

基于 BSCWEformer 的退火炉内分组式辊速序列预测^①



岳晓光¹, 石元博²

¹(辽宁石油化工大学 信息与控制工程学院, 抚顺 113001)

²(辽宁石油化工大学 人工智能与软件学院, 抚顺 113001)

通信作者: 石元博, E-mail: syb2011@yeah.net

摘要: 退火炉内带钢的长度受到温度、张力等因素的影响而变化, 导致辊的转速改变以及焊缝位置的不确定, 从而威胁生产安全. 为了准确预测辊的转速以计算焊缝的实时位置, 本文提出基于带状稀疏柯西自注意力的 BSCWEformer (banded sparse Cauchy weight enhanced Transformer) 模型. 模型采用带状稀疏的、使用基于相对位置计算的柯西分布权重值增强的自注意力结构, 在提高相邻输入序列的重要性的同时, 将自注意力的复杂度由二次方降低为线性. 通过实际生产数据进行实验, 并与 LogSparse Transformer、Transformer、RNMT+ 等模型进行对比, 得出本文所提出的 BSCWEformer 模型在退火炉内分组式辊速序列预测任务上具有较高的预测精度.

关键词: 退火炉; 带钢焊缝; 带状稀疏; 柯西分布权重; 时间序列

引用格式: 岳晓光, 石元博. 基于 BSCWEformer 的退火炉内分组式辊速序列预测. 计算机系统应用, 2024, 33(3): 213-219. <http://www.c-s-a.org.cn/1003-3254/9434.html>

Predicting Grouped Rollers Speed Series in Annealing Furnace Based on BSCWEformer

YUE Xiao-Guang¹, SHI Yuan-Bo²

¹(School of Information and Control Engineering, Liaoning Petrochemical University, Fushun 113001, China)

²(School of Artificial Intelligence and Software, Liaoning Petrochemical University, Fushun 113001, China)

Abstract: The length of the strip steel in the annealing furnace is affected by temperature, tension, and other factors, resulting in changes in roller speed and uncertainty in weld position and threatening production safety. To accurately predict roller speed, this study proposes the banded sparse Cauchy weight enhanced Transformer (BSCWEformer) model. The model adopts a banded sparse self-attention structure enhanced by Cauchy distribution weight values calculated from relative positions, which improves the importance of adjacent input sequences and reduces the complexity of self-attention from quadratic to linear. Through experiments with actual production data and comparison with LogSparse Transformer, Transformer, RNMT+, and other models, the BSCWEformer model shows higher accuracy in predicting grouped roller speed series.

Key words: annealing furnace; strip steel weld seam; banded sparse; Cauchy distribution weight; time series

在带钢热镀锌生产线中, 受到退火炉内各段温度、张力的影响, 带钢的长度会不断变化, 这对获取退火炉内各辊的转速造成了很大困扰, 也因此难以计算两卷带钢之间的焊缝到达气刀处的准确时间, 而这会

威胁气刀的安全. 为了得到焊缝的实时位置, 王超提出使用优化支持向量机^[1]、极限学习机算法^[2]对带钢进入退火炉后的延伸量进行预测以计算焊缝的位置, 冯兴隆^[3]分别提出多元线性回归、支持向量机和改进支

① 基金项目: 辽宁省教育厅科研项目 (LJKMZ20220737)

收稿时间: 2023-09-24; 修改时间: 2023-10-20; 采用时间: 2023-10-25; csa 在线出版时间: 2024-01-17

CNKI 网络首发时间: 2024-01-19

持向量机对焊缝通过退火炉的时长进行预测. 但这些作者都对问题进行了简化, 没有直接使用原始数据, 而是将焊缝进出退火炉这段时间内测得的一组原始数据求平均值后压缩成一条记录. 基于压缩后的数据提出的模型虽然在退火炉的设定参数变化不大的情况下表现良好, 但却无力应对实际生产过程中退火炉的设定参数有较大调整的情况. 与此同时, 区别于 Autoformer^[4]、FEDformer^[5]和 Informer^[6]等模型适用的连续性长时间序列, 带钢焊缝原始数据是分组式的, 时间上不连续且长度不大, 因此一般性的时间序列预测模型难以直接应用于退火炉内分组式辊速序列预测.

原始 Transformer 模型存在容易忽略局部依赖关系、时间和空间复杂度高的缺点^[7]. 为了增强自注意力捕获局部依赖关系的能力, Gao 等人^[8]提出基于自注意力的时变网络 SATVNN 模型, 并分别测试了在自注意力中加入基于相对位置计算的高斯分布权重值、拉普拉斯分布权重值、柯西分布权重值, 用于提高相邻输入序列的重要程度, 帮助模型更快地捕捉时间序列的局部依赖关系. 而为了降低 Transformer 模型中自注意力的复杂度, Ainslie 等人^[9]提出的 ETC 模型、Zaheer 等人^[10]提出的 Big Bird 模型利用自注意力矩阵在大多数数据点上都非常稀疏的特点, 采用了随机注意、滑动窗口注意、全局注意相结合的方式^[11], 将自注意力的复杂度由二次方降低为线性, 克服了 Transformer

模型时间和空间复杂度高的缺点.

本文结合带钢生产数据的实际特点, 在 Transformer 模型的基础上提出了使用带状稀疏的柯西自注意力的 BSCWEformer 模型, 成功地将 Transformer 模型应用于分组式辊速序列预测任务. BSCWEformer 模型仅使用 Transformer 模型的解码器部分, 并在自注意力中加入基于相对位置计算的柯西分布权重值, 增强了模型捕捉时间序列动态变化的能力. 与此同时, 根据生产数据具有天然的局部相关性的特点, 模型采用了带状稀疏的自注意力, 在不影响模型预测准确度的前提下, 极大地降低了模型的时间和空间复杂度, 并且实现简单, 避免了 ETC 等模型需要特殊的库支持、实现复杂且在分组式辊速序列预测任务中的有效性难以估计的缺点.

1 退火炉内辊速预测问题简介

退火炉的结构可分为预热段 (PHS)、还原加热段 (HS)、还原均热段 (SS)、冷却段 (RCS)、均衡段 (ES)、锌锅 (POT) 等, 如图 1 所示, 锌锅 (POT) 后 1.7 m 左右为气刀所在位置, 用于控制带钢表面镀锌的厚度, 平时关闭, 在带钢焊缝经过时打开^[12,13]. 带钢进入退火炉后, 在各段温度和张力作用下, 其长度会发生变化, 导致 2BR、3BR、4BR、5BR 处各辊的转速发生改变, 带钢焊缝也将难以准确定位, 从而无法确定焊缝到达气刀处的准确时间, 可能造成气刀的损坏.

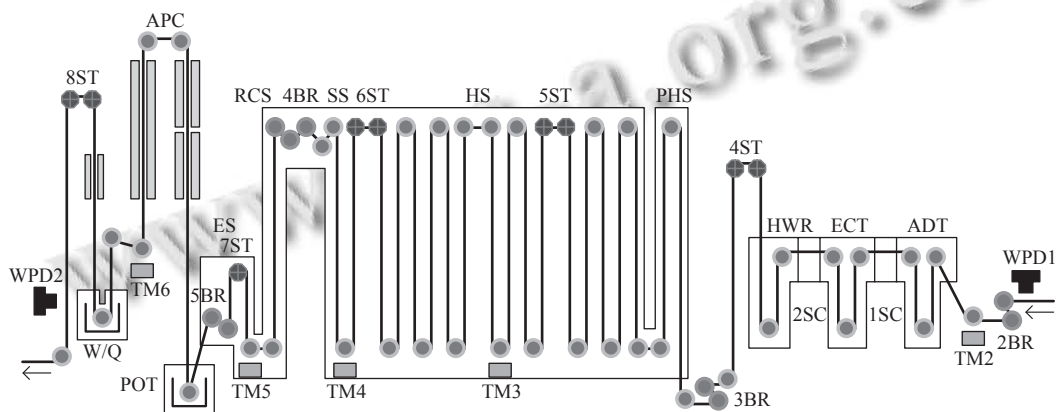


图 1 退火炉内带钢热镀锌工艺流程图

退火炉内的设备记录了带钢焊缝从 WPD1 处到 WPD2 处这段时间内各段温度、张力的时序数据. 需要根据这一组时序数据预测图 1 中 2BR、3BR、4BR、5BR 处的辊的转速, 从而计算带钢焊缝经过 WPD1 后多久到达气刀处.

可以将各段温度、张力数据视为输入序列 $X = (x_1, x_1, \dots, x_l)$, 其中 l 表示 X 的长度, 即每组时序数据的序列个数, 其大小不定, 且 $x_i \in \mathbb{R}^{d_x}$, d_x 表示输入序列 X 的属性个数. 将需要预测的 2BR、3BR、4BR、5BR 处的辊的转速视为目标序列 $Y = (y_1, y_2, \dots, y_l)$, 其中输

入序列 X 和目标序列 Y 的长度相等, $y_i \in \mathbb{R}^{d_y}$, d_y 表示目标序列 Y 的属性个数, 且 $d_x \neq d_y$. 为了计算焊缝到达气刀处的准确时间, 需要构建一个模型将输入序列 X 映射成目标序列 Y , 并且还要考虑退火炉内多个时刻不同温度、张力条件下的累积作用对带钢长度的影响.

2 基于 BSCWEformer 的预测方法

为了能够将温度、张力等分组式输入序列映射成对应的辊速目标序列, BSCWEformer 模型仅采用 Transformer 模型的解码器部分, 如图 2 所示, 并且使用带状稀疏的柯西自注意力代替 Transformer 模型中的自注意力. 通过在自注意力中加入基于输入序列之间的相对位置计算的柯西分布权重值, 增强模型捕捉输入序列动态变化的能力, 同时通过对原始自注意力进行带状的稀疏化, 降低模型的时间和空间复杂度.

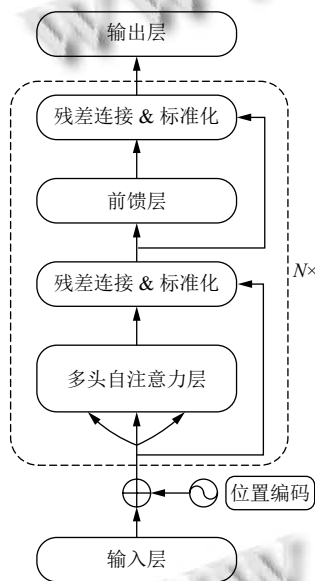


图2 模型的总体结构

2.1 自注意力实现带状稀疏的过程

与 Transformer 模型的自注意力不同, 带状稀疏的自注意力限制了目标序列预测需要关注的输入序列的范围. 在保证模型预测准确度的前提下, 通过将关注范围限制在较小的范围内, 避免了计算全部自注意力矩阵的时空开销, 从而降低自注意力的复杂度. 其实现方法是将原始自注意力中的查询、键和值先按规则进行分块, 然后再逐块计算分块自注意力, 避免了在 Transformer 模型中每个自注意力得分的计算都需要所有输入序列参与的复杂性, 其具体过程如下.

步骤 1. 将温度、张力序列 $X \in \mathbb{R}^{L \times d_x}$ 零填充对齐为固定长度后得到 $X' \in \mathbb{R}^{L \times d_x}$, 其中 L 表示固定的最大序列长度且 $L \geq l$, d_x 表示属性个数.

步骤 2. 设置自注意力层多头数量为 h , X' 经过 3 个不同的全连接层转换得 $Q \in \mathbb{R}^{L \times d_h}$ 、 $K \in \mathbb{R}^{L \times d_h}$ 、 $V \in \mathbb{R}^{L \times d_h}$, 其中 $d_h = h \times d_x$ 表示多头数量与属性个数的乘积, 方便之后多头的划分.

步骤 3. 将 Q 、 K 、 V 划分多头后, 分别得到 h 个 $Q_h \in \mathbb{R}^{L \times d_x}$ 、 $K_h \in \mathbb{R}^{L \times d_x}$ 、 $V_h \in \mathbb{R}^{L \times d_x}$, 准备分块.

步骤 4. 设置带状稀疏自注意力的分块大小为 w , 即辊速序列的预测需要关注的温度、张力序列的范围, 其大小为 $w+1 \sim 2w$, 如图 3 所示.

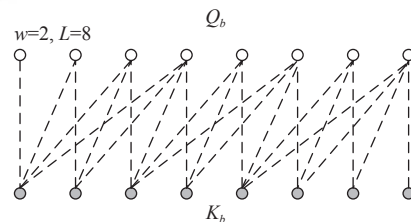


图3 分块后的 Q 、 K 关联示意

步骤 5. 将 Q_h 沿长度 L 的维度分块得到 $n = \text{ceil}(L/w)$ 个 $Q_b \in \mathbb{R}^{w \times d_x}$, 最后一个块可能需要零填充对齐, 此时 $n \times w \geq L$.

步骤 6. 与 Q_h 分块稍有不同, K_h 沿长度 L 的维度分块得到 n 个 $K'_h \in \mathbb{R}^{w \times d_x}$, 并对最后一个块零填充对齐后, 需要复制并连接相邻的两个 K'_h 得到 n 个 $K_b \in \mathbb{R}^{2w \times d_x}$, 以实现当前辊速序列的预测可以关注前 $w+1 \sim 2w$ 个温度、张力序列. 此外, K'_h 中的首个块需要在之前连接一个零填充的块以对齐.

步骤 7. 与 $K_h \rightarrow K_b$ 的过程类似, 转换 V_h 得到 n 个 $V_b \in \mathbb{R}^{2w \times d_x}$.

与原始自注意力中直接计算 Q_h 和 K_h 的乘积不同, BSCWEformer 模型带状稀疏的自注意力将计算 Q_h 和 K_h 分块后所得的 n 个 Q_b 和 K_b 的乘积, 只获取原始自注意力得分矩阵中的部分结果用于自注意力的计算. 因为不需要得到整个原始自注意力得分矩阵 $A \in \mathbb{R}^{L \times L}$, 因此带状稀疏的自注意力能够将自注意力的时间和空间复杂度从 $O(n^2)$ 降低为 $O(n)$, 其时间和空间复杂度与 Transformer 模型的自注意力的比例约为 w/L .

2.2 柯西自注意力

为了克服原始自注意力不易捕获局部依赖关系的

缺点, BSCWEformer 模型在自注意力中加入了柯西权重值, 提高了相邻序列的重要性, 称为柯西自注意力. 通过将温度、张力序列的相对距离信息加入到自注意力中, 提高最近的温度、张力序列的信息比重, 即越新的数据越重要, 保留的信息越多. 输入序列的重要性也会随着相对距离的增加而缓慢降低, 使模型能够更好地捕捉时序数据的动态变化.

柯西权重矩阵 $C \in \mathbb{R}^{w \times 2w}$ 如下:

$$C = \begin{pmatrix} c_{11} & \dots & c_{1(2w)} \\ \vdots & \ddots & \vdots \\ c_{w1} & \dots & c_{w(2w)} \end{pmatrix}$$

其中, c_{ij} 的值为:

$$c_{ij} = \frac{\gamma}{1 + \lambda|i - j|^2}$$

其中, γ 、 λ 为可配置超参数.

BSCWEformer 模型柯西自注意力的计算过程如下:

$$S_1 = \frac{Q_b(K_b)^T}{\sqrt{d_x}} \quad (1)$$

其中, $S_1 \in \mathbb{R}^{w \times 2w}$ 表示自注意力初始得分. 然后, 加入柯西权重矩阵 C :

$$S_2 = S_1 + C \quad (2)$$

其中, $S_2 \in \mathbb{R}^{w \times 2w}$ 表示加入柯西权重矩阵后的自注意力得分. 然后, 加入前向掩码 $M \in \mathbb{R}^{w \times 2w}$ 以遮挡预测的辊速序列之后未来的温度、张力序列:

$$S_3 = S_2 + M \quad (3)$$

其中, $S_3 \in \mathbb{R}^{w \times 2w}$ 表示最终的自注意力得分. 此时, 自注意力的输出为:

$$O = \text{Softmax}(S_3) V_b \quad (4)$$

其中, $O \in \mathbb{R}^{w \times d_x}$. 合并 n 个分块和 h 个多头后, 完整的自注意力输出为 $O' \in \mathbb{R}^{L \times d_h}$. 最后再经过一个全连接层的转换得到最终的自注意力层的输出:

$$O'' = \text{Linear}(O') \quad (5)$$

其中, $O'' \in \mathbb{R}^{L \times d_x}$.

3 实验

3.1 数据预处理

观察生产数据的分布规律, 正常的一组生产数据中应该包含数百条连续的记录, 并且只有两个卷号的交替, 而且两组之间的时间间隔十几分钟或几十分钟

不等, 然后通过以下步骤预处理数据并划分训练集和测试集.

- 1) 清除时间上孤立的异常数据, 即一段时间内只有一条或几条记录的数据.
- 2) 对数据进行分组, 只保留包含两个卷号的正常分组.
- 3) 剔除与预测任务无关的特征, 如卷号等.
- 4) 将所有分组数据随机打乱先后顺序后, 按照 7:3 的比例划分训练集和测试集.

经过上述步骤处理后, 得到训练集数据 949 组, 测试集数据 407 组, 每组序列长度 300-700 不等, 如图 4 所示. 输入序列 X 包括退火炉内各段温度、张力等 10 个属性, 输出序列 Y 包括各个辊的转速共 11 个属性, 如表 1 所示.

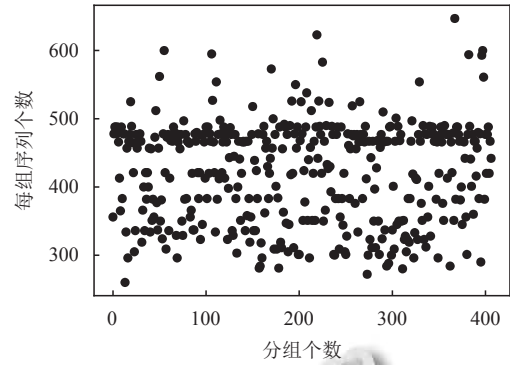


图 4 测试集各组序列的长度

表 1 生产数据整理后的属性

输入序列属性	属性描述	目标序列属性	属性描述
x_1	焊缝经过WPD1后的计时 (s)	y_1	2BR辊组第1个辊的转速
x_2	机组速度	y_2	2BR辊组第2个辊的转速
x_3	SS段张力	y_3	3BR辊组第1个辊的转速
x_4	HS段张力	y_4	3BR辊组第2个辊的转速
x_5	RCS段张力	y_5	4BR辊组第1个辊的转速
x_6	HS段温度	y_6	4BR辊组第2个辊的转速
x_7	SS段温度	y_7	4BR辊组第3个辊的转速
x_8	RCS段温度	y_8	5BR辊组第1个辊的转速
x_9	ES段温度	y_9	5BR辊组第2个辊的转速
x_{10}	TOP段温度	y_{10}	6BR辊组第1个辊的转速
—	—	y_{11}	6BR辊组第2个辊的转速

3.2 参数设置与模型对比

BSCWEformer 模型的超参数设置: 除输入、输出层外, 共有 $N = 3$ 层自注意力层, 每层的多头数量 $h = 3$; 每个自注意力层中的前馈层的输入神经元个数为 48,

输出神经元个数为 $d_x = 10$, 激活函数为 ELU; 带状稀疏的分块大小 $w = 10$; 柯西权重矩阵的超参数为 $\gamma = 1$ 、 $\lambda = 0.1$; 使用分组常数学习率, 前 400 次迭代学习率为 0.001, 后 200 次迭代学习率为 0.000 1, 共 600 次迭代, 批量大小为 59; 采用可学习的嵌入式位置编码. 温度、张力序列 X 经过最大最小归一化后输入模型, 得到辊速序列 Y , 使用平均绝对误差 MAE 作为损失函数, 模型的训练损失收敛曲线如图 5 所示.

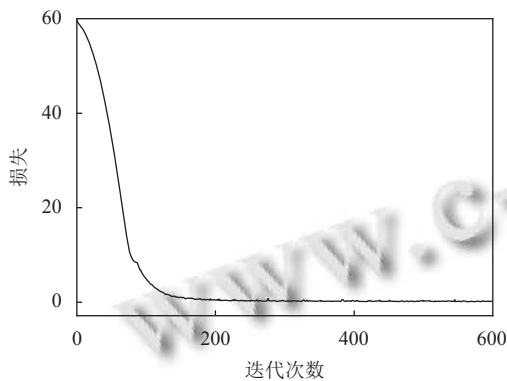


图 5 损失收敛曲线

与此同时, 本文选择了几个适用于分组式时间序列预测任务的模型作为基准模型, 包括前馈神经网络 (FNN)、LSTM^[14]、LSTMa^[15]、RNMT+^[16]、Transformer^[17]、LogSparse Transformer^[18]. 其中, LSTMa 模型使用 Luong 注意力; RNMT+ 由 LSTM 和多头自注意力堆叠而成, 常用于机器翻译; LogSparse Transformer 模型使用因果卷积获取查询和键, 简称为 LogTrans, Transformer 简称为 Trans.

此外, 本文比较了所有模型的时间和空间复杂度,

以及在生产过程中实际预测必需的迭代次数, 如表 2.

得益于带状稀疏的自注意力, BSCWEformer 模型拥有线性的时间和空间复杂度, 对比原始 Transformer 等模型复杂度大大降低. 并且在实际生产过程中, FNN、RNMT+、Transformer 模型需要完全实时地根据每一条输入序列预测辊速输出序列以计算焊缝在退火炉内的位置, 因此实际预测时的时间复杂度为 n ; 而 LSTM、LSTMa 模型在积累够所需步长的数据后, 也需要完全实时地逐条预测辊速序列, 其时间复杂度也为 n , 而预测开始时缺少的辊速序列可用预测的第 1 条辊速序列补齐; LogSparse Transformer、BSCWEformer 模型不需要递归地依赖前一步的预测值, 模型可以在焊缝到达气刀处之前的某个时间点一次性地预测得出之前所有的辊速序列, 然后再实时地逐条跟踪预测, 因此只需少量迭代后即可保障焊缝安全通过气刀处, 其时间复杂度为 1.

表 2 各模型复杂度对比

模型名称	时间复杂度	空间复杂度	预测迭代次数
FNN	n	n	n
LSTM	n	n	n
LSTMa	n	n	n
RNMT+	n^2	n^2	n
Trans	n^2	n^2	n
LogTrans	$n \log n$	n^2	1
BSCWEformer	n	n	1

3.3 结果与分析

通过 407 组测试数据对所有模型进行测试, 并将平均绝对误差 (MAE)、平均绝对百分比误差 (MAPE) 作为评价指标, 各模型在所有分组测试数据上的 MAE 损失如图 6 所示.

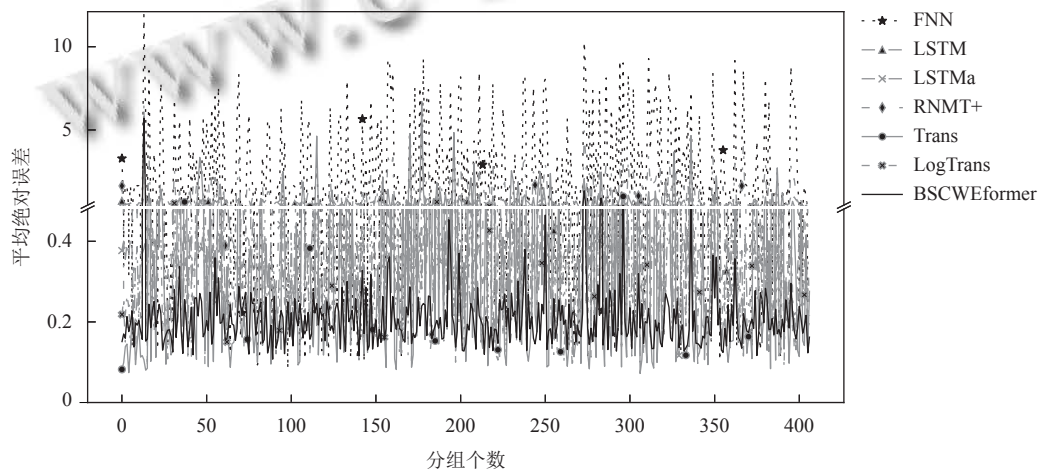


图 6 所有模型在测试集上每组测试数据的 MAE

图6中较大的波动,主要由于某组温度、张力序列中的机组速度、温度、张力等属性的突然变化,导致预测损失的剧烈波动.从图6可以看出,FNN模型在测试集上的MAE预测损失波动最为剧烈,其次是RNMT+和Transformer模型.而LSTM、LSTMa、LogTrans、BSCWEformer模型的MAE预测损失明显更加平稳,并且LSTMa和LogTrans、BSCWEformer模型的MAE预测损失始终处于LSTM之下.其中,BSCWEformer模型的MAE预测损失显著低于其他模型,其折线始终处于最底端,并且波动不大,证明了模型的准确性和稳定性.

各模型在所有测试数据上的总体MAE损失如图7所示.与图6显示的结果相对应,BSCWEformer模型的总体MAE最低,在分组式辊速序列预测任务上的表现最好.各模型在所有评价指标上的测试结果,如表3所示.

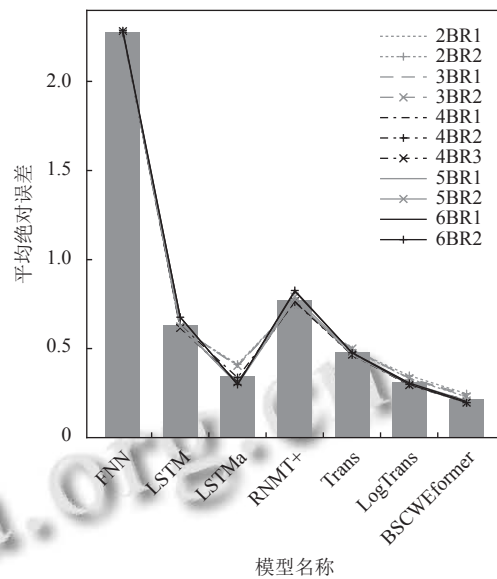


图7 各模型总体 MAE

表3 各模型在所有测试数据上的详细 MAE 和 MAPE

辊	FNN		LSTM		LSTMa		RNMT+		Trans		LogTrans		BSCWEformer	
	MAE	MAPE	MAE	MAPE	MAE	MAPE	MAE	MAPE	MAE	MAPE	MAE	MAPE	MAE	MAPE
2BR1	2.274	2.168	0.611	0.621	0.406	0.411	0.748	0.708	0.495	0.506	0.331	0.333	0.229	0.233
2BR2	2.271	2.166	0.611	0.625	0.413	0.417	0.747	0.708	0.493	0.504	0.348	0.349	0.246	0.249
3BR1	2.269	2.168	0.617	0.632	0.401	0.404	0.747	0.709	0.494	0.505	0.334	0.334	0.241	0.246
3BR2	2.269	2.167	0.619	0.633	0.403	0.405	0.747	0.709	0.499	0.511	0.334	0.337	0.231	0.236
4BR1	2.277	2.166	0.623	0.631	0.309	0.323	0.771	0.719	0.464	0.473	0.296	0.293	0.211	0.212
4BR2	2.274	2.162	0.633	0.644	0.337	0.352	0.761	0.712	0.464	0.474	0.305	0.301	0.196	0.201
4BR3	2.282	2.167	0.616	0.624	0.311	0.325	0.766	0.715	0.466	0.476	0.295	0.291	0.197	0.199
5BR1	2.277	2.167	0.633	0.649	0.297	0.315	0.777	0.724	0.465	0.477	0.298	0.295	0.195	0.198
5BR2	2.274	2.157	0.625	0.635	0.301	0.315	0.791	0.736	0.476	0.487	0.307	0.305	0.208	0.208
6BR1	2.277	2.156	0.677	0.691	0.298	0.303	0.821	0.765	0.477	0.487	0.307	0.307	0.203	0.204
6BR2	2.285	2.157	0.676	0.687	0.298	0.303	0.826	0.768	0.476	0.485	0.299	0.294	0.197	0.198
均值	2.276	2.164	0.631	0.643	0.343	0.352	0.773	0.725	0.479	0.489	0.314	0.313	0.214	0.217

从表3中的测试结果可以看出,前馈神经网络(FNN)在完全不关注之前多个温度、张力序列对当前辊速序列预测的累积作用的情况下的预测结果最差;而结构复杂、参数众多的RNMT+模型的预测损失虽然相对于FNN大幅降低,但预测结果也比较差,说明复杂的模型未必适合于分组式辊速序列的预测;LSTM模型的MAE预测损失相较于RNMT+模型降低约18%,加上Luong式点积注意力后MAE预测损失再次降低了45%;Transformer模型的MAE预测损失虽然相对于LSTM降低了24%,但内存占用高,只能以较小的批量进行训练,训练时间长;LogSparse Transformer模型的预测效果相对于Transformer模型提升明显,其MAE预测损失降低了34%,但是其采用的是对数形式的稀疏自注意力,在分组式辊速序列预测任务中节省

的空间很有限,却需要复杂的额外实现.本文提出的BSCWEformer模型预测效果最好,其MAE预测损失相对于Transformer模型降低了55%,且内存占用少,可以设置较大的批量进行训练.

4 结论与展望

本文提出的BSCWEformer模型成功地将Transformer模型应用于分组式辊速序列预测,具有模型结构简单、复杂度低、善于捕获局部依赖关系的优点.经过实际生产数据的测试,对比其他多种模型,验证了本文提出的BSCWEformer模型的性能和准确度.针对生产现场无法安装焊缝检测设备的实际情况,BSCWEformer模型利用现有设备的测量数据,能够准确地预测连续热镀锌工艺过程中退火炉内各个辊的转速,为

计算焊缝到达气刀处的准确时间提供了依据,从而可以精准地控制气刀的开合,对提高带钢镀锌的成品率并保障气刀的安全具有重要意义。

参考文献

- 1 王超,王建辉,顾树生,等.基于优化支持向量机的带钢延伸量软测量研究.东北大学学报(自然科学版),2015,36(8):1084-1088.[doi:10.3969/j.issn.1005-3026.2015.08.005]
- 2 Wang C, Wang JH, Gu SS, *et al.* Elongation prediction of steel-strips in annealing furnace with deep learning via improved incremental extreme learning machine. *International Journal of Control, Automation and Systems*, 2017, 15(3): 1466-1477. [doi: 10.1007/s12555-015-0463-7]
- 3 冯兴隆.镀锌生产线退火炉焊缝跟踪的数学模型研究[硕士学位论文].沈阳:东北大学,2016.
- 4 Wu HX, Xu JH, Wang JM, *et al.* Autoformer: Decomposition Transformers with auto-correlation for long-term series forecasting. *Proceedings of the 35th International Conference on Neural Information Processing Systems*. 2021. 22419-22430.
- 5 Zhou T, Ma ZQ, Wen QS, *et al.* FEDformer: Frequency enhanced decomposed Transformer for long-term series forecasting. *Proceedings of the 2022 International Conference on Machine Learning*. Baltimore: *Proceedings of Machine Learning Research*, 2022. 27268-27286.
- 6 Zhou HY, Zhang SH, Peng JQ, *et al.* Informer: Beyond efficient Transformer for long sequence time-series forecasting. *Proceedings of the 35th AAAI Conference on Artificial Intelligence*. Palo Alto: AAAI Press, 2021. 11106-11115. [doi: 10.1609/aaai.v35i12.17325]
- 7 Huang SZ, Tang EH, Li S, *et al.* Hardware-friendly compression and hardware acceleration for Transformer: A survey. *Electronic Research Archive*, 2022, 30(10): 3755-3785. [doi: 10.3934/era.2022192]
- 8 Gao CX, Zhang N, Li YR, *et al.* Self-attention-based time-variant neural networks for multi-step time series forecasting. *Neural Computing and Applications*, 2022, 34(11): 8737-8754. [doi: 10.1007/s00521-021-06871-1]
- 9 Ainslie J, Ontanon S, Alberti C, *et al.* ETC: Encoding long and structured inputs in Transformers. *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing*. Punta Cana: Association for Computational Linguistics, 2020. 268-284.
- 10 Zaheer M, Guruganesh G, Dubey A, *et al.* Big Bird: Transformers for longer sequences. *Proceedings of the 34th International Conference on Neural Information Processing Systems*. Vancouver: Curran Associates Inc., 2020. 1450.
- 11 Lin TY, Wang YX, Liu XY, *et al.* A survey of Transformers. *AI Open*, 2022, 3: 111-132. [doi: 10.1016/j.aiopen.2022.10.001]
- 12 王超.极限学习机算法及其在热镀锌生产过程建模中的应用研究[博士学位论文].沈阳:东北大学,2016.
- 13 刘明达.基于模糊自适应的锌层厚度预测控制研究[硕士学位论文].沈阳:东北大学,2015.
- 14 王菲.LSTM循环神经网络的研究进展与应用[硕士学位论文].哈尔滨:黑龙江大学,2021.
- 15 Luong T, Pham H, Manning CD. Effective approaches to attention-based neural machine translation. *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*. Lisbon: Association for Computational Linguistics, 2015. 1412-1421.
- 16 Chen MX, Firat O, Bapna A, *et al.* The best of both worlds: Combining recent advances in neural machine translation. *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics*. Melbourne: Association for Computational Linguistics, 2018. 76-86. [doi: 10.18653/v1/P18-1008]
- 17 Vaswani A, Shazeer N, Parmar N, *et al.* Attention is all you need. *Proceedings of the 31st International Conference on Neural Information Processing Systems*. Long Beach: Curran Associates Inc., 2017. 6000-6010.
- 18 Li SY, Jin XY, Xuan Y, *et al.* Enhancing the locality and breaking the memory bottleneck of Transformer on time series forecasting. *Proceedings of the 33rd International Conference on Neural Information Processing Systems*. Vancouver: Curran Associates Inc., 2019. 471.

(校对责编:牛欣悦)