

DOI: 10.19666/j.rlfed.202503051

基于FFCM-MHDA-iTransformer的汽轮 发电机定子铁芯温度预测模型

侯一凡¹, 袁和金¹, 孙帆²

(1.华北电力大学计算机系, 河北保定 071003;

2.国网宁夏电力有限公司电力科学研究院 宁夏银川 750002)

[摘要] 发电机定子铁芯温度异常可能导致绝缘材料老化、绕组短路等严重问题, 进而影响发电机的整体性能和寿命。提出一种基于FFCM-MHDA-iTransformer的汽轮发电机定子铁芯温度预测模型, 充分利用改进的Transformer架构, 即倒置Transformer (iTransformer) 模型, 使用倒置的时间序列编码方式, 解决Transformer在处理多元变量相关性方面的局限; 采用融合傅里叶卷积混合器 (fused Fourier convolution mixer, FFCM) 实现对时序数据的局部特征提取和增强; 使用多头差分注意力机制 (multi-head differential attention, MHDA) 替换传统自注意力, 有效消除注意力噪声, 促使模型关注关键信息。经过训练和验证, 与其他主流预测模型相比, 该模型具有更高的预测精度, 有助于及时发现潜在故障, 避免停机检修, 对保障汽轮发电机稳定运行具有重要应用价值, 有效地提升了温度预测技术的准确性与实用性。

[关键词] 汽轮发电机; 定子铁芯温度预测; iTransformer模型; 融合傅里叶卷积混合器; 多头差分注意力机制

[引用本文格式] 侯一凡, 袁和金, 孙帆. 基于FFCM-MHDA-iTransformer的汽轮发电机定子铁芯温度预测模型[J]. 热力发电, 2025, 54(12): 67-75. HOU Yifan, YUAN Hejin, SUN Fan. Prediction model for stator core temperature of turbo generator based on FFCM-MHDA-iTransformer[J]. Thermal Power Generation, 2025, 54(12): 67-75.

Prediction model for stator core temperature of turbo generator based on FFCM-MHDA-iTransformer

HOU Yifan¹, YUAN Hejin¹, SUN Fan²

(1.Department of Computer Science, North China Electric Power University, Baoding 071003, China;
2.Electric Power Science Research Institute, State Grid Ningxia Electric Power Co., Ltd., Yinchuan 750002, China)

Abstract: Abnormal stator core temperatures in generators can lead to serious issues such as aging of insulating materials and winding shorts, thereby affecting the overall performance and lifespan of the generator. This study presents a stator core temperature prediction model for turbo generators based on FFCM-MHDA-iTransformer. It leverages an improved Transformer architecture, namely the inverted Transformer (iTransformer) model, which adopts an inverted time-series encoding approach to address the limitations of the standard Transformer in handling multivariate variable correlations. The model employs fused Fourier convolution mixer (FFCM) to enhance and extract local features from time-series data. Furthermore, the model replaces conventional self-attention with multi-head differential attention (MHDA), effectively reducing attention noise and directing the model's focus towards critical information. After training and validation, the proposed model demonstrates higher prediction accuracy compared to other mainstream prediction models. It facilitates timely detection of potential faults, preventing shutdowns for maintenance, and holds significant application value for ensuring stable operation of turbo generators. This approach effectively enhances the accuracy and practicality of temperature prediction technology.

收稿日期: 2025-03-27

基金项目: 宁夏自然科学基金项目 (2024AAC03758)

Supported by: Natural Science Foundation of Ningxia (2024AAC03758)

第一作者简介: 侯一凡 (2000), 女, 硕士研究生, 主要研究方向为时间序列预测, 15847608588@163.com.

Key words: turbo generator; stator core temperature prediction; iTransformer model; fused Fourier convolution mixer; multi-head differential attention

“双碳”目标推动新能源发电快速发展，但风电与光伏的间歇波动性使其难以像传统电源一样精准控制，我国电力系统正面临灵活性资源需求增长与供给不足的矛盾^[1-2]。展望未来，煤电仍需扮演重要角色，不仅需保障基础供电，还需承担系统调节的重要功能^[3-4]。开展火电机组健康状态预测，能提前识别潜在缺陷，提升机组健康性和可靠性。发电机是机组的关键设备，而定子铁芯是其重要组成部件。一旦定子铁芯因绝缘层受损而局部异常升温，将严重威胁定子线棒安全，进而对发电机的稳定运行构成重大威胁^[5]。新旧发电机在相同工况下的定子铁芯温度存在显著差异，随着设备老化，铁芯温度可能升高，从而加剧设备磨损和故障风险^[6-7]。鉴于此，对定子铁芯温度进行预测显得尤为重要。

汽轮发电机定子铁芯温度预测基于数学模型或仿真技术，结合运行参数和结构特性，实时监测与预测铁芯温度。传统方法以热力学理论和多物理场耦合为核心，构建温度场模型。王芳等^[8]通过高精度三维计算流体力学与多物理场耦合仿真，系统揭示了汽轮发电机定子铁芯温度分布规律。胡磊^[9]对定子铁芯建立等效热网络得到了整个计算域内的温度分布。但此类方法存在计算耗时，难以实时预测的局限性。近年来，随着深度学习（deep learning, DL）的发展，考虑多源变量的数据分析监测模型受到广泛关注。Cai 等人^[10]提出了一种基于伪孪生嵌套长短时记忆（pseudo siamese nested long short-term memory, PSNLSTM）神经网络的模型，实现了对定子铁芯、定子齿和定子绕组的温度预测。常玉红等^[11]提出了一种基于时间卷积神经网络（temporal convolutional network, TCN）的模型，以历史监测数据训练模型，显著提升了抽水蓄能机组定子温度监测的精度与实时性。姚驰宇等^[12]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。温建民等^[13]基于长短时记忆（long short-term memory, LSTM）神经网络预测变压器绕组温度，选取绕组温度、环境温度、运行功率、运行电流作为输入变量进行训练得到温度预测模型。

具有较强非线性处理能力的卷积神经网络

（convolutional neural network, CNN）在一定程度上难以准确捕捉数据中的时间依赖关系，这可能导致其在处理时序数据时预测精度存在一定局限性。而循环神经网络（recurrent neural network, RNN）、LSTM 神经网络等时序数据模型对邻近输入依赖性较高，当序列数据存在急剧变化的现象时，其预测效果可能会出现一定误差。

因此，本文建立火电机组健康状态预测整体框架并提出一种基于 FFCM-MHDA-iTransformer 的预测模型，使用该模型预测汽轮发电机定子铁芯温度。针对火力发电机定子铁芯温度的时序性特征，引入倒置 Transformer（iTransformer）模型，采用倒置的时间序列编码策略，突破处理复杂多元相关性时的局限；通过融合傅里叶卷积混合器（fused Fourier convolution mixer, FFCM）深入挖掘数据的全局趋势和局部波动。使用多头差分注意力机制（multi-head differential attention, MHDA），有效降低注意力噪声，进一步提升预测的精度，旨在及时发现潜在故障，保障发电机稳定运行。

1 方法理论

1.1 基于 iTransformer 模型的时序预测架构

传统 Transformer 在处理多变量时间序列时面临性能退化、变量嵌入策略的不合理性以及语义信息匮乏等问题。为此，研究者提出了一种倒置 Transformer 架构的 iTransformer 模型^[14]，模型具体如图 1 所示。模型首先将同一变量的整个时间序列映射至高维特征空间，独立且精确地捕捉每个变量所承载的历史动态信息，运用注意力机制处理建模变量间的复杂相关性；同时，前馈神经网络在时间维度上逐层编码历史观测的特征，进而映射至未来的预测值。

iTransformer 模型在设计上进行了简化，由嵌入层（embedding layer）、层归一化（layer normalization, LN）、投影层（projector layer）以及若干可堆叠的 Transformer 模块（TrmBlock）构成。整体模型结构如图 2 所示。

本文构建模型采用 iTransformer 模型中的反转时间序列编码层作为数据嵌入层并以前馈神经网络（feed-forward neural network, FNN）作为中间层和最终投影层，同时对 TrmBlock 自注意力模块进

行改进。

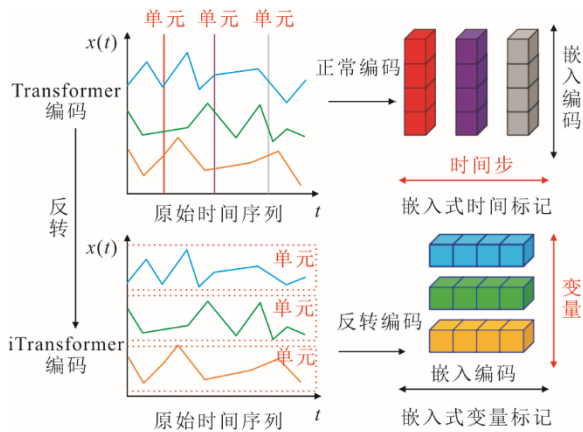


图 1 iTransformer 应用于序列的反转形状上
 Fig.1 Applying iTransformer to reversed shape of the sequence

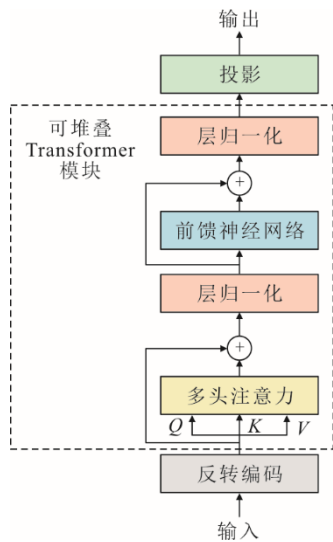


图 2 iTransformer 的整体模型结构
 Fig.2 Overall model architecture of iTransformer

1.1.1 嵌入层

对于一个时间跨度为 T 、包含 N 个变量的多维时间序列 $\mathbf{X}=\{x_1, \dots, x_T\} \in \mathbf{R}^{T \times N}$ 可以采用 2 种方式来表达:一是通过使用 $\mathbf{X}_{t,:}$ 表示在同一时间点上所有变量的状态,即捕捉到瞬间的系统快照;二是通过序列 $\mathbf{X}_{:,n}$ 来追踪单个变量随时间流转的完整轨迹。时间序列数据往往蕴含着丰富的语义信息,并且其各变量通常具有完全不同的物理含义,即使语义相同,其测量单位也可能完全不同。相较于传统方法中直接对原始时间序列数据 $\mathbf{X}_{t,:}$ 进行特征嵌入, iTransformer 模型创新性地修改了嵌入层,将时间序列数据的维度进行反转,使得每个变量成为一个时间序列,对每个 $\mathbf{X}_{:,n}$ 独立地进行特征映射,获得 N 个变量的特征表示, $\mathbf{H}=\{\mathbf{h}_1, \dots, \mathbf{h}_N\} \in \mathbf{R}^{N \times D}$, 其中

$\mathbf{h}_i \in \mathbf{R}^{N \times D}$ 蕴含了对应变量的过去时间段 T 内的时序变化趋势和特征信息。

1.1.2 前馈神经网络

iTransformer 利用 FNN 处理每个变量 token 的完整序列表示,使得 FNN 能够提取复杂且全面的时间序列特征,从而克服了信息错位和局部化限制的问题。通过堆叠转置模块, iTransformer 有效地对观测到的时间序列进行编码,并利用密集的非线性连接解码未来序列的表示。iTransformer 中的 FNN 作为预测表征学习器更为高效,其神经元经过训练能够捕捉时间序列的内在属性,如振幅、周期性和频谱特征,这些特征在不同时间序列中可共享。

1.2 融合傅里叶卷积混合器

直接应用 Transformer 处理时间序列数据存在局部特征提取不足的问题。有研究提出结合时空卷积网络与注意力机制^[15],融合卷积神经网络局部特征提取优势和 Transformer 全局建模能力,提高预测准确性。本文采用融合傅里叶卷积混合器技术^[16],实现对时序数据的局部特征提取和增强,其结构如图 3 所示。

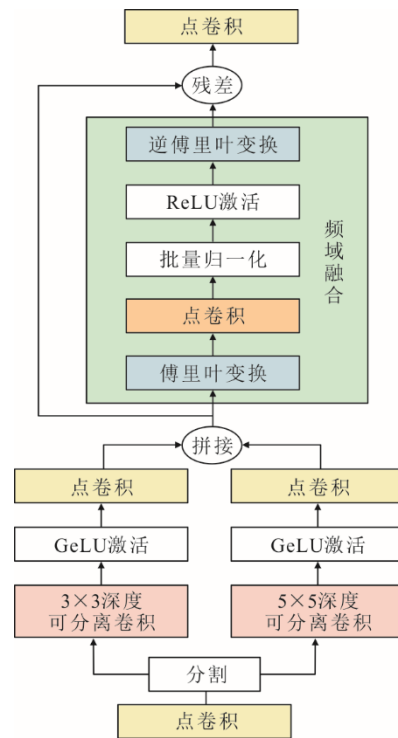


图 3 傅里叶卷积混合器
 Fig.3 Fused Fourier convolution mixer

具体而言,在空间域操作中,使用点卷积提升输入特征 $\mathbf{X} \in \mathbf{R}^{N \times W \times C}$ 并将其分成 2 组以提取多尺度局部特征,最终获得生成空间域特征图 $\mathbf{X}_{\text{spatial}} \in$

$\mathbf{R}^{N \times W \times 2C}$, 形式如下:

$$\mathbf{X}_{\text{spatial}} = \text{PConv}(\sigma \cdot \text{ConvMulti} \cdot \text{Split}(\text{PConv}(\mathbf{X}))) \quad (1)$$

式中: ConvMulti 为空间域中不同核大小的深度卷积; PConv 为点卷积; σ 为 GeLU 激活。

在频域操作中, $\mathbf{X}_{\text{spatial}}$ 经离散傅里叶变换 (discrete Fourier transform, DFT) 为实部和虚部分量, 将空间域特征图转换至频域, 生成蕴含数据全局信息的特征图, 特别是频率特性。连接的分量经过 1×1 核大小的卷积操作, 进一步提取和增强全局特征。调制后, 实部和虚部分量被分割, 离散傅里叶逆变换 (inverse discrete Fourier transform, IDFT) 将频域特征转换回空间域:

$$\mathbf{X}_R, \mathbf{X}_I = \text{DFT}(\mathbf{X}_{\text{spatial}}) \quad (2)$$

$$\hat{\mathbf{X}}_R, \hat{\mathbf{X}}_I = \sigma \cdot \text{BN}(\text{PConv}(\text{Concat}(\mathbf{X}_R, \mathbf{X}_I))) \quad (3)$$

$$\mathbf{X}_{\text{FFCM}} = \text{PConv}(\mathbf{X}_{\text{spatial}} + \mathbf{X}_{\text{frequency}}) \quad (4)$$

最后, 通过残差结构和 PConv 获得 FFCM 的输出:

$$\mathbf{X}_{\text{FFCM}} = \text{PConv}(\mathbf{X}_{\text{spatial}} + \mathbf{X}_{\text{frequency}}) \quad (5)$$

频域特征与原始空间域特征通过残差连接进行融合, 生成包含丰富全局和局部信息的输出特征图, 充分利用频域和空域的特征信息, 以提升时间序列预测的性能。

1.3 多头差分注意力机制

经典 Transformer 模型用注意力机制捕捉序列全局依赖, 但在时间序列预测中会产生注意力噪声。因处理海量信息时, 模型易关注无关信息, 复杂序列又增大区分难度。故减少注意力噪声、增强关键时间特征捕捉能力, 是当前研究亟待解决的问题。本文将 iTransformer 模型中的多头自注意力机制替换成多头差分注意力机制^[17], 以相邻时间步差异关注信息变化, 构建差分编码器消除噪声, 引导模型聚焦关键信息。多头差分注意力机制结构如图 4 所示。

具体来说, 给定输入 \mathbf{X} , 首先将它投射成查询 (query)、键 (key) 和值 (value), 即 \mathbf{Q}_1 、 \mathbf{Q}_2 、 \mathbf{K}_1 、 \mathbf{K}_2 、 \mathbf{V} 。然后差分注意力算子 DiffAttn 通过以下方式计算输出:

$$[\mathbf{Q}_1; \mathbf{Q}_2] = \mathbf{X}\mathbf{W}_Q, [\mathbf{K}_1; \mathbf{K}_2] = \mathbf{X}\mathbf{W}_K, \mathbf{V} = \mathbf{X}\mathbf{W}_V \quad (6)$$

$$\text{DiffAttn}(\mathbf{X}) = \left(\text{softmax}\left(\frac{\mathbf{Q}_1\mathbf{K}_1^T}{\sqrt{d}}\right) - \lambda \text{softmax}\left(\frac{\mathbf{Q}_2\mathbf{K}_2^T}{\sqrt{d}}\right) \right) \mathbf{V} \quad (7)$$

式中: \mathbf{W}_Q 、 \mathbf{W}_K 、 \mathbf{W}_V 为用于训练模型的参数矩阵; d

为向量维度; λ 为可学习的标量; softmax 为归一化函数。为了同步学习动态, 将标量 λ 重新参数化为:

$$\lambda = \exp(\lambda_{q1} \cdot \lambda_{k1}) - \exp(\lambda_{q2} \cdot \lambda_{k2}) + \lambda_{\text{init}} \quad (8)$$

式中: λ_{q1} 、 λ_{k1} 、 λ_{q2} 、 λ_{k2} 为可学习的向量; $\lambda_{\text{init}} \in (0, 1) \in (0, 1)$ 为用于初始化 λ 的常数。

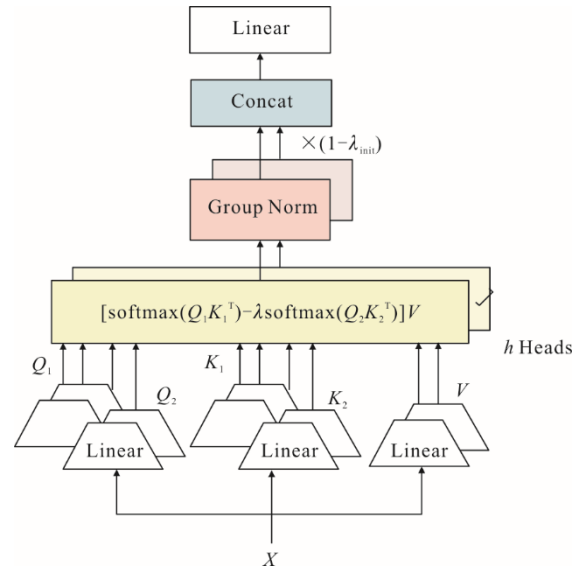


图 4 多头差分注意力机制

Fig.4 The multi-head differential attention mechanism

差分注意力机制的关键在于计算 2 个 softmax 注意力图之间的差异。模型首先利用查询和键向量生成两组不同的注意力图, 并通过相减的方式得到差分结果, 能够使模型更加集中于与当前任务紧密相关的上下文信息, 减少对不相关信息的注意。

h 表示注意力头 (head) 的数量, 多头差分注意力机制对各个头使用不同的投影矩阵 $\mathbf{W}_{V,i}$ 、 $\mathbf{W}_{K,i}$ 、 $\mathbf{W}_{Q,i}$, $i \in [1, h]$ 。标量 λ 在同一层内的头之间共享。然后对头输出执行归一化, LN 独立应用于每个 head, 由于差分注意力往往具有更稀疏的模式, 因此头之间的统计信息更加多样化。为了改进梯度的统计情况, LN 算子会在连接操作之前对每个头进行归一化。最终投射成最终结果, 如下所示:

$$\text{head}_i = \text{DiffAttn}(\mathbf{X}; \mathbf{W}_{Q,i}, \mathbf{W}_{K,i}, \mathbf{W}_{V,i}, \lambda) \quad (9)$$

$$\overline{\text{head}_i} = (1 - \lambda_{\text{init}}) \cdot \text{LN}(\text{head}_i) \quad (10)$$

$$\text{Multihead}(\mathbf{X}) = \text{Concat}(\overline{\text{head}_1}, \dots, \overline{\text{head}_h}) \mathbf{W}_O \quad (11)$$

式中: \mathbf{W}_O 为可学习的投影矩阵; LN 为对每个头使用均方根归一化; Concat 的作用是沿通道维度将头连接在一起。使用一个固定乘数 $(1 - \lambda_{\text{init}})$ 作为 LN 的缩放尺度, 以使梯度与 Transformer 对齐。

2 基于 FFCM-MHDA-iTransformer 的火电机组健康状态指标预测模型

2.1 火电机组健康状态指标预测整体框架

本文建立的火电机组健康状态指标预测整体框架如图 5 所示。具体过程为：从火电机组的各类监测系统和信息平台收集运行数据、历史故障记录、环境监测信息等，并进行数据清洗等预处理工作；利用相关性分析方法，从预处理后的数据中提取与待预测指标密切相关的特征变量；构建 FFCM-MHDA-iTransformer 预测模型，利用历史数据对模型进行训练，得到最优的预测模型参数；将特征变量输入训练好的预测模型，得到火电机组未来一段时间内的健康状态指标预测结果。

对预测结果进行分析，可以识别机组潜在的故障风险，为制定科学合理的维护计划提供支持。

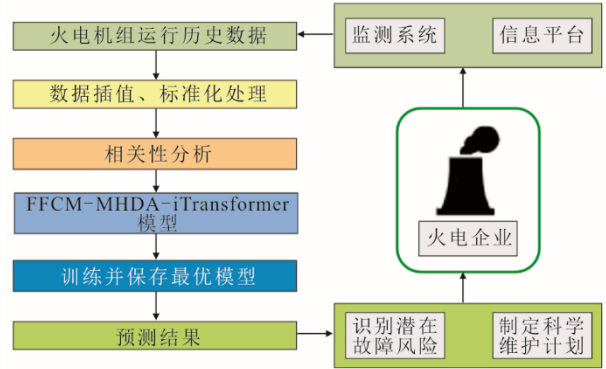


图 5 火电机组健康状态指标预测整体框架
Fig.5 The overall framework for predicting health status indicators of thermal power units

2.2 FFCM-MHDA-iTransformer 模型构建

本文以 FFCM、iTransformer 倒置时间序列编码层、MHDA、FFN 构建 FFCM-MHDA-iTransformer 预测模型，模型结构如图 6 所示。

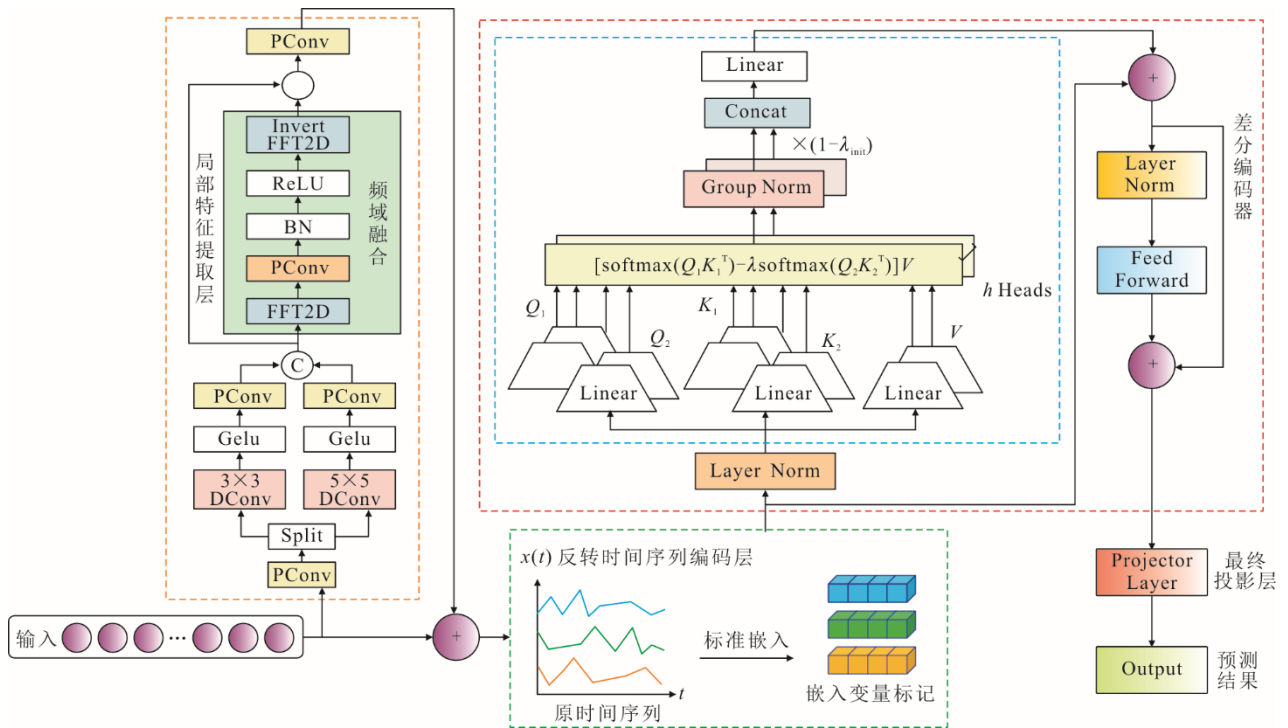


图 6 FFCM-MHDA-iTransformer 预测模型
Fig.6 The FFCM-MHDA-iTransformer prediction model

该模型首先通过 FFCM 实现对时序数据的局部特征提取和增强，并通过残差连接原始输入数据和局部特征增强后的数据，然后采用反转的时间序列嵌入层将同一变量的整个时间序列映射至高维特征空间，形成变量标记，编码后的数据通过多头差分注意力模块减少时间序列预测中的注意力噪声。最终经过前馈神经网络映射得到预测结果。

3 基于 FFCM-MHDA-iTransformer 的发电机定子铁芯温度预测

3.1 数据集和数据处理

本文的数据来源于宁夏某火电机组，机组的历史设备运行数据经传感器收集，并被保存在电厂的数据采集与监视控制系统 (supervisory control and

data acquisition, SCADA) 中, 包含汽轮机、锅炉、发电机三大设备各类测点数据 180 余种, 采样间隔为 15 min, 选取其中 2023 年 1 月 1 日至 2023 年 2 月 28 日的 5 664 组数据作为数据集。

首先从现场采集的海量数据中初步筛选出与预测指标定子铁芯温度紧密相关的数据子集。采用四分位法判断并处理数据中的离群值, 运用线性插值的方法补全缺失的数据; 然后使用相关性分析的方法, 深入挖掘并确定价值密度高、与预测目标高度相关、适合作为模型训练输入的数据, 将原始数据集进行归一化得到多特征数据集; 最后按照 7:2:1 的比例划分为训练集、验证集、测试集。输入变量的维度为 6, 时间跨度即输入数据的窗口大小为 36, 即使用前 9 h 的数据预测下一时刻的铁芯温度。具体方法为将某一时刻的多特征时间序列数据 $[X_i, X_{i+1}, \dots, X_{i+35}]$ 作为模型输入, 将下一个测点的数据 X_{i+36} 作为模型输出。

3.1.1 发电机定子铁芯温度预测数据选取

在预测火电机组发电机定子铁芯温度的研究

中, 除了直接相关的指标定子铁芯温度历史数据外, 本文选择以下关键数据作为初步数据子集: 母线电压、励磁电压、励磁电流、定子机端电压、定子机端电流、主变绕组温度以及主变油面温度。

火电机组中, 发电机作为重要的能量转换设备, 其定子铁芯温度受到多种因素的影响。母线电压和定子机端电压反映了发电机的电气运行状态, 参数变化可能导致发电机内部的电磁场分布变化, 从而影响定子铁芯的损耗和温度; 励磁电压及励磁电流则直接决定了发电机磁场的强弱, 进而影响定子铁芯中的磁通密度和涡流损耗, 涡流损耗是定子铁芯发热的主要原因之一, 因此励磁参数与定子铁芯温度密切相关; 定子机端电流反映了发电机的输出电流情况, 而电流的大小和变化可能导致定子绕组中的铜耗增加, 进而通过热传导影响定子铁芯的温度; 此外, 主变绕组温度及主变油面温度作为变压器油冷却系统的重要参数, 其变化可以反映变压器的负载情况和散热效率, 也会对发电机定子铁芯温度产生影响。初步数据子集信息见表 1。

表 1 发电机定子铁芯温度预测相关参数
Tab.1 Relevant parameters for predicting the stator core temperature of the generator

序号	参数	序号	参数
1	定子铁芯温度/℃	5	定子机端电压/V
2	母线电压/V	6	定子机端电流/A
3	励磁电压/V	7	主变绕组温度/℃
4	励磁电流/A	8	主变油面温度/℃

3.1.2 发电机定子铁芯温度预测数据处理

在构建指标预测模型时, 全面纳入特征参数常因高维度和冗余性导致预测精度下降。特征筛选是降维的有效方法, 但火电机组历史数据的高维度、非线性及强相关性使筛选复杂化。本文引入距离相关系数法综合评估线性和非线性相关性, 为模型构建提供精确特征选择依据, 以克服传统 Pearson 相关系数对线性关系和正态分布的依赖。

设 $\{(x_i, y_i), i=1, 2, \dots, n\}$ 为总体 (X, Y) 的随机样本, 首先计算每对观测值之间的欧氏距离, 形成距离矩阵 A 和 B 。

$$a_{ij} = \|x_i - x_j\|, b_{ij} = \|y_i - y_j\| \quad i, j = 1, 2, \dots, n \quad (12)$$

然后对距离矩阵 A 和 B 进行双重中心化处理, 即从每个元素中减去其行平均值、列平均值, 并加上整体平均值。

$$A_{ij} = a_{ij} - \bar{a}_i - \bar{a}_j + \bar{a}, B_{ij} = b_{ij} - \bar{b}_i - \bar{b}_j + \bar{b} \quad (13)$$

距离协方差和距离方差的计算公式如下:

$$dCov^2(X, Y) = \frac{1}{n^2} \sum_{i,j=1}^n A_{ij} B_{ij} \quad (14)$$

$$dVar^2(X) = \frac{1}{n^2} \sum_{i,j=1}^n A_{ij}^2 \quad (15)$$

$$dVar^2(Y) = \frac{1}{n^2} \sum_{i,j=1}^n B_{ij}^2 \quad (16)$$

代入距离相关系数的公式, 得到两随机变量的距离相关样本估计值, 距离相关系数越大, 说明相关性越强, 特征参数越重要。

$$dCor(X, Y) = \frac{dCov^2(X, Y)}{\sqrt{dVar(X)dVar(Y)}} \quad (17)$$

计算出发电机定子铁芯温度特征参数之间的相关系数, 绘制出相关性系数热力图如图 7 所示。

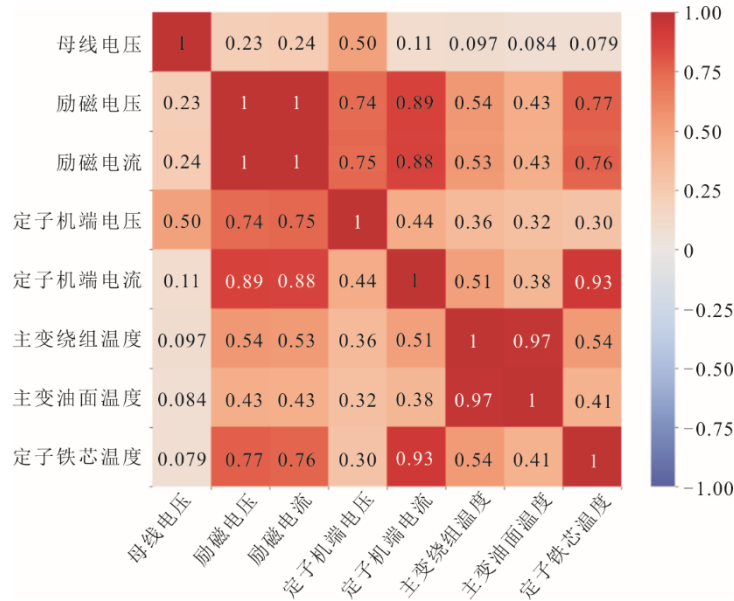


图 7 发电机定子铁芯温度相关系数热力图

Fig.7 Correlation coefficient heatmap for stator core temperature of the generator

认为距离相关系数大于 0.4 的特征与发电机定子铁芯温度数据变化有较大相关性，发电机定子铁芯温度预测模型输入数据为定子铁芯温度、励磁电压、励磁电流、定子机端电流、主变绕组温度以及主变油面温度。

3.2 评价指标

为了评价 FFCM-MHDA-iTransformer 模型的预测精度，本文采用平均绝对误差 (mean absolute error, MAE) δ_{MAE} 、均方根误差 (root mean squared error, RMSE) δ_{RMSE} 、平均绝对百分比误差 (mean absolute percentage error, MAPE) δ_{MAPE} 以及拟合优度 R_2 作为模型的评价指标。

各评价指标计算公式为：

$$\delta_{MAE} = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (18)$$

$$\delta_{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (19)$$

$$\delta_{MAPE} = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \times 100\% \quad (20)$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (21)$$

式中： n 为样本数量； y_i 为第 i 个样本的真实值； \hat{y}_i 为第 i 个样本的预测值； \bar{y} 为真实值的平均值。

3.3 超参数和模型设置

实验使用 Python3.8 编程语言，基于 Pytorch

1.10.2 框架搭建模型，并借助 NVIDIA GeForce RTX 3080 GPU 加速计算，使用 Adam 优化器，训练轮次为 100。为提高训练效率，本文采用动态调整学习率策略及早停机制。表 2 总结了本文模型的关键超参数设置。

表 2 参数设置

Tab.2 Parameter settings

序号	参数名称	设置值
1	批次大小	64
2	初始学习率	0.001
3	Dropout	0.1
4	多头注意力头数	8
5	编码器数量	2
6	解码器数量	1
7	早停机制耐心值	20
8	可变学习率	20 epoch/0.5

3.4 实验结果

3.4.1 消融实验

为验证改进方法的有效性，本文使用相同的实验环境、实验数据、模型参数等进行消融实验。在整体模型的基础上提出以下 4 种模型：第 1 组为不使用 iTransformer 反转时间序列编码，而使用传统 Transformer 嵌入编码的模型，记为 no-iTransformer；第 2 组为不使用多头差分注意力机制 MHDA 模块而使用普通自注意力机制的模型，记为 no-MHDA；第 3 组为不使用傅里叶卷积混合器 FFCM 的模型，记为 no-FFCM；第 4 组为本文提出的使用反转时间序列编码、FFCM 及 MHDA 的完整模型。消融实验

结果见表3, 可得如下结论:

1) 预测效果最好的模型是包含全部模块的FFCM-MHDA-iTransformer模型, 各项指标都优于其他模型;

2) 分别消除iTransformer时间序列编码、傅里叶卷积混合模块及多头差分注意力机制模块后, 模型误差有所增大。

表3 消融实验评价指标对比

Tab.3 Comparison of evaluation indicators in ablation experiments

模型	R_2	δ_{MAE}	δ_{RMSE}	δ_{MAPE}
no-iTransformer	0.918 7	0.931 6	1.239	1.56
no-MHDA	0.928 2	0.867 0	1.164	1.46
no-FFCM	0.917 1	0.949 3	1.251	1.60
FFCM+MHDA+iTransformer	0.931 8	0.840 9	1.135	1.42

消融实验中发电机定子铁芯温度预测结果如图8所示, 实验间隔为15 min。

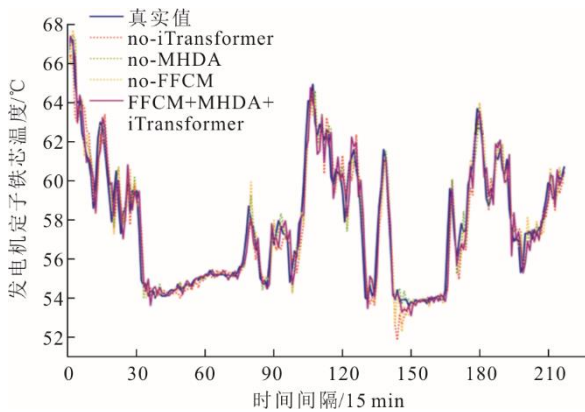


图8 消融实验发电机定子铁芯温度预测结果

Fig.8 The prediction results for generator stator core temperature in ablation experiments

由图8可知:

1) 加入iTransformer时间序列编码后, 输入序列的每个变量被独立地嵌入为变量标记(token), 反转操作使得嵌入的标记能够更好地捕捉时间序列的全局特征, 实现对非线性模式的准确捕捉;

2) 加入FFCM傅里叶卷积混合模块后, 有效地捕捉到了时间序列中的周期性成分和趋势变化, 从而提升了模型对时间序列数据的处理能力, 在变化趋势明显的的数据上拟合情况较好;

3) 加入MHDA多头差分注意力机制模块后, 模型能够更加专注于时间序列中的关键时间特征并抑制预测任务不直接相关的信息, 在连续波动、变化趋势不明显的的数据上拟合较好;

4) 当反转编码方式、FFCM和MHDA同时应用于模型时, 模型能够实现对时间序列数据的全面

理解和准确预测, 结果充分证明反转编码方式、FFCM和MHDA在预测任务中的协同作用。

消融实验证明, FFCM-MHDA-iTransformer预测模型中的每个模块都展现出了独特的优势。

3.4.2 对比实验

为验证本文模型FFCM-MHDA-iTransformer在发电机定子铁芯温度预测中的有效性, 设置以下几种经典的时间序列预测模型CNN、SVR、GRU、LSTM、CNN-LSTM作为对照算法。为消除其他因素对各模型结果造成干扰, 各算法使用相同的实验环境、训练集、测试集, 以及训练轮次、优化器等参数。对比实验结果见表4。

表4 对比实验评价指标对比

Tab.4 Comparison of evaluation indicators in comparative experiments

模型	R^2	δ_{MAE}	δ_{RMSE}	δ_{MAPE}
CNN	0.912 9	0.985 9	1.282	1.68
SVR	0.912 9	1.066 2	1.282	1.81
GRU	0.910 3	0.998 4	1.301	1.70
LSTM	0.918 1	0.936 7	1.243	1.58
CNN-LSTM	0.917 6	0.963 5	1.247	1.63
FFCM+MHDA+iTransformer	0.931 8	0.840 9	1.135	1.42

对比实验中发电机定子铁芯温度预测结果如图9所示。由图9可知, 本文所提模型在每种评价指标下均取得了最优的结果。相比于CNN、SVR、GRU、LSTM、CNN-LSTM模型, FFCM+MHDA+iTransformer模型在发电机定子铁芯温度预测中, δ_{MAPE} 分别降低了15.48%、21.55%、16.47%、10.13%、12.88%, δ_{RMSE} 分别降低了11.47%、11.47%、12.76%、8.69%、10.91%, δ_{MAE} 分别降低了14.71%、21.13%、15.78%、10.23%、12.72%, 拟合系数均高于其他模型, 表明本文方法相比传统预测方法有更好的预测效果。

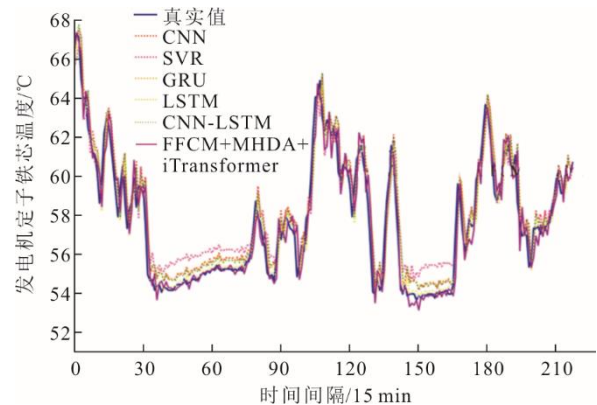


图9 对比实验发电机定子铁芯温度预测结果

Fig.9 The prediction results for generator stator core temperature in comparative experiments

4 结 论

本文建立火电机组健康状态预测的整体框架,提出一种基于FFCM-MHDA-iTransformer的预测模型,实现了火力发电机定子铁芯温度的准确预测,为汽轮发电机状态乃至火电机组的健康状态预测提供了新的思路和方法,主要研究结论如下。

1) 基于FFCM-MHDA-iTransformer的预测模型能够挖掘多维时序数据的非线性关系,与CNN、LSTM等模型进行比较,每种评价指标均取得了最优结果。

2) 针对火力发电机定子铁芯温度的时序性特征,引入iTransformer时间序列编码方式,能够更敏锐地捕捉时间序列的全局特性;通过结合傅里叶变换与卷积操作,可以捕捉时间序列的局部波动,提高预测的准确性。

3) 针对注意力噪声过大的问题,使用多头差分注意力机制能够有效抑制与预测任务不相关的信息,提升预测准确性。

[参 考 文 献]

- [1] 杨贤东,袁旭峰,熊炜,等.考虑源荷不确定性的风光火储系统低碳经济调度[J].智慧电力,2022,50(8):22-29.
YANG Xiandong, YUAN Xufeng, XIONG Wei, et al. Low-carbon economic dispatch of wind-solar-fired-storage system considering source-load uncertainty[J]. Smart Power, 2022, 50(8): 22-29.
- [2] 张智刚,康重庆.碳中和目标下构建新型电力系统的挑战与展望[J].中国电机工程学报,2022,42(8):2806-2819.
ZHANG Zhigang, KANG Chongqing. Challenges and prospects for constructing the new-type power system towards a carbon neutrality future[J]. Proceedings of the CSEE, 2022, 42(8): 2806-2819.
- [3] 苏鹏,王文君,杨光,等.提升火电机组灵活性改造技术方案研究[J].中国电力,2018,51(5):87-94.
SU Peng, WANG Wenjun, YANG Guang, et al. Research on the technology to improve the flexibility of thermal power plants[J]. Electric Power, 2018, 51(5): 87-94.
- [4] 杜小泽,张璇,庞力平,等.燃煤发电机组灵活调峰下机炉安全状态监测与调控研究综述[J].中国电机工程学报,2024,44(18):7178-7200.
DU Xiaozhe, ZHANG Xuan, PANG Liping, et al. Review on the monitoring and regulation of boiler and turbine safety states in coal-fired power generation units under flexible peak shaving[J]. Proceedings of the CSEE, 2024, 44(18): 7178-7200.
- [5] 刘祝鸿.汽轮发电机定子铁芯损耗及温升试验案例分析[J].机电信息,2016(18):22-23.
LIU Zhuhong. Case analysis on stator core loss and temperature rise test of turbo-generator[J]. Mechanical and Electrical Information, 2016(18): 22-23.
- [6] KANG J B, LEE J Y, NGUYEN T, et al. Effects of

- material deformation due to aging of electrical steel on the brushless wound-field synchronous generator[J]. Electronics, 2023, 12: 2279.
- [7] 武钰.定子匝间短路故障下发电机损耗特性分析[D].北京:华北电力大学,2022:1-10.
WU Yu. Analysis of generator loss characteristics under stator interturn short circuit[D]. Beijing: North China Electric Power University, 2022: 1-10.
- [8] 王芳,高斯博,汤文侠,等.大型水氢氢冷却汽轮发电机定子温度分布[J].哈尔滨理工大学学报,2015,20(1):26-30.
WANG Fang, GAO Sibao, TANG Wenxia, et al. Temperature distribution of large water-hydrogen-hydrogen cooled turbo-generator stator[J]. Journal of Harbin University of Science and Technology, 2015, 20(1): 26-30.
- [9] 胡磊.1100 MW级核电水氢氢汽轮发电机定子铁芯温度场研究[J].电机技术,2019(5):10-16.
HU Lei. Research on temperature field of stator core of 1100 MW nuclear water-hydrogen-hydrogen turbine generator[J]. Electrical Machine Technology, 2019(5): 10-16.
- [10] CAI Y P, CEN Y F, CEN G, et al. Temperature prediction of PMSMs using pseudo-siamese nested LSTM[J]. World Electric Vehicle Journal, 2021, 12: 57.
- [11] 常玉红,吴月超,何铮,等.基于MIC-TCN-Attention的抽水蓄能机组发电电动机定子温度预警方法研究[J].中国农村水利水电,2021(7):125-131.
CHANG Yuhong, WU Yuechao, HE Zheng, et al. Research on the generator stator temperature early warning method of pumped storage unit based on MIC-TCN-Attention[J]. China Rural Water and Hydropower, 2021(7): 125-131.
- [12] 姚驰宇,桂晶,李泼,等.基于Transformer语义迁移的定子热故障诊断[J].电气技术,2025,26(3):59-64.
YAO Chiyu, GUI Jing, LI Po, et al. Transformer-based semantic transfer for stator thermal fault diagnosis[J]. Electrical Engineering, 2025, 26(3): 59-64.
- [13] 温建民,何斌,王开康,等.基于LSTM神经网络的地铁变压器绕组温度预测[J].都市轨道交通,2023,36(3):77-81.
WEN Jianmin, HE Bin, WANG Kaikang, et al. Temperature prediction of subway transformer windings based on LSTM[J]. Urban Rapid Rail Transit, 2023, 36(3): 77-81.
- [14] LIU Y, HU T, ZHANG H, et al. iTransformer: inverted Transformers are effective for time series forecasting[C]. International Conference on Learning Representations (ICLR 2024), Kigali, Rwanda, 2024.
- [15] 张艳红,郭伟彤,张斌.基于多视图时空卷积的交通流量预测模型[J].公路,2024,69(1):255-262.
ZHANG Yanhong, GUO Weitong, ZHANG Bin. Traffic flow prediction model based on multi-view spatio-temporal convolution[J]. Highway, 2024, 69(1): 255-262.
- [16] GAO N, JIANG X Y, ZHANG X H, et al. Efficient frequency-domain image deraining with contrastive regularization[C]. European Conference on Computer Vision (ECCV 2024), Milan, Italy, 2024.
- [17] YE T Z, LI D, XIA Y Q, et al. Differential transformer[C]. International Conference on Learning Representations (ICLR 2025), Singapore, 2025.

(责任编辑 邓玲惠)