



GPT4PF:一种微调预训练LLM的光伏功率预测模型

王德文, 苗庆健, 李成浩, 孙媛, 赵文清

引用本文:

王德文, 苗庆健, 李成浩, 等. GPT4PF:一种微调预训练LLM的光伏功率预测模型[J]. 智能系统学报, 2026, 21(4): 864-875.

WANG Dewen, MIAO Qingjian, LI Chenghao, et al. GPT4PF: fine-tuning pre-trained LLM for photovoltaic power prediction[J]. *CAAI Transactions on Intelligent Systems*, 2026, 21(4): 864-875.

在线阅读 View online: <https://dx.doi.org/10.11992/tis.202507023>

您可能感兴趣的其他文章

融合共现推理的Faster R-CNN输电线路金具检测

Integrating co-occurrence reasoning for Faster R-CNN transmission line fitting detection

智能系统学报. 2021, 16(2): 237-246 <https://dx.doi.org/10.11992/tis.202012023>

样本仿真结合迁移学习的声呐图像水雷检测

Detection of underwater mine target in sidescan sonar image based on sample simulation and transfer learning

智能系统学报. 2021, 16(2): 385-392 <https://dx.doi.org/10.11992/tis.202101030>

融合多层次特征的中文语义角色标注

Chinese semantic role labeling with multi-level linguistic features

智能系统学报. 2020, 15(1): 107-113 <https://dx.doi.org/10.11992/tis.201910012>

结合谱聚类的标记分布学习

Label distribution learning based on spectral clustering

智能系统学报. 2019, 14(5): 966-973 <https://dx.doi.org/10.11992/tis.201809019>

基于支持向量的最近邻文本分类方法

The nearest neighbor text classification method based on support vector

智能系统学报. 2018, 13(5): 799-807 <https://dx.doi.org/10.11992/tis.201711007>

快速双非凸回归算法及其电力数据预测应用

Fast double nonconvex regression algorithm for forecast of electric power data

智能系统学报. 2018, 13(4): 665-672 <https://dx.doi.org/10.11992/tis.201708033>

DOI: 10.11992/tis.202507023

网络出版地址: <https://link.cnki.net/urlid/23.1538.TP.20260422.1210.002>

GPT4PF: 一种微调预训练 LLM 的光伏功率预测模型

王德文^{1,2}, 苗庆健¹, 李成浩¹, 孙媛¹, 赵文清^{1,3}

(1. 华北电力大学 计算机系, 河北 保定 071003; 2. 河北省能源电力知识计算重点实验室, 河北 保定 071003; 3. 复杂能源系统智能计算教育部工程研究中心, 河北 保定 071003)

摘要: 准确的光伏功率预测对于保障电力系统安全、提高太阳能利用效率至关重要。本文提出一种基于微调预训练大语言模型的光伏功率预测模型。设计了一种高适应性且高效的三阶段微调策略, 在提升预测精度的同时, 训练参数量仅占总参数量的 0.49%。构建了一种嵌入时间信息的输入编码层, 以增强大语言模型对时间信息的建模能力, 消融实验表明平均绝对误差下降约 1%。提出了一种光伏数据切片标记方式, 解决单时间点信息有限、难反映光伏数据整体特征规律的问题。实验结果表明, 本模型相较 DLinear、PatchTST、LLM4TS, 平均绝对误差分别下降 18.9%、4.1%、5.4%。

关键词: 预训练大语言模型; 光伏功率预测; 大语言模型; 时间编码; 多阶段微调; 时序切片; 输入编码; 线性探测

中图分类号: TP183; TM615 文献标志码: A 文章编号: 1673-4785(2026)04-0864-12

中文引用格式: 王德文, 苗庆健, 李成浩, 等. GPT4PF: 一种微调预训练 LLM 的光伏功率预测模型 [J]. 智能系统学报, 2026, 21(4): 864-875.

英文引用格式: WANG Dewen, MIAO Qingjian, LI Chenghao, et al. GPT4PF: fine-tuning pre-trained LLM for photovoltaic power prediction[J]. CAAI transactions on intelligent systems, 2026, 21(4): 864-875.

GPT4PF: fine-tuning pre-trained LLM for photovoltaic power prediction

WANG Dewen^{1,2}, MIAO Qingjian¹, LI Chenghao¹, SUN Yuan¹, ZHAO Wenqing^{1,3}

(1. Department of Computer, North China Electric Power University, Baoding 071003, China; 2. Hebei Key Laboratory of Knowledge Computing for Energy & Power, Baoding 071003, China; 3. Engineering Research Center of Intelligent Computing for Complex Energy Systems Ministry of Education, Baoding 071003, China)

Abstract: Accurate photovoltaic power forecasting is crucial for ensuring the security of power systems and improving the efficiency of solar energy utilization. This study proposed a photovoltaic power forecasting model named GPT4PF, which was developed by fine-tuning a pre-trained large language model (LLM). A highly adaptive and efficient three-stage fine-tuning strategy was proposed, which enhanced forecasting accuracy while limiting the number of training parameters to merely 0.49% of the total model parameters. An input encoding layer embedded with temporal information was constructed to strengthen the LLM's capability of modeling temporal features. Ablation experiments showed that the mean absolute error (MAE) was reduced by approximately 1%. In addition, the issue that a single time point carries limited information and struggles to reflect the overall characteristics and patterns of photovoltaic data was addressed through time-series segmentation technology. Experimental results showed that compared with DLinear, PatchTST, and LLM4TS, the proposed model achieved a reduction in MAE by 18.9%, 4.1%, and 5.4% respectively.

Keywords: pre-trained large language model; photovoltaic forecast; large language model; time encoding; multi-stage fine-tuning; time-series segmentation; input encoding; linear probing

随着全球能源危机与环境污染的日益加剧, 可再生能源的开发利用变得愈发紧迫。截至

2024 年底, 全国太阳能发电装机容量达 8.9 亿千瓦, 同比增长 45.2%^[1]。太阳能凭借其清洁、无排放的特性, 在“碳达峰、碳中和”战略目标实施过程中, 扮演着日益重要的角色。然而, 光伏发电的输出功率受太阳辐射强度、季节及温度等因素

收稿日期: 2025-07-21. 网络出版日期: 2026-04-22.

基金项目: 国家自然科学基金项目 (62371188).

通信作者: 王德文. E-mail: wde@ncepu.edu.cn.

©《智能系统学报》编辑部版权所有

的影响,呈现较大的波动性,因此,准确的光伏功率预测对提高太阳能利用效率至关重要。

光伏功率预测模型主要分为3类,即传统方法、机器学习方法与深度学习方法。传统方法包括物理方法和统计方法。物理方法运用数值天气预报模型^[2]来预测光伏发电的输出功率。常见的统计方法包括时间序列分析^[3]、回归分析^[4]、差分自回归移动平均模型^[5]。然而,面对复杂多变的气象状况,传统方法的预测能力显示出局限性。机器学习模型擅长从输入中直接映射非线性复杂特征至输出,常用技术涵盖支持向量机^[6]、随机森林^[7]等。这些技术能够高效处理历史数据,进而实现精准预测,为电网提供更为精确的电力调度策略。随着光伏发电系统所采集数据的复杂度不断提升,机器学习算法在应对复杂特征关联及非线性模型时,难以精确捕捉数据中的深层次特征。相比之下,深度学习凭借其更为强大的表征学习能力,在处理大规模数据时,有着更为出色的表现^[8]。因此,在解决复杂的光伏功率预测任务时,深度学习展现出了明显的优势。循环神经网络(recurrent neural network, RNN)^[9]在捕获顺序依赖关系方面具有较高性能,已被广泛应用于光伏预测领域。基于注意力的模型由于其显著的性能而在时间序列预测中越来越受欢迎。文献^[10]提出一种线性分解和周期增强 Informer 的地表太阳辐射短临预报方法,大幅提高模型短临预报效果。

预训练的基础模型在计算机视觉、自然语言处理(natural language processing, NLP)领域取得了令人印象深刻的进步,研究人员正在探索如何将大模型应用到时间序列预测中^[11]。GPT4TS(GPT for time series)^[12]首次将大语言模型应用到时间序列预测领域,利用一个冻结的预训练语言模型,通过微调其中的位置编码模块和层归一化模块,在各种主要类型的时间序列分析任务中实现了SOTA(state-of-the-art)或相当水平的性能。Sun等^[13]通过构建编码器并将时间序列数据标记化,使用冻结的大语言模型(large language model, LLM)实现时间序列任务,结果表明 LLM 在时序任务中表现良好。Time-MoE^[14]由一系列解码器模块组成,采用自回归方式运行,支持不同输入上下文长度下的灵活预测范围,首次将时间序列基础模型规模扩展至24亿参数,大幅提升预测精度。

上述研究证明了将 LLM 应用到时间序列预测任务的可行性,然而如何充分发挥其优势并应用到光伏预测任务中仍存在如下挑战。

1) LLM 在预训练阶段主要基于海量文本语

料,以离散的单词、字符为基本处理单位,模型通过学习离散符号间的概率分布和语义关联实现语言理解与生成。而光伏数据不仅包含数值信息,更蕴含复杂的时间依赖关系、趋势变化和周期性特征,这与 LLM 擅长处理的离散序列差异显著,导致 LLM 难以满足时间序列分析与预测的需求。

2) 由于注意力机制具有二次复杂性,受计算和记忆约束,限制了 LLM 处理时间序列数据的能力。此外光伏数据易受云层遮挡、传感器误差等噪声影响,模型在处理含噪声的时序数据时易过拟合,同时现有的微调方法在提升预测精度与适应性上仍有不足,存在计算开销大、模型收敛慢等问题。

针对以上问题和研究现状,本文提出一种基于微调预训练 LLM 的光伏功率预测模型 GPT-4PF(GPT for photovoltaic forecast),主要贡献总结为

1) 通过时间序列切片技术将光伏数据标记化,在丰富信息的同时,解决了单个时间点携带信息有限,且难以反映光伏数据整体特征与规律的问题。

2) 提出了一种高适应性且高效的三阶段微调策略,提升预测准确度的同时增强模型的适应性。首先将模型与时间序列数据对齐,然后使用线性探测训练一个新的输出层,最后对模型进行全微调,让 LLM 逐步适应光伏预测任务。

3) 通过参数选择与低秩适配,降低了微调过程中需要更新的参数数量和计算开销,使参与训练的参数仅占总参数数量的0.49%,同时减小光伏数据中存在的噪声在训练过程中对过拟合的影响。

4) 考虑到 LLM 缺乏对光伏数据中时间信息显式建模的能力,设计了一种嵌入时间信息的输入编码层,对时间信息进行编码并将其嵌入每个切片,解决了 LLM 无法对时间信息显式建模的问题,使模型可以更好地理解时间顺序与周期性。

1 相关工作

1.1 深度学习方法

在光伏功率预测领域,深度学习模型的应用日益广泛。基于 RNN 开发的长短时记忆网络(long short-term memory, LSTM)^[15],优化了递归神经网络存在的梯度消失和梯度爆炸问题。Kumari等^[16]使用 LSTM-CNN(convolutional neural network)模型提取时空特征,结果表明,模型在不同气候、季节和天空条件下具有良好的预测精度。文献^[17]提出了一种基于二次分解方法和改进粒子群算法的 CNN-BiLSTM(CNN-bidirectional LSTM)光伏功

率预测模型,采用自适应噪声完备集合经验模态分解对关键气象因素和功率进行二次分解,充分挖掘时序信息并降低数据的不稳定性。但由于RNN模型参数共享和多次连乘的特性,在处理长序列时难以捕捉长期依赖关系。

基于注意力的模型由于其显著的性能而在时间序列预测中越来越受欢迎^[18]。Informer^[19]通过改进自注意力机制,在大幅减少内存占用与计算成本的前提下,强化了对长序列数据中关键时序模式的提取能力。Autoformer^[20]将时序数据分解为趋势项与周期项分别建模,显著提升了预测结果的稳定性与准确性。PatchTST(patch time series Transformer)^[21]利用了计算机视觉中基于补丁的技术,将单点的词元(Token)变为子序列级别的Patch Token,被证明用于时序预测是有效的。然而自注意力机制在建模长序列时易捕捉过多噪声而导致过拟合,导致难以实现准确预测。

1.2 大语言模型方法

LLM具有强大的表征能力及可迁移性,预训练LLM在时序预测任务中已展现出良好的跨模态适配潜力与实用价值。目前针对时间序列预测的LLM研究可分为以下3类:

1)通过提示设置实现预训练的LLM直接进行预测^[22]。受聊天机器人的启发,Time-LLM^[23]开发了一个重编程预测框架,将LLM用于时间序列数据,其中输入用文本原型重新编程,以使源数据和自然语言的模态保持一致。

2)通过将来自不同领域的时间序列数据聚合到一个大模型中,创建一个适用于广泛场景的模型。TimeGPT-1^[24]基于Transformer的架构开发,该模型使用的数据集包含了来自金融、经济、天气、能源、交通和银行业务等多个领域的超过1000亿个数据点。这使得TimeGPT-1能够学习到不同的时间模式,从而更好进行泛化。Rasul等^[25]提出的Lag-Llama,在Monash时序仓库等多个数据集上进行了训练,并在未见过的数据集上进行了零样本测试,实验结果优于基线模型。

3)第3种侧重于对LLM进行微调。LLM4TS^[26]提出了一种多级微调方法,首先对骨干网进行微调,使时间序列和文本数据对齐,然后对下游预测层进行训练。BERT(bidirectional encoder representation from Transformers)4ST^[27]利用了BERT模型的双向注意力网络,通过重新设计时空编码来捕捉风电数据的时空依赖性,相较主流模型具有出色的性能。Wu等^[28]提出一种时空增强预训练大语言模型(spatio-temporal enhanced pre-trained

large language model, STELLM),为增强模型的时空建模能力,引入了时空提示,之后利用自回归微调策略将LLM与时间序列数据对齐,实验显示STELLM显著优于基线方法。ChatTS^[29]是首个以多元时间序列为输入进行理解和推理的时间序列大型语言模型,该模型仅在合成数据集上进行微调,模型在对齐与推理任务上具有突出优势。

2 本文方法

本文模型整体结构如图1所示,提出了一种高适应性且高效的三阶段微调策略,先将LLM与时间序列数据对齐,然后使用初始线性探测训练一个新的输出层,最后对模型进行全微调。考虑到LLM输入编码层并不适配于时间序列数据且无法对时间信息建模,构建了一种嵌入时间信息的输入编码层,以增强LLM对时间信息的建模能力。此外,设计了一种光伏数据切片标记方式,解决单个时间点携带信息有限,且难以反映光伏数据整体特征与规律的问题。

2.1 相关性分析

准确选取与光伏功率紧密相关的关键变量是构建高效预测模型的基础。光伏功率受多种因素影响,包括太阳辐射强度、温度、湿度、降水、风速等,因此需要选取对光伏功率影响程度较大的因素参与预测,从而提高预测的准确性。由于各影响因素与光伏功率之间可能存在非线性关系,传统的Pearson相关系数难以准确衡量此类关系。因此,本研究采用Spearman相关系数进行变量间相关性分析。Spearman相关系数通过计算变量秩次的线性相关性,能够有效捕捉变量间单调变化的趋势,无论这种变化是线性还是非线性,从而为关键变量的筛选提供可靠依据。其计算公式为

$$\rho_s = 1 - \frac{6 \sum_{i=1}^n d_i^2}{n(n^2 - 1)} \quad (1)$$

式中: n 为样本数量, d_i 表示第 i 个样本中两个变量秩次的差值。 ρ 取值范围为 $[-1,1]$,其绝对值越接近1,表明变量间的相关性越强。

对DKASC数据集,计算了光伏功率与其他变量的Spearman相关系数,并绘制热力图如图2所示。根据图中计算结果,可以看出光伏功率与总倾斜辐射、总水平辐射、漫射倾斜辐射、漫射水平辐射、温度、湿度、风向相关性最高,分别为0.97、0.92、0.80、0.75、0.75、-0.96、-0.99,因此选择功率与以上7个参数参与预测。对SRRL数据集采取同样的方式进行处理。

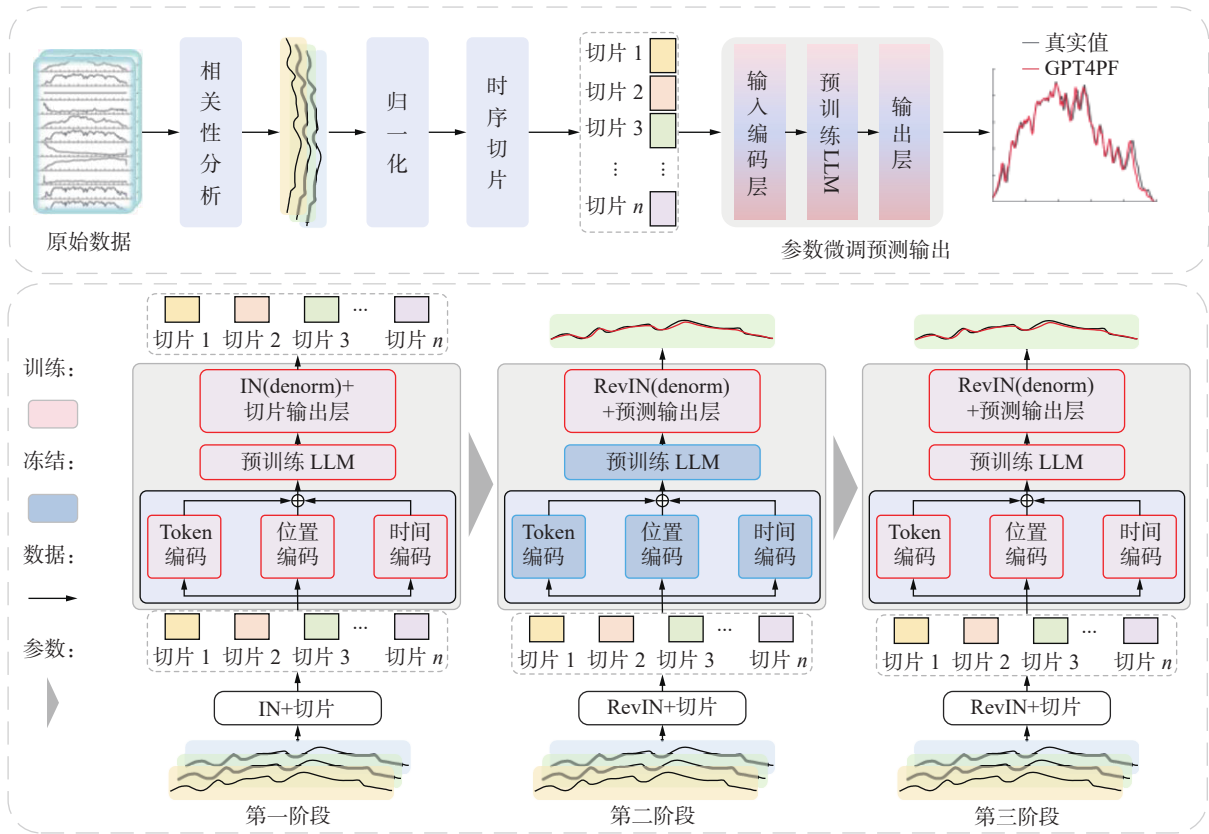


图 1 GPT4PF 总体结构

Fig. 1 Overall structure of GPT4PF

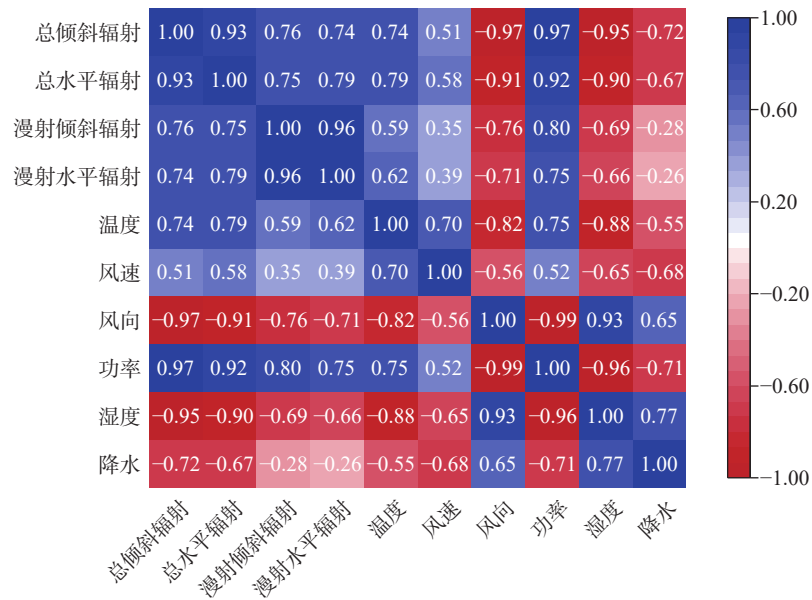


图 2 Spearman 相关系数热图

Fig. 2 Spearman correlation coefficient heatmap

2.2 主干模型选取

本文选择 GPT2 作为主干模型, GPT2 是基于 Transformer 预训练的语言模型, 不同于早期依赖循环神经网络或浅层自注意力机制的模型, GPT2 通过自回归机制进行大规模无监督学习。区别于 BERT 的双向上下文建模, GPT2 采用单向

的自注意力机制, 逐层生成文本序列, 这使得它在文本生成任务中表现出更强的连贯性和创造性。GPT2 模型主干由堆叠的 Transformer 解码器层构成, 每层包含掩码自注意力模块和前馈神经网络, 模型的输入通过标记嵌入和位置嵌入的叠加实现文本表示, 位置嵌入采用可学习参数捕捉

序列顺序关系。在预训练阶段, GPT2 利用掩码自注意力机制, 确保每个位置的预测仅依赖于左侧历史信息, 从而严格遵循文本生成的因果依赖关系。在训练完成后, GPT2 在预训练权重基础上采用微调策略, 在模型顶部添加特定任务输出层, 并使用标注数据对全模型参数进行有监督训练, 从而让 GPT2 适配下游任务。

2.3 时间序列切片

与 NLP 不同, 在时间序列预测中单个时间步长不像一个单词一样具有语义信息, 在传统的时间序列处理中, 单个时间点所携带的信息量有限, 难以充分反映整体数据特征与规律。

对于输入长度为 L 的单变量时间序列, 设切片长度为 P , 步长为 S , 经切片操作后, 输入切片的数量从 L 减少到 N , 其中 $N = \lfloor (L - P) / S \rfloor + 2$, 由于注意力机制的内存使用和计算复杂度与输入数量呈二次关系, 切片数量的大幅减少使计算开销显著降低。通道独立是针对多变量时间序列任务提出的一种处理方式, 每个变量即一个通道, 通道独立将多变量序列拆分成多个单变量序列, 每个序列独立进入模型进行训练, 并共享权重。在多通道数据中, 不同通道之间可能存在冗余信息, 通道独立可以避免通道间的干扰, 减少模型对冗余特征的依赖, 从而降低过拟合风险。对于一个长度为 L , C 个通道的多变量时间序列样本 $\mathbf{x} = (x_1, x_2, \dots, x_L) \in \mathbb{R}^{C \times L}$, 模型将输入分为 C 个单变量序列 $\mathbf{x}^i \in \mathbb{R}^{(1 \times L)}, i = 1, 2, \dots, C$, 再对每个单变量序列 $\mathbf{x}^i$ 应用切片操作, 生成切片序列 $\mathbf{s} \in \mathbb{R}^{C \times N \times P}$ 。

2.4 输入编码层

适当的编码方式对于将时间序列数据输入到 GPT2 中至关重要, 预训练 GPT2 的输入编码层并不适配于时间序列数据且无法对时间信息建模, 为此本文重新设计了模型的输入层, 同时嵌入时间信息, 以增强模型对时间信息的建模能力。在 GPT2 模型中, 通过使用两个可训练的权重表词嵌入表与位置编码表, 将每个输入词元映射到高维空间, 然而这种方式仅适用于标量数据。因此本文不使用 GPT2 中的词元嵌入层, 而采用一维卷积层代替, 卷积层能够更好得提取时间序列数据中局部数据间的关联信息, 即:

$$\mathbf{e}_{\text{segment}} = \text{Conv1d}(\mathbf{s}) \quad (2)$$

此外本文还重新设计了位置编码层, 它用于标记同一批次输入中各个切片的顺序, 帮助模型捕捉时间序列中的顺序信息和长期依赖关系。切片处理改变了数据的顺序和结构, 本文通过一个可训练的查找表来映射每个切片的位置信息, 即:

$$\mathbf{e}_{\text{position}} = \text{Embed}(n, d_{\text{model}}) \quad (3)$$

式中: n 表示切片的序数, d_{model} 表示嵌入维度, Embed 为映射函数。

由于 GPT2 缺乏对时间信息显式建模的能力, 为此本文设计了一个时间编码层, 如图 3 所示, 将单个时间戳的月、日、时、分 4 个分量分别映射到与模型输入维度 d_{model} 相同的维度, 再对映射后的 4 个向量求和得到最终的时间编码, 即:

$$\mathbf{e}_{\text{temporal}} = \text{Embed}(\text{month}, d_{\text{model}}) + \text{Embed}(\text{day}, d_{\text{model}}) + \text{Embed}(\text{hour}, d_{\text{model}}) + \text{Embed}(\text{min}, d_{\text{model}}) \quad (4)$$

最后将这 3 种编码相加, 作为骨干模型的输入:

$$\mathbf{e}_{\text{input}} = \mathbf{e}_{\text{segment}} + \mathbf{e}_{\text{position}} + \mathbf{e}_{\text{temporal}} \quad (5)$$

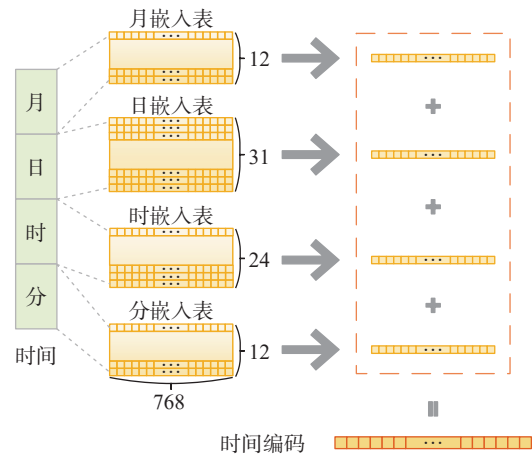


图 3 时间嵌入
Fig. 3 Time embedding

2.5 参数高效微调

大语言模型在海量的数据中学习了丰富的知识, 但对于预训练数据涉及较少的领域, 大模型的表现并不理想。参数高效微调旨在减少微调过程中需要更新的参数数量和计算开销, 从而提高微调大语言模型的效率。参数选择方法选择性的对原始模型中的部分参数进行微调, 在垂直领域的任务中, 模型中通常只有部分参数起决定性作用, 通过微调这些参数, 可以在减少计算开销的同时提升模型性能。低秩适配方法 (low-rank adaptation, LoRA) 通过引入低秩矩阵来替代传统的全参数微调。在传统方法中为适配下游任务, 对参数矩阵 $\mathbf{W}_0 \in \mathbb{R}^{d \times k}$ 通过更新矩阵 $\Delta \mathbf{W}$ 进行更新, 即 $\mathbf{W} = \mathbf{W}_0 + \Delta \mathbf{W}$, 这需要占用大量的 GPU 内存。如图 4 所示, LoRA 将 $\Delta \mathbf{W}$ 分解为两个低参数量的矩阵 $\mathbf{B} \in \mathbb{R}^{d \times r}$ 和 $\mathbf{A} \in \mathbb{R}^{r \times k}$, 使更新过程变为 $\mathbf{W} = \mathbf{W}_0 + \alpha \mathbf{B} \mathbf{A}$, α 是缩放因子, 用于控制 LoRA 权重的大小。这种方法将原始的参数矩阵低秩分解为两个小矩阵, 在微调时冻结原始矩阵仅训练这两个小矩阵, 从而大幅减少计算成本。

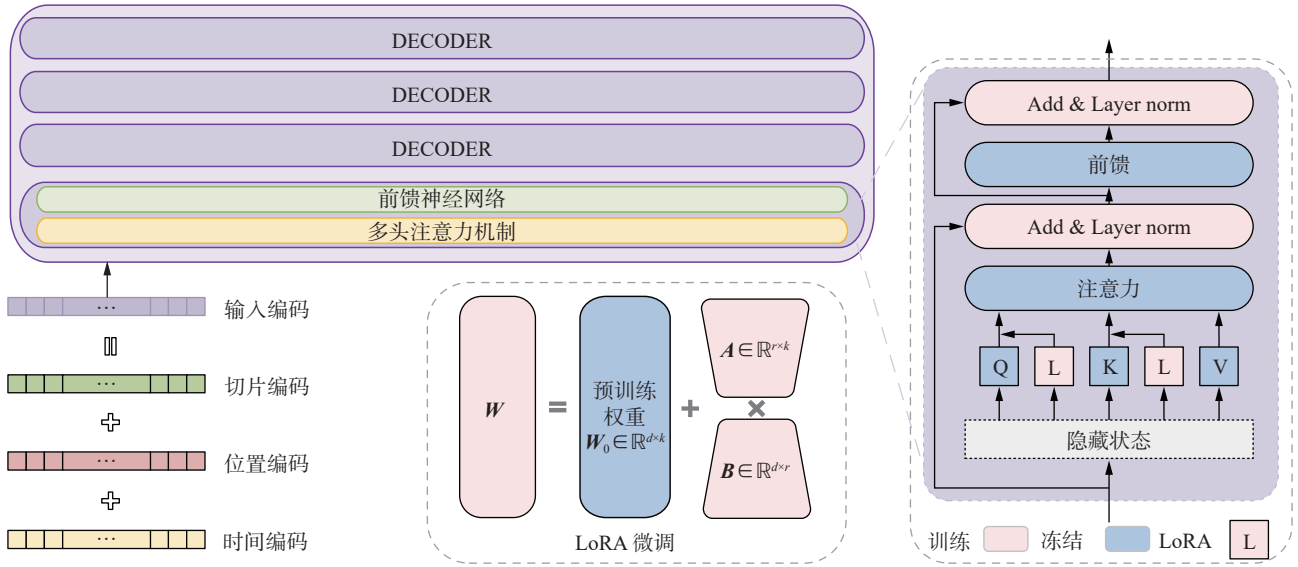


图4 模型输入与微调

Fig. 4 Model input and fine-tuning

如图4所示,对GPT2中解码模块的Query矩阵、Key矩阵以及全连接前馈层进行训练。将LoRA应用于对自注意力机制中的Query和Key矩阵的更新,通过使用两个低秩矩阵的乘积来更新原始的权重矩阵,从而在降低显存占用的同时显著提高训练效率。GPT4PF采用GPT2的前4层作为主干模型,模型的总参数量约为68 MB,通过参数选择与低秩适配,微调训练参数仅占总参数量的0.39%,降低了微调中需要更新的参数量和计算开销。

2.6 三阶段微调

2.6.1 时间序列对齐阶段

时间序列对齐阶段是将时间序列数据与LLM内部表征空间进行语义和结构匹配的关键环节。由于GPT2主干模型预训练于语言数据,对时间序列数据适应性较差,此阶段需通过自回归训练实现GPT2与光伏数据特征的对齐,使其有效捕捉时序动态特征与逻辑关系。基于2.2节的时间序列切片方法,将光伏数据切分为固定长度的时序切片序列,自回归训练的预测目标设为根据前 $N-1$ 个时序切片,预测第 N 个时序切片,损失函数采用均方误差。此外在时间序列对齐阶段采用实例归一化(instance normalization, IN)对输入数据进行处理,帮助模型更稳定地学习数据分布,从而加速训练过程并提高模型的性能。

2.6.2 输出层微调阶段

在这一阶段对模型的结构进行了调整,采用了新的归一化层——可逆实例归一化(RevIN)。RevIN是由Kim等^[30]为时间序列预测专门设计的一种归一化方法,它通过可逆的归一化与反归

一化操作,解决时间序列数据中分布随时间变化的核心问题,显著提升预测模型的鲁棒性。在归一化阶段,对输入的时间序列数据 x ,RevIN像实例归一化一样,对每个样本的每个通道单独计算均值 μ 和标准差 σ ,然后,对输入数据进行归一化处理,使其均值为0、方差为1,即:

$$\text{RevIN}_{\text{norm}}(x) = \frac{x - \mu}{\sigma + \varepsilon} \quad (6)$$

其中 ε 是一个很小的常数,用于防止除以零。在反归一化阶段,模型输出预测结果后,RevIN利用之前计算的均值 μ 和标准差 σ ,对预测结果进行反归一化操作,将其恢复到原始数据的尺度,即:

$$\text{RevIN}_{\text{denorm}}(y) = y \times \sigma + \mu \quad (7)$$

其中 y 是模型的预测输出。

先进行线性探测再进行全微调(linear probing followed by full fine-tuning, LP-FT)是一种经典的微调策略,特别适用于目标数据集较小或计算资源有限的情况。Tomihari等^[31]的研究证明了这种方法的有效性。LP-FT先冻结预训练模型中除最后一个线性层之外的所有参数,仅对最后一个线性层进行训练,这个过程在不改变预训练模型主体特征的基础上,找到一个适合当前下游任务的线性变换,初步探索出如何利用预训练特征来解决下游任务。之后在完成线性探测的基础上,对整个模型进行全微调。

因此,在输出层微调阶段本文结合RevIN与线性探测,冻结其余全部参数,仅对模型的输出层进行训练。

2.6.3 全微调阶段

在时间序列对齐阶段与输出层微调阶段完成

之后,模型对光伏预测任务已初步适配,在此基础上进行全微调以实现模型性能的深度优化。在全微调阶段,参与微调的可训练参数全部解冻,模型针对光伏预测任务进行全面优化。

3 实验仿真及结果分析

3.1 数据集选取

本研究选择了两个数据集进行实验分析,数据集一选取澳大利亚沙漠知识太阳能中心(The Desert Knowledge Australia Solar Centre, DKASC)爱丽丝泉的实测数据,数据集采集时间为2015年9月—2016年9月,采集时间间隔为5 min;数据集二选取美国太阳能辐射研究实验室(NREL Solar Radiation Research Laborator, SRRL)的实测数据,数据采集时间为2024年1月—2024年12月,采集时间间隔为5 min。数据均涵盖太阳辐照强度、温度、湿度、风速等气象参数,为光伏功率预测提供了丰富的信息基础。为保证预测的准确性,对数据集中的缺失值采用线性插值法进行填充,对超出光伏数据正常物理范围的值同样通过线性插值法替换,以保证数据的连续性。数据在输入前均经归一化处理。

3.2 实验设置

为验证所提方法的可行性与有效性,在两个具有代表性的数据集DKASC与SRRL上,设计并实施了对比试验与消融实验。采用平均绝对误差(mean absolute error, MAE)、均方根误差(root mean square error, RMSE)来评估模型的性能,评估指标定义为

$$E_{MA} = \frac{1}{n} \sum_{i=1}^n |\hat{y}_i - y_i| \quad (8)$$

$$E_{RMS} = \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2} \quad (9)$$

其中: n 表示光伏功率样本数量, y_i 表示第 i 个光伏功率的真实值, $\hat{y}_i$ 表示第 i 个光伏功率的预测值。

本文选择GPT2作为主干模型,GPT2由12层Transformer解码器堆叠而成,模型维度为768,在实验中仅使用前4层作为主干模型。实验中将窗口长度设置为256,切片大小为24,预测步长为12,隐藏状态的维度同样设置为768,与GPT2的输入维度保持一致。两个数据集均按照7:2:1的比例分为训练集、验证集和测试集。实验环境为Ubuntu22.04LTS、PyTorch、Intel Core i5-13600kf、Nvidia 4070。使用Adam优化器对网络进行训练,MSELoss作为损失函数。学习率设置为0.0001,批大小设置为64。学习率随训练轮数动态下降,并引入早停机制,当超过30轮网络的损失不再下降时停止训练,以避免过度拟合。

3.3 对比实验

为验证本文模型的性能,将GPT4PF与多种模型对比,包括时间序列模型与基于LLM的时序模型。其中LSTM是经典的时间序列预测模型。Informer、Autoformer和PatchTST均为基于Transformer模型的时序预测模型。与Transformer模型不同,DLinear是一种通过序列分解和线性映射实现的单层线性模型。这些模型在时序预测任务上均取得了较好的效果。此外,还将GPT4PF与基于LLM的时序模型进行对比,包括GPT4TS与LLM4TS,它们均是基于GPT2的时间序列模型。

3.3.1 与经典时间序列预测模型的对比实验

与经典时间序列预测模型的对比结果如表1所示,可以看出本文所提方法相较主流模型在两个数据集上均取得最优结果,证明了本文所提方法的有效性。在DKASC数据集上,本文模型相较Informer与Autoformer,在1 h与8 h预测长度上,MAE均下降超30%,说明其对噪声的泛化能力较弱。得益于Patch机制,PatchTST在降低计算量的同时保留全局依赖,从而表现出优异性能。DLinear凭借单层线性结构规避了复杂注意力计算带来的过拟合风险,但在2 h以上预测长度中表现不佳,说明简单架构难以捕捉非线性动态特征。

表 1 与经典时间序列预测模型的对比实验结果

Table 1 Comparison experiment results with classical time series prediction models

数据集	方法	1 h		2 h		4 h		8 h	
		MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE
DKASC	LSTM	<u>0.303</u>	1.834	0.490	3.042	0.650	3.514	0.758	4.001
	Informer	0.413	2.103	0.515	3.260	0.650	3.497	0.749	3.912
	Autoformer	0.418	2.207	0.540	3.277	0.660	3.679	0.758	3.892
	DLinear	0.356	1.841	0.489	3.001	0.642	3.506	0.723	3.914
	PatchTST	0.304	<u>1.829</u>	<u>0.429</u>	<u>2.679</u>	<u>0.431</u>	<u>2.703</u>	<u>0.514</u>	<u>3.271</u>
	GPT4PF	0.291	1.716	0.402	2.429	0.429	2.676	0.487	3.150

续表 1

数据集	方法	1 h		2 h		4 h		8 h	
		MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE
SRRL	LSTM	0.210	0.377	0.321	0.520	0.461	0.623	0.546	0.721
	Informer	0.257	0.479	0.359	0.622	0.460	<u>0.607</u>	0.549	0.729
	Autoformer	0.260	0.489	0.361	0.618	0.467	0.613	0.549	0.730
	DLiner	0.227	0.440	0.329	0.583	0.449	0.615	0.534	0.714
	PatchTST	<u>0.202</u>	<u>0.370</u>	<u>0.311</u>	<u>0.512</u>	<u>0.410</u>	0.621	<u>0.487</u>	<u>0.694</u>
	GPT4PF	0.198	0.369	0.294	0.491	0.373	0.570	0.485	0.690

注: 加粗表示结果在该列最好, 下划线表示结果在该列次好。

为验证性能差异的统计可靠性, 本文选取性能最接近的 PatchTST 模型, 针对 1 h 与 8 h 预测长度, 选择 DKASC 数据集, 在固定条件下完成 10 次独立重复实验, 经独立样本 t 检验: 1 h 预测任务

中 MAE 差异 $p=0.0188 (<0.05)$, 8 h 任务中 $p<0.0001$, 表明 GPT4PF 的性能优势具有统计显著性。

3.3.2 与 LLM 时间序列预测模型的对比实验

与基于 LLM 时序模型对比结果如表 2 所示。

表 2 与基于 LLM 时间序列预测模型的对比实验结果
Table 2 Comparison experiment results with LLM-based time series prediction models

数据集	方法	1 h		2 h		4 h		8 h	
		MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE
DKASC	GPT4TS	0.373	2.197	0.507	3.312	0.565	3.217	0.701	3.839
	LLM4TS	0.309	1.788	0.434	2.574	0.459	2.900	0.507	3.339
	GPT4PF	0.291	1.716	0.402	2.429	0.429	2.676	0.487	3.150
SRRL	GPT4TS	0.263	0.499	0.413	0.629	0.620	0.892	0.842	1.157
	LLM4TS	0.212	0.394	0.311	0.533	0.384	0.621	0.512	0.725
	GPT4PF	0.198	0.369	0.294	0.491	0.373	0.570	0.485	0.694

注: 加粗表示结果在该列最好。

由表 2 可以看出本文所提方法在两个数据集上均取得最优结果。GPT4TS 首次将 LLM 应用到时间序列预测领域, 表现出较强的性能, 但未针对光伏数据周期性、数据分布等设计中间优化环节。得益于两阶段微调方法, LLM4TS 预测精度相较 GPT4TS 显著提升, 但仍未解决通用 LLM 到光伏任务的适配断层问题。GPT4PF 采用三阶段微调逐步适配光伏任务并结合参数高效微调技术进行训练, 此外本文模型设计了新的输入编码层, 解决对时间信息显式建模的问题, 增强模型在较长时序预测中的准确度。训练效率方面, 由于采用多阶段微调策略且新增模块引入了额外参数, GPT4PF 的训练时长较 GPT4TS 大幅增加, 较 LLM4TS 增加约 19%; 推理速度上, GPT4P 较 GPT4TS 提升约 9%, 与 LLM4TS 基本持平, 表明新增模块未对推理效率造成显著影响, 模型在保证预测精度优势的同时, 仍维持了较好的推理高效性。所提模型在两个数据集的各预测长度均展现出性能优势, 其 MAE 与 RMSE 较 LLM4TS 普遍降低 3%~8%。为

验证性能差异的统计可靠性, 本文选取性能最接近的 LLM4TS, 针对 1 h 与 8 h 预测长度, 选择 SRRL 数据集, 在固定条件下完成 10 次独立重复实验, 经独立样本 t 检验: 1 h 预测任务中 MAE 差异 $p=0.0001 (<0.05)$, 8 h 任务中 $p<0.0001$, 表明 GPT4PF 的性能优势具有统计显著性。在 DKASC 和 SRRL 上的可视化实验结果如图 5~8 所示。

从图 5~8 可以看出, 在 1 h 预测长度上基于 LLM 的时间序列预测模型均实现了较好的预测结果, 在 8 h 预测长度上, LLM4TS 与 GPT4PF 均有较高的预测精度, 二者性能优势印证了多阶段微调策略的有效性。从图 6 和图 8 可以看出, 在 8 h 预测长度中, 除 PatchTST 外的经典时序模型均有较大偏移, 反观本文模型 GPT4PF 预测结果均与真实值高度贴合, 这得益于时间编码与时序切片技术, 从而在长时预测任务中保持稳定精度。

为验证所提方法的有效性, 本文设计并开展了两个消融实验。首先验证多阶段微调带来的性能提升, 之后对时间编码的有效性进行验证。

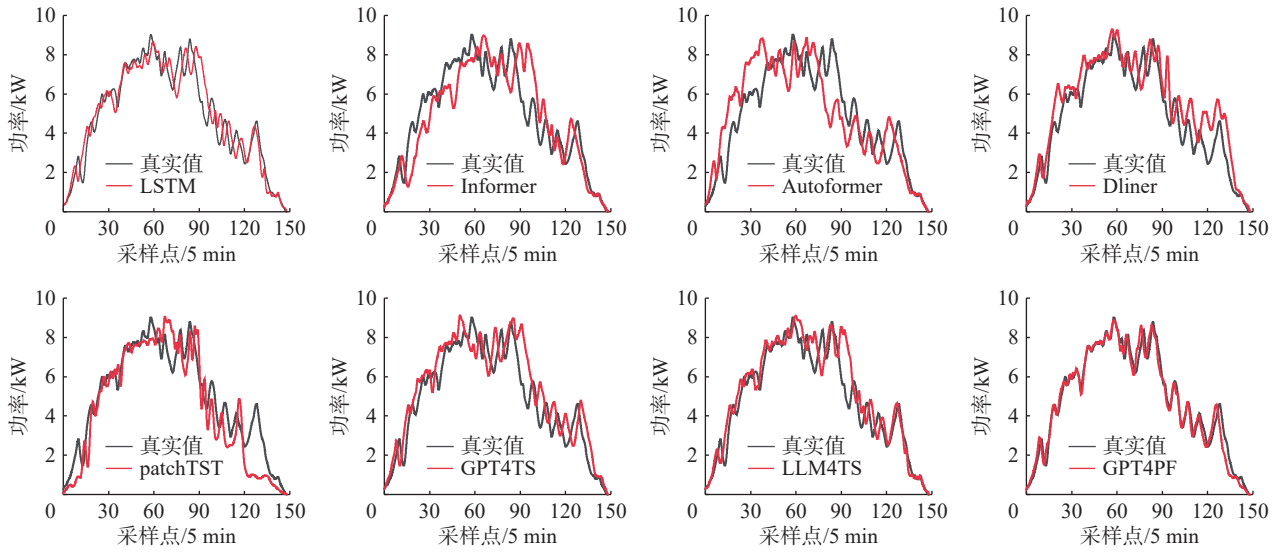


图 5 DKASC 数据集上 1 小时预测长度结果可视化

Fig. 5 Visualization of 1-hour prediction results on the DKASC dataset

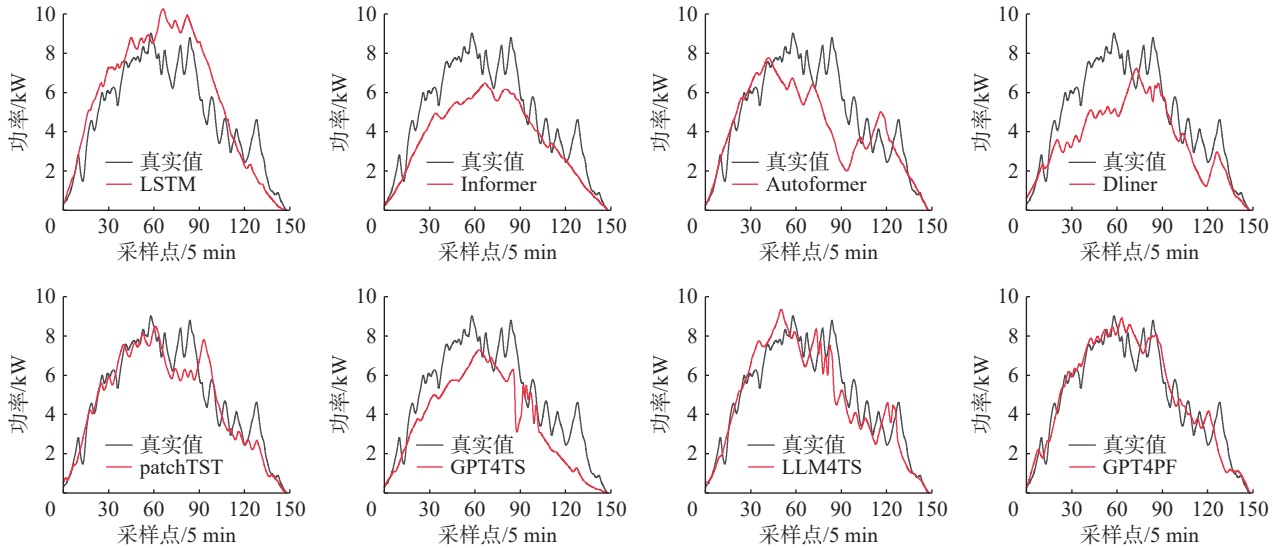


图 6 DKASC 数据集上 8 小时预测长度结果可视化

Fig. 6 Visualization of 8-hour prediction results on the DKASC dataset

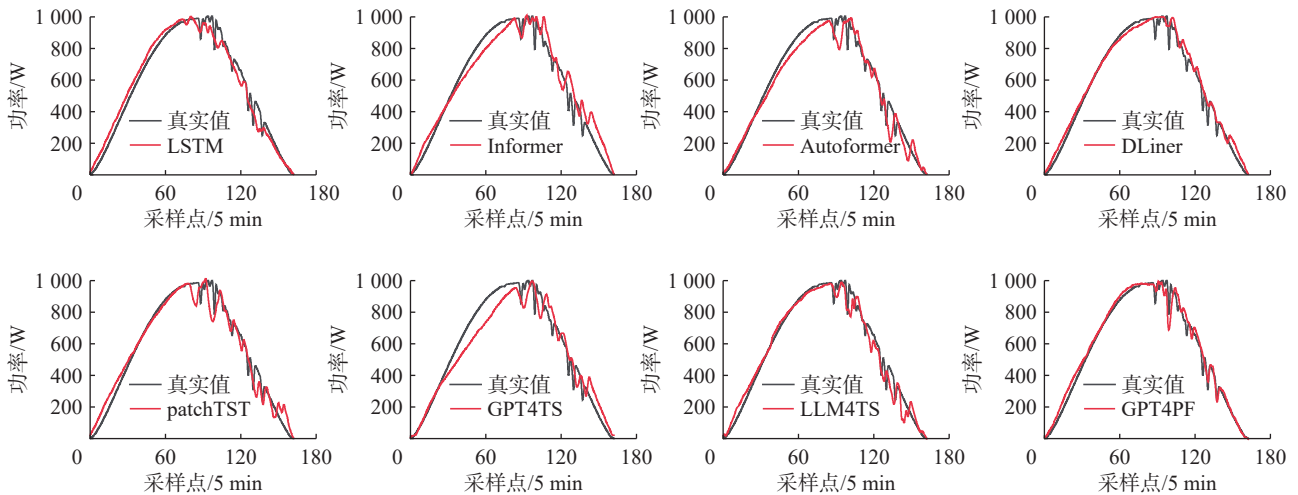


图 7 SRRL 数据集上 1 小时预测长度结果可视化

Fig. 7 Visualization of 1-hour prediction results on the SRRL dataset

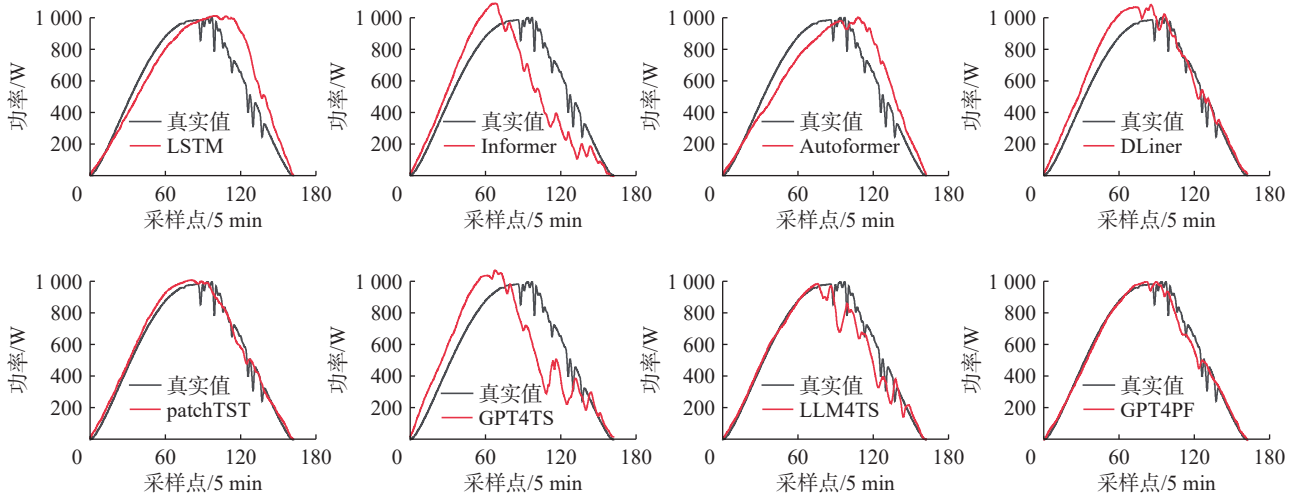


图8 SRRL数据集上经典时序模型8小时预测长度结果可视化

Fig. 8 Visualization of 8-hour prediction results on the SRRL dataset

3.4 消融实验

3.4.1 多阶段微调消融实验

GPT4PF 采用三阶段微调方法进行训练, 模型第1阶段进行自监督训练将时间序列与 LLM 对齐, 第2阶段采用可逆实例归一化并通过线性探测训练新的输出层, 第3阶段进行全微调。将仅经过第1阶段微调后的模型命名为 GPT4PF-1s,

将经过前两阶段微调后的模型命名为 GPT4PF-2s。由表3可知, 在两个数据集的各预测长度下, GPT4PF-2s 的 MAE 与 RMSE 均显著低于 GPT4PF-1s, 而 GPT4PF 的指标较 GPT4PF-2s 进一步降低 1%~2%, 这一阶梯式性能提升表明, 多阶段微调的递进式优化, 增强了模型对光伏预测任务的适应性, 为模型的性能提升提供了关键支撑。

表3 多阶段微调消融实验结果

Table 3 Ablation study results of multi-stage fine-tuning

数据集	方法	1 h		2 h		4 h		8 h	
		MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE
DKASC	GPT4PF-1s	0.301	1.767	0.407	2.491	0.441	2.686	0.497	3.211
	GPT4PF-2s	0.296	1.750	0.403	2.479	0.433	2.681	0.488	3.174
	GPT4PF	0.291	1.716	0.402	2.429	0.429	2.676	0.487	3.150
SRRL	GPT4PF-1s	0.202	0.378	0.301	0.501	0.376	0.586	0.491	0.706
	GPT4PF-2s	0.198	0.374	0.298	0.501	0.375	0.581	0.488	0.703
	GPT4PF	0.198	0.369	0.294	0.491	0.373	0.570	0.485	0.694

注: 加粗表示结果在该列最好。

3.4.2 时间编码消融实验

考虑到 LLM 缺乏对光伏数据中时间信息的显式建模能力, 本文重新设计了模型的输入编码层, 将每个切片对应的时间映射到与模型输入相同的维度, 得到时间编码, 将时间编码、切片编码和位置编码求和得到模型的输入。由表4可知, 在两个数据集的各预测长度下, 引入时间编码的模型其 MAE 与 RMSE 均低于未采用时间编码的版本, 这一结果表明, 所提时间编码能够有效弥补 LLM 对光伏数据时序规律的捕捉能力, 减少因时间信息缺失导致的预测偏差, 验证了时间编码设计的合理性与有效性。

3.5 不同太阳辐射条件下的模型鲁棒性验证实验

为验证模型在不同辐射条件下的泛化能力与鲁棒性, 本节在 DKASC 数据集上设置对比实验, 将数据集按光伏辐射强度划分为强辐射、中辐射、弱辐射三组场景, 预测长度为 1 h 与 8 h, 以 MAE 与 RMSE 为评价指标, 将本文模型与 PatchTST、LLM4TS 进行跨场景性能对比, 实验结果如表5所示。预测长度为 1 h, 在强辐射与中辐射条件下, 三者性能差距较小, GPT4PF 的 MAE 较 PatchTST 与 LLM4TS 降低 1%~3%; 在弱辐射条件下, GPT4PF 的优势显著扩大, 其 MAE 较 PatchTST、LLM4TS 降低超 5%。预测长度为 8 h, 在中辐射与弱

辐射场景下,基于LLM的模型表现优于PatchTST, GPT4PF的MAE较PatchTST降低超5%,较LLM4TS降低2%~4%。这一结果表明:对比模型在辐射强度波动时易出现性能衰减,而GPT4PF

得益于时间编码对辐射周期性特征的精准捕捉,以及多阶段微调对复杂数据分布的适配能力,在不同辐射条件下均保持稳定预测精度,充分证明了模型在光伏辐射场景变化下的泛化能力与鲁棒性。

表4 时间编码消融实验结果
Table 4 Ablation study results of time embedding

数据集	方法	1 h		2 h		4 h		8 h	
		MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE
DKASC	无时间编码	0.293	1.729	0.404	2.444	0.432	2.693	0.492	3.196
	GPT4PF	0.291	1.716	0.402	2.429	0.429	2.676	0.487	3.150
SRRL	无时间编码	0.200	0.372	0.296	0.496	0.377	0.578	0.490	0.703
	GPT4PF	0.198	0.369	0.294	0.491	0.373	0.570	0.485	0.694

注:加粗表示结果在该列最好。

表5 不同辐射条件下实验结果
Table 5 Experimental results under different radiation conditions

数据集	方法	1 h		8 h	
		MAE	RMSE	MAE	RMSE
强辐射	PatchTST	0.220	1.494	0.392	2.797
	LLM4TS	0.221	1.472	0.391	2.824
	GPT4PF	0.210	1.407	0.380	2.709
中辐射	PatchTST	0.359	1.981	0.590	3.502
	LLM4TS	0.362	2.006	0.585	3.453
	GPT4PF	0.355	1.922	0.560	3.361
弱辐射	PatchTST	0.413	2.399	0.678	4.236
	LLM4TS	0.416	2.409	0.653	4.112
	GPT4PF	0.393	2.265	0.639	4.023

注:加粗表示结果在该列最好。

4 结束语

本文提出了一种光伏功率预测模型GPT4PF。针对单个时间点携带信息有限,且难以反映光伏数据整体特征与规律的问题,设计了时间序列切片方法;为增强LLM对时间信息的建模能力,构建了嵌入时间信息的输入编码层;为提高模型的适应性与预测精度,提出了一种高适应性且高效的三阶段微调策略。对比实验结果表明,本文模型在两个数据集上的预测精度具有明显优势。消融实验证明本文所提的三阶段微调方法与时间编码的有效性。

但改进的模型仍有不足,当前使用GPT2作为主干模型,在处理超长时间序列时计算开销较高,且对极端天气下的突发数据鲁棒性有待提升;其次时间编码层虽有效补充时序建模能力,但未与气象数据进行动态关联优化,未能充分挖掘“时间-气象”协同特征对预测的增益。未来将探索轻量化主干模型,在降低计算成本的同时,构建“时间-

气象”融合编码机制,进一步提升模型泛化能力。

参考文献:

- [1] 国家能源局. 国家能源局发布 2024 年全国电力工业统计数据[EB/OL]. (2025-01-21)[2025-07-08]. <http://www.nea.gov.cn/20250121/097bfd7c1cd3498897639857d86d5dac/c.html>. National Energy Administration. National energy administration releases 2024 national statistical data on the electric power industry[EB/OL]. (2025-01-21)[2025-07-08]. <http://www.nea.gov.cn/20250121/097bfd7c1cd3498897639857d86d5dac/c.html>.
- [2] MURATA A, OHTAKE H, OOZEKI T. Modeling of uncertainty of solar irradiance forecasts on numerical weather predictions with the estimation of multiple confidence intervals[J]. *Renewable energy*, 2018, 117: 193–201.
- [3] HIRATA Y, AIHARA K. Improving time series prediction of solar irradiance after sunrise: comparison among three methods for time series prediction[J]. *Solar energy*, 2017, 149: 294–301.
- [4] AGOUA XG, GIRARD R, KARINIOTAKIS G. Short-term spatio-temporal forecasting of photovoltaic power production[J]. *IEEE transactions sustainable energy*, 2017, 9(2): 538–546.
- [5] 李承皓, 杨永标, 宋嘉启, 等. 基于IMVMD和BiLSTM-SARIMA组合模型的台区光伏短期发电功率预测[J]. *太阳能学报*, 2025, 46(2): 433–440. LI Chenghao, YANG Yongbiao, SONG Jiaqi, et al. Short term photovoltaic power generation prediction based on imvmd and BiLSTM-SARIMA combination model in station area[J]. *Acta energiae solaris sinica*, 2025, 46(2): 433–440.
- [6] 朱琼锋, 李家腾, 乔骥, 等. 人工智能技术在新能源功率预测的应用及展望[J]. *中国电机工程学报*, 2023, 43(8): 3027–3047. ZHU Qiongfeng, LI Jiateng, QIAO Ji, et al. Application and prospect of artificial intelligence technology in renewable energy forecasting[J]. *Proceedings of the CSEE*, 2023, 43(8): 3027–3047.
- [7] 张赞宁, 魏广军. 考虑特征选择的短期光伏功率组合预测模型[J]. *电力系统及其自动化学报*, 2024, 36(8): 122–132. ZHANG Yunning, WEI Guangjun. Combined prediction model for short-term photovoltaic power considering feature selection[J]. *Proceedings of the CSU-EPSA*, 2024, 36(8): 122–132.

- [8] 马世龙, 乌尼日其格, 李小平. 大数据与深度学习综述[J]. 智能系统学报, 2016, 11(6): 728–742.
MA Shilong, WUNIRI Qiqige, LI Xiaoping. Deep learning with big data: state of the art and development[J]. *CAAI transactions on intelligent systems*, 2016, 11(6): 728–742.
- [9] LI Gangqiang, WANG Huaizhi, ZHANG Shengli, et al. Recurrent neural networks based photovoltaic power forecasting approach[J]. *Energies*, 2019, 12(13): 2538.
- [10] 姚蕊, 刘小芳. 线性分解和周期增强 Informer 的太阳辐射短临预报研究[J]. 太阳能学报, 2025, 46(2): 505–510.
YAO Rui, LIU Xiaofang. Linera decomposition and periodic enhancement in short-term solar irradiance forecasting with informer[J]. *Acta energiae solaris sinica*, 2025, 46(2): 505–510.
- [11] 陈嘉俊, 刘波, 林伟伟, 等. 基于 Transformer 的时间序列预测方法综述[J]. 计算机科学, 2025, 52(6): 96–105.
CHEN Jiajun, LIU Bo, LIN Weiwei, et al. Survey of Transformer-based time series forecasting methods[J]. *Computer science*, 2025, 52(6): 96–105.
- [12] ZHOU Tian, NIU Peisong, WANG Xue, et al. One fits all: Power general timeseries analysis by pretrained lm[J]. *Advances in neural information processing systems*, 2023, 36: 43322–43355.
- [13] SUN Chenxi, LI Hongyan, LI Yaliang, et al. TEST: text prototype aligned embedding to activate LLM's ability for time series [C]//The Twelfth International Conference on Learning Representations. Vienna: ICLR, 2024.
- [14] SHI Xiaoming, WANG Shiyu, NIE Yuqi, et al. Time-MoE: billion-scale time series foundation models with mixture of experts[C]//The Thirteenth International Conference on Learning Representations. Singapore: ICLR, 2025.
- [15] CHAI Minkang, XIA Fei, HAO Shuotao, et al. PV power prediction based on LSTM with adaptive hyperparameter adjustment[J]. *IEEE Access*, 2019, 7: 115473.
- [16] KUMARI P, TOSHNIWAL D. Long short term memory-convolutional neural network based deep hybrid approach for solar irradiance forecastin-g[J]. *Appl energy* 2021: 295.
- [17] 王德文, 焦天媛. 基于二次分解的不同太阳辐射下光伏功率预测[J]. 太阳能学报, 2024, 45(9): 360–368.
WANG Dewen, JIAO Tianyuan. Prediction of photovoltaic power under different solar radiation based on secondary decomposition[J]. *Acta energiae solaris sinica*, 2024, 45(9): 360–368.
- [18] 范航舟, 梅红岩, 赵勤, 等. 融合双注意力机制的 GNN 多维时间序列预测[J]. 智能系统学报, 2024, 19(5): 1277–1286.
FAN Hangzhou, MEI Hongyan, ZHAO Qin, et al. Multivariate time series forecasting with a graph neural network and dual attention mechanism[J]. *CAAI transactions on intelligent systems*, 2024, 19(5): 1277–1286.
- [19] ZHOU Haoyi, ZHANG Shanghang, PENG Jieqi, et al. Informer: beyond efficient transformer for long sequence time-series forecasting[C]//The Thirty-Fifth AAAI Conference on Artificial Intelligence. Online: AAAI, 2021: 11106.
- [20] WU Haixu, XU Jiehui, WANG Jianmin, et al. Auto-former: decomposition transformers with auto-correlation for long-term series forecasting[J]. *Advances in neural information processing systems*, 2021, 34: 22419.
- [21] NIE Yuqi, NGUYEN Nam H, SINTHONG Phanwadee, et al. A time series is worth 64 words: long-term forecasting with transformers[EB/OL]. (2022–11–27)[2025–07–21]. <http://arxiv.org/abs/2211.14730>.
- [22] XUE Hao, SALIM Flora D. Promptcast: a new prompt-based learning paradigm for time series forecasting[J]. *IEEE transactions on knowledge and data engineering*, 2023, 36(11): 6851–6864.
- [23] JIN Ming, WANG Shiyu, MA Lintao, et al. Time-LLM: time series forecasting by reprogramming large language models [C]//The Twelfth International Conference on Learning Representations. Vienna: ICLR, 2024.
- [24] GARZA A, CHALLU C, MERGENTHALER C M. Time-GPT-1[EB/OL]. (2023–10–05)[2025–07–21]. <https://arxiv.org/abs/2310.03589>.
- [25] RASUL K, ASHOK A, WILLIAMS A R, et al. Lag-Llama: towards foundation models for time series forecasting[EB/OL]. (2023–10–12)[2025–07–21]. <https://arxiv.org/abs/2310.08278>.
- [26] CHANG Ching, PENG W C, CHEN T F. LLM4TS: two-stage fine-tuning for time-series forecasting with pre-trained LLMs[EB/OL]. (2023–08–16)[2025–07–21]. <https://arxiv.org/abs/2308.08469v1>.
- [27] LAI Zefeng, WU Tangjie, FEI Xihong, et al. BERT4ST: fine-tuning pre-trained large language model for wind power forecasting[J]. *Energy conversion and management*, 2024, 307: 118331.
- [28] WU Tangjie, LING Qiang. STELLM: spatio-temporal enhanced pre-trained large language model for wind speed forecasting[J]. *Applied energy*, 2024, 375: 124034.
- [29] XIE Zhe, LI Zeyan, HE Xiao, et al. ChatTS: aligning time series with LLMs via synthetic data for enhanced understanding and reasonin[EB/OL]. (2024–12–04)[2025–07–21]. <https://arxiv.org/abs/2412.03104>.
- [30] KIM T, KIM J, TAE Y, et al. Reversible instance normalization for accurate time-series forecasting against distribution shift[C]//The Ninth International Conference on Learning Representations. Online: ICLR, 2021.
- [31] TOMIHARI A, SATO I. Understanding linear probing then fine-tuning language models from ntk perspective [EB/OL]. (2024–05–27)[2025–07–21]. <https://arxiv.org/abs/2405.16747v2>.

作者简介:



王德文, 副教授, 主要研究方向为时间序列预测。主持或参与国家自然科学基金项目 4 项; 获省科技进步奖 3 项; 以第一完成人获得国家专利授权 3 项; 发表学术论文 50 余篇。E-mail: wde@ncepu.edu.cn。



苗庆健, 硕士研究生, 主要研究方向为综合能源系统负荷预测与光伏预测。E-mail: mjgncepu@163.com。



李成浩, 硕士研究生, 主要研究方向为光伏预测。E-mail: patricklee@163.com。