

# 基于 Transformer 语义迁移的定子热故障诊断

姚驰宇<sup>1</sup> 桂晶<sup>1</sup> 李泼<sup>1</sup> 王伟<sup>1</sup> 陈聪<sup>2,3</sup>

(1. 淮南平圩发电有限责任公司, 安徽 淮南 232089;

2. 中电华创电力技术研究有限公司, 上海 200086;

3. 中电华创(苏州)电力技术研究有限公司, 江苏 苏州 215123)

**摘要** 汽轮发电机定子冷水系统须保持良好运行状态, 以保障发电机的可靠性和安全性。通常采用停机检修或温差阈值的方法进行热故障检测, 但无法在发电机运行状态下及时有效地检出故障。为了更精确地发现定子热故障, 本文提出基于 Transformer 架构的温度预测算法, 并基于多测点的温度预测结果对未来温差进行估计, 建立定子热故障诊断模型。为了缓解故障运行数据较少的问题, 本文使用不同核函数的高斯过程生成多种模式的时间序列, 并与原数据进行组合, 极大地扩充了训练样本空间。最后, 利用已有的测试数据开展实验。结果表明, 本文所提预测算法的预测效果优于传统自回归积分滑动平均 (ARIMA) 模型和长短期记忆 (LSTM) 算法, 基于本文所提预测算法建立的故障诊断模型, 对运行状态的识别准确率达到 91.9%, 且具有较高的精度和召回率, 从而确保了较低的误报警率和漏报警率。

**关键词:** 定子热故障; 出水温度; Transformer; 高斯过程

## Transformer-based semantic transfer for stator thermal fault diagnosis

YAO Chiyu<sup>1</sup> GUI Jing<sup>1</sup> LI Po<sup>1</sup> WANG Wei<sup>1</sup> CHEN Cong<sup>2,3</sup>

(1. Anhui Huainan Pingwei Power Generation Co., Ltd, Huainan, Anhui 232089;

2. China Power Hua Chuang Electricity Technology Research Co., Ltd, Shanghai 200086;

3. China Power Hua Chuang (Suzhou) Electricity Technology Research Co., Ltd, Suzhou, Jiangsu 215123)

**Abstract** The stator cooling water system of a turbine generator must maintain optimal operating conditions to ensure the reliability and safety of the generator. Typically, thermal faults are detected using methods such as shutdown maintenance or temperature difference thresholds, but these methods cannot effectively detect faults in real time while the generator is in operation. To more accurately identify stator thermal faults, this paper proposes a temperature prediction algorithm based on the Transformer architecture. Using the predicted temperatures from multiple measurement points, the future temperature difference is estimated, and a diagnosis model for stator thermal faults is established. To address the issue of limited fault operation data samples, this paper utilizes Gaussian processes with different kernel functions to generate various types of time series, which are then combined with the original data, significantly expanding the training sample space. Finally, experiments are conducted using existing test data. The results indicate that the predictive algorithm proposed in this paper outperforms traditional autoregressive integrated moving average (ARIMA) and long short term memory (LSTM) algorithms. Moreover, the diagnostic model based on this predictive algorithm achieves an accuracy rate of 91.9% in identifying operational states, while also maintaining high precision and recall rates, ensuring low false alarm and missed alarm rates.

**Keywords:** stator thermal faults; outlet-water temperature; Transformer; Gaussian process

## 0 引言

水冷汽轮发电机定子绕组及其内冷水系统在电力生产中起着至关重要的作用<sup>[1]</sup>。定子绕组内的冷却系统需要保持良好的运行状态,以确保发电机在高负荷条件下的可靠性和安全性。若冷却系统出现故障,可能导致定子绕组过热,进而引发设备损坏或停机事故,严重影响电力供应的稳定性<sup>[2]</sup>。定子绕组冷却系统的传统检测手段主要依赖停机检修时的超声波法和热水流法<sup>[3-4]</sup>。这些方法虽然能在一定程度上评估冷却系统的性能,但是存在显著的局限性,如:无法在发电机运行状态下实时监测冷却水的流动情况,限制了对动态热故障的诊断能力;测量精度易受现场条件(绝缘引水管的长度等)制约,导致检测结果的准确性不足;难以有效识别轻微堵塞或传感器位置略有偏差的情况。

为了提高故障预警的准确性和及时性,在较早阶段发现定子绕组冷水系统的潜在问题,防止严重事故的发生,应引入先进的传感技术和数据分析方法。部分研究人员尝试通过研究温度分布规律<sup>[5-8]</sup>,建立基于机理的故障诊断模型,但由于影响因素较多,在复杂多变的工况下模型容易失效。近两年,将神经网络技术应用于电力设备参数预测和故障诊断领域的研究越来越受到重视<sup>[9-12]</sup>。其中,梁郑秋等提出基于卷积神经网络的核电多相无刷励磁系统旋转整流器故障诊断方法,展示了神经网络技术在复杂系统故障检测中的高准确率和应用潜力<sup>[13]</sup>。衣思彤等提出一种基于贝叶斯优化、卷积神经网络和双向长短期记忆神经网络的锂电池健康状态评估方法,通过结合多健康特征和优化算法,实现了对电池健康状态的高精度预测,为复杂系统的故障诊断提供了有效参考<sup>[14]</sup>。文献[15]提出的通过组内传感器最大相关性分组和快速训练自编码器长短期记忆(long short term memory, LSTM)模型,在空气冷却冷水机组的传感器偏差故障检测中表现优异。此外,文献[16]通过稀疏自动编码器(stacked auto-encoder, SAE)和 LSTM 网络的结合,利用滑动窗口方法检测过热故障,建立了定子绕组过热故障预警方法。以上研究表明,将数据驱动的方法用于解决此类问题具备一定的可行性。

同时,大语言模型的相关研究表明,Transformer 系列模型比 LSTM 等时间序列预测模型在序列识别和理解问题方面具备更大潜力<sup>[17-19]</sup>。尽管现阶段

Transformer 算法主要应用于文本类的相关语义问题中<sup>[20-21]</sup>,但已有研究表明,其在数值类型的时间序列预测问题中也表现较优<sup>[18, 22-23]</sup>。本文通过将连续数值类型的时间序列转化为离散类型,把 Transformer 模型应用于定子绕组冷水管出水温度的时间序列预测中,并结合高斯过程(Gaussian process, GP)实现语义迁移学习,降低对样本数量的要求,然后基于预测结果建立热故障诊断模型。将该模型与传统的时间序列模型如自回归积分滑动平均(autoregressive integrated moving average, ARIMA)模型<sup>[24]</sup>和 LSTM<sup>[25-26]</sup>模型进行对比,以证明该模型的优越性。

## 1 方法原理

### 1.1 归一化值离散化

本文基于出水温度时间序列的预测建立定子绕组冷水系统热故障诊断模型,在预测环节采用 Transformer 算法。为了采用与自然语言相似的模型训练方式,首先对数据进行归一化,然后离散化为固定数量的离散值。

对于时间序列向量  $\mathbf{X}$ ,采用如式(1)所示的归一化处理方式。

$$\mathbf{X}' = \frac{\mathbf{X} - \mathbf{X}_{\min}}{\mathbf{X}_{\max} - \mathbf{X}_{\min}} \quad (1)$$

式中: $\mathbf{X}'$ 为归一化后的时间序列; $\mathbf{X}_{\min}$ 和 $\mathbf{X}_{\max}$ 分别为 $\mathbf{X}$ 的最小值和最大值。

将归一化后的时间序列 $\mathbf{X}'$ 离散化为固定数量的离散值,如式(2)所示。

$$q(x) = \begin{cases} 1 & -\infty < x < b_1 \\ 2 & b_1 \leq x < b_2 \\ \vdots & \vdots \\ B & b_{B-1} \leq x < \infty \end{cases} \quad (2)$$

式中: $q(x)$ 为离散化后的值; $b_1, b_2, \dots, b_{B-1}$ 为离散化过程中确定的阈值点; $B$ 为离散区间的数量。

### 1.2 Transformer 算法

Transformer 算法架构如图 1 所示,采用图中的编码器-解码器结构,其中编码器的输入为输入序列,即历史温度序列,解码器的输入为偏置后的输出序列,即待预测的温度序列,输出为下一时刻的概率分布。推理过程为自回归形式,每一步计算得到的结果会被置于解码器输入环节的偏置输出序列中。

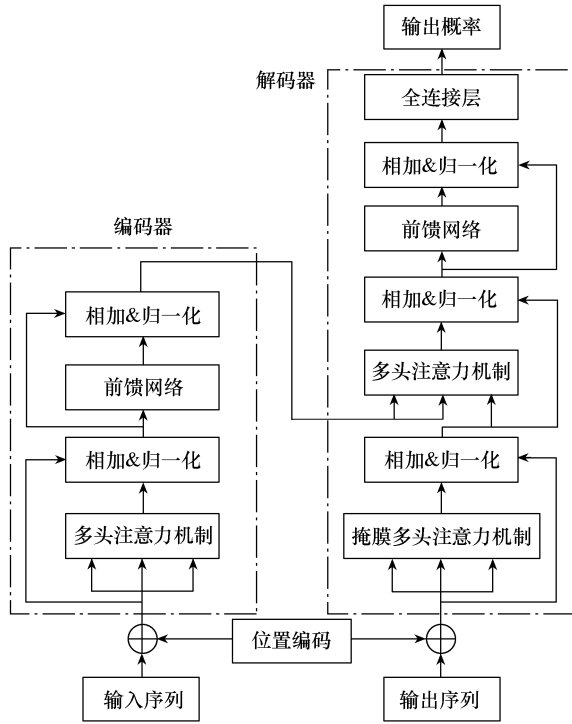


图1 Transformer 算法架构

为了让模型能够捕捉序列中元素的相对位置和顺序信息，引入位置编码 (positional encoding)。位置编码为每个输入元素提供位置信息，使模型能够感知序列中元素的顺序，从而更好地理解输入数据。Transformer 使用固定的正弦和余弦函数生成位置编码。由于输入为单测点温度序列，可以简化为

$$P_E(p_{os}) = \sin \frac{p_{os}}{10000} \quad (3)$$

式中： $P_E(p_{os})$  为位置编码函数； $p_{os}$  为输入序列中元素的位置。

编码向量被加到输入向量中，从而将位置信息嵌入序列数据中。

多头注意力机制 (multi-head attention) 是 Transformer 架构的核心组件之一，主要思想是在处理输入序列时，允许模型从多个“头 (head)” 中捕捉不同的上下文信息。这些“头”能够并行计算不同的注意力分布，帮助模型从多个角度理解输入数据的依赖关系。传统的注意力机制只能关注一个注意力分布，而多头注意力机制则通过将输入数据分别传入多个注意力头中，使模型能够学习不同的特征和模式，进而提高对序列信息的捕捉能力。每个注意力头可以看作是在不同的子空间中独立地执行注意力操作，最终这些子空间的结果被组合起来，以更好地表示输入数据。多头注意力机制如图 2 所示。

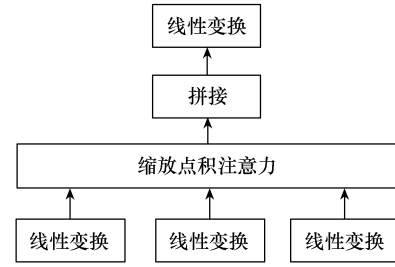


图2 多头注意力机制

输入序列为单测点温度序列，记为  $X$ ，序列长度为  $n$ 。为了计算注意力，每个注意力头都从输入中生成三个不同的向量： $Q$  (query)、 $K$  (key) 和  $V$  (value)。这些向量通过线性变换得到。

$$\begin{cases} Q = XW^Q \\ K = XW^K \\ V = XW^V \end{cases} \quad (4)$$

式中， $W^Q$ 、 $W^K$ 、 $W^V$  为参数矩阵。通过式 (5) 可计算注意力分数。

$$\text{Attention}(Q, K, V) = \text{softmax} \left( \frac{QK^T}{\sqrt{d_k}} \right) V \quad (5)$$

式中： $\text{Attention}(\cdot)$  为注意力函数； $\text{softmax}(\cdot)$  为向量归一化操作，将向量转化为概率分布向量； $d_k$  为注意力头中  $Q$ 、 $K$ 、 $V$  向量的维度，以  $1/\sqrt{d_k}$  作为缩放因子，目的是避免点积值过大导致的梯度消失问题。

### 1.3 语义迁移

在实际应用场景中，相对 Transformer 模型庞大的参数量，定子绕组冷水系统的出水温度时间序列数据样本量仍显得十分不足。尽管传统的数据增强方法，比如添加噪声、掩码等，一定程度上可以缓解过拟合问题，但增加数据集数量仍然是最佳选择。时间序列通常拆分成趋势、季节性、随机误差等成分，对于一些更加复杂的时间序列，拆分粒度更加复杂。为了生成大量多样化模式的时间序列数据，本文使用高斯过程采用不同的核函数进行采样，生成不同的时间序列模式。

高斯过程是一种用于定义函数分布的非参数贝叶斯方法<sup>[27]</sup>。高斯过程由其均值函数  $m(t)$  和协方差函数  $\kappa(t, t')$  定义，表示为

$$f(t) \sim \text{GP}(m(t), \kappa(t, t')) \quad (6)$$

式中： $m(t)$  为高斯过程在点  $t$  处的期望值； $\kappa(t, t')$  为点  $t$  和  $t'$  之间的协方差。通过选择不同的核函数，

可以捕捉不同的时间序列模式。各核函数生成的时间序列如图3所示。

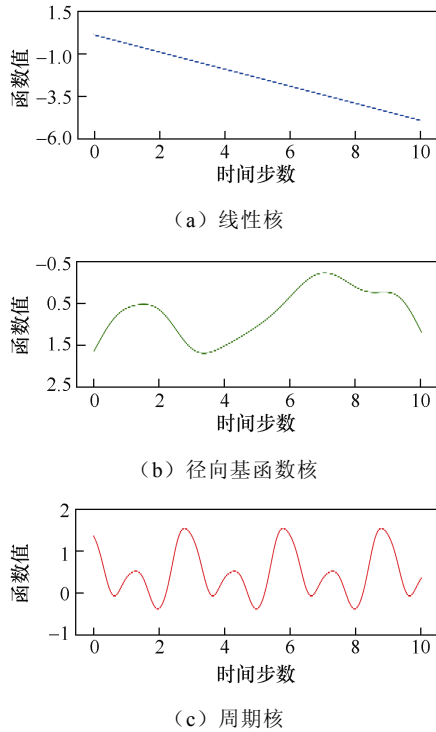


图3 各核函数生成的时间序列  
线性核定义为

$$\kappa(t, t') = \sigma^2 (t - c)(t' - c) \quad (7)$$

式中： $\sigma$  为标准差； $c$  为时间偏移量。线性核适用于捕捉线性趋势的时间序列。

径向基函数 (radial basis function, RBF) 核定义为

$$\kappa(t, t') = \sigma^2 \exp\left(-\frac{(t - t')^2}{2l^2}\right) \quad (8)$$

式中， $l$  为局部变化的尺度，影响曲线平滑度。RBF核适用于捕捉平滑的局部变化。

周期核定义为

$$\kappa(t, t') = \sigma^2 \exp\left(-\frac{2\sin^2 \frac{\pi |t - t'|}{p}}{l^2}\right) \quad (9)$$

式中， $p$  为周期长度。周期核适用于捕捉季节性或周期性模式。

这些核函数的组合可以生成具有复杂模式的合成时间序列，将这些时间序列与现有的温度时间序列进行组合，从现有和生成数据集中进行采样，并

取其加权组合作为新的序列，即

$$\tilde{\mathbf{x}}_{\text{mix}} = \sum_{i=1}^k \lambda_i \tilde{\mathbf{x}}^{(i)} \quad (10)$$

式中： $\tilde{\mathbf{x}}^{(i)}$  为第  $i$  个时间序列； $\lambda_i$  为第  $i$  个时间序列的权重； $k$  为参与加权计算的时间序列的数量。

## 2 故障诊断方法

基于 Transformer 语义迁移的定子热故障诊断流程如图4所示。

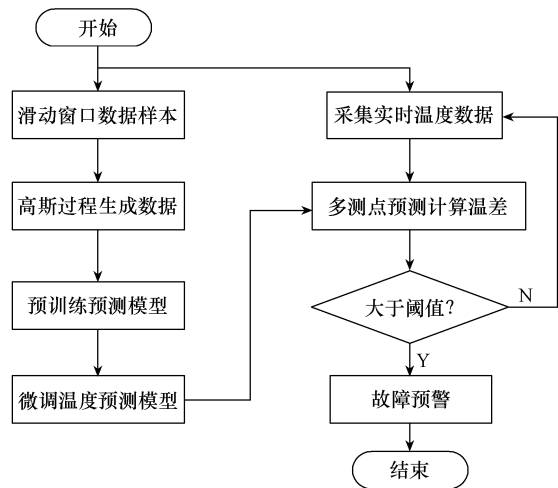


图4 基于 Transformer 语义迁移的定子热故障诊断流程

1) 从历史温度时间序列数据中提取滑动窗口样本。假设温度时间序列为  $\mathbf{X} = (x_1, x_2, \dots, x_n)$ ，滑动窗口的大小为  $w$ ，则窗口样本可以表示为  $\mathbf{X}_i = (x_i, x_{i+1}, \dots, x_{i+w-1})$ ，其中  $i$  为滑动窗口的起始索引。

2) 使用高斯过程的不同核函数生成大量的模拟数据，并添加在原始的时间序列数据中，作为新的数据样本  $\hat{\mathbf{X}}_i$ ，新的数据集样本数量扩充为原始数据样本量的 1 000 倍。

3) 使用扩充后的数据集  $\{\hat{\mathbf{X}}_i\}$ ，对预测模型进行预训练。

4) 对预训练后的模型在原始数据上进行微调。

5) 系统实时采集多个温度测点的当前温度数据，记为  $\mathbf{X}_n = (x'_1, x'_2, \dots, x'_M)$ ，其中  $M$  为测点数。

使用训练好的模型对未来  $N$  个时刻的温度进行预测，预测结果为  $\hat{\mathbf{X}} = \{\hat{\mathbf{X}}_1, \hat{\mathbf{X}}_2, \dots, \hat{\mathbf{X}}_j, \dots, \hat{\mathbf{X}}_N\}$ ，其中  $\hat{\mathbf{X}}_j = (\hat{x}_{1j}, \hat{x}_{2j}, \dots, \hat{x}_{Mj})$  为在第  $j$  个时刻各测点的预测温度。

对每个预测时刻  $j$ ，计算各测点之间的最大温差为

$$\Delta X_j = \max(\hat{x}_{i,j}) - \min(\hat{x}_{i,j}) \quad (11)$$

式中， $i \in \{1, 2, \dots, M\}$ 。从而得到未来  $N$  个时刻的温差序列  $(\Delta X_1, \Delta X_2, \dots, \Delta X_N)$ ，然后计算温差序列中的最大值为

$$\Delta X_{\max} = \max(\Delta X_1, \Delta X_2, \dots, \Delta X_N) \quad (12)$$

6) 将  $\Delta X_{\max}$  与预设的故障阈值  $\Delta X_{\text{threshold}}$  进行比较，如果  $\Delta X_{\max} > \Delta X_{\text{threshold}}$ ，则判定为可能出现热故障，系统发出故障预警，否则继续监测实时温度数据，重复以上步骤。

### 3 实验与分析

#### 3.1 预测算法实验

本文采用安徽某厂的汽轮发电机 2021—2023 年的运行数据作为训练数据，总计 42 个测点，数据采样周期为 1min。对训练数据进行滑窗分段，窗口长度为 100min，步长为 20min。每个温度测点时间序列样本长度为 100。

使用高斯过程生成相同长度的数据样本，数据量为现有数据样本数量的 1 000 倍，然后与现有数据进行加权叠加，并使用新生成的数据对 Transformer 模型进行预训练。预训练完成后，利用原训练数据样本进行模型微调。

同时，训练 ARIMA 和 LSTM 模型，并对训练结果进行对比分析。本文统计 235 组测试数据的预测值与真实值之间的方均根误差 (root mean square error, RMSE) 平均值，RMSE 越低，说明算法对数据的表达能力越强。不同预测算法的 RMSE 结果见表 1。

表 1 不同预测算法 RMSE 结果

算法	预测值与真实值 RMSE 平均值
ARIMA	0.052 3
LSTM	0.073 9
本文算法	0.048 4

#### 3.2 故障诊断分析

本文以 8K 作为阈值，即以经典阈值法的标准判据作为温度序列是否为故障序列的基准判据。如果预测的温度时间序列中，某一时刻下的最大温差超过了阈值，则认为发生了热故障。这需要预测算

法能够准确地预测温度序列的未来走向，从而更加准确地计算温差。真实与预测温差分布如图 5 所示，两者关系越接近线性且斜率越接近 1，则说明预测越精准。从图 5 可以看出，真实温差和预测温差整体呈线性相关，但当真实温差达到 8K 及以上时，预测出现相对较大的偏差。这一现象与数据特性有关，温差越大说明数据变化越剧烈，此时模型的预测能力有所下降，但总体来说，模型仍然具有较优的预测能力。

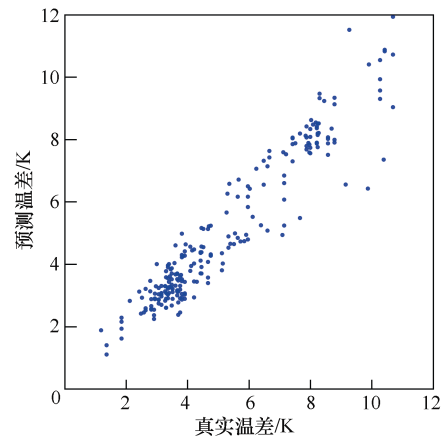


图 5 真实与预测温差分布

在已有的 235 组测试数据集上进行热故障预警测试，正常序列标签为 0，故障序列标签为 1。其中，对于故障序列，使用温差阈值达到传统报警点之前的温度序列进行模型推理，与工程实践保持一致。模型对于两种序列采用相同的预测方法，若提前预测出温差将超过 8K，则发出预警信息。之后，统计模型对两种序列的报警情况，以进一步分析模型在故障诊断中的表现。故障诊断混淆矩阵见表 2。通过混淆矩阵，计算得到故障诊断结果指标见表 3。

表 2 故障诊断混淆矩阵

类别	预测标签 0	预测标签 1
真实标签 0	186	8
真实标签 1	11	30

表 3 故障诊断结果指标

指标	数值/%
准确率	91.9
精度	79.0
召回率	73.2
$F_1$ 分数	76.0

表 3 中, 准确率综合体现对正常和故障情况的预测效果, 为 91.9%, 说明诊断方法总体效果较优。同时, 对于故障数据样本较少的情形, 通常也关注精度和召回率。精度高意味着误报警少, 召回率高意味着漏报警少。本文方法的精度和召回率均高于 70%, 表明其在故障样本数据较少的情况下的故障诊断性能优异。在上述预警数据测试评估中发现, 本文方法平均可提前 14.3min(发出预警时刻与阈值法检测到故障时刻的时间间隔)发出热故障预警, 有利于缩短故障响应时间。

#### 4 结论

本文所提基于 Transformer 语义迁移的定子热故障诊断模型, 在数据预测环节优于传统的 ARIMA 和 LSTM 模型, 同时在故障数据样本较少的情况下, 本文模型的故障诊断结果达到了 91.9%的准确率, 且兼具较高的精度和召回率, 性能表现优异, 能够比传统的阈值法更早发现热故障隐患, 可为实际定子运行过程中的热故障预警提供参考。

#### 参考文献

- [1] 吴笃超. 汽轮发电机定子线棒层间温度和出水温度异常分析与处理[J]. 通信电源技术, 2021, 38(6): 44-46.
- [2] KABOLI S, ORAEE H. Heat trap detection in the stator bar of large turbogenerators during the fault of cooling water channels blockage[C]//IECON 2016-42nd Annual Conference of the IEEE Industrial Electronics Society, Florence, Italy, 2016: 1501-1506.
- [3] 钱锋. 发电机水冷定子线圈局部温度异常的故障诊断分析[J]. 大电机技术, 2004(5): 24-27.
- [4] 吕伟, 肖爱国. 发电机定子线棒温度异常分析及诊断[J]. 河南科技, 2021, 40(2): 51-53.
- [5] RIOUAL M. A thermohydraulic modelling for the stator bars of large turbogenerators: development, validation by laboratory and on site tests[J]. IEEE Transactions on Energy Conversion, 1997, 12(1): 1-9.
- [6] 刘振兴, 尹项根, 张哲, 等. 300MW 汽轮发电机定子绕组温度在线监测与故障诊断[J]. 电力系统自动化, 2002, 26(5): 32-35.
- [7] BIAN Xu, BAO Shijie. Influence of waterway blockage fault on temperature distribution of stator transposition winding[C]//2022 IEEE 5th International Electrical and Energy Conference (CIEEC), Nanjing, China, 2022, 689-694.
- [8] 钱锋. 发电机水冷定子线圈局部温度异常的故障诊断分析[J]. 大电机技术, 2004(5): 24-27.
- [9] 严璐晗, 林培杰, 程树英, 等. 基于增量学习的 CNN-LSTM 光伏功率预测[J]. 电气技术, 2024, 25(5): 31-40.
- [10] 王步华, 朱劲璇, 熊浩清, 等. 基于长短期记忆神经网络的检修态电网暂态稳定评估方法[J]. 电气技术, 2023, 24(1): 29-35.
- [11] 李英顺, 阚宏达, 郭占男, 等. 基于数据预处理和 VMD-LSTM-GPR 的锂离子电池剩余寿命预测[J]. 电工技术学报, 2024, 39(10): 3244-3258.
- [12] 尹杰, 刘博, 孙国兵, 等. 基于迁移学习和降噪自编码器-长短时间记忆的锂离子电池剩余寿命预测[J]. 电工技术学报, 2024, 39(1): 289-302.
- [13] 梁郑秋, 郝亮亮, 周艳真, 等. 基于卷积神经网络的核电多相无刷励磁系统旋转整流器故障诊断[J]. 电工技术学报, 2023, 38(20): 5458-5472.
- [14] 衣思彤, 刘雅浓, 马耀沔, 等. 基于贝叶斯优化-卷积神经网络-双向长短期记忆神经网络的锂电池健康状态评估[J]. 电气技术, 2024, 25(5): 1-10.
- [15] GAO Long, LI Donghui, LI Ding, et al. An improved LSTM based sensor fault diagnosis strategy for the air-cooled chiller system[C]//2019 Chinese Control Conference (CCC), Guangzhou, China, 2019: 4990-4995.
- [16] YANG Yulei, ZHANG Shuming, SU Kaisen, et al. Early warning of stator winding overheating fault of water-cooled turbogenerator based on SAE-LSTM and sliding window method[J]. Energy Reports, 2023, 9: 199-207.
- [17] CHUNG H W, HOU L, LONGPRE S, et al. Scaling instruction-finetuned language models[EB/OL]. 2022: arXiv: 2210.11416. <http://arxiv.org/abs/2210.11416>.
- [18] XUE H, SALIM F D. PromptCast: a new prompt-based learning paradigm for time series forecasting[J]. IEEE Transactions on Knowledge and Data Engineering, 2024, 36(11): 6851-6864.
- [19] CHOWDHERY A, NARANG S, DEVLIN J, et al. PalM: scaling language modeling with pathways[EB/OL]. 2022: arXiv: 2204.02311. <http://arxiv.org/abs/2204.02311>.
- [20] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[EB/OL]. 2017: arXiv preprint arXiv: 1706.03762. <http://arxiv.org/abs/1706.03762v2>.