第 52 卷 第 7 期：3075-3085 高电压技术 Vol.52, No.7: 3075-3085
2026 年 7 月 31 日 High Voltage Engineering July 31, 2026
DOI: 10.13336/j.1003-6520.hve.20251101

基于海上风电功率数据的多任务符号序列生成统一建模框架

刘浩锋，于佳豪，王　晶，刘　青，何　敏，秦　亮
（武汉大学电气与自动化学院，武汉 430072）

摘　要：随着海上风电装机规模持续扩大，基于同源数据衍生出的多类任务显著增加，为实现任务间的信息共享与协同优化，该文提出一种多任务符号序列生成框架(symbolic sequence generative framework, SSGF)，以统一建模范式对多类任务进行协同表示与建模。该框架设计了一套通用的符号系统，将任务类型、输入数据与目标输出统一编码为标准化的 token 序列，并引入“下一个 token 预测”的训练机制，将不同任务转化为统一的序列生成问题。基于此，构建了一个 3 层的 decoder-only 模型作为底层模型，具备参数完全共享、推理路径一致的能力，在同一神经网络中可实现多任务协同训练与推理。通过将所提出的方法应用于海上风电功率数据，实现了风电功率预测与异常数据填补两项任务的统一建模，且风电功率预测任务的均方误差较单独训练下降了 3.18%，而异常数据填补任务的填补精度提高了约 1.72%。该方法不仅消除了传统方法中的结构割裂问题，还显著降低了模型设计与部署成本，为构建统一、高效、可扩展的智能感知框架提供了可行路径。

关键词：符号序列生成框架；多任务建模；功率预测；异常填补；海上风电

Unified Modeling Framework for Multi-task Symbolic Sequence Generation Based on Offshore Wind Power Data

LIU Haofeng, YU Jiahao, WANG Jing, LIU Qing, HE Min, QIN Liang
(School of Electrical Engineering and Automation, Wuhan University, Wuhan 430072, China)

Abstract：With the continuous expansion of offshore wind power installed capacity, the number of diverse tasks derived from homogeneous data has significantly increased. To facilitate information sharing and mutual enhancement across tasks, this paper proposes a symbolic sequence generative framework (SSGF) that uses a unified modeling paradigm to represent and model multiple tasks. The framework designs a universal symbolic system, encoding task types, input data, and target outputs into standardized token sequences, and introduces a “next token prediction” training mechanism to transform diverse tasks into a unified sequence generation problem. Based on this, a three-layer decoder-only model is constructed as the underlying architecture, featuring fully shared parameters and consistent inference paths, enabling multi-task collaborative training and inference within the same neural network. By applying the proposed method to offshore wind power data, this paper achieves unified modeling for both wind power prediction and abnormal data imputation tasks,. The mean squared error of the wind power prediction task is decreased by 3.18% compared to training separately, and the imputation accuracy of the abnormal data imputation task is increased by approximately 1.72%. This method can be used to not only eliminate structural fragmentation issues in conventional methods, but also significantly reduce model design and deployment costs, providing a feasible pathway for building a unified, efficient, and scalable intelligent sensing framework.

Key words：symbol sequence generation framework; multi-task modeling; power prediction; anomaly filling; offshore wind power

基金资助项目：国家重点研发计划(海上风电并网系统远程监测与故障诊断技术)(2023YFB2406900)。
Project supported by National Key R&D Program of China (Remote Monitoring and Fault Diagnosis Technology for Offshore Wind Power Grid Connected Systems) (2023YFB2406900).

0 引言

随着海上风电装机容量不断扩大，其在新能源体系中的地位持续提升[1]，系统运行数据日益庞大，对智能化感知与管理提出更高要求。人工智能技术已广泛应用于风电功率预测[2-4]、智能检测[5]、故障诊断等任务，有效提升了自动化水平。然而，随着运维任务不断增多，基于同源数据衍生的预测、监测、修复、评估等任务持续增加，现有方法大多各自建模，导致数据重复使用、任务隔离、模型冗余，难以共享信息、协同优化。面对任务多样、结构相似但模型分散的挑战，亟需构建统一、高效、可扩展的多任务建模框架，以支撑海上风电系统智能化发展。

海上风电系统在实际运行中涉及多种任务，通常需构建风电功率预测、负荷预测[6]、异常检测[7]、数据填补[8]等多种模型来加以实现。然而，这种单一模型单独训练的方法忽视了不同任务在数据来源、时间尺度和输入特征上存在高度相似性和关联性，导致重复建模严重，增加了系统的训练成本和资源消耗；同时，任务之间缺乏知识的有效迁移，限制了系统的整体性能和泛化能力。

为缓解多个独立模型带来的冗余问题，部分研究者开始尝试利用任务之间的关联性，从而实现多任务的协同建模。其中，多任务学习(multi-task learning，MTL)方法成为主流思路之一。例如，Brealy等人提出一种基于多任务学习的风电场功率建模方法，构建了一个层次贝叶斯模型，实现了对不同风机功率的联合预测[9]。Fu 等人提出基于动态图交互神经网络(DGINet)的多任务学习方法，实现了风速和时间变化风切变系数的联合预测，提升了垂直风廓线构建精度[10]。Navarin 等人构建 DNN-MTL 模型，实现了多风场功率爬坡事件的同步预测，提升了模型在高波动场景下的整体表现[11]。然而，这种“共享特征层+多任务输出头”的架构在建模方式上仍存在明显局限。该结构需要人为划分共享模块，不同任务的输入必须完全一致，限制了模型的扩展性，若依然采用独立的训练目标与策略，缺乏统一的训练机制，导致训练流程复杂、协同优化困难，难以充分挖掘潜在的任务间关联，限制了模型整体的表达与泛化能力[12-13]。

随着 DeepSeek、GPT、Claude 等大语言模型(large language models, LLM)取得显著进展，在多种异构任务之间展现出统一建模的能力[14]。这类模型普遍采用基于自回归机制的生成式结构，并通过统一的训练方式，在无需显式区分任务模块的情况下完成翻译、问答、摘要等复杂任务。在同源数据的基础上实现不同任务间的高效协同与参数共享，充分展现跨任务统一建模的强大潜力与广泛适应性。

受到大语言模型统一建模范式的启发，本文尝试将其通用表示与自回归生成机制引入到风电系统的多任务建模中，以应对当前方法存在的结构分裂、训练不一致和扩展性差等问题。现有大语言模型的训练目标和预训练语料主要面向自然语言处理，与时序风电数据在数据形态、任务逻辑和评价指标上完全不同，直接调用难以保证预测结果的可靠性和生成过程的可控性。另一方面，大语言模型参数规模庞大，对计算资源和存储空间有较高要求，而风电系统的应用场景更强调建模效率与部署可行性，在模型推理的响应速度与资源消耗方面均难以满足实际需求。

为此，本文结合大语言模型“提示词+回答”的思想，构建了一个在设定符号体系下的人工语言模型，并提出一种符号序列生成框架(symbolic sequence generative framework, SSGF)。本文的主要贡献在于：针对传统方法无法有效传递任务信息且难以实现高效信息共享的问题，提出了 token 化统一表示框架，将原始连续数据离散化并设计符号系统引入任务信息，共同构成 token 序列以模拟大语言模型的提示词工程，从而在输入中同时编码任务语义与数据特征；针对预测、填补等多类任务之间输入输出形式差异、训练方式不统一的问题，提出统一的序列化表示方法，将任务符与离散数据按特定顺序拼接，采用自回归生成机制和“下一个 token 预测”的训练方式实现一致的多任务生成与解析流程；最后，构建了 3 层轻量化 decoder-only 模型，以交叉熵损失为目标函数，实现了完全共享参数前提下不同任务的统一训练与推理。

1 多任务一体化框架基础

在电力系统中，诸如海上风电功率预测与异常数据填补等多类任务，尽管面向不同的运行目标与建模需求，但均基于相同的数据源展开，具有多任务统一建模的潜力。近年来，DeepSeek[15]等大语言模型在多任务统一建模方面取得了显著成果，表明

利用统一的共享网络处理多种任务不仅具备可行性，还能够在共享数据的基础上提升整体的效率与灵活性，为电力系统多任务建模提供了有益的启示。

1.1　大语言模型的多任务处理能力

大语言模型在应对多任务问题中展现出强大的建模能力，尤其在任务类型多样、目标复杂的场景中具备显著优势。该类模型通常采用统一的参数体系与训练范式，能够在无需调整模型结构的情况下同时处理多种异构任务，通过高效的信息编码与共享表达机制，提升整体建模效率与系统适应性。在自然语言处理领域，大语言模型可覆盖文本生成、语言翻译、情感分析等多类任务，表现出良好的通用性与泛化能力。其结构的高度统一性不仅显著降低了系统开发与维护成本，也为构建具备跨任务处理能力的通用模型架构提供了有效范式[16]。图 1 展示了 DeepSeek 通过统一的生成机制逐步输出相应的任务结果。

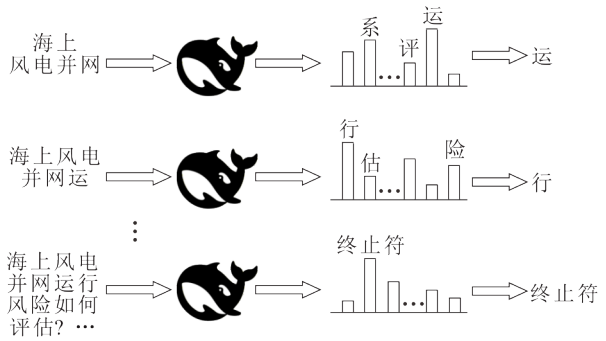


图 1　大语言模型生成过程示意图

Fig.1　Illustration of the LLM generation process

尤其值得关注的是，大语言模型在实现任务集成的同时，在多个基准任务中取得了优于传统单任务模型的结果。以 GPT-3 为例，该模型在自然语言理解与生成领域的多项任务中表现优异，如在 SQuAD 问答和 WebText 文本生成等任务中均实现领先性能[17]。这表明在保持结构统一的前提下，通过参数完全共享与上下文建模机制，模型能够有效融合不同任务的语义信息，从而提升整体的泛化能力与学习效率。这一能力特性表明，大语言模型的多任务处理范式具备跨领域迁移的可行性。

对于电力系统中的多任务场景，构建统一的模型架构不仅可实现多种高度关联任务的一体化建模，如风电功率预测、异常数据填补与故障检测，还能够显著降低模型冗余与部署成本。同时，统一建模有助于促进任务间的信息共享与潜在关联挖掘，为提升系统整体感知能力与建模效率提供了有效路径。图 2 展示了本文在借鉴大语言模型后，提出的多任务建模方法与传统方法以及一般多任务方法的对比，进一步阐明了该统一建模架构在处理复杂电力系统任务时的优势。

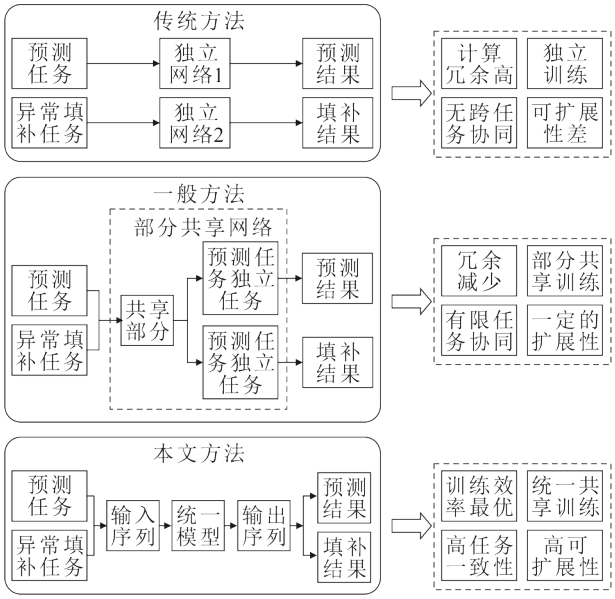


图 2　多任务学习方法比较

Fig.2　Comparison of multi-task learning methods

1.2　符号系统在多任务学习中的作用

在海上风电系统的多任务处理场景中，若单纯依赖原始数据进行建模，往往难以有效表达任务意图，导致模型在面对如功率预测、异常修复、故障检测等任务时难以准确区分各自的需求。

为提升模型对多任务语义的识别能力，引入能够显式标注任务类型与数据状态的符号系统显得尤为关键。在自然语言处理领域，T5 模型[18]便通过添加任务前缀(如“translate English to French:”)的方式，引导模型完成特定语义任务，显著增强了多任务学习框架下的适应能力。Unified-IO 模型[19]通过在多模态输入中插入如“<image>”、“<text>”等符号，实现了不同输入类型与任务语义的一致表示，有效支持了图文问答、描述生成等多种复杂任务的统一处理。尽管此类机制在自然语言处理与多模态建模中已被广泛验证有效，但在电力系统的时序数据场景下，其应用尚处于探索阶段，相关研究成果仍较为有限[20-21]。

符号系统不仅有助于明确任务指令，还能通过语义引导增强模型对任务目标的理解力。通过为不同任务类型分配独立的符号，模型在编码阶段即能

捕捉到任务需求，有助于降低语义歧义，从而实现精确建模与针对性学习。例如，风电功率数据可通过附加不同的符号来指示其用于功率预测还是异常填补，便于模型理解当前任务意图。另一方面，符号还可作为分隔标记，用于标示任务内部不同的数据状态。在异常填补任务中，模型可通过识别缺失标记、参考填补片段等信息，准确定位待修复区域，从而提升填补的精度与适应能力。更为关键的是，符号系统为多任务统一序列建模奠定了基础。借助符号化表达，不同任务可统一组织为类语言式的token序列，使模型能够通过自回归机制在统一的符号体系下完成多任务输出。训练过程中，模型逐步学习各类符号在不同任务语境下的语义作用，从而在推理时灵活切换任务，生成符合不同任务需求的输出结果。在海上风电系统这一多任务协同需求强烈的场景中，引入符号系统有望实现任务识别、信息编码与目标建模的协同统一，为构建具备高适应性和智能性的统一建模框架提供坚实支撑。

2 海上风电系统的符号序列生成框架

本章提出了一种面向海上风电系统多任务建模的符号序列生成框架，旨在在统一结构下协同完成功率预测与异常数据填补等任务。该方法将符号系统与生成式建模机制相结合，通过统一的序列化输入方式，实现多任务在同一模型架构中的共享建模与联合训练。其整体结构与工作流程如图3所示。

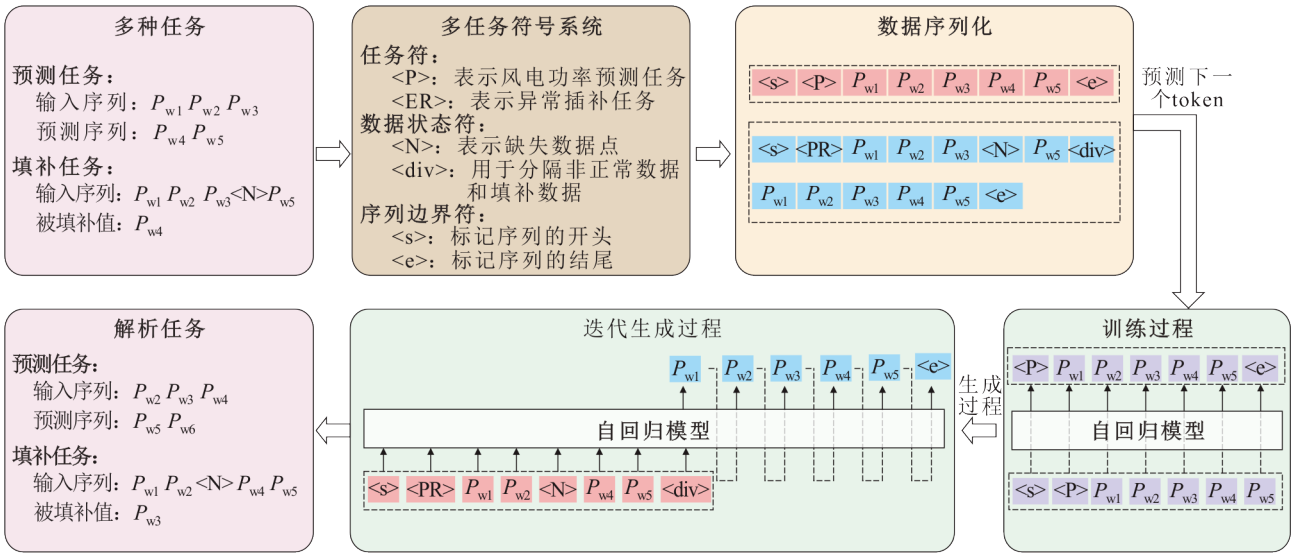


图3 符号序列生成框架整体结构示意图

Fig.3 Overall structure of the symbolic sequence generative framework

2.1 符号系统与数据序列化设计

为实现多任务学习中的统一表达与协同处理，本文提出了一套符号化与序列化表示方法。该方法通过构建符号系统、设计统一的序列化输入方式以及一致的数据编码机制，使不同任务类型及数据状态得以在同一模型框架下以一致的格式表示与高效处理，进而提升多任务建模的通用性与可扩展性。

1）符号系统设计

本文构建了一套符号系统，用于将任务类型与数据状态以序列化形式进行显式编码，从而在统一模型下实现多任务的协同建模与处理。该符号系统包含3类符号，分别用于任务识别、数据状态标示以及序列边界定义，具体如下所述。

(1)任务标识符(用于指示不同的任务类型)

<P>：表示风电功率预测任务(Prediction)；

<ER>：表示异常数据填补任务(Error Repair)。

(2)数据状态符号(用于标示序列中数据的特殊状态)

<N>：表示缺失数据点，需由模型进行填补；

：用于分隔异常数据与填补后数据片段。

(3)序列边界符号(用于标记序列的起始与结束位置)

<s>：表示序列起始；

<\s>：表示序列结束。

该符号系统的引入在多任务统一建模中具有重要意义。一方面，它使原始输入数据在保持序列结构特性的同时，嵌入了明确的任务语义信息；另一方面，符号机制通过将任务类型、数据状态与序

列结构显式编码为 token，降低了模型设计的复杂性，避免了传统多任务方法中需要针对每个任务定制模型结构的繁复过程。

2）数据序列化方式

在符号系统的基础上，本文将各类任务数据序列化为统一的输入格式，以便在同一模型框架内实现多任务的协同处理。该过程将任务类型标识、数据状态信息与原始观测值进行结构化拼接，构成具有一致语义格式的输入表达。通过这一方式，可将风电功率预测、异常数据填补等任务纳入同一建模框架，实现多任务输入的格式标准化与训练流程的一体化。

预测任务：对于风电功率预测任务，设原始风电功率数据序列为 $P_{w1}P_{w2}P_{w3}P_{w4}P_{w5}P_{w6}P_{w7}$，为了使模型明确该任务为预测问题，在数据序列首部添加任务标识符<P>，并用<s>和<\s>分别标记序列的起始与结束，从而得到统一的序列化输入形式为：<s><P> $P_{w1}P_{w2}P_{w3}P_{w4}P_{w5}P_{w6}P_{w7}$ <\s>。

异常填补任务：对于异常填补任务，假设存在缺失数据的原始序列为 $P_{w1}P_{w2}P_{w3}P_{w4}$ <N> $P_{w6}P_{w7}$，其对应的完整序列为 $P_{w1}P_{w2}P_{w3}P_{w4}P_{w5}P_{w6}P_{w7}$。则可以构造输入序列：<s><ER> $P_{w1}P_{w2}P_{w3}P_{w4}$ <N> $P_{w6}P_{w7}$ $P_{w1}P_{w2}P_{w3}P_{w4}P_{w5}P_{w6}P_{w7}$ <\s>。其中<ER>表明当前任务为异常填补， 用于区分待修复序列与参考填补结果片段。模型在训练中根据缺失值<N>的上下文以及后续信息学习修复策略，进而实现准确的异常数据还原。

3）数据编码方式

在完成序列化之后，构造好的序列还需进一步编码，以满足模型输入格式的要求。采用如下编码方法对序列中的符号与数值进行统一映射与处理。

特殊符号编码：首先对任务标识符与数据状态符号进行编码，为每个符号分配唯一的 token 编号。例如，<P>和<ER>分别被编码为 0 和 1；序列边界符<s>、<\s>与分隔符 分别编码为 2、3 和 4；缺失数据标识符<N>被编码为 5。

连续数据编码：对于连续型数值数据，采用区间划分的方法进行离散化，并为每个量化区间分配唯一的编码编号。例如：风电功率数据中的连续取值被量化为若干离散等级，并分别映射到从 6 至 L+5 的编号区间，其中 L 表示量化等级的总数。具体的量化过程如式(1)所示：

$$P_{wi}=\left\lfloor \frac{P_{wi,ori}-P_{w,min}}{P_{w,max}-P_{w,min}+\varepsilon}\cdot L \right\rfloor \tag{1}$$

式中：$\lfloor\cdot\rfloor$ 表示向下取整操作，即取不大于括号；$P_{wi,ori}$ 为第 i 个时间步的原始连续风功率值；$P_{w,min}$ 和 $P_{w,max}$ 分别为数据中的最小值和最大值；ε 取 10^{-3}，用于防止分母为 0。

本文提出的“提示词+回答”的核心思想可以推广到多种任务上，通过直接或间接方式将各任务转化为具有特定语义标识的提示词+答案的形式。在处理其他任务时，仅需引入新的语义符号，并将任务信息与数据内容统一表示为 token 序列，通过自回归的训练方式对不同任务进行统一学习。在推理阶段，模型将通过自回归机制生成对应任务的输出。整个过程无需调整任何网络结构，依然保持参数共享和统一的生成机制，保证了模型在扩展至多任务时的平滑过渡。本文以功率预测和缺失值填补作为示例任务，已经满足多任务学习的基本范式。

2.2　模型构建与训练方法

2.2.1　模型架构设计

本文提出的模型采用了基于自回归机制的生成式网络结构，采用 3 层的 decoder-only 模型，利用掩码多头自注意力机制来提取历史序列特征，并通过因果掩码机制实现逐步预测[22]。相较于传统的 encoder-decoder 架构的 Transformer 模型，本文采用更为简洁高效的 decoder-only 模型，适用于具有非线性特征和跳跃变化的时间序列任务，具体由 3 个主要部分组成：嵌入层、掩码自注意力机制层和输出层。模型整体结构及输入输出流程如图 4 所示。

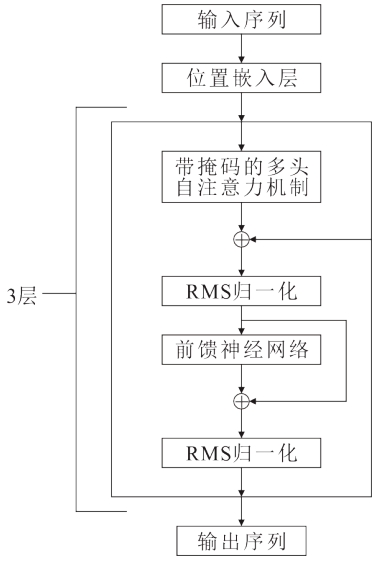


图 4　本文提出的 3 层 decode-only 模型结构示意图

Fig.4　Structure of the proposed three-layer decode-only model

1）嵌入层

输入序列中的每个 token 都会被映射为一个固定维度的高维嵌入向量[23]。该过程使模型能够在统一的向量空间中处理不同类型的输入数据，为后续模块提供了语义信息与位置信息兼容的输入表示，有助于提升模型对多任务序列的建模能力与时序感知能力。对于第 i 个 token $\boldsymbol{w}_i$，其对应的嵌入 $\boldsymbol{e}_i$ 如式(2)所示：

$$\boldsymbol{e}_i = \boldsymbol{E}\boldsymbol{w}_i + \boldsymbol{p}_i \tag{2}$$

式中：$\boldsymbol{E}$ 为嵌入矩阵；$\boldsymbol{w}_i$ 表示输入序列中的所有 token(包括特殊符号和离散化后的风电功率数值 P_{wi})；$\boldsymbol{p}_i$ 为位置编码项[23]，用于引入位置信息。

2）掩码自注意力机制层

模型的中间结构由 3 层堆叠的掩码自注意力模块构成，通过多层掩码自注意力机制与前馈网络的交替组合，在保持自回归生成因果性的前提下，实现对输入序列深层次的特征提取与上下文建模。每一层由带掩码的自注意力机制 (masked self-attention)[24]、前馈神经网络以及均方根层归一化 (root mean square layer normalization, RMSNorm)3 部分组成。

(1) 带掩码的自注意力机制

该机制保证模型在进行序列生成时，仅能关注当前位置之前的信息，从而严格保留自回归生成过程中的因果性约束，避免出现“偷看”未来位置信息的情况。具体的注意力分数计算如式(3)、式(4)所示：

$$\text{Attention}(\boldsymbol{Q},\boldsymbol{K},\boldsymbol{V}) = \text{softmax}\left(\frac{\boldsymbol{Q}\boldsymbol{K}^{\mathrm{T}}}{\sqrt{d_k}} + \boldsymbol{M}\right)\boldsymbol{V} \tag{3}$$

$$\boldsymbol{M} = \begin{cases} 0, & i \geq j \\ -\infty, & i < j \end{cases} \tag{4}$$

式中：$\boldsymbol{Q}$、$\boldsymbol{K}$、$\boldsymbol{V}$ 分别为查询(Query)、键(Key)和值(Value)矩阵，通常由输入经过不同线性映射后得到；Attention($\boldsymbol{Q}$,$\boldsymbol{K}$,$\boldsymbol{V}$) 表示由查询、键和值计算得到的注意力输出；softmax(·)为归一化指数函数，用于将注意力分数转换为权重分布；d_k 是键向量的维度，用于归一化注意力分数；$\boldsymbol{M}$ 为掩蔽矩阵，其中 i 和 j 分别表示注意力分数矩阵中的查询位置索引和键/值位置索引。通过对未来位置进行掩蔽，使得当前时间步 t 只能关注过去的时间步，以此屏蔽未来信息的干扰，确保模型在训练与推理阶段始终遵循单向生成的逻辑。

(2) 前馈神经网络

在每个掩码自注意力机制后接前馈神经网络子层，用于对每个序列位置的输出进行非线性特征变换与维度映射。该子层通常由两层全连接网络组成，输入首先通过一层线性映射及非线性激活(如 ReLU 函数)，随后再经过第二层线性变换以恢复原始维度。该结构不仅增强了模型的非线性表示能力，还有助于捕获局部特征与跨位置信息的交互。具体计算式为：

$$F(\boldsymbol{x}) = \boldsymbol{W}_2 \cdot \sigma(\boldsymbol{W}_1\boldsymbol{x} + \boldsymbol{b}_1) + \boldsymbol{b}_2 \tag{5}$$

式中：$\boldsymbol{W}_1$、$\boldsymbol{W}_2$ 分别为权重矩阵；$\boldsymbol{b}_1$、$\boldsymbol{b}_2$ 为偏置项；$\boldsymbol{x}$ 表示前馈神经网络子层的输入；$\sigma(\cdot)$ 表示非线性激活函数，如 ReLU 函数。

(3) RMSNorm

为了加速收敛并提升训练的稳定性，模型在每个子层之后均采用残差连接(residual connection)与 RMSNorm 操作。RMSNorm 通过对每个样本的特征维度计算均方根值，进行归一化处理，确保网络保持良好的表达能力。RMSNorm 具体式如(6)、式(7)所示，ε 含义同式(1)，该机制有效缓解了梯度消失或爆炸问题，有助于深层网络结构的稳定训练。

$$r(\boldsymbol{x}) = \frac{\boldsymbol{x}}{f(\boldsymbol{x})} \cdot \gamma \tag{6}$$

$$f(\boldsymbol{x}) = \sqrt{\frac{1}{d}\sum_{i=1}^{d} \boldsymbol{x}_i^{\,2} + \varepsilon} \tag{7}$$

式中：$\boldsymbol{x}$ 表示 RMSNorm 层的输入；$f(\boldsymbol{x})$ 为特征向量的均方根值；γ 为缩放因子；d 为特征维度；$\boldsymbol{x}_i$ 为第 i 个特征分量。

3）输出层

输出层将掩码自注意力机制层的最终隐藏状态映射为 token 空间上的概率分布，用于确定当前时间步应生成的输出 token。

2.2.2 训练过程与损失函数

本模型采用“下一个 token 预测”作为统一的训练方式，即在每一步中基于已生成的 token 序列，预测下一个最可能的输出 token。这种自回归训练机制贯穿所有任务类型，无需针对每类任务设计专门的结构或损失函数。通过该机制，模型能够在共享参数的前提下，同时完成风电功率预测与异常数据填补等多种任务，显著提升训练效率与任务间的协同能力。

1）训练数据构造

每一条训练样本序列均按照“起始符-任务符-数据序列-终止符”的方式构建，使模型能够逐步学习依

次生成每个 token。以风电功率预测序列任务为例，若输入序列为<s><P> $P_{w1}P_{w2}P_{w3}P_{w4}P_{w5}P_{w6}P_{w7}$ <\s>，模型首先学习从起始符号<s>预测任务标识符<P>，再从<P>预测第一个功率值 P_{w1}，依此类推，直至完整生成整个序列。这种训练范式在所有任务中保持一致，确保模型在统一机制下完成多任务学习。

2）损失函数

模型采用式(8)所示交叉熵损失函数，来度量预测 token 与实际 token 之间在每个序列位置上的差异。损失函数会在所有位置上进行逐项计算并累加，得到该序列的总损失，作为当前训练样本的优化目标。训练过程中，模型通过多轮迭代不断最小化该损失，以提升序列生成精度和多任务学习效果。

$$L=-\sum_{i}\ln(p(\boldsymbol{w}_i\mid\boldsymbol{w}_{1:i-1})) \tag{8}$$

式中：$p(\boldsymbol{w}_i\mid\boldsymbol{w}_{1:i-1})$ 表示在已知前序 token 序列 $\boldsymbol{w}_{1:i-1}$ 的条件下，模型预测第 i 个位置的 token 为 $\boldsymbol{w}_i$ 的条件概率；p 表示模型输出的条件概率；$\boldsymbol{w}_i$ 为目标序列在第 i 个位置的真实 token，$\boldsymbol{w}_{1:i-1}$ 表示模型在当前时间步前所已生成(或已提供)的 token 序列，作为当前预测的上下文条件。

2.2.3　自回归的序列生成机制

在完成训练后，模型采用自回归方式进行 token 序列的生成，通过逐步输出 token 构建完整序列，并结合后续解析过程提取各类任务的最终预测结果。该生成机制适用于所有任务类型，其整体流程主要包括以下几个步骤。

1）初始化

序列生成的起始阶段基于预设的初始输入序列，或由起始符号<s>与任务标识符组成的基础提示构成。在模型推理过程中，尤其是在测试阶段，任务类型与输入内容均被明确提供，确保模型具备清晰的生成目标，以有效避免模型在初始阶段出现歧义或随机行为。任务描述作为引导信息，能够明确指定模型所需完成的任务类型，从而提升输出的准确性和相关性。

2）迭代生成

模型在每一步中根据当前已生成 token 序列，预测下一个最可能的 token，并将其追加至当前输入序列的末尾。更新后的序列再次送入模型，用于后续 token 的生成。该过程以自回归方式不断迭代，直到生成终止符号<\s>，或达到预设的最大序列长度。整个生成过程保持上下文持续更新，有效捕捉序列中的时序关系和任务依赖特性。

3）最终输出

当 token 序列生成完毕后，模型将依据预定义的解析规则对输出序列进行处理，从中提取出当前任务对应的有效预测结果。根据任务标识符与结构符号进行切片操作，精确定位预测目标或填补片段，完成从 token 序列到数值结果的转换。这种统一的生成与解析流程适用于多种任务，确保了模型在跨多任务过程中的一致性和通用性。

3　算例分析

3.1　数据集构建

本研究选取位于中国江苏省沿海的某海上风电场的实测功率数据，验证所提出的符号序列生成框架在多任务处理中的有效性。实验数据覆盖时间范围为 2020 年 1 月 15 日至 2022 年 3 月 22 日，采样周期为 5 min，持续记录风电出力的时间序列。首先对原始数据进行统计分析与可视化展示，以挖掘其分布特征与波动规律。数据中出现的负值主要源于使用风电场系统输出功率作为衡量指标，即总发电量扣除风电场运行期间的内部用电功耗。为全面呈现实验背景，图 5 展示了该海上风电场 2022 年部分风电功率波形，表 1 则列出了该数据集的主要统计特性，用于支持后续模型训练与评估工作。

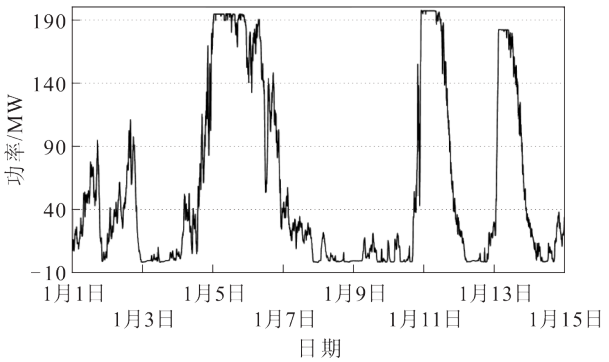


图 5　风电场部分功率波形示意图

Fig.5　Ilustration of partial power output waveforms of wind farm

表 1　原始功率统计特性

Table 1　Statistical properties of raw power output

参数	最大值/MW	最小值/MW	均值/MW	方差	极差/MW
取值	195.50	–5.28	54.26	3 234.30	200.78

在数据预处理阶段，首先剔除了风电系统在限电、检修等非正常运行期间的长时间停运数据，以

保障后续建模所用数据的有效性与代表性。在数据清洗完成后，对数据中存在的缺失值进行了处理，采用特殊符号<N>进行标记，使模型在训练过程中能够识别并处理这些缺失点。为实现统一的符号序列建模框架，需将原始的连续数值型数据进行离散化编码。本文采用等间距划分策略，将数据范围划分为 100 个离散区间，并对每个区间分配独立的token编号，从而完成数值数据到token空间的映射。其具体编码方式如式(1)所示。

本研究构建了两个数据集，分别用于海上风电功率预测与异常数据填补任务。其中，功率预测任务的数据集通过对原始风电功率数据直接采样生成，保留了原始时间序列的连续性，为模型提供充分的历史信息用于训练与预测。而在异常填补任务中，则通过随机移除部分数据点模拟缺失情况，构造带缺失数据的序列用于评估模型的填补能力。随后将数据集按照时间顺序以 8:1:1 的比例划分为训练集、验证集与测试集。其中，前 80%的样本用于模型训练，中间 10%的样本用于超参数调整与早停策略验证，最后 10%的则用于评估模型在未见数据上的泛化性能。

在两个任务中均引入统一的符号系统，用以区分任务类型与数据状态。最终，每个样本均被组织为序列化格式，实现多任务数据在输入结构上的一致性与规范化。为全面评估模型在风电功率预测与异常数据填补两类任务中的性能表现，本文采用以下 4 种常见的误差评估指标：

1）准确率(accuracy)

模型通过自回归方式生成token序列，准确率 P 用于衡量模型生成token序列中，预测token与实际token匹配的比例，用于评价在统一离散化设定下模型对序列生成的吻合度。具体计算式为：

$$P=\frac{N_{\text{correct}}}{N_{\text{total}}} \tag{9}$$

式中：N_{correct} 表示正确预测的 token 数量；N_{total} 表示总的 token 数量。

2）均方误差(mean squared error，MSE)

由于模型生成的是离散token，需映射为连续功率值后进行误差计算，因此采用均方误差来衡量数值还原后的预测精度 E。具体计算式为：

$$E=\frac{1}{n}\sum_{i=1}^{n}(y_i-\hat{y}_i)^2 \tag{10}$$

式中：y_i 表示第 i 个时刻的真实值；$\hat{y}_i$ 表示第 i 个时刻的预测值；n 表示样本总数。

3）均方根误差(root mean squared error，RMSE)

RMSE 是 MSE 的平方根，保留与原数据一致的量纲，能够更直观地解释数值预测误差 R 的大小。具体计算式为：

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

4）平均绝对误差(mean absolute error，MAE)

MAE 用于衡量预测值与真实值之间绝对误差的平均水平 A，具体计算式为：

$$A=\frac{1}{n}\sum_{i=1}^{n}\left|y_i-\hat{y}_i\right| \tag{12}$$

3.2 算例分析

3.2.1 各模型多任务性能对比实验

本实验围绕海上风电功率预测与异常数据填补两个任务展开，采用 3 种神经网络结构(包括三层decoder-only 模型、LSTM[25]和 GRU)，对比分析模型在单任务训练与多任务混合训练下的性能表现。所有模型均在相同的数据集上分别进行两种训练方式：单独训练指模型仅使用对应任务的数据集进行训练，混合训练则指模型同时使用所有任务的数据集进行联合训练。实验指标涵盖均方误差(MSE)、平均绝对误差(MAE)、均方根误差(RMSE)及 token 序列生成准确率，旨在全面评估所提出的符号序列生成框架在多任务场景下的适应性与泛化能力。

实验结果表明，本文模型在显存占用和训练时长上均未引入额外负担。显存占用方面，本文模型单独训练时的 GPU 峰值显存分别为 34.1 MB 和 40.3 MB，而混合训练时为 40.3 MB，未出现任务叠加导致的显存膨胀。单独训练的总训练时长为 225.22 s，而混合训练时长为 220.11 s，几乎与二者之和一致，说明多任务整合未增加额外的训练开销。同时，由于模型在同一损失函数和训练流程下联合优化，也节省了单独调参所需的时间与资源。

实验误差指标如表 2 所示，验证了所提出的符号序列生成框架在多任务建模中的有效性。在如图 6 所示风电功率预测任务中，3 种不同神经网络模型均在保证预测精度的同时有效实现了多任务处理能力。以本文提出的 3 层 decoder-only 模型为例，混合训练下的 MSE 为 15.929 5，MAE 为 2.636 8，RMSE 为 3.991 2，且 token 序列生成的准确率达到 87.39%，与单独训练结果(MSE 为 16.436 0，MAE 为 2.766 8，RMSE 为 4.054 1，准确率 86.60%)相比，

在回归误差减小的前提下，准确率提升了 0.79%。而使用 SSGF 框架的 LSTM 和 GRU 模型在混合训练下分别达 MSE 17.998 8、MAE 2.719 7、RMSE 4.242 5、准确率 85.11%与 MSE 18.722 3、MAE 2.837 4、RMSE 4.326 9、准确率 84.67%。在相同离散化设定下，本文模型的准确率与 3 项回归指标整体优于 SSGF 框架下的 LSTM 和 GRU 模型，且这些结果均优于单独训练模式，进一步验证了所提出方法在多任务场景下的适应性与有效性。此外，从与传统模型的对比可以看出，引入符号体系后的 LSTM 与 GRU 相较于传统 LSTM[26]与传统 GRU[27]，误差指标差距并不显著；本文提出的模型与未使用符号化的版本相比，同样保持了相近的回归性能。这说明本文在引入符号体系将连续数值离散化之后进行训练的效果，并未较原有直接利用连续数值训练的效果产生明显劣化，从而验证了所提出符号化建模策略的合理性与有效性。

如图 7 所示在异常数据填补任务中，所提出的符号序列生成框架同样展现出较优的综合性能。混合训练条件下，本文提出的 3 层 decoder-only 模型的 MSE、MAE 与 RMSE 分别为 10.904 7、2.134 8 和 3.302 2，token 生成准确率达到 91.31%，与单独训练的结果(MSE 10.934 6，MAE 2.174 0，RMSE 3.306 7，准确率 89.59%)相比，各项指标均有所提升。使用 SSGF 框架的 LSTM 与 GRU 模型在混合训练下也分别实现了 MSE 11.552 7、MAE 2.313 0、RMSE 3.398 9、准确率 90.09%和 MSE 12.517 8、MAE 2.439 8、RMSE 3.538 1、准确率 88.67%，均优于单任务训练下的对应指标。同时，与传统模型对比可见，引入符号体系后的 LSTM、GRU 及本文模型与各自未引入符号体系的版本差距并不大，进一步说明离散化训练并未削弱模型性能。

实验结果表明，依托符号系统与统一自回归生成机制构建的 SSGF 框架，在同一模型与参数共享的条件下可以同时完成功率预测与异常填补多种任务；与单任务训练相比，联合训练在 MSE 等回归指标上略有改善，准确率亦同步提升；与传统 LSTM、GRU 及未符号化的模型对比，引入符号体系后的各类模型在误差指标上差距并不大，说明离散化训练并未削弱原有连续数值建模的效果。这表明统一建模能够更充分地利用任务间可迁移的信息，增强异常填补与整体泛化表现，并为后续更多风电相关任务的扩展提供了实证依据。

表 2　不同模型的性能对比

Table 2　Performance comparison of different models

任务类型		模型类型	MSE	MAE	RMSE	准确率/%
预测任务	混合训练	LSTM(SSGF 框架)	17.998 8	2.719 7	4.242 5	85.11
		GRU(SSGF 框架)	18.722 3	2.837 4	4.326 9	84.67
		本文提出的模型	15.929 5	2.636 8	3.991 2	87.39
	单独训练	LSTM(SSGF 框架)	19.636 7	3.052 6	4.431 3	74.41
		LSTM(传统方式)	18.302 9	2.938 1	4.278 2	—
		GRU(SSGF 框架)	20.552 8	3.193 4	4.533 5	72.56
		GRU(传统方式)	19.213 6	3.087 2	4.382 2	—
		本文提出的模型	16.436 0	2.766 8	4.054 1	86.60
		本文模型未符号化	15.883 7	2.589 7	3.985 4	—
填补任务	混合训练	LSTM(SSGF 框架)	11.552 7	2.313 0	3.398 9	90.09
		GRU(SSGF 框架)	12.517 8	2.439 8	3.538 1	88.67
		本文提出的模型	10.904 7	2.134 8	3.302 2	91.31
	单独训练	LSTM(SSGF 框架)	12.665 3	2.363 6	3.558 8	85.39
		LSTM(传统方式)	12.285 3	2.341 4	3.505 0	—
		GRU(SSGF 框架)	13.212 3	2.465 7	3.634 8	84.60
		GRU(传统方式)	12.815 9	2.391 4	3.579 9	—
		本文提出的模型	10.934 6	2.174 0	3.306 7	89.59
		本文模型未符号化	10.606 6	2.149 5	3.256 8	—

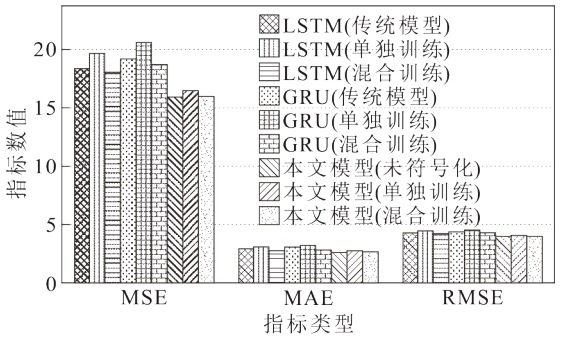


图 6　预测任务指标对比

Fig.6　Metric comparison of prediction task

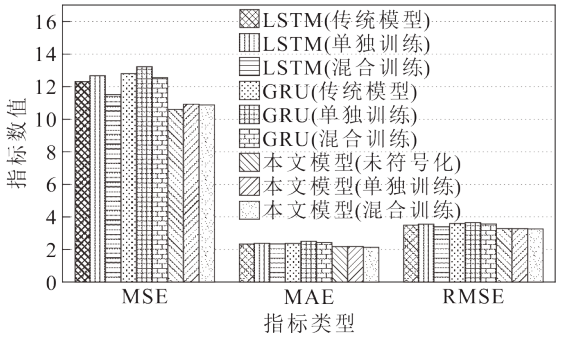


图 7　缺失值填补任务指标对比

Fig.7　Metric comparison of missing value imputation task

3.2.2　符号使用逻辑实验

为了进一步验证所提出的符号序列生成框架在生成过程中是否符合预期语义逻辑，本文随机生

成了 1 000 条以<s>开头的序列，并且对这些序列进行了统计分析。符号正确使用的判定标准包括：<s>作为序列的首个符号且仅出现一次；<\s>为序列的最后一个符号且仅出现一次；<P>和<ER>应紧随<s>出现，并且要符合任务语义逻辑；在包含<ER>的序列中，必须含有<N>，且<N>应出现在 之前； 应该只出现在含有<ER>的序列中，且不应连续重复出现。

各类符号的使用率如表 3 所示，本文提出的 3 层 decoder-only 模型在任务执行过程中表现出较高的符号准确性与语义一致性，所有关键符号的使用准确率均超过 96%，验证了所构建框架在统一编码规则与语义引导下具备良好的生成逻辑一致性与任务区分能力。

表 3　符号准确率统计

Table 3　Statistical summary of symbol accuracy

符号	正确使用数	总使用数	使用准确率/%
<P>	468	481	97.30
<ER>	515	518	99.42
<N>	515	535	96.26
	514	519	99.04
<s>	1 000	1 000	100.00
<\s>	1 000	1 000	100.00

综合实验结果可见，所提出的符号序列生成框架在多任务建模中展现出良好的综合性能。通过多组对比实验，验证了该框架在风电功率预测与异常数据填补两类任务中均具备较强的精度表现和稳定的训练过程。同时，符号使用准确性分析结果表明，模型在自动生成过程中能够正确识别与应用各类任务相关的符号信息，进一步印证了符号机制在统一建模中的有效性与可控性。

整体而言，本文所提出的方法在提升模型通用性的同时，实现了多任务间的协同优化，为电力系统同源数据不同任务的统一建模提供了一种可行、高效的解决方案。

4　结论

1）本文提出了符号化的多任务输入机制，通过为任务类别与数据状态引入显式的符号标记，构建了一套通用符号系统，使不同任务在输入端即可被准确区分与引导。该机制不仅提升了模型对任务语义的感知能力，也为多任务协同奠定了结构基础。

2）建立了统一的序列生成方法：借助“起始符-任务符-数据序列-终止符”的结构化表示形式，所有任务在训练与推理中统一为一种自回归序列生成任务。该方式大幅简化了数据结构，避免了多模型部署与特征割裂的问题。

3）构建了共享参数的多任务生成模型：在本文构建的 3 层 decoder-only 模型下，所有任务通过统一的训练流程学习表征与预测。在功率预测与异常填补两类代表性任务上，该模型在保持预测精度的同时，实现了异常填补性能的明显提升，具备良好的跨任务泛化能力与实际应用潜力。

本文提出的符号化序列生成框架突破了传统“任务独立建模”的范式限制，在保持模型简洁性的同时，实现了多任务间的信息流通与联合优化，为电力系统中多任务处理提供了一种高效可扩展的建模思路。未来研究可进一步拓展该框架至更多任务类型与异构数据场景，并结合边缘计算环境，探索其在实际工程中的部署适应性与泛化能力。

参考文献　References

[1] 李岩，冯俊杰，卢毓欣，等. 大容量远海风电柔性直流送出关键技术与展望[J]. 高电压技术，2022，48(9)：3384-3393.
LI Yan, FENG Junjie, LU Yuxin, et al. Key technologies and prospects of VSC-HVDC for large-capacity and long-distance offshore wind power transmission[J]. High Voltage Engineering, 2022, 48(9): 3384-3393.

[2] 董雪，赵宏伟，赵生校，等. 基于二次分解和多目标优化的超短期海上风电功率预测[J]. 高电压技术，2022，48(8)：3260-3270.
DONG Xue, ZHAO Hongwei, ZHAO Shengxiao, et al. Ultra-short-term offshore wind power forecasting based on secondary decomposition and multi-objective optimization[J]. High Voltage Engineering, 2022, 48(8): 3260-3270.

[3] 肖小刚，吕东晓，彭利鸿，等. 基于 LightGBM-Seq2Seq 的异常天气下的风电功率预测[J]. 电力信息与通信技术，2024，22(9)：62-69.
XIAO Xiaogang, LYU Dongxiao, PENG Lihong, et al. Wind power forecasting based on LightGBM-Seq2Seq model under abnormal weather[J]. Electric Power Information And Communication Technology, 2024, 22(9): 62-69.

[4] 李国栋，徐明扬. 基于 KCR-Informer 的长期风电功率预测研究[J]. 电力信息与通信技术，2024，22(4)：55-62.
LI Guodong, Xu Mingyang. Research on long-term wind power prediction based on KCR-Informer[J]. Electric Power Information And Communication Technology, 2024, 22(4): 55-62.

[5] 何敏，秦亮，刘开培，等. 变电站电力仪表智能检测算法研究[J]. 高电压技术，2024，50(7)：2942-2954.
HE Min, QIN Liang, LIU Kaipei, et al. Research on intelligent detection algorithms for substation power instruments[J]. High Voltage Engineering, 2024, 50(7): 2942-2954.

[6] 沈赋，刘思蕊，徐潇源，等. 基于多尺度特征提取的 IES 多元负荷短期联合预测[J]. 高电压技术，2024，50(7)：2918-2930.
SHEN Fu, LIU Sirui, XU Xiaoyuan, et al. Multi-load short-term joint forecasting of integrated energy system based on multi-scale feature extraction[J]. High Voltage Engineering, 2024, 50(7): 2918-2930.

[7] MESTAV K R, WANG X, TONG L. A deep learning approach to

anomaly sequence detection for high-resolution monitoring of power systems[J]. IEEE Transactions on Power Systems, 2023, 38(1): 4-13.

[8] KIM H J, KIM M K. An unsupervised data-mining and generative-based multiple missing data imputation network for energy dataset[J]. IEEE Transactions on Industrial Informatics, 2024, 20(11): 13429-13440.

[9] BREALY S M, BULL L A, BELTRANDO P, et al. Inter-turbine modelling of wind-farm power using multi-task learning[J]. Mechanical Systems and Signal Processing, 2025, 241: 113540.

[10] FU K, REN Z R. Multi-task prediction of wind speed and time-varying wind shear coefficient using dynamic graph interactive neural network[J]. Information Fusion, 2026, 125: 103478.

[11] DORADO-MORENO M, NAVARIN N, GUTIéRREZ P A, et al. Multi-task learning for the prediction of wind power ramp events with deep neural networks[J]. Neural Networks, 2020, 123: 401-411.

[12] DASH S, SAHOO N C. A multi-task deep learning approach for non-intrusive load monitoring of multiple appliances[J]. IEEE Transactions on Smart Grid, 2024, 15(3): 3337-3340.

[13] WU Z, LI Q, XIA X H. Multi-timescale forecast of solar irradiance based on multi-task learning and echo state network approaches[J]. IEEE Transactions on Industrial Informatics, 2021, 17(1): 300-310.

[14] 马超，李武峰，陈羽飞，等. ChatGPT 类大语言模型赋能电力标准数字化转型的核心技术、技术特征及应用展望[J]. 高电压技术，2025，51(4)：1727-1746.
MA Chao, LI Wufeng, CHEN Yufei, et al. Core technologies, technical characteristics, and application prospects of chatGPT class big language models on the digital transformation of standards in electric power field[J]. High Voltage Engineering, 2025, 51(4): 1727-1746.

[15] REN Z Z, SHAO Z H, SONG J X, et al. DeepSeek-prover-V2: advancing formal mathematical reasoning via reinforcement learning for subgoal decomposition[EB/OL]. (2025-04-30).[2025-07-10]. https://arxiv. org/abs/2504.21801.

[16] GUO D Y, YANG D J, ZHANG H W, et al. DeepSeek-R1 incentivizes reasoning in LLMs through reinforcement learning[J]. Nature, 2025, 645(8081): 633-638.

[17] BROWN T B, MANN B, RYDER N, et al. Language models are few-shot learners[C]//Proceedings of the 34th International Conference on Neural Information Processing Systems. Vancouver, Canada: Curran Associates Inc., 2020: 1877-1901.

[18] RAFFEL C, SHAZEER N, ROBERTS A, et al. Exploring the limits of transfer learning with a unified text-to-text transformer[J]. Journal of Machine Learning Research, 2020, 21(140): 1-67.

[19] LU J, CLARK C, ZELLERS R, et al. Unified-IO: a unified model for vision, language, and multi-modal tasks[C]//Proceedings of the 11th International Conference on Learning Representations. Kigali, Rwanda: ICLR, 2023: 1-34.

[20] 陈艳波，方哲，张宁，等. 基于大语言模型绿电预测和绿电交易的园区综合能源系统集群多目标协同运行方法[J]. 高电压技术，2024，50(7)：2849-2863.
CHEN Yanbo, FANG Zhe, ZHANG Ning, et al. Multi-objective collaborative operation method for park-level integrated energy system cluster based on large language model for green electricity prediction and trading[J]. High Voltage Engineering, 2024, 50(7): 2849-2863.

[21] 余涛，王艺澎，罗庆全，等. 基于大语言模型的新型电力系统用户特性立体应用模式初探[J]. 高电压技术，2024，50(7)：2833-2848.
YU Tao, WANG Yipeng, LUO Qingquan, et al. Preliminary exploration for a large language model-based user characteristics stereoscop-ic application paradigm in new power system[J]. High Voltage Engineering, 2024, 50(7): 2833-2848.

[22] 董俊，刘瑞，束洪春，等. 基于 BIRCH 聚类的 L-Transformer 分布式光伏短期发电功率预测[J]. 高电压技术，2024，50(9)：3883-3893.
DONG Jun, LIU Rui, SHU Hongchun, et al. Short-term distributed photovoltaic power generation prediction based on BIRCH cluster-ing and l-transformer[J]. High Voltage Engineering, 2024, 50(9): 3883-3893.

[23] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[C]//Proceedings of the 31st International Conference on Neural Information Processing Systems. Long Beach, USA: Curran Associates Inc., 2017: 6000-6010.

[24] RADFORD A, WU J, CHILD R, et al. Language models are unsupervised multitask learners[J]. OpenAI Blog, 2019, 1(8): 9.

[25] SAEED A, LI C S, DANISH M, et al. Hybrid bidirectional LSTM model for short-term wind speed interval prediction[J]. IEEE Access, 2020, 8: 182283-182294.

[26] ZELIOS V, MASTOROCOSTAS P, KANDILOGIANNAKIS G, et al. Short-term electric load forecasting using deep learning: a case study in Greece with RNN, LSTM, and GRU Networks[J]. Electronics, 2025, 14(14): 2820.

[27] LI F D, GAN Y J, LI X L. The application of wind power prediction based on the NGBoost–GRU fusion model in traffic renewable energy system[J]. Sustainability, 2025, 17(14): 6405.



刘浩锋
1999—，男，博士生
研究方向为人工智能在电力系统中的应用
E-mail: 2018302070013@whu.edu.cn

LIU Haofeng
Ph.D. candidate



秦　亮(通信作者)
1980—，男，博士，副教授，博导
主要从事电力电子化电力系统的分析与控住、柔性直流系统，人工智能技术在电力系统中的应用、可再生能源发电方面的研究工作
E-mail: qinliang@whu.edu.cn

QIN Liang
Ph.D.
Associate professor
Corresponding author

收稿日期　2025-07-22　修回日期　2025-12-15　　编辑　卫李静