

DOI:10.13296/j.1001-1609.hva.2026.08.008

基于注意力—门控循环—时间卷积网络的变压器油温预测

张京娥¹, 程中杰², 刘 宝²

(1. 甘肃钢铁职业技术学院信息工程系, 甘肃 嘉峪关 735100; 2. 西安科技大学电气与控制工程学院, 西安 710054)

摘要: 顶层油温是评价电力变压器热特性的重要指标, 准确预测顶层油温变化趋势对评估变压器的负载能力和及时发现潜在的内部热故障具有重要作用。针对顶层油温预测中精度低、特征选择过程繁琐和信息使用错误的问题, 提出了一种基于时间卷积网络(TCN)、注意力机制(attention mechanism)和门控循环单元(GRU)的变压器顶层油温预测方法(Attention-GRU-TCN)。首先, 在卷积中引入TCN, 保证预测过程中的因果性, 减少未来数据对预测过程的干扰。其次, 注意力机制的引入, 赋予关键数据特征更高的权重, 取代复杂耗时的数据筛选过程, 提高预测精度。最后, 在某公司与传统的变压器顶层油温预测算法(例如, RNN、LSTM、GRU、BiLSTM、Transformer等)进行实验实测。结果表明, 与RNN模型、LSTM模型、GRU模型、BiLSTM模型、Transformer模型相比, Attention-GRU-TCN模型的平均绝对误差(e_{MAE})分别下降23.69%、15.88%、7.77%、29.35%、14.51%; 均方误差(e_{MSE})分别下降45.79%、31.34%、12.31%、43.32%、38.85%; 决定系数(R^2)上升6.0%、3.10%、0.79%、12.30%、4.57%。说明该方法对变压器顶层油温的预测精度更高, 有助于为潜在热故障预警提供参考并分析变压器负载能力。

关键词: 变压器; 顶层油温预测; 注意力机制; 门控循环单元; 时间序列预测

Transformer Oil Temperature Prediction Based on Attention-gated Recurrent-temporal Covolutional Network

ZHANG Jing'e¹, CHENG Zhongjie², LIU Bao²

(1. Department of Information Engineering, Gansu Iron and Steel Vocational Technical College, Gansu Jiayuguan 735100, China;

2. College of Electrical and Control Engineering, Xi'an University of Science and Technology, Xi'an 710054, China)

Abstract: Top oil temperature is an important indicator for evaluating thermal characteristics of power transformer. Accurately predicting the variation trend of the top oil temperature plays a crucial role in assessing the load capacity of the transformer and promptly identifying potential thermal faults. In view of such issues as low accuracy, cumbersome feature selection process, and incorrect use of information in top oil temperature prediction, a transformer top oil temperature prediction method (Attention-GRU-TCN) is proposed based on temporal convolutional network (TCN), attention mechanism, and gated recurrent unit (GRU). First, TCN is introduced into the convolution to ensure causality in the prediction process and reduce the interference of future data on the prediction process. Then, the introduction of the attention mechanism assigns higher weights to key data features, replacing the complex and time-consuming data filtering process, thereby improving prediction accuracy. Finally, the proposed method is experimentally validated at a company against traditional top oil temperature prediction models (e.g., RNN, LSTM, GRU, BiLSTM, and Transformer, ect). The results show that compared with the RNN model, LSTM model, GRU model, BiLSTM model and Transformer model, the Attention-GRU-TCN model achieves a decrease in mean absolute error (e_{MAE}) by 23.69%, 15.88%, 7.77%, 29.35%, and 14.51% respectively, a decrease in mean squared error

收稿日期: 2026-04-30; 修回日期: 2026-06-18

基金项目: 国家自然科学基金面上项目(52074213); 中煤西安设计工程有限责任公司项目(XMSHT*KC-KY-SMGF-00-24-020)。

Project Supported by National Natural Science Foundation of China(52074213), China Coal Xi'an Design Engineering Co., Ltd. (XMSHT*KC-KY-SMGF-00-24-020).

(e_{MSE}) by 45.79%, 31.34%, 12.31%, 43.32%, and 38.85% respectively; and an increase in the coefficient of determination (R^2) by 6.0%, 3.10%, 0.79%, 12.30%, and 4.57% respectively. It is shown that the method has higher prediction accuracy for transformer top oil temperature, which is conducive to providing a reference for early warning of potential thermal faults and analyzing the load capacity of transformers.

Key words: transformer; top oil temperature prediction; attention mechanism; gated recurrent unit; time series prediction

0 引言

变压器作为电力系统的核心输变电设备,其应用十分广泛。然而,变压器故障所引发的系统运行事故也日益增多。据统计,在变压器的所有故障类型中,过热故障所占比例最大。其中,局部过热(如绕组热点、接触不良等)因初期难以直接检测,往往演变为严重事故;而整体过热则会加速绝缘老化,缩短设备寿命。顶层油温是当前工程实践中评估变压器整体热状态的主要指标,且测量难度较低。通过准确预测顶层油温的变化趋势,可以掌握变压器的整体热负荷水平及散热能力,为判断是否存在异常温升(尤其是整体性过热倾向)提供依据^[1-4]。

虽然变压器顶层油温是可直接测量的物理量,但单纯的实时测量仅能反映当前热状态,无法回答“未来顶层油温将如何演变”这一运维决策中的核心问题。顶层油温变化具有显著的热惯性,从内部热点温升到顶层油温响应存在明显的时间滞后。因此,仅依赖实时测量值进行告警,当温度达到阈值时,绝缘损伤可能已经发生。相比之下,顶层油温预测能够实现从被动告警向主动预警的跨越:通过预测未来数小时乃至一天内的顶层油温趋势,运维人员可在温度实际超标前启动冷却装置、调整负载或转移负荷,从而避免过热故障。因此,开展顶层油温精准预测,对于提前识别异常热累积趋势、辅助判断潜在故障、优化运行维护策略具有重要意义。

深度学习技术如:循环神经网络(recurrent neural net, RNN)、长短期记忆(long short-term memory, LSTM)、门控循环单元(gate recurrent unit, GRU)、双向长短期记忆网络(bidirectional long short-term memory, BiLSTM)等在处理数据方面具有广泛应用^[5-12],等。董雪情等选取环境温度以及负载数据对应的有功和无功功率作为特征参数,采用LSTM实现对顶层油温的预测^[13]。赵军等提出了一种基于多输出策略的GRU神经网络多步预测模型^[14]。通过改变模型的结构超参数和训练超参数来研究超参数之间的耦合关系,采用多目标灰狼优化算法对具有不同预测结果倾向的GRU模型的超参数进行

优化。谭志超等使用集成经验模态分解和小波将浓度信号分解为几个子信号,然后应用CNN-GRU模型^[15]来预测浓度信号。卫永鹏等将Radam优化器和门控循环单元相结合,提出了一种改进的GRU预测模型^[16],用于变压器油中的气体浓度。周春阳等基于LSTM网络^[17]预测了某±800 kV换流站的油温趋势,并通过油温预警验证了该方法的实用性。孟衡等考虑了外部环境对负载的影响以及相邻节点之间存在的空间耦合特性,并在预测模型中加入了Transformer,提高了变压器负载预测的准确性^[18]。樊清川等使用粒子群优化算法来优化BiLSTM网络^[19]的超参数以提高模型的准确性和鲁棒性。然而,BiLSTM网络的随机生成的超参数增加了模型的不稳定性,容易陷入局部最优。上述方法加入负载和天气条件等与顶层油温相关的影响因素,为变压器顶层油温预测提供了思路。

针对上述问题,文中提出一种结合时间卷积网络、注意力机制和门控循环单元的变压器顶层油温预测方法(Attention-GRU-TCN)。首先,采用因果卷积和扩张卷积的TCN网络对输入功率数据进行时序特征提取,然后,引入具有自学习能力的注意力机制,对TCN输出的特征序列进行自适应加权,其次,利用门控循环单元对加权后的特征序列进行长期时序依赖建模,捕捉顶层油温变化的长期热惯性及滞后响应关系;最后,经全连接层输出顶层油温预测值。

1 理论基础

1.1 GRU 模块

GRU网络^[20]的工作原理如下:

$$r_t = \sigma(W_r h_{t-1} + W_r x_t) \quad (1)$$

$$u_t = \sigma(W_z h_{t-1} + W_z x_t) \quad (2)$$

$$\tilde{h}_t = \tanh[r_t \odot (W'_z h_{t-1}) + W'_x x_t] \quad (3)$$

$$h_t = u_t \odot h_{t-1} + (1 - u_t) \odot \tilde{h}_t \quad (4)$$

$$\tanh = \frac{e^x - e^{-x}}{e^x + e^{-x}} \quad (5)$$

式(1)-(5)中: σ 为sigmoid函数,函数的输出范

围(0,1)是作为门控信号,用于控制信息遗忘或保留的比例; W_r 、 W_z 、 W'_h 、 W'_x 分别为可训练的权重; r_t 为重置门; u_t 为更新门; x_t 为时间 t 的输入数据; h_t 为时间 t 的输出; h_{t-1} 为前一刻的隐藏状态; $\tilde{h}_t$ 为基于当前输入和经重置门筛选后的历史; $\odot$ 为 Hadamard 乘法; $\tanh$ 为双曲正切函数。GRU 网络结构见图 1。

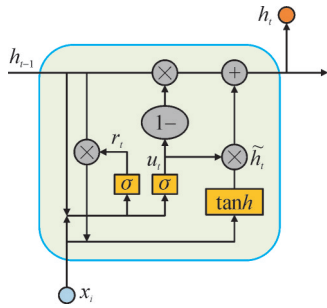


图1 GRU 模型结构图

Fig. 1 GRU model structure diagram

隐藏状态 h_{t-1} 的物理映射: 变压器顶层油温不能瞬变, 具有显著的热惯性, 且无极端异常。当前时刻顶层油温不仅取决于当前功率, 更强烈依赖于前一时刻的温度及历史功率累积效应。在顶层油温预测中, h_{t-1} 可以理解为综合热状态。它不仅包含上一时刻的顶层油温, 还编码了历史功率累积热效应等无法直接观测的状态变量。

重置门 r_t 的作用: 当功率突然骤降, r_t 会趋近于 0。模型会“遗忘”部分高功率下的历史热状态, 让 $\tilde{h}_t$ 基于当前较低功率 x_t 重新计算升温趋势。这符合物理规律: 当热源消失, 顶层油温不再受高功率历史的主导, 开始向环境温度衰减。

更新门 u_t 的作用: u_t 决定了顶层油温变化的平滑程度。由于油温变化时间常数(几分钟到几小时)远大于数据采样间隔(如分钟级), u_t 在稳态下通常接近 0, 使得 $h_t \approx h_{t-1}$, 模拟热惯性; 在功率突变时, u_t 增大, 允许模型吸收新的热输入。

1.2 注意力机制

受人类注意力启发, 专家学者提出了注意力机制^[21]。在数据预测领域, 注意力机制有效解决了传统神经网络的一些问题, 如随着输入长度增加而导致系统性能下降以及特征提取能力不足^[22]。自注意力机制^[23]是注意力机制的一种, 其结构示意图见图 2。自注意力工作原理如下。

$$\text{Attention}(Q, K, V) = \text{Soft max} \left(\frac{QK^T}{\sqrt{D_k}} \right) V \quad (6)$$

式(6)中: Q 为查询向量; K 为键向量; V 为值

向量。

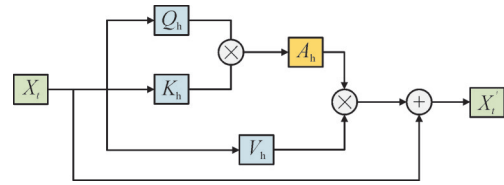


图2 自注意力机制模型结构图

Fig. 2 Self-attention structure diagram

$x_t \in R^6$ 为 t 时刻的输入特征向量。自注意力机制的查询向量 Q 、键向量 K 、值向量 V 均由输入功率特征 X_t 经不同可学习权重矩阵线性变换得到。 Q 源自当前时刻的特征, 代表“当前工况下, 需要关注什么历史经验”。 K 源自所有历史时刻的特征, 代表“每个历史时刻的工况标签”。 V 源自所有历史时刻的特征, 代表“每个历史时刻的实际热状态信息”。

1.3 TCN 模型

时域卷积网络(temporal convolutional network)简称 TCN^[24], 主要由扩张卷积、残差连接块、因果卷积、Dropout 组成, 扩张卷积结构见图 3。

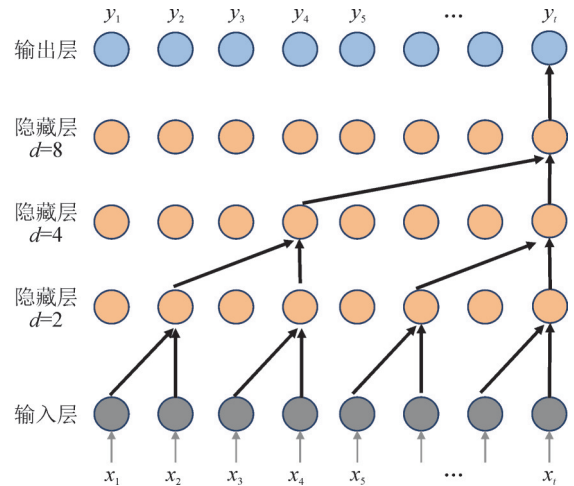


图3 因果扩张卷积结构图

Fig. 3 Causal expansion convolution structure of TCN

图 3 中: x_t 分别为 0 到 t 时刻输入的特征; d 为扩张系数, 根据式(7)得到 t 时刻的输出特征 y_t 为

$$y_t = \sum_{j=0}^m l_j (x_0^{s-d} - x_t^{s-d}) \quad (7)$$

式(7)中: m 为卷积核数; l_j 为第 j 个卷积核的滤波器大小; s 为步长; x_0^{s-d} 为 0 时刻感受野边界的重构特征; x_t^{s-d} 为 t 时刻感受野边界的重构特征; $s-d$ 是过去的方向。

TCN 模型结构图见图 4。图 4 展示了 TCN 残差块用跨层连接的方式结合因果扩张卷积及残差块,

这方便模型处理具有长期历史依赖的时序数据,有助于解决梯度消失^[25-26]或梯度爆炸^[27]问题。

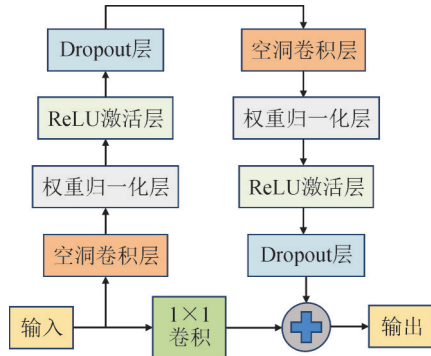


图4 TCN模型结构图

Fig. 4 TCN structure diagram

2 模型构建

2.1 整体模型

Attention-GRU-TCN的模型结构见图5。该结构由输入层、TCN层、Attention层、GRU层、输出层组成,采用“TCN-Attention-GRU”的串行级联结构,网络结构具体工作流程如下:

1)输入层接收长度为 i 的时间窗口样本,每个时间步包含6维功率特征,记为 $X=\{x_1, x_2, \dots, x_i\}$ 。

2)TCN层对输入序列 x 进行时序卷积操作。TCN的因果卷积确保 t 时刻的输出仅依赖于 t 及之前时刻的输入;扩张卷积在不增加参数量的前提下扩大感受野。TCN输出特征序列 $H^{\text{TCN}}=\{h_1^{\text{TCN}}, h_2^{\text{TCN}}, \dots, h_i^{\text{TCN}}\}$ 。

3)Attention层接收 H^{TCN} 通过自注意力机制计算每个时间步的注意力权重 α_i ,对 H^{TCN} 进行加权求和,输出加权后的特征序列 $H^{\text{Att}}=\{\alpha_1 h_1^{\text{TCN}}, \alpha_2 h_2^{\text{TCN}}, \dots, \alpha_i h_i^{\text{TCN}}\}$ 。

4)GRU层将 H^{Att} 作为输入,按时间步递推更新隐藏状态 $h_t^{\text{GRU}}=\text{GRU}(h_{t-1}^{\text{GRU}}, \alpha h_t^{\text{TCN}})$,最终输出最后一个时间步的隐藏状态 h_T^{GRU} 。

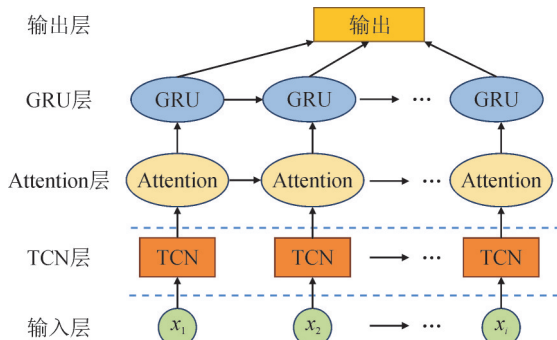


图5 Attention-GRU-TCN模型结构图

Fig. 5 Attention-GRU-TCN model structure diagram

变压器功率数据中的短期波动会引发顶层油温的短期动态响应。然而,油温作为整体热状态量,具有显著的热惯性,不会发生局部突变。TCN通过扩张因果卷积,在不引入未来信息的前提下,高效提取功率序列中不同时间尺度的波动特征,精准捕捉由功率突变引发的顶层油温短期变化趋势,从而适配顶层油温“长期热惯性主导、短期功率扰动调制”的物理特性。GRU通过更新门与重置门有效缓解了梯度消失问题,能够稳定建模顶层油温的长期时序依赖与热惯性,精准捕捉功率与顶层油温之间的滞后响应关系。注意力机制通过自适应权重分配,自动为对顶层油温贡献更大的时间步与功率维度赋予更高权重,抑制无关或冗余信息的干扰,实现关键特征的聚焦。

三者形成了“局部特征提取到长期时序建模再到关键信息聚焦”的完整链路,与变压器顶层油温数据特性高度匹配。这种设计可以发挥各模块的优势,最终实现更精准、更鲁棒的顶层油温预测。此外,变压器顶层油温受电流与环境温度共同影响。文中所使用的数据集未包含环境温度记录,因此模型仅以功率数据作为输入。这一局限性意味着模型无法捕捉环境温度突变对顶层油温的影响,预测精度在环境温度剧烈波动时可能下降。然而,在功率主导的工况下(如大负荷波动频繁的工业用户变压器),电流仍是顶层油温变化的主要驱动因素;同时,文中提出的多模块耦合结构具备较强的特征提取能力,可在一定程度上从功率序列的历史变化中隐式学习环境温度的长期趋势(如季节性规律)。此外,文中的主要贡献在于验证Attention-GRU-TCN结构在顶层油温预测中的有效性,环境温度的缺失不影响各模型之间的相对性能比较。

2.2 预测流程

基于Attention-GRU-TCN模型的预测流程框图见图6。具体实现步骤如下。

步骤1),数据处理。模型获取变压器的特性数据,搜索缺失值,通过线性插值法对数据进行缺失值填补。为提升模型训练的收敛速度与稳定性,笔者采用Min-Max归一化方法对样本数据进行预处理,保证网络的输入样本中没有特殊样本,并确保特征值的范围在0到1之间,与GRU门控单元中sigmoid函数的敏感工作区间匹配,有助于缓解梯度饱和问题,计算公式为

$$X^* = \frac{X - X_{\min}}{X_{\max} - X_{\min}} \quad (8)$$

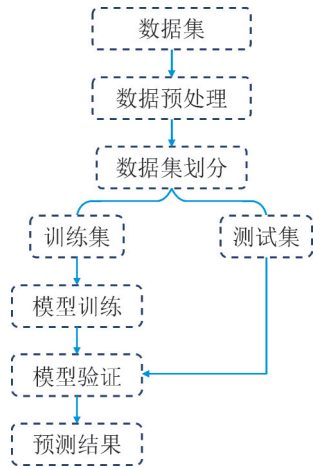


图6 预测流程框图

Fig. 6 Prediction flow chart

式(8)中: X^* 为归一化样本数据; X 为原始样本数据; $X_{\min}$ 为样本的最小数据; $X_{\max}$ 为样本的最大数据。

步骤2), 构建 Attention-GRU-TCN 模型及其超参数。将数据集的 80% 作为训练集, 剩余 20% 作为验证集。使用训练集对整个模型进行训练, 计算实际结果与预测结果之间的评价指标。通过多次实验确定最佳模型参数。

步骤3), 确定理想模型后, 使用预试样本进行测试。模型得到相应时刻具有物理意义的顶层油温值。模型将其与真实值进行比较, 并计算误差评估指标, 以验证模型的拟合准确性和泛化能力。

3 实验结果与分析

3.1 数据集

变压器的某些运行参数与顶层油温具有较强的相关性, 可以准确反映变压器顶层油温的变化。某公司每小时从两个变电站的两台变压器收集数据, 共收集了一年的真实数据, 建立了变压器数据集, 用于研究和预测变压器的顶层油温, 该变压器基本参数见表1。数据集中每个数据包含 17 421 条数据, 每个数据点由 8 个特征组成, 包括收集日期、顶层油温, 以及以下 6 种功率数据: 高压侧有效功率、高压侧无效功率、中压侧有效功率、中压侧无效功率、低压侧有效功率、低压侧无效功率。实验以这 6 维功率数据作为输入特征, 以顶层油温作为预测目标。实验以高、中、低的有效功率数据和无效功率数据共 6 种功率数据为输入, 顶层油温为输出。笔者在该公司进行实测以验证方法有效性。

表1 变压器基本参数

Table 1 Basic parameters of transformer

项目	特征参数
产品型号	SFSZ11-10000/35
额定容量/kVA	10 000
额定频率/Hz	50
额定电压/kV	35±3×2.5%/10.5/6.3
额定电流/A	165/550/916
铁心形式	三相三柱
冷却方式	油浸自冷

3.2 实验环境及参数

本实验平台硬件配置包括 AMD Ryzen 75800H 中央处理器和 NVIDIA GTX 3060 显卡, 操作系统为 Windows 11 64 位专业版。模型编程语言选择 Python 3.9, 同时采用 PyCharm 编写运行代码, 采用 Pytorch 深度学习框架进行神经网络训练和验证, 经过实验对比和分析, 确定了模型的具体参数配置, 相关参数见表2。

表2 参数设定

Table 2 Hyperparameter settings

参数	数值
学习率	0.001
优化器	Adam
训练轮数	200
卷积核大小	3
TCN 扩张系数	4
GRU 隐藏层数量	2
GRU 隐藏层神经元	128
批次大小	16

GRU 模块使用 Xavier 均匀初始化对权重矩阵进行初始化, 偏置项初始化为零。该初始化方式能够保持前向与反向传播过程中的方差一致性, 有效

缓解梯度消失问题, 适用于顶层油温预测中需捕捉长时热惯性的任务场景。在模型预实验中发现, 若将更新门的偏置初始化为正值, 可引导模型初期倾向于保留历史状态, 与顶层油温热惯性物理先验相符, 能够将收敛 epoch 数减少约 40%, 但对最终预测精度的提升有限。因此, 为保持方法的一般性, 文中仍采用标准 Xavier 初始化。

3.3 评价指标

使用均方误差 e_{MSE} (mean square error)、平均绝对误差 e_{MAE} (mean absolute error)、决定系数 R^2 (R-Square) 来评估各种方法之间的预测结果。

1) 均方误差计算公式为

$$e_{\text{MSE}} = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (9)$$

式(9)中: y_i 为真实值; $\hat{y}_i$ 为预测值; n 为样本数量。 e_{MSE} 的值范围是 $[0, +\infty)$ 。该值预测数据和原始数据之间相应点的误差平方和的平均值, 越小说明预测值越接近实际值。顶层油温数据中存在非高斯噪声, 且高温区域对变压器绝缘寿命的危害呈指数级增长, 预测误差在高温时应被赋予更高代价。当预测值远离真实值时, e_{MSE} 梯度线性增大。在变压器顶层油温安全阈值附近, 微小误差可能意味着“正常”与“报警”的误判。 e_{MSE} 的二次惩罚会在阈值附近产生强梯度, 迫使模型在关键区域精确拟合。

2) 平均绝对误差计算公式

$$e_{\text{MAE}} = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (10)$$

式(10)中: y_i 为真实值; $\hat{y}_i$ 为预测值; n 为样本数量。 e_{MAE} 的值范围是 $[0, +\infty)$ 。它是绝对误差的平均值, 反映了预测值与真实值误差的实际情况。该值越小说明预测值越接近实际值。

3) 决定系数计算公式

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \bar{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (11)$$

式(11)中: y_i 为真实值; $\hat{y}_i$ 为预测值; n 为样本数量; $\bar{y}_i$ 为真实值的均值。 R^2 的值范围是 $(-\infty, 1)$ 。它反映了模型拟合效果的质量。越接近 1 表示模型拟合更完美, 预测效果更好。

3.4 消融实验

为了验证 Attention-GRU-TCN 模型的有效性, 进行消融实验, 实验结果见表 3。

表 3 消融实验结果

Table 3 Ablation experimental results

模型	e_{MAE}	e_{MSE}	R^2
GRU	1.55	5.02	0.92
Attention-GRU	1.45	4.51	0.93
TCN-GRU	1.46	4.75	0.92
Attention-GRU-TCN	1.43	4.41	0.93

相比于 GRU 模型, Attention-GRU 模型、TCN-GRU 模型、Attention-GRU-TCN 模型的 e_{MAE} 值分别减少了 6.4%、5.81%、7.74%; e_{MSE} 值分别减少了 10.16%、5.38%、12.15%, 其中 Attention-GRU-TCN 模型性能最优。消融实验结果表明, 加入 TCN 卷积和注意力机制使得 GRU 模型的各项指标有所改善。TCN 通过膨胀卷积扩大感受野, 能够同时捕捉功率数据中的短期波动与长期趋势。将原始功率序列先经 TCN 提取特征, 再输入 GRU, 相当于为 GRU 提供了更丰

富的热动态先验信息, 减轻了 GRU 仅依靠原始序列学习长程依赖的负担。注意力模块通过学习每个时间步的权重, 使模型能够重点关注对顶层油温变化起关键作用的时刻, 例如功率突增、长时间高负荷运行等场景。当注意力机制与 GRU 结合时, 它让 GRU 在更新隐藏状态时更聚焦于相关历史热状态, 抑制无关噪声的干扰。从消融结果来看, 单独移除 TCN 或注意力机制均导致性能下降, 验证了两者与 GRU 之间具有互补的协同作用。

3.5 对比实验

为了进一步验证提出的模型在变压器顶层油温预测的准确性, 其他预测算法先后进行了两次对比实测, 得到不同算法的结果, 见表 4、5。

表 4 第一次对比实验结果

Table 4 Results of the first comparative experiment

模型	e_{MAE}	e_{MSE}	R^2	预测耗时/ms
RNN	1.88	8.12	0.87	3.12
LSTM	1.71	6.41	0.90	6.54
GRU	1.55	5.02	0.92	5.03
BiLSTM	2.03	7.76	0.83	8.87
Transformer	1.68	7.19	0.89	16.34
Attention-GRU-TCN	1.43	4.40	0.93	9.67

表 5 第二次对比实验结果

Table 5 Results of the second comparative experiment

模型	e_{MAE}	e_{MSE}	R^2	预测耗时/ms
RNN	3.21	18.24	0.86	4.75
LSTM	2.52	12.96	0.91	7.32
GRU	2.16	9.32	0.93	6.06
BiLSTM	2.76	15.77	0.88	10.44
Transformer	2.51	12.15	0.91	15.91
Attention-GRU-TCN	2.01	8.78	0.93	10.28

选取了测试集数据作为展示样本, 绘制了该算法和其他算法的预测曲线图, 见图 7。在第一次对比实验中, 各模型性能差异显著: RNN 因梯度消失问题难以捕捉顶层油温长期热惯性, e_{MAE} 为 1.88、 e_{MSE} 为 8.12、 R^2 为 0.87, 表现靠后; BiLSTM 引入的双向时序信息对单向热累积的顶层油温无物理意义, 反而引入噪声干扰, e_{MAE} 为 2.03、 e_{MSE} 为 7.76、 R^2 为 0.83, 性能最差; LSTM 与 Transformer 表现中等, LSTM 通过门控结构缓解梯度消失, 可部分建模时序依赖, e_{MAE} 为 1.71、 e_{MSE} 为 6.41、 R^2 为 0.90, Transformer 依靠自注意力机制关注全局特征, 但易过度关注无关时间步, e_{MAE} 为 1.68、 e_{MSE} 为 7.19、 R^2 为 0.89; GRU 简化门控结构, 高效适配顶层油温慢变平稳特性, e_{MAE} 为 1.55、 e_{MSE} 为 5.02、 R^2 为 0.92, 性能较优; 而 Attention-GRU-TCN 模型通过 TCN 提取局部

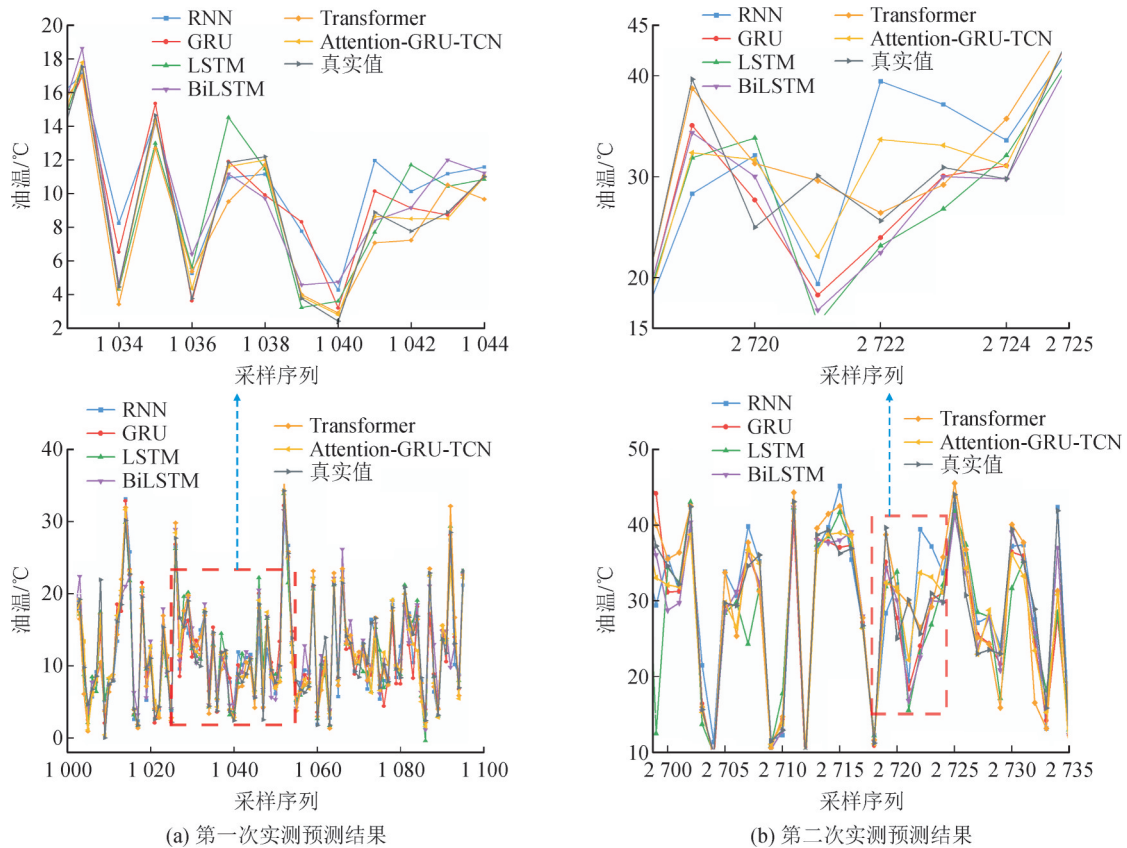


图7 预测曲线图

Fig. 7 Predictive curve

热累积特征、GRU 建模长期热惯性、注意力机制聚焦关键功率信息,形成贴合顶层油温数据特性的完整预测链路,最终 e_{MAE} 降至 1.43、 e_{MSE} 降至 4.40、 R^2 提升至 0.93,实现最优性能。

在第二次对比实验中,各模型性能均有所下降,但相对趋势与第一次实验完全一致:RNN、BiLSTM 性能降幅最大,进一步验证其对顶层油温热惯性与局部波动的建模能力不足;LSTM、Transformer 虽有一定鲁棒性, e_{MAE} 分别升至 2.52、2.51,但仍难以应对复杂耦合场景;GRU 降幅相对较小, e_{MAE} 升至 2.16;而 Attention-GRU-TCN 模型凭借层级化耦合机制,仍保持最优性能, e_{MAE} 为 2.01、 e_{MSE} 为 8.78、 R^2 为 0.93,证明其对顶层油温数据特性的适配性更强,鲁棒性更优。

在两次实验中,使用预测方法耗时体现模型复杂程度。模型推理速度的排序完全相同:RNN 最快,其次为 GRU、LSTM、BiLSTM、Attention-GRU-TCN,Transformer 最慢。这符合各模型的理论计算复杂度:RNN 结构最简单,GRU 比 LSTM 参数量更少且门控更精简,BiLSTM 需双向计算导致延迟更高,Transformer 的自注意力和大量全连接层使其成为最重模块,而 Attention-GRU-TCN 由于多模块串联,延迟介于 BiLSTM 和 Transformer 之间。文中提出的 Attention-GRU-TCN 模型在两次实验中的平均推理

耗时分别为 9.67、10.28 ms,远低于 1 s 的工程上限。即使考虑到数据预处理、通信延迟等额外开销,总响应时间仍在数十 ms 量级,完全满足在线实时预测的要求。相比之下,Transformer 模型耗时最高,但仍属毫秒级,所有模型均符合实时性约束。因此,在满足实时性的前提下,应优先选择预测精度最高的模型,即文中所提的 Attention-GRU-TCN。

从图 7 可以看出,各个模型整体上展现出了不错的拟合效果,主要差别在波峰与波谷,这一点从局部放大图中可以更为清晰的看到。

RNN、LSTM 和 Transformer 模型稍显不足,预测结果相比 Attention-GRU-TCN 不太准确。波峰波谷预测差异定量分析见表 6。

表6 波峰波谷预测差异定量分析表

Table 6 Quantitative analysis of prediction errors at peaks and troughs

模型名称	波峰		波谷	
	e_{MAE}	e_{MSE}	e_{MAE}	e_{MSE}
RNN	1.89	4.50	1.76	4.04
LSTM	1.62	3.42	1.53	3.03
GRU	1.47	2.82	1.39	2.46
BiLSTM	1.78	4.12	1.69	3.69
Transformer	1.59	3.28	1.51	2.96
Attention-GRU-TCN	1.21	1.74	1.15	1.32

从误差统计结果可见,RNN与BiLSTM因无法有效捕捉短期热冲击或存在冗余时序信息,在两类工况下误差最大;LSTM与Transformer虽能部分建模时序依赖,但对局部波动细节与长期热惯性的刻画仍有不足,表现中等;GRU凭借高效门控结构适配顶层油温慢变特性,误差进一步降低。文中所提Attention-GRU-TCN模型在波峰与波谷工况下的指标均为所有模型中最低,既通过TCN精准捕捉了功率突变引发的短期热累积特征,又通过GRU稳定建模了顶层油温冷却的长期热惯性,同时借助注意力机制聚焦关键功率信息,充分验证了其在变压器顶层油温极值点预测中的优越性与工程实用性。

4 结论

为了进一步提升变压器顶层油温预测的准确度,笔者提出了一种对顶层油温时序数据趋势进行预测建模的神经网络,采用了 e_{MSE} 、 e_{MAE} 、 R^2 等评价指标对方法的有效性进行了验证。研究结论如下:

1)针对变压器顶层油温预测精度不高的问题,笔者提出Attention-GRU-TCN模型。通过实测得出,Attention-GRU-TCN模型提高了变压器顶层油温温度的预测效率和精度。

2)通过实测得出,TCN卷积在扩大感受野的同时保证了预测任务的因果性。注意力机制可以增强模型对重要特征的捕捉能力。

3)Attention-GRU-TCN模型通过优化模型处理信息的能力,实现了较高的预测精度和拟合能力,可为变压器超前预警、安全功率控制与全生命周期管理提供技术支撑,对于研究控制变压器顶层油温温度具有一定的意义。

4)文中方法目前仅以功率数据作为输入,未考虑环境温度等外部因素。后续研究将纳入环境温度、风速、日照冷却器状态等多源数据,构建更全面的顶层油温预测模型,并探索基于注意力机制的多源时序融合模型,以进一步提升模型的泛化能力。

5)理论上,对每个模型进行更细致的超参数优化可能进一步提升其预测精度,甚至缩小与Attention-GRU-TCN模型的差距。然而,受限于计算资源和时间成本,文中未对所有对比模型进行极致调优。即便如此,文中的实验设置已保证了模型比较的相对公平性,且Attention-GRU-TCN在所有对比实验中均一致地取得最优结果,验证了其结构在变压器油温预测任务上的优势文中的对比结果基于有限超参数调优,未来可探索更自动化的调优方法

以评估各模型的极限性能。

参考文献:

- [1] 霍伟锋,李应光,陈斌,等.基于油浸式电力变压器顶层油温预测模型的多参数灵敏度分析[J].变压器,2021,58(11):16-20.
Huo Weifeng, Li Yingguang, Chen Bin, et al. Multi-parameter sensitivity analysis based on prediction model of top-oil temperature of oil-immersed power transformer[J]. Transformer, 2021, 58(11): 16-20.
- [2] 张若愚,石峥,郭建,等.基于数据质量提升的变压器顶层油温预测模型[J].山东电力技术,2025,52(12):113-122.
Zhang Ruoyu, Shi Zheng, Guo Jian, et al. Transformer top oil temperature prediction model based on data quality improvements[J]. Shandong Electric Power, 2025, 52(12): 113-122.
- [3] 王飞,蔡庆娜.基于AIoT云平台与CNN-LSTM的变压器油温预测研究[J].变压器,2025,62(2):21-26.
Wang Fei, Cai Qingna. Research on transformer oil temperature prediction based on AIoT cloud platform and CNN-LSTM[J]. Transformer, 2025, 62(2): 21-26.
- [4] 郭鹏鸿,唐治平,张光昊,等.基于双重注意力机制WOA-CNN-BiLSTM模型的变压器油温预测方法[J].变压器,2025,62(1):41-50.
Guo Penghong, Tang Zhiping, Zhang Guanghao, et al. Transformer oil temperature prediction based on WOA-CNN-BiLSTM model with dual attention mechanism[J]. Transformer, 2025, 62(1): 41-50.
- [5] Liu Jingrui, Hou Zhiwen, Liu Bowei, et al. Mathematical and machine learning innovations for power systems: predicting transformer oil temperature with beluga whale optimization-based hybrid neural networks[J]. Mathematics, 2025, 13(11): 1785-1819.
- [6] 张帆,俞永军,蔡重凯,等.基于多时间尺度的干式空心电抗器匝间短路故障演化与寿命预测模型[J].高压电器,2025,61(8):99-107.
Zhang Fan, Yu Yongjun, Cai Chongkai, et al. Fault evolution and life prediction model for inter-turn short circuit of dry air core reactor based on multiple time scales[J]. High Voltage Apparatus, 2025, 61(8): 99-107.
- [7] 朱夏,陈颂,袁明瀚,等.基于CNN-LSTM风光荷预测的主动配电网双层扩展规划方法[J].高压电器,2025,61(5):218-227.
Zhu Xia, Chen Song, Yuan Minghan, et al. Double layer expansion planning method for active distribution network based on CNN-LSTM wind, solar and load prediction[J]. High Voltage Apparatus, 2025, 61(5): 218-227.
- [8] 汤健,侯慧娟,王劭菁,等.基于策略梯度优化的变压器油中溶解气体预测模型[J].高压电器,2025,61(7):91-100.
Tang Jian, Hou Huijuan, Wang Shaojing, et al. Prediction model for dissolved gas in transformer oil based on policy gradient optimization[J]. High Voltage Apparatus, 2025, 61(7): 91-100.
- [9] 胡雪莹,李琳.基于全寿命周期成本和综合故障率的油浸式变压器经济寿命评估[J].智慧电力,2024,52(1):47-54.
Hu Xueying, Li Lin. Economic life evaluation of oil-immersed transformer based on life cycle cost and comprehensive failure rate[J]. Smart Power, 2024, 52(1): 47-54.

- [10] 林苑,赵晋斌,孙明琦,等. 新型电力系统下基于物理信息LSTM网络的电力变压器状态评估方法[J]. 电力系统保护与控制, 2025, 53(14): 133-141.
Lin Yuan, Zhao Jinbin, Sun Mingqi, et al. Power transformer condition assessment method based on physics-informed LSTM network in the context of new power systems[J]. Power System Protection and Control, 2025, 53(14): 133-141.
- [11] 张晓华,吕志瑞,孙云生,等. 基于修正热路模型的油浸式变压器绕组热点温度计算研究[J]. 电测与仪表, 2025, 62(3): 78-84.
Zhang Xiaohua, Lyu Zhirui, Sun Yunsheng, et al. Research of hot spot temperature calculation in oil-immersed transformer based on modified thermal model[J]. Electrical Measurement & Instrumentation, 2025, 62(3): 78-84.
- [12] 潘文霞,陈星池,赵坤,等. 基于温度场计算的油浸式变压器热点温度仿真分析方法[J]. 电测与仪表, 2025, 62(3): 71-77.
Pan Wenxia, Chen Xingchi, Zhao Kun, et al. Simulation analysis method of hot spot temperature of oil-immersed transformer based on temperature field calculation[J]. Electrical Measurement & Instrumentation, 2025, 62(3): 71-77.
- [13] 董雪倩,荆澜涛,田瑞,等. 基于LSTM模型的变压器顶层油温预测方法[J]. 电力学报, 2023, 38(1): 38-45.
Dong Xueqing, Jing Lantao, Tian Rui, et al. Prediction method of transformer top oil temperature based on LSTM model[J]. Journal of Electric Power, 2023, 38(1): 38-45.
- [14] 赵军,高树国,何瑞东,等. 变压器油中乙炔门控循环单元网络多步预测超参数优化方法[J]. 高压电器, 2024, 60(7): 163-172.
Zhao Jun, Gao Shuguo, He Ruidong, et al. Hyper-parameters optimization method for multi-step prediction of acetylene in power transformer oil by gated cyclic unit network[J]. High Voltage Apparatus, 2024, 60(7): 163-172.
- [15] 谭志超,范竞敏,冯陆滔,等. 基于模态分解和混合式CNN-GRUT的变压器油中溶解气体预测方法[J]. 电工电能新技术, 2024, 43(7): 80-90.
Tan Zhichao, Fan Jingmin, Feng Lutao, et al. Prediction method of dissolved gas in transformer oil based on modal decomposition and hybrid CNN-GRUT[J]. Advanced Technology of Electrical Engineering and Energy, 2024, 43(7): 80-90.
- [16] 卫永鹏,苏益辉,王胜利,等. 基于改进门控循环单元的变压器油中气体浓度预测[J]. 电气技术, 2022, 23(2): 55-60.
Wei Yongpeng, Su Yihui, Wang Shengli, et al. Prediction of gas concentration in transformer oil based on improved gated recurrent unit[J]. Electrical Engineering, 2022, 23(2): 55-60.
- [17] 周春阳,李亚锦,邓光武,等. 基于多维度分析的换流变预警技术研究与应用[J]. 电气工程学报, 2021, 16(4): 151-158.
Zhou Chunyang, Li Yajin, Deng Guangwu, et al. Multi-dimensional analysis and early warning model of converter transformer[J]. Journal of Electrical Engineering, 2021, 16(4): 151-158.
- [18] 孟衡,张涛,王金,等. 基于多尺度时空图卷积网络与Transformer融合的多节点短期电力负荷预测方法[J]. 电网技术, 2024, 48(10): 4297-4305.
Meng Heng, Zhang Tao, Wang Jin, et al. Multi-node short-term power load forecasting method based on multi-scale spatiotemporal graph convolution network and transformer[J]. Power System Technology, 2024, 48(10): 4297-4305.
- [19] 樊清川,于飞,宣敏. 基于优化Bi-LSTM模型的电力变压器故障诊断[J]. 计算机仿真, 2022, 39(11): 136-140.
Fan Qingchuan, Yu Fei, Xuan Min. Power transformer fault diagnosis based on optimized Bi-LSTM model[J]. Computer Simulation, 2022, 39(11): 136-140.
- [20] 李腾,廖军,樊培培,等. 基于自注意力机制与改进循环神经网络混合模型的特高压换流变压器顶层油温预测方法[J]. 现代电力, 2024, 41(6): 1167-1175.
Li Teng, Liao Jun, Fan Peipei, et al. Forecasting method for ultra-high voltage converter transformer top-oil temperature based on self-attention mechanism and improved recurrent neural network hybrid model[J]. Modern Electric Power, 2024, 41(6): 1167-1175.
- [21] 高芬,苏依拉,牛向华,等. 基于Transformer的蒙汉神经机器翻译研究[J]. 计算机应用与软件, 2020, 37(2): 141-146.
Gao Fen, Su Yila, Niu Xianghua, et al. Mongolian-Chinese neural machine translation based on transformer[J]. Computer Applications and Software, 2020, 37(2): 141-146.
- [22] 周磊,竺筱晶. 基于MA-CNN-LSTM和自注意力机制的单变量短期电力负荷预测[J]. 科学技术与工程, 2024, 24(22): 9408-9416.
Zhou Lei, Zhu Xiaojing. Univariate short-term electrical load based on MA-CNN-LSTM-self attention[J]. Science Technology and Engineering, 2024, 24(22): 9408-9416.
- [23] Hao Siyu, Lee D H, Zhao De. Sequence to sequence learning with attention mechanism for short-term passenger flow prediction in large-scale metro system[J]. Transportation Research Part C - Emerging Technologies, 2019(107): 287-300.
- [24] 李梦阳,陈柳,史蒙,等. 基于聚合模态分解和TCN-BiGRU的光伏功率预测模型[J]. 太阳能学报, 2026, 47(2): 91-99.
Li Mengyang, Chen Liu, Shi Meng, et al. Photovoltaic power forecasting model based on aggregate mode decomposition and TCN-BiGRU[J]. Acta Energetica Solaris Sinica, 2026, 47(2): 91-99.
- [25] 李泓钢,缪旻. 基于ATT-LSTM网络的多核系统封装温度预测[J]. 电子器件, 2025, 48(5): 1074-1081.
Li Honggang, Miao Min. Temperature prediction of multi-core system-in-package based on ATT-LSTM network[J]. Chinese Journal of Electron Devices, 2025, 48(5): 1074-1081.
- [26] Baron N. Vanishing gradients in an artificial deep neural network: the impact of activation functions and optimization techniques[J]. Journal of Computer and Communications, 2025, 13(11): 128-162.
- [27] 陈聪,候磊,李乐乐,等. 基于GRU改进RNN神经网络的飞机燃油流量预测[J]. 科学技术与工程, 2021, 21(27): 11663-11673.
Chen Cong, Hou Lei, Li Lele, et al. Prediction of aircraft fuel flow based on recurrent neural network[J]. Science Technology and Engineering, 2021, 21(27): 11663-11673.

张京城(1981—),女,硕士,副教授,研究方向为电气自动化(E-mail: 13830791196@163.com)。

刘宝(1983—),男,博士,副教授,研究方向为目标跟踪、信息融合、图像处理(通信作者)(E-mail: xiaobei0077@163.com)。