

DOI:10.19651/j.cnki.emt.2415495

基于差分非平稳 Transformer 的液压支架立柱压力预测^{*}

杨艺^{1,2} Aimen Malik^{1,2} 袁瑞甫³ 王科平^{1,2}

(1. 河南理工大学电气工程与自动化学院 焦作 454000; 2. 河南理工大学河南省煤矿装备智能检测与控制重点实验室 焦作 454000; 3. 河南理工大学能源科学与工程学院 焦作 454000)

摘要: 液压支架立柱压力预测是回采工艺决策的重要依据,也是确保围岩稳定的基础信息之一。然而,液压支架立柱压力虽然具有一定的规律性,却无法用简单的数学模型进行预测;且在回采过程中,支架不接顶、顶板破碎、传感器检测误差等带来大量的随机噪声,使得压力数据劣化为非平稳时间序列,给压力的预测带来的很大的困难。本文在Transformer基础上,提出一种差分非平稳Transformer模型,在Transformer的编码器和解码器中分别引入差分归一化和反归一化操作,以提升序列的平稳性。同时,在Transformer中采用去平稳注意力机制,计算序列元素之间的关联关系,以增强模型的预测能力。在真实的煤矿支架立柱数据集上的对比实验表明,本文提出的差分非平稳Transformer的预测效果达到0.674,表现明显优于LSTM、Transformer和非平稳Transformer模型。

关键词: 煤矿安全;工作面;液压支架;压力预测;非平稳Transformer;差分

中图分类号: TP391 文献标识码: A 国家标准学科分类代码: 510.4030

Hydraulic support pressure prediction based on non-stationary differencing Transformer

Yang Yi^{1,2} Aimen Malik^{1,2} Yuan Ruifu³ Wang Keping^{1,2}(1. School of Electrical Engineering and Automation, Henan Polytechnic University, Jiaozuo 454000, China;
2. Henan Key Laboratory of Intelligent Detection and Control of Coal Mine Equipment, Jiaozuo 454000, China;
3. School of Energy Science and Engineering, Henan Polytechnic University, Jiaozuo 454000, China)

Abstract: Hydraulic support pillar pressure prediction has been a pivotal basis for decision-making in the mining process. It has been one of the fundamental pieces of information for ensuring the stability of the surrounding rock. However, although the pressure of hydraulic support pillars followed certain patterns, it couldn't be predicted using simple mathematical models. Additionally, during the mining process, issues such as the support detaching the roof, roof fragmentation, and sensor detection errors introduced a significant amount of random noise, turning the pressure data into a non-stationary time series. This significantly complicated the pressure prediction. Based on the Transformer model, this paper proposed a differencing non-stationary Transformer model, which introduced differencing normalization and de-normalization operations in the Transformer's Encoder and Decoder, respectively, to enhance the stationarity of the series. At the same time, a de-stationary attention mechanism was deployed within the Transformer to calculate the correlations between sequence elements, which thereby enhanced the model's predictive capabilities. Comparative experiments on a real coal mine support pillar dataset showed that the differencing non-stationary Transformer model proposed in this paper achieved a prediction performance of 0.674, which was significantly better than LSTM, Transformer, and non-stationary Transformer models.

Keywords: coal mine safety; work face; hydraulic support; pressure prediction; non-stationary Transformer; differencing

0 引言

煤炭作为我国经济的关键能源资源,对我国的能源结

构和经济发展有着不可忽视的重要性。然而,伴随着煤炭的广泛应用,煤炭安全问题逐渐凸显,涉及矿井安全、采煤工程和煤炭生产等多个方面,一直受到国家的高度关注^[1]。

收稿日期:2024-02-08

* 基金项目:国家重点研发计划项目(2018YFC0604502)、河南省科技攻关项目(232102210040)资助

在采煤过程中,工作面顶板控制对于确保矿井安全至关重要,而液压支架则扮演着支撑工作面顶板的重要角色。液压支架的压力状态是一个关键因素,过大的压力可能引发顶板垮塌、支架崩塌以及地面破坏等严重的事故^[2]。因此,对液压支架的压力进行准确预测成为确保煤矿安全生产的关键任务。

传统的液压支架压力预测方法主要依赖基于地质调查的回归模型,并结合岩层、煤层厚度、开采速度、以及裂缝量等因素建立经验公式。2016 年,Zhang 等^[3]通过回归和概率分析并揭示了浅埋煤层工作面压力、地层岩性和开采参数之间的关系,深入揭示了浅埋煤层工作面最大压力及其分布,从而有助于在液压支架选择和工作面顶板控制方面进行精确估算和决策。2013 年,Hosseini 等^[4]采用被动地震层析成像和地质统计估计,用以研究长壁压力分布。然而,这些实证结果仅适用于特定的地质条件和开采状态,研究范围相对有限。

基于数据的模型采用了支持向量机(support vector machine,SVM)^[5]、灰色神经网络^[6]和遗传算法-反向传播(genetic algorithm-back propagation,GA-BP)神经网络^[7],具有较好的稳定性且误差相对较低。2018 年,Cheng 等^[8]提出了盾构与顶板状态智能感知系统,研究了基于决策树的盾构-顶板相互作用。2022 年,Gao 等^[9]提出了基于多时间尺度特征提取的神经网络,用以预测液压支架压力,从而降低矿井顶板水害风险。但是,这些方法在训练速度和预测精度方面仍有待改进。

近年来,随着新兴技术的发展,传统煤矿正在向智慧矿山^[10]方向发展。为解决切割煤初始状态压力演化趋势的问题,多采用基于统计的时间序列模型,如 ARIMA、SARIMA^[11]。此类模型虽能够处理非平稳时间序列,但其准确性仅适用于较小的时间区间。随着 Transformer^[12-13]模型的引入,序列建模领域(包括时间序列预测)发生了重大变革,成功地解决了长短期记忆网络(long short term memory,LSTM)^[14]在序列建模中遇到的梯度弥散问题^[15]。近年来,基于 Transformer 模型时间序列预测模型出现多种改进,如 Informer^[16]、Reformer^[17]、Pyraformer^[18]和 FEDformer^[19]等,用以解决不同的问题。然而,Transformer 及其改进算法在处理非平稳数据方面仍存在挑战。

确保压力预测的准确性,关键在于消除时间序列的非平稳性^[20]。虽然归一化技术^[21-23]可以使数据趋向平稳,但这可能会削弱真实数据的本来特性。即使在反归一化的过程中,这些固有特性也难以完全复原。为了避免过度平稳化(即过度归一化),非平稳 Transformer(non-stationary transformer,NST)^[24]引入了一种新的去平稳注意力机制。在计算注意力时,该机制考虑了数据失去的固有特性和时间依赖性。通过采用 z-score 归一化^[25],原始的非平稳 Transformer 成功地使非平稳的压力数据获得了显著的平

稳序列。这种方法旨在消除非平稳性的同时保留数据的重要特性,从而提高模型的泛化能力。

本研究提出了非平稳差分 Transformer (non-stationary differencing transformer,NSDT)模型,旨在更高程度地平稳化时间序列数据。该模型采用了差分法^[26]对 NST 进行更高程度的平稳化。首先,一阶差分通过应用滑动窗口从当前数据中减去滞后值,显著提升了序列的平稳性。这种新的平稳性有助于 NSDT 模型更有效地收敛并提高泛化性能。其次,为了克服过度平稳化的问题,NSDT 模型还引入了一种新的去平稳注意力机制。在模型的解码阶段,通过应用反归一化模块,将滞后值重新融入模型输出,同时通过去差分操作将所有统计信息还原到初始状态。最终,在真实的煤矿支架立柱数据集上进行实验,结果表明相较于 LSTM、Transformer 和原始非平稳 Transformer 模型,NSDT 模型在捕获压力预测的短期和长期趋势方面表现最佳。

1 非平稳差分 Transformer 机理

1.1 Transformer 基础架构

在时序序列预测过程中,Transformer 能够捕获序列各数值之间的依赖关系,能更好的建立序列的内生趋势模型,从而使得时间序列的短时预测和长期趋势估计更加精准。Transformer 的核心是时间序列的多头自注意力(multi-head attention mechanism),标准 Transformer 的架构如图 1 所示。

设序列为 x , 经过输入嵌入(input embedding),将序列 x 规范为长度为 n 的词向量 x_e 。然后对 x_e 乘以权重,分别获取各个词向量的查询向量 Q (Query)、键向量 K (Key)和值向量 V (Value),并根据式(1)计算自注意力。

$$\text{Attn}(Q, K, V) = \text{Softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (1)$$

式中: d_k 表示第 k 个词向量的维度。通常情况下,所有词向量的维度相同。

当时间序列具备平稳特性时,即便数据较长,Transformer 依旧可以较好地建立数据间的关联模型,从而准确的完成数据预测。然而,工程实际中存在大量非平稳信号,各个数据间的关联模型呈现为更加复杂的非线性特征。特别是煤矿工作面回采过程中,液压支架降柱和顶板来压等产生的压力突变、以及支架擦顶移架过程中带来的压力扰动,使得支架压力信号的概率分布参数发生较大改变。因此,支架的压力时间序列具备典型的非平稳性。此时,简单的神经网络结构和自注意力机制难以表征数据间的复杂非线性关系,致使时间序列的预测困难。

1.2 非平稳 Transformer 运行机理

2022 年,Liu 等^[24]为以传统的 Transformer 模型为基础,针对非平稳时间序列的预测预测问题提出了非平稳 Transformer (non-stationary Transformer, NST) 模型,如

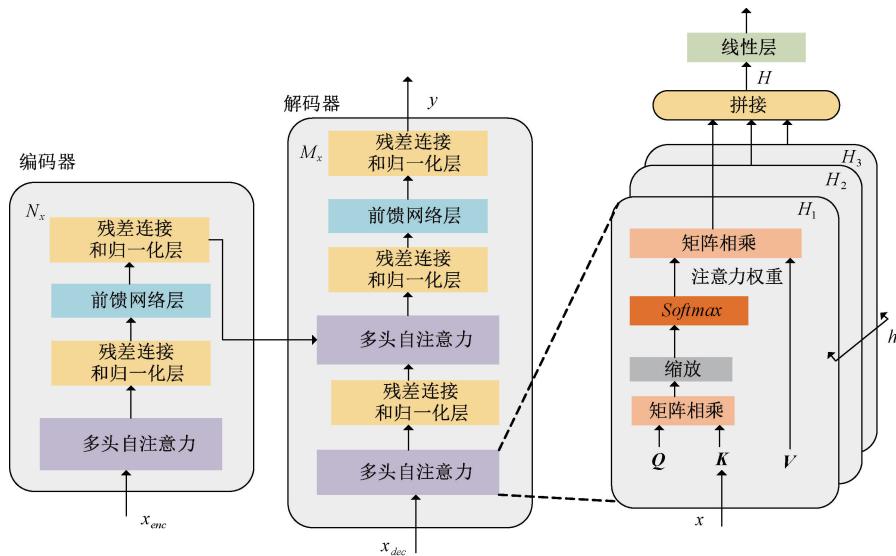


图 1 标准 Transformer 架构

图 2 所示。该模型在传统的 Transformer 的基础上引入了序列平稳化模块和去平稳注意力模块 (de-stationary attention),从而较好地的缓解时间序列的非平稳特性。

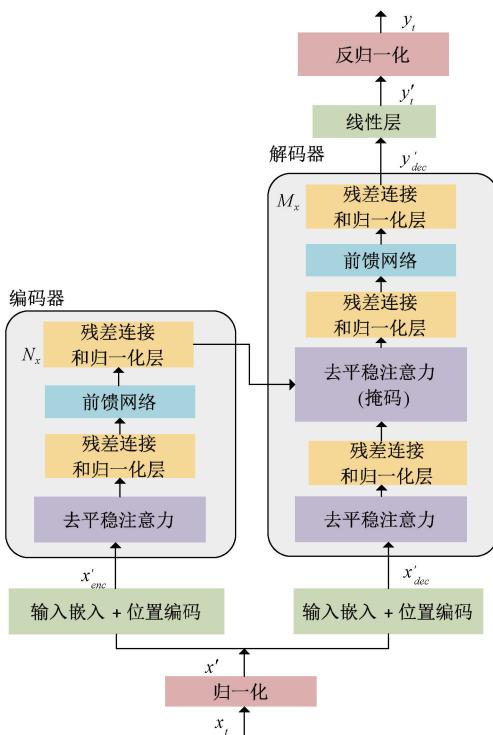


图 2 非平稳 Transformer 结构

序列平稳化模块包含信息归一化 (Normalization) 和反归一化 (De-Normalization) 两个部分。其中, 归一化模块提取时间序列的均值 μ_x 和方差 σ_x , 并重新规范序列的概率分布, 以削弱时间序列的非平稳性。反归一化则是通过将序列的均值 μ_x 和方差 σ_x 重新引入序列中, 以还原真实

信息。

去平稳注意力模块是将时间序列的均值 μ_x 和方差 σ_x 引入自注意力计算中, 以防止时间序列的过度平稳化。使用两个多层感知机, 分别从 (x, μ_x) 和 (x, σ_x) 信息中提取平稳化因子 Δ 和 τ , 参与时间序列的自注意力学习。

2 基于差分非平稳 Transformer 的支架压力预测

时间序列中的数值跳变、外部扰动等引起的非平稳性, 本身是序列中的重要信息。非平稳 Transformer 通过引入时间序列的均值和方差信息来缓解序列非平稳性, 使得非平稳时间序列满足建立预测模型的基础条件。同时, 又采用去平稳注意力模块来避免重要信息因平稳化而丢失。因此, 非平稳 Transformer 中的序列平稳化模块和非平稳注意力模块在功能上存在一定对抗性, 这就意味着需要在二者之间找到最佳平衡点, 才能使非平稳时间序列的预测达到最优。

然而, 平稳化模块和非平稳注意力模块均以时间序列的均值和方差作为依据, 且非平稳因子是通过多层感知机学习而得。这导致无法建立优选模型使上述两个模块在功能对抗上达到最佳平衡。为此, 本文提出一种差分非平稳 Transformer 架构, 建立时间序列中信息间的差分信息模型, 从而缓解非平稳 Transformer 中平稳化模块和非平稳注意力模块的功能对抗, 提高时间序列的预测精度。差分非平稳 Transformer 结构如图 3 所示。

差分非平稳 Transformer 整体由编码器和解码器部分构成。Encoder 主要对输入信号提取注意力等特征, Decoder 则主要依据输入信息和提取的注意力特征, 对时间序列进行预测。

设 t 个采样时刻的时间序列为 x_t , 有:

$$x_t = \{a_1, a_2, \dots, a_s\}, t \geq S \quad (2)$$

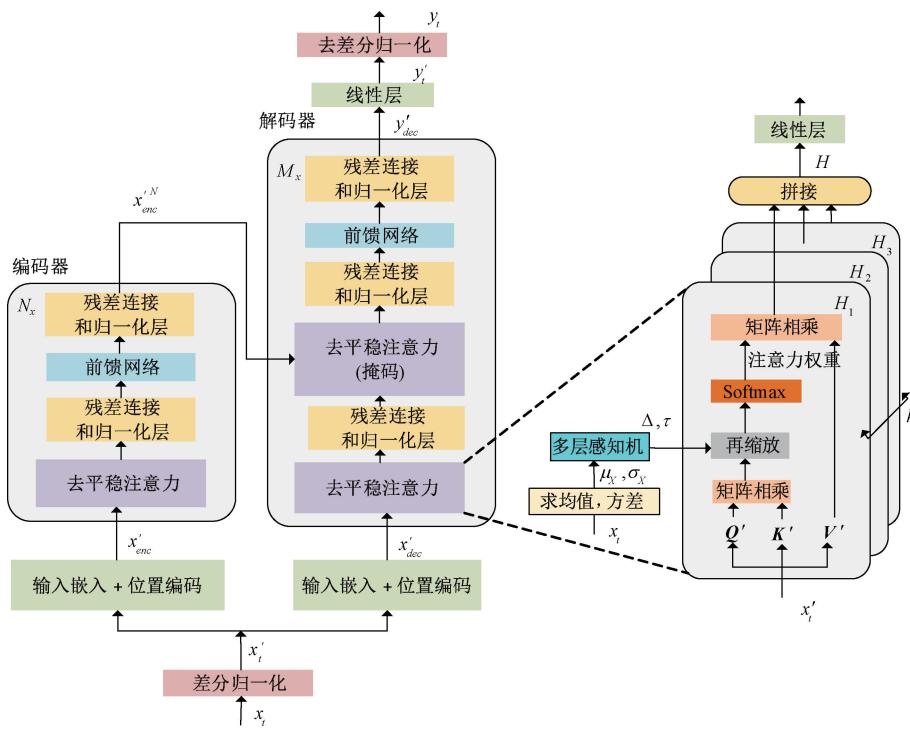


图 3 差分非平稳 Transformer 结构

式中: $a_i \in \mathbb{R}^D, i = \{1, 2, \dots, S\}$ 表示第 i 个数据, D 表示第 i 个数据的维度, s 表示数据长度。可将其视为矩阵, 则:

$$\mathbf{x}_t = [a_1, a_2, \dots, a_s]^T \in \mathbb{R}^{s \times D} \quad (3)$$

2.1 x_t 的差分归一化

x_t 的差分归一化 (Differencing) 用于替代非平稳 Transformer 中的 Normalization 归一化模块。将前一时刻的时间序列 x_{t-1} 仍然当作矩阵处理, 则 $\mathbf{x}_{t-1} = [a_0, a_1, \dots, a_{s-1}]^T \in \mathbb{R}^{s \times D}$, 差分归一化的计算过程如下:

$$\mathbf{x}'_t = \mathbf{x}_t - \mathbf{x}_{t-1} = [a_1 - a_0, a_2 - a_1, \dots, a_s - a_{s-1}]^T \in \mathbb{R}^{s \times D} \quad (4)$$

当 $t = S$ 时, $a_0 = [0, \dots, 0] \in \mathbb{R}^D$ 。

在非平稳 Transformer 中, 归一化模块主要通过去除序列的均值和方差, 将发散的数据重新聚拢形成清晰的概率分布。这相当于对数据的原始分布进行了修正, 从而破坏了原始有效信息。而差分归一化采用式(4)所示的差分信息, 本质上是求取序列的变化状态, 在保留数据原始信息的同时, 依然能改善数据的发散状况, 使数据向变化的概率分布聚拢, 减轻数据的非平稳特性。

本文采用序列的 τ 和 Δ 作为非平稳因子, 以评价序列的非平稳性, 非平稳因子计算公式如下:

$$\begin{cases} \log \tau = \text{MLP}(\sigma_x, \mathbf{x}_t) \\ \Delta = \text{MLP}(\mu_x, \mathbf{x}_t) \end{cases} \quad (5)$$

式中: $\text{MLP}(\cdot)$ 表示多层感知机, $\tau \in \mathbb{R}^{1 \times 1}, \Delta \in \mathbb{R}^{s \times 1}; \mu_x$ 和 σ_x 由下式分别计算。

$$\begin{cases} \mu_x = \frac{1}{S} \sum_{t=0}^s \mathbf{x}_t \\ \sigma_x^2 = \frac{1}{S} \sum_{t=0}^s (\mathbf{x}_t - \mu_x)^2 \end{cases} \quad (6)$$

式中: $\mu_x \in \mathbb{R}^{1 \times D}, \sigma_x^2 \in \mathbb{R}^{1 \times D}$ 。

2.2 去平稳注意力

去平稳注意力模块采用非平稳 Transformer 相同的结构, 需要对差分后的信息进行输入嵌入和位置编码 (positional encoding)

$$\mathbf{x}'_{enc} = \text{linear}(\mathbf{x}') + PE(\mathbf{x}'), \quad (7)$$

式中: $\text{linear}(\cdot)$ 表示全连接神经网络, $PE(\cdot)$ 表示位置编码。则有 $\mathbf{x}'_{enc} \in \mathbb{R}^{s \times d_k}$, d_k 表示数据扩展后的维度, 即 $\text{linear}(\cdot)$ 的输出维度。

在编码器和解码器中, 去平稳注意力模块的结构相同, 只是由于在解码器中需要加入逐步顺序加入时间序列信号, 用以预测数据, 所以需要对输入信息采用掩码操作。

1) 注意力基础值计算

输入去平稳注意力模块的基础值与 Transformer 相同, 分别为查询向量 \mathbf{Q}' 、键向量 \mathbf{K}' 和值向量 \mathbf{V}' , 有:

$$\begin{aligned} \mathbf{Q}' &= \mathbf{x}'_{enc} \times \mathbf{W}_q \\ \mathbf{K}' &= \mathbf{x}'_{enc} \times \mathbf{W}_k \\ \mathbf{V}' &= \mathbf{x}'_{enc} \times \mathbf{W}_v \end{aligned} \quad (8)$$

式中: $\mathbf{W}_q \in \mathbb{R}^{d_k \times d_k}, \mathbf{W}_k \in \mathbb{R}^{d_k \times d_k}$ 和 $\mathbf{W}_v \in \mathbb{R}^{d_k \times d_k}$ 分别是查询向量、键向量和值向量的权重, 且有 $\mathbf{Q}' \in \mathbb{R}^{s \times d_k}, \mathbf{K}' \in \mathbb{R}^{s \times d_k}, \mathbf{V}' \in \mathbb{R}^{s \times d_k}$ 。当采用多头注意力机制时, 设注意力

分头计算的数量为 h 。则按多头计算注意力基础值时, 第 j 个注意力头的基础值计算为

$$\begin{aligned} \mathbf{Q}'_j &= x'_{enc} \times \mathbf{W}_q^j \\ \mathbf{K}'_j &= x'_{enc} \times \mathbf{W}_k^j \\ \mathbf{V}'_j &= x'_{enc} \times \mathbf{W}_v^j \end{aligned} \quad (9)$$

式中: $j = \{1, 2, \dots, h\}$, $\mathbf{W}_q^j \in \mathbb{R}^{d_k \times \frac{d_k}{h}}$, $\mathbf{W}_k^j \in \mathbb{R}^{d_k \times \frac{d_k}{h}}$, $\mathbf{W}_v^j \in \mathbb{R}^{d_k \times \frac{d_k}{h}}$ 。各注意力头以并行的方式对输入信息计算注意力, 得 $\mathbf{Q}'_j \in \mathbb{R}^{S \times \frac{d_k}{h}}$, $\mathbf{K}'_j \in \mathbb{R}^{S \times \frac{d_k}{h}}$, $\mathbf{V}'_j \in \mathbb{R}^{S \times \frac{d_k}{h}}$ 。

2) 注意力计算

按照多头注意力计算方式, 各头注意力基础值 \mathbf{Q}'_j , \mathbf{K}'_j , \mathbf{V}'_j 和非平稳因子 τ 、 Δ , 经过矩阵相乘 (MatMul)、再缩放 (Rescale)、Softmax 之后, 得到去平稳注意力:

$$H_j = Attn(\mathbf{Q}'_j, \mathbf{K}'_j, \mathbf{V}'_j) = \text{Softmax}\left(\frac{\tau \mathbf{Q}'_j \mathbf{K}'_j^\top + 1 \Delta^\top}{\sqrt{d_k}}\right) \mathbf{V}'_j \quad (10)$$

式中: $1 \in \mathbb{R}^{S \times 1}$, $H_j \in \mathbb{R}^{S \times \frac{d_k}{h}}$ 。

最后将所有单头注意力 H_j 拼接在一起, 形成序列注意力 H :

$$H = \text{concat}(H_1, H_2, \dots, H_h) \quad (11)$$

式中: $H \in \mathbb{R}^{S \times d_k}$ 。

2.3 差分非平稳 Transformer 数值预测

差分非平稳 Transformer 的编码器和解码器分别由 N_x 和 M_x 层去平稳注意力计算单元构成。在解码器的最后一层输出的去平稳注意力基础上, 经过一前馈网络和归一化层后, 得到 $y'_{dec}, y''_{dec} \in \mathbb{R}^{O \times d_k}$ 。其中 O 表示输出预测长度。

预测长度为 O 的差分序列 x_{t-1} 的最后值被加入到解码器输出 y'_t 中。最后, 采用去差分归一化 (De-Differencing) 复原预测信号。

$$y_t = y'_t + x_{t-1}[S - O + 1:] \quad (12)$$

式中: y_t 表示第 t 个采样时刻得预测值, $y_t \in \mathbb{R}^{O \times D}$ 。

差分非平稳 Transformer 算法的伪代码如算法 1 所示。

算法 1: 差分非平稳 Transformer 算法伪代码如下:

输入: 序列长度 S ; 特征变量 D ; 模型维度 d_k ; 编码层 N ; 解码层 M .

输出: 预测序列长度 O ;

1. x_t 经过差分归一化得到 $x'_t \in \mathbb{R}^{S \times D}$
2. 求取 x_t 的均值 $\mu_x \in \mathbb{R}^{1 \times D}$ 和方差 $\sigma_x^2 \in \mathbb{R}^{1 \times D}$
3. 通过多层感知机得到非平稳因子 $\log \tau \in \mathbb{R}^{1 \times 1}$ 和 $\Delta \in \mathbb{R}^{S \times 1}$

4. $x'_{enc} = x'_t, x'_{dec} = \text{concat}(x'_{t(\frac{S}{2}:S)}, \text{zeroes}(O, C))$ 将 $x'_t \in \mathbb{R}^{S \times D}$

作为序列长度为 S 的编码器输入 $x'_{enc} \in \mathbb{R}^{S \times D}$, 将 $x'_t \in \mathbb{R}^{S \times D}$ 的后半部分序列与长度为 O 的序列进行拼接作为解码器的输入 $x'_{dec} \in \mathbb{R}^{(S/2+O) \times D}$

5. 对 $x'_{enc} \in \mathbb{R}^{S \times D}$ 进行输入嵌入和位置编码得到 $x'_{enc} \in \mathbb{R}^{S \times d_k}$

6. 在编码器中, $x'_{enc} \in \mathbb{R}^{S \times d_k}$ 送入去平稳注意力层、前馈网络残差连接和归一化层分别进行处理, 得到编码器的输出 $x'_{enc}^N \in \mathbb{R}^{S \times d_k}$

7. 对 $x'_{dec} \in \mathbb{R}^{(S/2+O) \times D}$ 进行输入嵌入和位置编码得到 $x'_{dec} \in \mathbb{R}^{(S/2+O) \times d_k}$

8. 在解码器中, $x'_{dec} \in \mathbb{R}^{(S/2+O) \times d_k}$ 首先送入去平稳注意力层, 然后经过残差链接和归一化层处理, 随后与编码器得到的 $x'_{enc} \in \mathbb{R}^{S \times d_k}$ 一起送入带有掩码的去平稳注意力层, 再送入前馈网络、残差连接和归一化层得到 $x'_{dec}^M \in \mathbb{R}^{(S/2+O) \times d_k}$

9. 将 x'_{dec}^M 中最后一个长度为 O 的预测序列 $y'_{dec} \in \mathbb{R}^{O \times d_k}$ 送入线性层以转换特征维度得到 $y'_t \in \mathbb{R}^{O \times D}$

10. y'_t 经过去差分归一化得到 $y_t \in \mathbb{R}^{O \times D}$

11. 得到结果 y_t

3 实验及结果分析

3.1 液压支架压力预测数据集

液压支架压力数据来自郑州煤机液压电控公司部署在某矿综采工作面的实时数值。该工作面配备 227 个液压支架, 压力数据采样间隔为 1 s, 并以连续 7 天的数据制作了数据集。数据集较为准确地反映了支架降柱、移架、升柱等动作过程, 涵盖了支架触顶、脱顶、擦顶等与压力密切相关的支架姿态。因此, 该数据集对一般煤矿的压力预测具有一定的泛化能力。但需要注意的是, 数据集尚未涵盖坚硬顶板、易碎顶板等特殊地质条件下的煤矿压力数据。因此, 本文方法对这类工作面的压力预测方法需要进一步研究。

由于该数据属于初步采集状态, 须经过数据预处理和清洗才能确保其质量和准确性。图 4(a)中异常的峰值表明存在噪声干扰, 这类干扰通常为瞬时性质, 因此首先要进行剪切以去除噪声和异常值, 剪切后的数据将按比例转换为正常值, 如图 4(b)所示。

图 5 给出了压力数据的详细直观图, 图 5(a)为清洗后的数据, 图 5(b)为序列 [3 530:3 620] 的局部放大示意图, 从中可以观察到压力的敏感性。序列 [3 535:3 610] 描述了液压支架在煤矿开采过程中的工作循环, 这一循环可以细分为 4 个关键阶段。首先是初始支护阶段, 此阶段的主要任务是将支架提升至预设的初始支撑力度, 以确保煤层

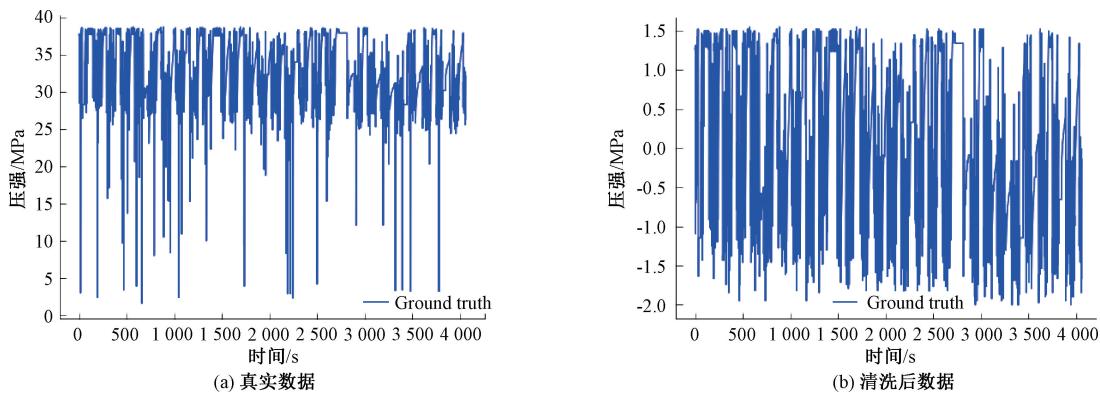


图 4 液压支架观测的原始压力数据和处理后压力数据

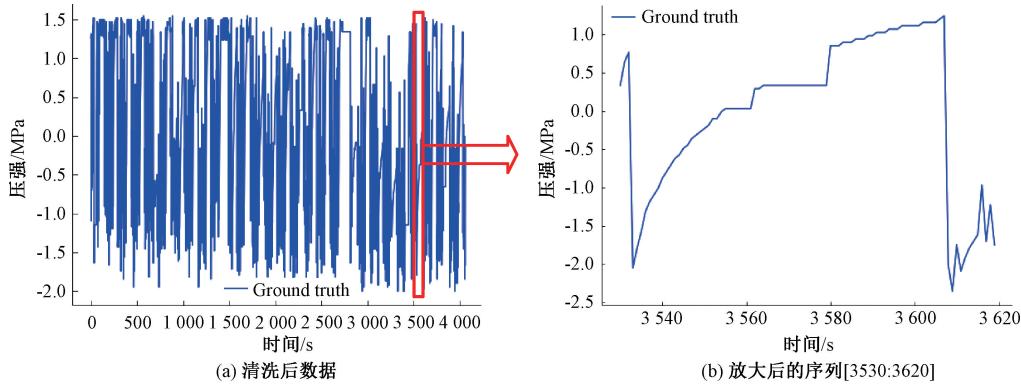


图 5 液压支架压力真实数据

的稳定性。随后进入承载阶段,此时随着采煤作业的推进,支架所承受的压力会逐渐增大,以适应煤层的动态变化。第3阶段涉及邻近支架的移动,这一过程中,部分压力会从移动的支架转移至当前支架,增加其负担。最后是降压阶段,此阶段的目的是迅速减轻支架压力,准备进行下一轮的支撑循环。这一连续的工作流程确保了煤矿开采的安全性和效率。

所有实验均基于 PyTorch 实现，并在 NVIDIA GeForce RTX 3060 上运行。数据集的 90% 作为训练集，10% 作为测试集。实验的部分参数设置如表 1 所示。

表 1 实验参数设置

模型参数	参数值
输入序列长度	$S=60$
输出序列长度	$O=1$
编码器层	$N=2$
解码器层	$M=1$
头	$h=8$
模型维度	$d_k=512$
批处理大小	16
优化器	Adam
损失函数	均方误差/均方根误差
学习率	0.000 1
LSTM 层	3
LSTM 隐藏层	120

3.2 算法对比实验

为对比各种压力预测模型的性能,采用均方误差(mean-square error,MSE)、均方根误差(root mean-square error,RMSE)和预测精度(Accuracy)对模型性能进行评价,均方误差、均方根误差以及预测精度的计算公式为:

$$MSE = \frac{1}{N} \sum_{i=0}^N (y_i - \hat{y}_i)^2 \quad (13)$$

$$RMSE = \sqrt{MSE} \quad (14)$$

$$Accuracy = 1 - \frac{1}{N} \sum_{i=0}^N |y_i - \hat{y}_i| \quad (15)$$

式中: y_i 为真实值、 \hat{y}_i 为预测值。

在所对比的算法中,本文将模型的输入序列长度均设置为 60,对应的输出序列长度设置为 1。

通过 z-score 归一化改进的 LSTM 模型的 $MSE = 0.399$, $Accuracy = 0.525$, 而应用相同归一化技术的 Transformer 模型的 MSE 稍低, 为 0.380 , $Accuracy = 0.522$ 。NST 模型的 $MSE = 0.322$, $Accuracy = 0.670$, 性能有所提高。此外, NSDT 模型的表现明显优于其他模型, MSE 最低, 仅为 0.314 , $Accuracy$ 最高, 达到了 0.674 。结果表明, 与 NST、LSTM 和 Transformer 模型相比, NSDT 模型能更有效地捕捉压力的基本模式。

因为 NSDT 和 NST 具有相同的模型参数,所以在推理时间方面,这两个模型的推理时间大致相同,而预测误

差却显著降低。液压支架压力测试数据应用不同模型获得的预测结果如表2所示。

表2 液压支架压力测试数据应用不同模型获得的预测结果

模型	均方误差	预测精度
LSTM	0.399	0.525
Transformer	0.380	0.522
NST	0.322	0.670
NSDT	0.314	0.674

各模型对测试数据进行压力预测的结果如图6所示,在这些模型中,NSDT模型的预测曲线与真实值数据的贴合度最高,其差分归一化的特点增强了Transformer模型的非平稳预测能力。NST模型统计特性的不一致导致了其较差的收敛性。尽管如此,NST模型的去静态化有助于克服过度静态化的问题。

将Transformer模型的预测与z-score归一化一起考虑会导致预测结果不佳。即便其与真实值的重叠度较高,但在预测压力峰值时常常偏离实际情况。液压支架的峰

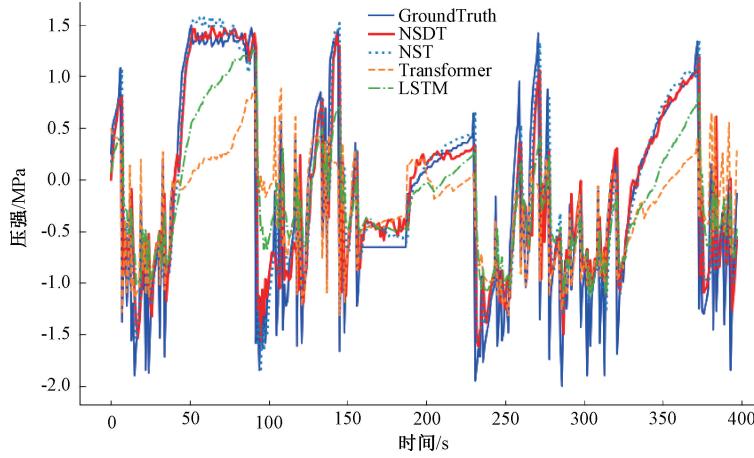


图6 测试数据的比较模型的预测结果

值压力是煤矿安全的首要条件,因此,Transformer在这一关键指标上的不稳定表现削弱了其可靠性。这种预测不准的根本原因在于过度静态化,这不仅损害了数据的内在信息,使其在模型中未能得到适当重视。

表3提供了选定测试子序列的MSE/RMSE指标分析。结果显示NSDT模型在这些指标上的损失最小,这表明它在精确性方面的表现最佳。原始NST模型在准确率评估中紧随其后,占据第二位。

表3 不同模型在液压支架压力数据上的预测准确度

模型	LSTM		Transformer		NST		NSDT	
	均方误差	均方根误差	均方误差	均方根误差	均方误差	均方根误差	均方误差	均方根误差
[50:105]	0.492	0.701	0.392	0.626	0.197	0.443	0.187	0.432
[80:135]	0.608	0.779	0.580	0.761	0.432	0.657	0.426	0.652
[245~300]	0.485	0.696	0.216	0.464	0.525	0.724	0.490	0.70
[320:375]	0.282	0.530	0.392	0.626	0.197	0.443	0.187	0.432

图7展示了放大视图的测试数据子序列,其中包含了多个比较模型的预测结果。在图7(a)、(b)和(d)中,可以观察到,NSDT模型的粗体预测曲线在所有比较模型中与真实值的吻合度最高,其MSE/RMSE损失值也是最低的,显示出卓越的预测准确性。NST模型的预测精度位列第二,紧随NSDT之后。相比之下,LSTM和Transformer

模型的预测曲线大多数情况下与真实值存在较大偏差。特别是在图7(c)中,尽管Transformer模型展示了最小的MSE/RMSE损失,但该模型却未能准确预测地面的峰值曲线。虽然Transformer在与真实值的重叠程度方面表现良好,但其在预测峰值压力方面的不足,其在可靠性方面有待提升。

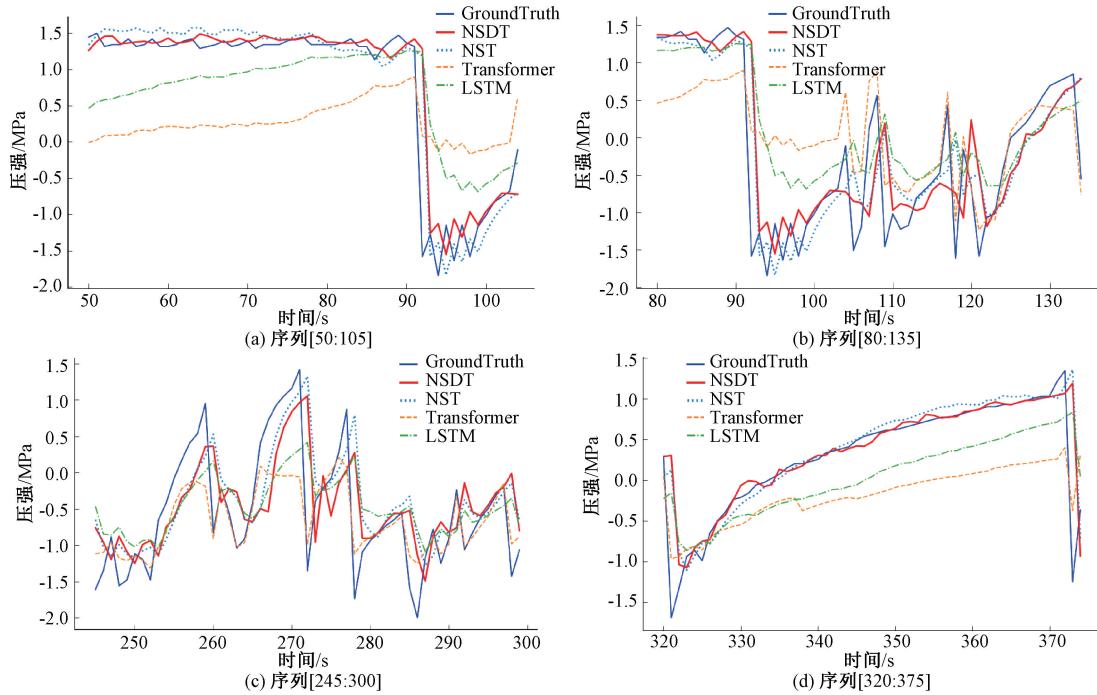


图 7 测试数据子序列与所有比较模型预测结果的可视化图

4 结 论

本文针对压力序列的非平稳特性,提出了一种差分非平稳 Transformer 模型,通过差分归一化和反归一化环节削减序列的非平稳性;同时采用去平稳注意力模块提升对序列的预测能力。真实数据的实验证明,差分非平稳 Transformer 的预测精度达到了 0.674,明显优于其他模型。

支架的降移升过程本身是引起压力数据大幅跳变的关键因素之一,在后续的研究过程中,将在模型中引入支架的动作序列,与压力时间序列同步训练,以提升模型预测精度。

参考文献

- [1] 王国法,刘峰,庞义辉,等. 煤矿智能化——煤炭工业高质量发展的核心技术支撑[J]. 煤炭学报,2019,44(2):349-57.
- [2] WANG H, XUE S, JIANG Y, et al. Field investigation of a roof fall accident and large roadway deformation under geologically complex conditions in an underground coal mine[J]. Rock Mechanics and Rock Engineering, 2018, 51: 1863-1883.
- [3] ZHANG T, ZHAO Y, ZHU G, et al. A multi-coupling analysis of mining-induced pressure characteristics of shallow-depth coal face in Shendong mining area[J]. Journal of China Coal Society, 2016, 41(S2): 287-296.
- [4] HOSSEINI N, ORRAEE K, SHAHRIAR K, et al. Studying the stress redistribution around the longwall mining panel using passive seismic velocity tomography and geostatistical estimation[J]. Arabian Journal of Geosciences, 2013, 6: 1407-1416.
- [5] MAHMOUD A A, ELKATATNY S, MAHMOUD M, et al. Prediction of pore and fracture pressures using support vector machine [C]. 11th International Petroleum Technology Conference. European Association of Geoscientists & Engineers, 2019, 2019(3): 1-11.
- [6] WANG K, ZHUANG X, ZHAO X, et al. Roof pressure prediction in coal mine based on grey neural network[J]. IEEE Access, 2020, 8: 117051-117061.
- [7] TAN T, YANG Z, CHANG F, et al. Prediction of the first weighting from the working face roof in a coal mine based on a GA-BP neural network[J]. Applied Sciences, 2019, 9(19): 4159.
- [8] CHENG J, WAN Z, JI Y. Shield-roof interaction in longwall panels: Insights from field data and their application to ground control[J]. Advances in Civil Engineering, 2018, 2018(12): 1-18.
- [9] GAO T, LI Z, LI H, et al. A prediction network for hydraulic support pressure based on multimescale feature extraction [J]. Mathematical Problems in Engineering, 2022: 1-9, DOI: 10.1155/2022/8526744.
- [10] WANG G, REN H, ZHAO G, et al. Research and practice of intelligent coal mine technology systems in China[J]. International Journal of Coal Science &

- Technology, 2022, 9(1): 24.
- [11] PANG Y H, WANG H B, ZHAO J J, et al. Analysis and prediction of hydraulic support load based on time series data modeling[J]. Geofluids, 2020: 1-15, DOI: 10.1155/2020/8851475.
- [12] VASWANI A, SHAZEE N, PARAR N, et al. Attention is all you need [J]. Advances in Neural Information Processing Systems, 2017, 30, DOI: 10.48550/arXiv.1706.03762.
- [13] 李欣宇,孙传猛,魏宇,等.融合Transformer与残差通道注意力的恶劣场景水位智能检测方法[J].电子测量与仪器学报,2023,37(1):59-69.
- [14] 周建新,霍彤明.基于NMWOA-LSTM的卷取温度预测模型[J].电子测量技术,2023,46(18):60-66.
- [15] LE P, ZUIDEMA W. Quantifying the vanishing gradient and long distance dependency problem in recursive neural networks and recursive LSTMs[C]. Proceedings of the 1st Workshop on Representation Learning for NLP, 2016: 87-93.
- [16] ZHOU H, ZHANG S, PENG J, et al. Informer: Beyond efficient transformer for long sequence time-series forecasting [C]. Proceedings of the AAAI Conference on Artificial Intelligence, 2021, 35(12): 11106-11115.
- [17] KITAEV N, KAISER Ł, LEVSKAYA A. Reformer: The efficient transformer[C]. International Conference on Learning Representations, 2019.
- [18] LIU S, YU H, LIAO C, et al. Pyraformer: Low-complexity pyramidal attention for long-range time series modeling and forecasting [C]. International Conference on Learning Representations, 2021.
- [19] ZHOU T, MA Z, WEN Q, et al. Fedformer: Frequency enhanced decomposed transformer for long-term series forecasting[C]. International Conference on Machine Learning, 2022: 27268-27286.
- [20] AHUJA K, CABALERO E, ZHANG D, et al. Invariance principle meets information bottleneck for out-of-distribution generalization [J]. Advances in Neural Information Processing Systems, 2021, 34: 3438-3450.
- [21] SHAO X. Self-normalization for time series: A review of recent developments[J]. Journal of the American Statistical Association, 2015, 110(512): 1797-1817.
- [22] PASSAIS N, TEFAS A, KANNIAINEN J, et al. Deep adaptive input normalization for time series forecasting[J]. IEEE Transactions on Neural Networks and Learning Systems, 2019, 31(9): 3760-3765.
- [23] KIM T, KIM J, TAE Y, et al. Reversible instance normalization for accurate time-series forecasting against distribution shift[C]. International Conference on Learning Representations, 2021.
- [24] LIU Y, WU H, WANG J, et al. Non-stationary transformers: Exploring the stationarity in time series forecasting [J]. Advances in Neural Information Processing Systems, 2022, 35: 9881-9893.
- [25] ALI P J M, FARAJ R H, KOYA E, et al. Data normalization and standardization: A technical report[J]. Mach Learn Tech Rep, 2014, 1(1): 1-6.
- [26] MCCABE B P M, TREAYNE A R. Testing a time series for difference stationarity[J]. The Annals of Statistics, 1995, 23(6): 1015-1028.

作者简介

杨艺,博士,副教授,主要研究方向为深度学习、强化学习等。

E-mail:yangyi@hpu.edu.cn

Aimen Malik(通信作者),硕士研究生,主要研究方向为时间序列预测、智能控制。

E-mail:aimenmalik78665@gmail.com