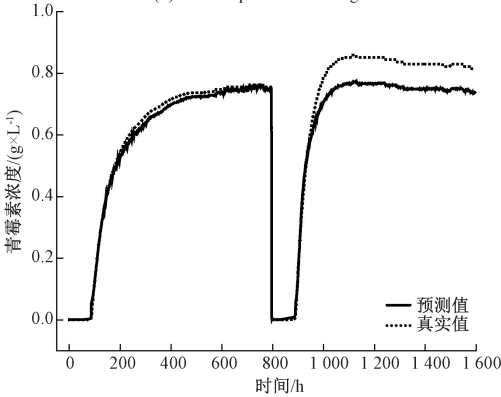
如图 9(a)~(f)所示,6 种神经网络模型的拟合曲线图,STCGAT 模型的预测精度高于 LSTM 模型、GSTAE 模型、SAE 模型和 STA-LSTM 模型,可以看出,所提 CNN-STA-DLSTM 模型在第 1 200 个采样点之前预测值和真实值基本相吻合,而且在第 1 200 个采样点之后,预测值与真实值之间具有小而稳定的误差。也可以通过如图 10 所示,进行更清晰的对比,其中黑线表示实际值等于预测值的情况,正方形表示 LSTM 的实际预测精度,圆形表示 GSTAE 的实际预测精度,正三角形表示 SAE 的实际预测精度,倒三角形表示 STA-LSTM 的实际预测精度,菱形代表 STCGAT 的预测精度,六边形黑点表示 CNN-STA-DLSTM 的实际预测精度,因为六边形黑点更接近黑线,说明 CNN-STA-DLSTM 模型的拟合误差较小,具有更好地预测性能,验证了该模型的有效性。



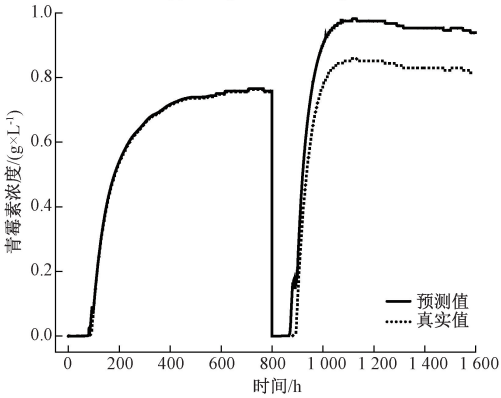
(a) LSTM预测拟合曲线
(a) LSTM prediction fitting curve



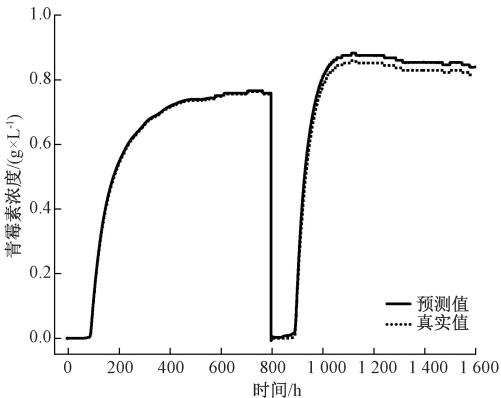
(b) GSTAE预测拟合曲线
(b) GSTAE prediction fitting curve



(c) SAE预测拟合曲线
(c) SAE prediction fitting curve



(d) STA-LSTM预测拟合曲线
(d) STA-LSTM prediction fitting curve



(e) STCGAT预测拟合曲线
(e) STCGAT prediction fitting curve

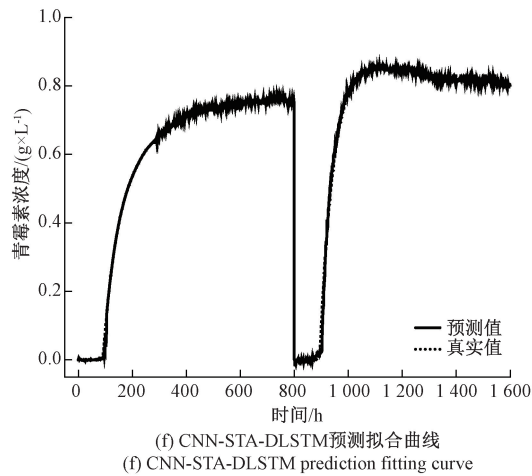


图 9 不同的模型对青霉素浓度的预测拟合曲线
Fig. 9 Prediction and fitting curves of penicillin concentration by different models

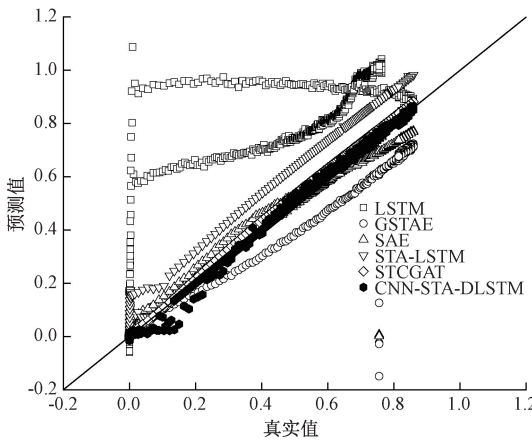


图 10 实际值与预测值的可视化
Fig. 10 Visualization of actual and predicted values

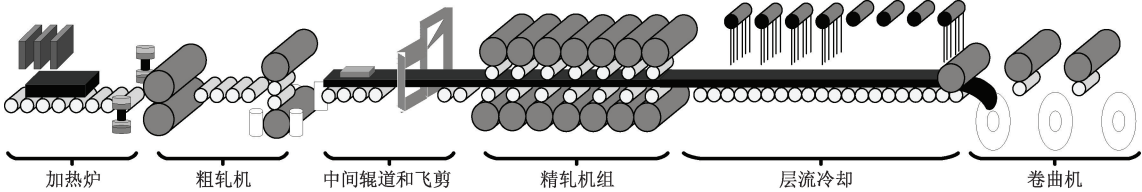


图 11 带钢热连轧过程工艺布置图
Fig. 11 Hot strip rolling process layout

表 4 过程变量及质量变量分配表
Table 4 Distribution table of process variables and quality variables

变量	类型	描述	单位
1~7	过程变量	7 个机架的平均辊缝	mm
8~14	过程变量	7 个机架的轧制力	MN
15~20	过程变量	7 个机架的弯辊力	MN
21	质量变量	精轧末机架出口厚度	mm

通过实验说明,所提出的 CNN-STA-DLSTM 模型能够更好的提取时间序列数据之间以及数据内部的特征,通过在构建的双层 LSTM 网络中引入时空注意力机制,获取了特征数据与质量变量之间潜在的时空相关性。通过表 2 所示,对 6 个模型的 RMSE、MAE 进行对比,CNN-STA-DLSTM 模型的 RMSE 和 MAE 都有明显的减少,说明模型的预测精度有了明显的提高,同时相比于其他 5 种模型,CNN-STA-DLSTM 模型 MSE 值最小,说明该模型的训练效果稳定性最好。

2)带钢热连轧过程

带钢热连轧过程的生产线由加热炉、粗轧机、中间延时辊道、飞剪、精轧机组、热输出辊道、层流冷却和卷曲机主要构成,是一种需要高投入、产出高质量产品且产生高效益的轧钢工艺。其基本布置如图 11 所示。

厚度为 100 ~ 200 mm 板坯通过加热炉加热至 1 200 ℃,接着对其炉后除磷并送入粗轧机。将板坯经过粗轧机的若干次轧制至 28 ~ 45 mm 的中间坯输出,中间坯在中间辊道上快速运行通过飞剪,带钢被剪切并传送到精轧机,精轧机将其连续精轧至满足厚度、宽度、板型及最终温度的实际要求的带钢,然后层流冷却对钢材的性能进行提升,最后由地下卷曲机卷成卷。带钢的厚度是产品的关键质量指标,通常难以在线实时检测。

以某钢厂 1 700 mm 带钢热连轧生产线为研究对象,采集生产厚度为 2.70 mm 成品带钢的现场数据进行实验验证。过程变量有 20 个,包括 7 个精轧机的 7 个机架辊缝、轧制力和弯辊力(第 1 机无弯辊),质量变量为精轧机的出口的带钢厚度,过程变量与质量变量的具体描述如表 4 所示^[32]。

为验证 CNN-STA-DLSTM 模型的预测效果,对带钢热连轧过程进行研究和数据预处理之后,通过 PLS 方法

计算其相关性并进行降维,得到 7 个相关性系数较大的变量作为预测模型的输入。如图 12 所示为输入变量的权重系数,其大小解释了输入变量与预测目标的相关性。

为进一步验证所提模型的有效性,以热连轧过程中的带钢厚度作为预测目标,与 LSTM、GSTAE、SAE、STA-LSTM、STCGAT 的质量预测模型进行实验比较,6 种预测模型的结构参数设置如表 2 所示。因神经网络具有随机

性,故不能用某一次产生的实验结果作为最终的实验结果,因此为了使结果更具有客观性,因此使用相同的数据集及模型参数在每个预测模型中进行 10 次实验,取平均值作为模型的对比评价指标,其结果如表 5 所示。

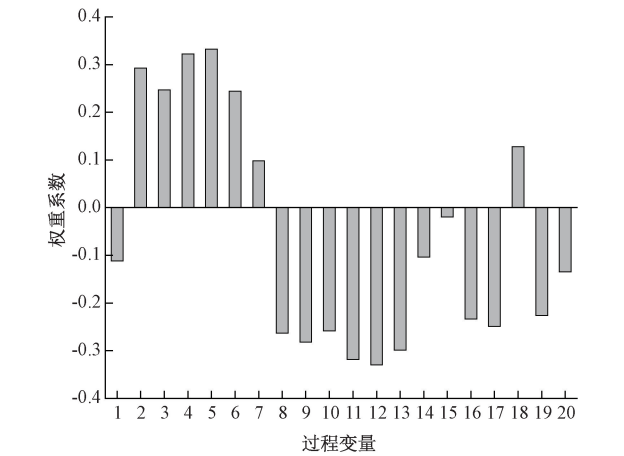


图 12 输入变量的权重系数

Fig. 12 Weight coefficients of input variables

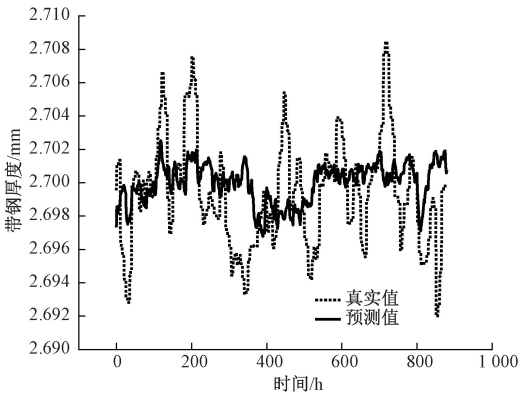
表 5 各网络对带钢厚度的预测结果

Table 5 Prediction results of strip thickness by each network

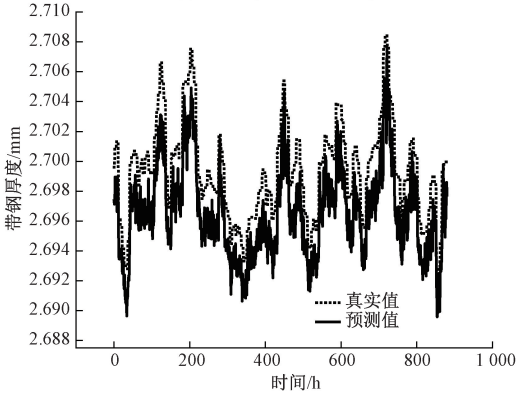
网络模型	RMSE	MAE	MSE
LSTM	0.004 1	0.003 4	1.086
GSTAE	0.002 9	0.002 7	0.101 2
SAE	0.001 1	0.000 9	0.081 9
STA-LSTM	0.001 0	0.000 9	0.057 7
STCGAT	0.000 9	0.000 7	0.051 3
CNN-STA-DLSTM	0.000 7	0.000 5	0.042 5

如表 5 所示,当 6 种预测模型的结构参数设置如表 2 所示, STA-LSTM、STCGAT、CNN-STA-DLSTM 相较于 LSTM、GSTAE、SAE 的 MAE、RMSE 和 MSE 都有明显的减小,且 CNN-STA-DLSTM 的拟合误差有阶跃性的减小,且 CNN-STA-DLSTM 的拟合误差最小,即真实值与预测值偏离程度最小,说明该模型性能相较于其他 5 种模型预测精度最高。

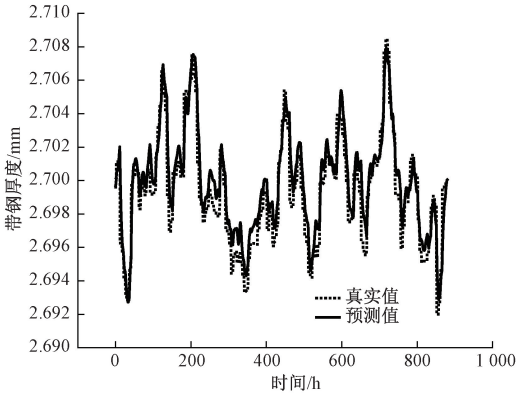
如图 13(a)~(f) 所示,6 种神经网络模型的拟合曲线图,可以看出,STCGAT 模型的预测精度高于 LSTM 模型、GSTAE 模型、SAE 模型和 STA-LSTM 模型,本文所提 CNN-STA-DLSTM 模型在预测的过程中预测值和真实值吻合度最高,而且其预测值与真实值之间具有小而稳定的误差。也可以如图 14 所示,进行更清晰的对比,其中黑线表示实际值等于预测值的情况,正方形表示 LSTM 的实际预测精度,圆形表示 GSTAE 的实际预测精度,正三角形表示 SAE 的实际预测精度,倒三角形表示 STA-LSTM 的实际预测精度,菱形代表 STCGAT 的预测



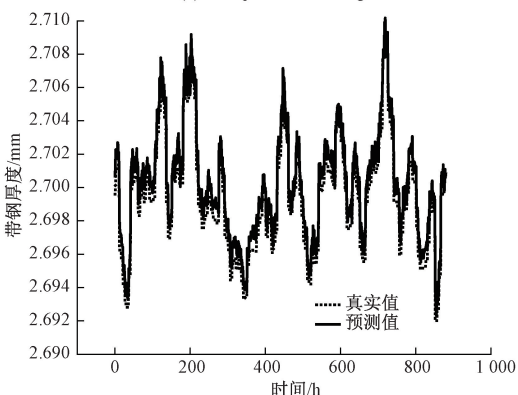
(a) LSTM 预测拟合曲线
(a) LSTM prediction fitting curve



(b) GSTAE 预测拟合曲线
(b) GSTAE prediction fitting curve



(c) SAE 预测拟合曲线
(c) SAE prediction fitting curve



(d) STA-LSTM 预测拟合曲线
(d) STA-LSTM prediction fitting curve

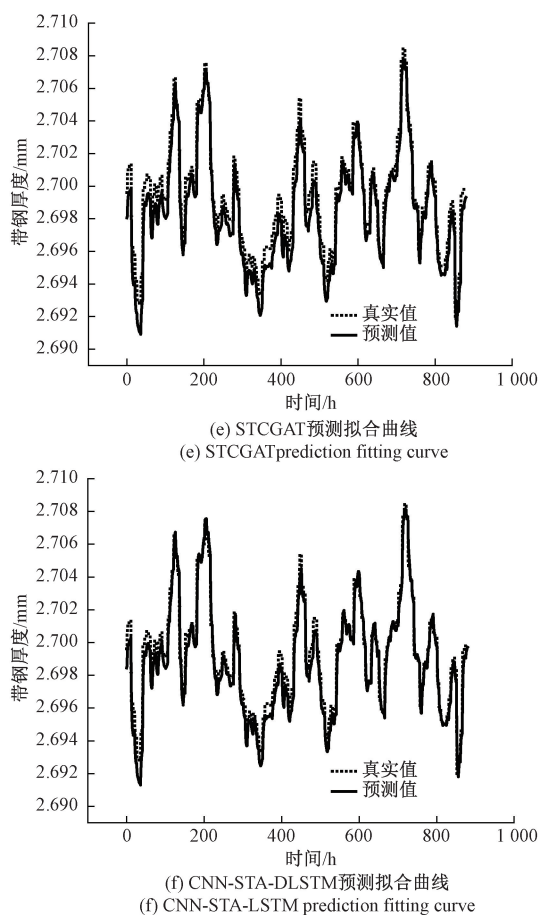


图 13 不同的模型对带钢厚度的预测拟合曲线

Fig. 13 Prediction and fitting curves of strip thickness by different models

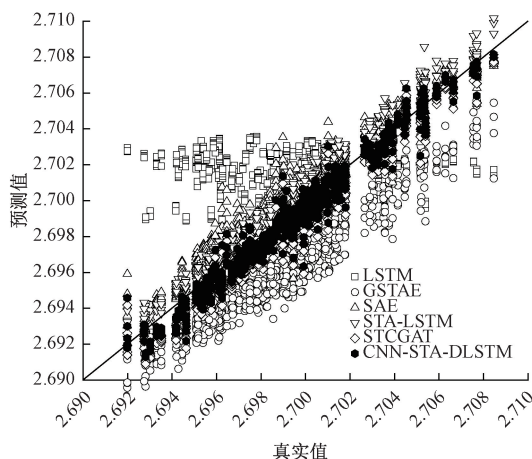


图 14 实际值与预测值的可视化

Fig. 14 Visualization of actual and predicted values

精度,六边形黑点代表 CNN-STA-DLSTM 的预测精度,因为六边形黑点更接近黑线,说明 CNN-STA-DLSTM 模型的拟合误差较小,具有更好地预测性能,进一步验证了该模型的有效性。

4 结 论

针对间歇过程变量深层特征提取困难,数据非线性特性和动态特性而导致的预测精度不高问题,提出了基于 CNN-STA-DLSTM 的间歇过程质量预测模型,该模型充分考虑了间歇过程数据特性导致预测精度不高的问题,利用长短期神经网络的记忆功能实现长时间序列的在线预测,提高间歇过程质量预测的效果。首先,对间歇过程数据按变量方向展开成二维矩阵,再对其进行 Max-Min 方法进行标准化处理,使用 CNN 对数据特征进行深层提取,充分提取数据之间的相关特征,然后建立基于时间和空间注意力机制的双层 LSTM 的预测模型,利用注意力机制自适应地学习时间步长的相关历史信息,以提高模型的长期记忆能力,并加强过程变量与质量变量之间的时空相关性,为更好的缩短模型的训练时间,引入随机-网格搜索法优化 CNN-STA-DLSTM 模型的超参数,确定超参数的最佳组合,建立预测模型。通过对青霉素发酵过程中对青霉素浓度和热连轧工艺中对带钢厚度预测的实验验证,与 5 种模型的预测效果对比分析,验证了所提模型对间歇过程的质量预测具有实用性和有效性。但是由于 CNN-STA-DLSTM 模型在训练和监测过程中需要大量的计算资源,未来将研究基于 CNN-STA-DLSTM 的轻量级网络。

参考文献

- [1] 张敏,袁毅,李贤均. 基于 KECA 和 MBA-NARX 的多阶段间歇过程质量预测[J]. 西南交通大学学报, 2021, 58(3): 685-695.
ZHANG M, YUAN Y, LI X J. Multi-stage quality prediction of batch process based on KECA and MBA-NARX [J]. Journal of Southwest Jiaotong University, 2023, 58(3): 685-695.
- [2] REN J, NI D. A batch-wise LSTM-encoder decoder network for batch process monitoring [J]. Chemical Engineering Research and Design, 2020, 164: 102-112.
- [3] 周新杰,王建林,艾兴聪,等. 基于 IDPC-RVM 的多模态间歇过程质量变量在线预测[J]. 化工学报, 2022, 73: 3120-3130.
ZHOU X, WANG J, AI X C, et al. IDPC-RVM based online prediction of quality variables for multimode batch processes [J]. CIESC Journal, 2022, 73 (7): 3120-3130.
- [4] RATO T, REIS M. An integrated multiresolution framework for quality prediction and process monitoring in batch processes [J]. Journal of Manufacturing Systems, 2020, 57: 198-216.
- [5] LEE S, KWAK M, TSUI K L, et al. Process monitoring

- using variational autoencoder for high-dimensional nonlinear processes [J]. *Engineering Applications of Artificial Intelligence*, 2019, 83: 13-27.
- [6] 石欣, 张夏恒, 朱雅亲, 等. 基于 VMD-NARX 的 MOSFET 剩余使用寿命预测方法[J]. *仪器仪表学报*, 2023, 44(9): 275-286.
- SHI X, ZHANG X H, ZHU Y Q, et al. Residual service life prediction method of MOSFET based on VMD-NARX [J]. *Chinese Journal of Scientific Instrument*, 2023, 44(9): 275-286.
- [7] 赵小强, 徐蓉蓉. 改进价值函数的多阶段间歇过程故障检测[J]. *电子测量与仪器学报*, 2022, 36(12): 45-54.
- ZHAO X Q, XU R R. Fault detection for multi-stage batch processes with improved cost function [J]. *Journal of Electronic Measurement and Instrumentation*, 2022, 36(12): 45-54.
- [8] 周新杰, 王建林, 艾兴聪, 等. 基于 IDPC-RVM 的多模态间歇过程质量变量在线预测[J]. *化工学报*, 2022, 73(7): 3120-3130.
- ZHOU X J, WANG J L, AI X C, et al. Online prediction of quality variables in multi-mode batch processes based on IDPC-RVM [J]. *Acta Chemical Sinica*, 2022, 73(7): 3120-3130.
- [9] 赵小强, 周文伟, 惠永永. 基于核熵投影的 CPLS 间歇过程监测及质量预测[J]. *高校化学工程学报*, 2018, 32(5): 1186-1193.
- ZHAO X Q, ZHOU W W, HUI Y Y. Monitoring and quality prediction of cpls batch process based on kernel entropy projection[J]. *Journal of Chemical Engineering of Chinese Universities*, 2018, 32(5): 1186-1193.
- [10] HAN H G, FU SH J, SUN H Y, et al. Data-driven model-predictive control for nonlinear systems with stochastic sampling interval[J]. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 2022, 53(5): 3019-3030.
- [11] LAURÍ D, LENNOX B. Expanded PLS algorithm: modeling of batch processes [J]. *Chemometrics and Intelligent Laboratory Systems*, 2014, 132: 111-123.
- [12] 彭开香, 马亮, 张凯. 复杂工业过程质量相关的故障检测与诊断技术综述[J]. *自动化学报*, 2017, 43(3): 349-365.
- PENG K X, MA L, ZHANG K. Review of quality-related fault detection and diagnosis techniques for complex industrial processes [J]. *Acta Automatica Sinica*, 2017, 43(3): 349-365.
- [13] 贾润达, 毛志忠, 王福利. 基于 KPLS 模型的间歇过程产品质量控制[J]. *化工学报*, 2013, 64(4): 1332-1339.
- JIA R D, MAO ZH ZH, WANG F L. Quality control of batch process products based on KPLS model [J]. *CIESC Journal*, 2013, 64(4): 1332-1339.
- [14] 王锡昌, 王普, 高学金, 等. 一种新的基于 MKPLS 的间歇过程质量预测方法[J]. *仪器仪表学报*, 2015, 36(5): 1155-1162.
- WANG X CH, WANG P, GAO X J, et al. A new batch process quality prediction method based on MKPLS[J]. *Chinese Journal of Scientific Instrument*, 2015, 36(5): 1155-1162.
- [15] YAO M, WANG H G, XU W L. Batch process monitoring based on functional data analysis and support vector data description[J]. *Journal of Process Control*, 2014, 24(7): 1085-1097.
- [16] 骆楠, 祁佳康, 罗娜. 基于双向门控循环单元神经网络的间歇过程最终产品质量预测[J]. *华东理工大学学报(自然科学版)*, 2020, 46(6): 807-814.
- LUO N, QI J K, LUO N. Product quality prediction in batch process based on bidirectional gated recurrent unit neural network[J]. *Journal of East China University of Science and Technology*, 2020, 46(6): 807-814.
- [17] 郭小萍, 马美卉, 李元. 基于 DAE-BP 神经网络的化工过程质量预测[J]. *计算机测量与控制*, 2023, 31(1): 181-186, 193.
- GUO X P, MA M H, LI Y. Chemical process quality prediction based on DAE-BP neural network [J]. *Computer Measurement and Control*, 2023, 31(1): 181-186, 193.
- [18] 袁桂丽, 吴振民, 刘骅骐, 等. 基于深度置信网络的短期风电功率预测[J]. *太阳能学报*, 2022, 43(2): 451-457.
- YUAN G L, WU ZH M, LIU H Q, et al. Short-term wind power prediction based on deep belief network [J]. *Journal of Solar Energy*, 2022, 43(2): 451-457.
- [19] YAO L, GE ZH Q. Deep learning of semisupervised process data with hierarchical extreme learning machine and soft sensor application[J]. *IEEE Transactions on Industrial Electronics*, 2017, 65(2): 1490-1498.
- [20] WANG K, GOPALUNI R, CHEN J, et al. Deep learning of complex batch process data and its application on quality prediction [J]. *IEEE Transactions on Industrial Informatics*, 2018, 16(12): 7233-7242.
- [21] 王雅欣, 徐宝昌, 徐朝农, 等. 基于工厂数据的注意力 LSTM 网络辨识方法[J]. *化工学报*, 2020, 71: 5664-5671.
- WANG Y X, XU B CH, XU CH N, et al. Attention LSTM network identification method based on factory data[J].

- CIESC Journal, 2020, 71(12): 5664-5671.
- [22] NIU ZH Y, ZHONG G Q, YU H. A review on the attention mechanism of deep learning [J]. Neurocomputing, 2021, 452: 48-62.
- [23] 周飞燕, 金林鹏, 董军. 卷积神经网络研究综述[J]. 计算机学报, 2017, 40(6): 1229-1251.
- ZHOU F Y, JIN L P, DONG J. A review of convolutional neural networks [J]. Chinese Journal of Computers, 2017, 40(6): 1229-1251.
- [24] 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.
- [25] YUAN L, YAN Y Y, TANG X CH. Online product quality prediction for multi-phase based on local model[J]. Journal of System Simulation, 2020, 28(4): 966-971.
- [26] 赵小强, 脱奔奔, 惠永永, 等. 基于 ISTA-LSTM 模型的间歇过程质量预测[J]. 控制与决策, 2023, 38(11): 3279-3289.
- ZHAO X Q, TUO B B, HUI Y Y, et al. Quality prediction of batch process based on ISTA-LSTM model [J]. Control and Decision, 2023, 38(11): 3279-3289.
- [27] YUAN X F, LI L, WANG Y L. Nonlinear dynamic soft sensor modeling with supervised long short-term memory network [J]. IEEE Transactions on Industrial Informatics, 2019, 16(5): 3168-3176.
- [28] SUN Q Q, GE ZH Q. Gated stacked target-related autoencoder: A novel deep feature extraction and layerwise ensemble method for industrial soft sensor application [J]. IEEE Transactions on Cybernetics, 2020, 52(5): 3457-3468.
- [29] WANG K, GOPALUNI R B, CHEN J H, et al. Deep learning of complex batch process data and its application on quality prediction [J]. IEEE Transactions on Industrial Informatics, 2018, 16(12): 7233-7242.
- [30] 刘云翔, 梁智超. 一种高效的连续时序图注意力网络的交通预测模型[J/OL]. 计算机工程, 1-14 [2024-09-01]. <https://doi.org/10.19678/j.issn.1000-3428.0069059>.
- LIU Y X, LIANG ZH CH. A highly efficient continuous

sequence diagram attention network traffic prediction model [J/OL]. Computer engineering, 1-14 [2024-09-01]. <https://doi.org/10.19678/j.issn.1000-3428.0069059>.

- [31] YAO H, ZHAO X, LI W, et al. Multi-stage fusion regression network for quality prediction of batch process[J]. The Canadian Journal of Chemical Engineering, 2023, 101(12): 6977-6994.
- [32] MA L, DONG J, PENG K, et al. A novel data-based quality-related fault diagnosis scheme for fault detection and root cause diagnosis with application to hot strip mill process [J]. Control Engineering Practice, 2017, 67: 43-51.

作者简介



惠永永 (通信作者), 2014 年于兰州理工大学获得学士学位, 2019 年于兰州理工大学获得博士学位, 现为兰州理工大学电气工程与信息工程学院自动化系副教授, 主要研究方向为基于数据驱动的复杂系统监测、故障诊断、关键性能指标预测及软测量等。

E-mail: hyylut@163.com

Hui Yongyong (Corresponding author) received his B. Sc. degree from Lanzhou University of Technology in 2014 and Ph. D. degree from Lanzhou University of Technology in 2019. Now he is currently an associate professor of Automation Department in School of Electrical Engineering and Information Engineering, Lanzhou University of Technology. His main research interests include data-driven complex system monitoring, fault diagnosis, key performance index prediction and soft measurement, etc.



孙凯文, 2022 年于华北科技学院获得学士学位, 现为兰州理工大学研究生, 主要研究方向为间歇过程的故障诊断、关键性能指标预测及软测量。

E-mail: 1404705893@qq.com

Sun Kaiwen received his B. Sc. degree in 2022 from North China University of Science and Technology. Now he is a M. Sc. candidate at Lanzhou University of Technology. Candidate at Lanzhou University of Technology. His main research interests include batch process fault diagnosis, key performance index prediction and soft measurement.