

基于混合注意力机制和 BiLSTM-CRF 的电力行业实体识别模型

高朋¹ 鲁斌¹ 王敦聪¹ 别同¹ 胡恒华¹
GAO Peng LU Bin WANG Duncong BIE Tong HU Henghua

摘要

电力领域文本存在长短期依赖并存、实体边界复杂及标签一致性不足等问题，传统命名实体识别（NER）方法难以有效应对。基于此，文章提出了一种结合预训练语言模型、混合注意力机制与 BiLSTM-CRF 的实体抽取方法。该模型在 RoBERT 的基础上，利用预训练表示增强特征表达能力，引入局部窗口与缩放点积注意力的混合机制，以门控方式融合局部与全局特征；随后通过 BiLSTM 建模顺序依赖，并在 CRF 层保证标签序列的一致性。实验结果表明，该方法能够更加准确地学习电力行业实体边界，且实体识别性能在电力行业 BIOES 标注数据集上显著优于多种基线模型。

关键词

命名实体识别；RoBERT；双向长短期记忆神经网络；条件随机场；注意力机制

doi: 10.3969/j.issn.1672-9528.2025.12.046

0 引言

随着新能源的快速发展，电力系统运行场景逐渐呈现出复杂化与多维化特征，系统规模不断扩大，电力调度所依赖的知识更为复杂^[1-2]。这使得仅依靠人工经验进行调度的难度显著增加，同时也凸显出构建以智能化为核心的新型电力系统的迫切需求^[3]。在此背景下，大模型技术的快速演进为电力系统智能化转型提供了重要机遇。“AI+ 电力”已成为能源系统转型的关键方向之一，依托大语言模型在语义理解、逻辑推理与内容生成方面的优势，可显著增强电力调度工作的知识支撑能力与决策水平^[4-7]。

命名实体识别（named entity recognition, NER）作为自然语言处理的重要基础任务，是推动“大模型 + 电力”应用的关键环节^[8]。NER 的核心目标在于识别文本中的实体，包括地名、人名、组织以及专业术语等，从而实现对文本数据更契合语义的 Token 化处理。在电力领域，由于数据安全性与专业性要求较高，公开可用的数据集极为有限，这在一定程度上制约了现有大模型在该领域的理解与推理能力。通过引入 NER 技术，能够有效提升大模型对电力行业专业术语的识别与学习能力，从而增强其对电力知识的建模与应用水平。

现有研究中，NER 方法大体可分为基于规则的方法^[9-10]与基于统计的方法。前者依赖人工构建识别规则与模式来进行实体抽取，但在电力系统这一高度复杂且专业化的场景中，人工规则难以覆盖复杂语境，且对人工依赖程度较高，效率

较低。相比之下，基于统计的方法将实体识别问题转化为分类预测任务，通过将文本表示为高维向量并将实体类别序列化，借助神经网络学习“文本 - 实体类别”的映射关系，实现更为自动化的实体抽取。值得注意的是，由于其在标注成本与处理效率上的相对优势，基于统计的方法近年来受到学术界的广泛关注，并取得了诸多重要进展^[11-14]。

基于统计的方法研究大多依赖于大模型的 Token 化能力，通过将字词 Token 化，将其从文本数据转为数值数据，从而对其进行处理。随着 BERT 模型的提出^[15]，其在 NLP 领域被广泛应用^[16-17]，尤其是生物医学领域，BERT 对于生物医学命名实体识别研究提供了基础模型^[18-21]。

随着电力行业数字化转型的深入，大量运行记录、事故报告和检修文档不断积累，如何从中自动抽取关键实体信息，如：设备名称、电压等级、事故事件、数值参数等，已成为构建电力知识图谱和智能问答系统的核心环节。然而，电力行业文本具有若干挑战：一是句子中同时存在长程依赖与短程搭配，常规方法难以兼顾；二是实体边界复杂，尤其是包含电压等级、设备名和数值单位的复合短语，容易被错误切分；三是存在同形异义和语义歧义，如地名与机构名难以区分，事件词既可能是动作也可能是实体；四是数字、量纲和专业术语密集，传统模型对其识别能力不足、行业标注样本有限，使得通用模型在直接迁移时效果不佳。

针对上述问题，本文提出了一种基于预训练模型的混合注意力 - BiLSTM-CRF 实体抽取模型，如图 1 所示。该模型在预训练编码器的上下文表示基础上，引入局部窗口多头自

1. 南京南瑞电智科技有限公司 江苏南京 211106

注意力与点积缩放注意力的混合注意力机制，通过门控融合同时捕获短程搭配和长程语义；随后利用 BiLSTM 结构增强顺序依赖与边界判别能力，并在输出层采用 CRF 约束保证标签序列合法性，从而有效降低实体边界识别错误。为了适配中文电力文本，模型设计了 BIOES 标注与子词对齐策略，在数字、符号等子词切分场景下保持标注一致性；同时针对电力领域特点进行了领域特征优化，提升对设备、时间、地点、事件和数值类实体的识别精度。实验结果表明，该方法在少样本条件下仍能显著提升实体抽取的准确率和召回率，为电力行业知识获取与智能应用提供了有力支撑。

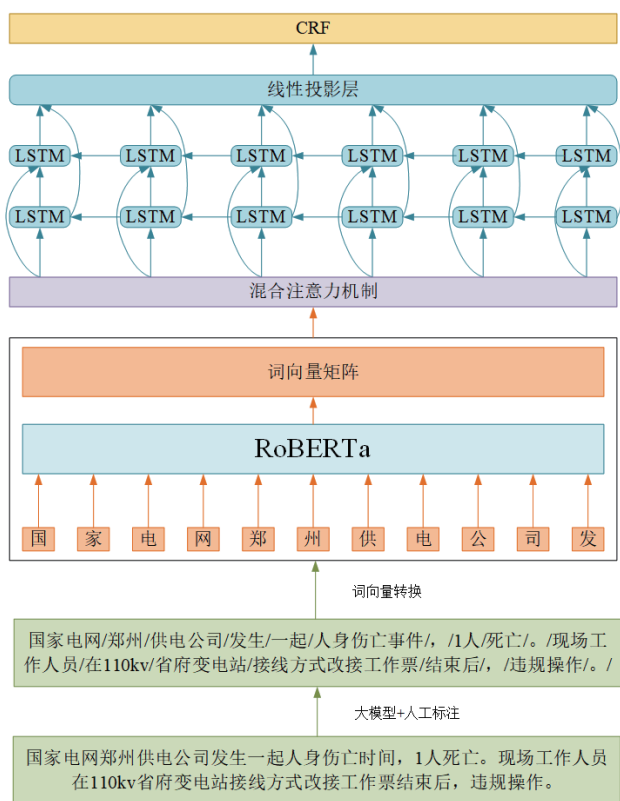


图1 模型结构

1 电力行业实体数据

数据集是人工智能领域研究的重要基础。电力行业由于行业的特殊性，具有较高的保密性，数据集获取难度较高。本文针对这一问题，基于国家标准、行业标准、行业报告等公开资料，建立了基于 BIOES 标准的电力行业实体数据集。其中，B 代表实体的开始；I 代表实体的内部；O 代表非类别的内容；E 代表实体的结束；S 代表单字符。通过“大模型+人工标注”的形式，对实体数据进行标注。设定的类别为9类，如表1所示。电力行业问题具有专业性较强的特征，相较于其他公开的实体识别数据集，例如 CLUENER2020、MSRA 等，具有较多的电力行业专有词汇组合，例如：“220 kV 变压器”“国家电网郑州供电公司”“C

相”“II 回线路”等，为准确识别电力行业实体边界，精准分割电力行业实体，本文基于缩放点积注意力机制和局部窗口多头注意力机制，提出混合注意力机制，在对电力行业文本数据的全局特征进行学习的同时，限制注意力的感受野，提高其对于局部特征的学习能力，为后续的 BiLSTM-CRF 提供更加全面的综合特征。

表1 实体类型及示例

实体类型	实体示例
TIME	2025年9月22日星期一
LOCATION	北京
ORG	国家电网
DEVICE	220 kV 变压器
ROLE	运维人员，调度员
EVENT	检查
VALUE	一起，10 亿
RULE	《电力安全工作规程》
AREA	带电区域，现场

2 模型设计

2.1 混合注意力机制

2.1.1 缩放点积注意力机制

注意力是自然语言处理（neural language processing, NLP）模型的核心内容之一。通过计算输入的词向量（Query）与键（Key）之间的相似性，学习文本的上下文信息，捕获文本的上下文关系，输出期望的值（Value）。缩放点积注意力机制是一种经典且高效的注意力机制，将输入通过线性投影层将输入投影到高维空间，实现注意力的学习，使得模型能够学习如何从输入序列中得到合适的 Query、Key 和 Value，同时通过缩放操作来避免因向量维度增加导致的梯度爆炸或梯度消失。

$$Q = XW_Q \quad (1)$$

$$K = XW_K \quad (2)$$

$$V = XW_V \quad (3)$$

$$\text{Atten}(Q, K, V) = \text{Softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (4)$$

式中： X 为输入的词向量； Q 、 K 、 V 为查询（Query）网络、键（Key）网络和价值（Value）网络，分别通过权重为 W_Q 、 W_K 、 W_V 的线性投影将输入的词向量映射到高维空间。

2.1.2 局部窗口多头注意力机制

缩放点积注意力能够通过计算注意力和缩放实现了对全局特征的学习，然而电力行业的实体模式为“数值+名称”，

例如“110 V 变电站”，使实体抽取模型不仅要学习句子的全局特征，更要学习局部特征。为此，本文基于时间滑动窗口，提出了局部窗口多头注意力机制（local-window multi-head self attention, LMSA）。

假设输入序列表示为 $X \in \mathbb{R}^{L \times d_{\text{model}}}$ ，通过 W_Q 、 W_K 、 W_V ，被投影到高维空间，得到 Query、Key 和 Value 矩阵。

为实现对局部特征的关注，LMSA 通过设置窗口大小 W_{size} ，对注意力机制的感受野进行了限制，用公式表示为：

$$\text{score}(i, j) = \begin{cases} \frac{Q_i K_j^T}{\sqrt{d_k}}, & |i - j| \leq W_{\text{size}} \\ -\infty, & |i - j| > W_{\text{size}} \end{cases} \quad (5)$$

式中： $\text{score}(i, j)$ 是当前词向量与下一个词向量之间的相关性得分。通过 W_{size} ，LMSA 将注意力限制在以 i 为中心、 W_{size} 为半径的范围内，从而实现局部特征的学习。

多头机制允许模型在不同子空间中并行捕捉多种局部依赖模式，从而增强了对实体内部结构和边界特征的建模能力，同时计算复杂度由 $O(L^2)$ 降至 $O(L \cdot W_{\text{size}})$ 。

2.1.3 混合注意力融合

为实现对于全局特征和局部特征的学习，将缩放点积注意力与局部窗口多头注意力进行结合，同时为实现两者的平衡，引入门控机制，提出基于门控的混合注意力融合策略，用公式表示为：

$$g_i = \sigma(W_{\text{hybrid}}[\text{concat}(h_i^{\text{global}}, h_i^{\text{local}})]) \quad (6)$$

$$h_i = g_i \odot h_i^{\text{global}} + (1 - g_i) h_i^{\text{local}} \quad (7)$$

式中： $\sigma(\cdot)$ 为 Sigmoid 激活函数； $\text{concat}(\cdot)$ 为拼接函数； g_i 为门控系数，通过可学习的权重 W_{hybrid} 与全局特征和局部特征的拼接得到。

在不同的语句情景下，基于门控的混合注意力机制融合策略能够自适应调整全局特征和局部特征的权重。当任务更依赖于局部连续特征时，模型倾向于增大局部注意力的比重；而在需要捕捉长程依赖时，模型则会自动加强全局注意力的作用。通过门控机制的动态融合，混合注意力机制不仅提升了实体边界识别的精度，还增强了模型在复杂语义场景下的鲁棒性。

2.2 RoBERTa 模型

RoBERTa (a robustly optimized BERT pretraining approach) 是 2019 年提出的一种基于 BERT 进行改进的预训练语言模型。相较于 BERT，其在保持原始 BERT 框架的基础上，采用了动态的掩码策略，通过在每次迭代中对输入数据重新掩码，从而提高模型对上下文信息学习的鲁棒性和泛化能力。同时，RoBERTa 相较于 BERT，具有更大规模的数据集和更

长的文本处理能力。

本文利用 RoBERTa 强大的文本编码能力，将 BIOES 标注的数据集输入其编码器中，将其转化为词向量，为后续的混合注意力机制、BiLSTM 等模块提供数据支撑。如图 2 所示。

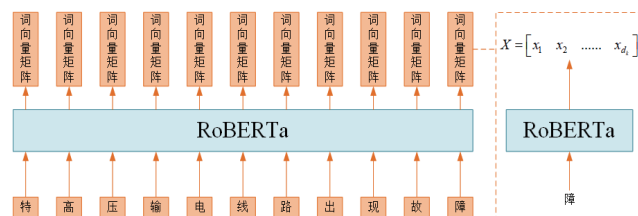


图 2 基于 RoBERTa 的词向量编码

2.3 BiLSTM

上下文特征的学习是 NER 任务的核心之一。为充分学习电力行业文本上下文信息，准确提取电力行业实体，采用 BiLSTM 作为上下文时序特征提取层。BiLSTM 通过双向 LSTM 的模式，扩展了时序模型对于特征学习的感受野，通过未来信息和历史信息的融合，提高模型对上下文语义的理解能力，从而增强实体边界识别的准确性。在前向和后向时序中分别捕捉词汇间的依赖关系，有效建模长距离依赖问题，进一步提升序列标注的连贯性与稳定性。

BiLSTM 是在 LSTM 的基础上，进行双向的特征学习。设输入为 $\{x_i\}_{i=1}^L$ ，每步的输入向量维度为 d_{in} ，隐藏维度为 d_{hid} ，LSTM 在 t 时刻的门控和状态更新为：

$$i_t = \sigma(W_{xi} x_t + W_{hi} h_{t-1} + b_i) \quad (8)$$

$$f_t = \sigma(W_{xf} x_t + W_{hf} h_{t-1} + b_f) \quad (9)$$

$$o_t = \sigma(W_{xo} x_t + W_{ho} h_{t-1} + b_o) \quad (10)$$

$$\tilde{c}_t = \sigma(W_{xc} x_t + W_{hc} h_{t-1} + b_c) \quad (11)$$

式中： i_t 、 f_t 、 o_t 、 $\tilde{c}_t$ 分别为第 t 时刻的记忆门值、遗忘门值、输出门值和临时状态值； b_i 、 b_f 、 b_o 、 b_c 分别为对应网络的偏置。

状态更新和最终的输出为：

$$c_t = f_t \odot c_{t-1} + i_t \odot \tilde{c}_t \quad (12)$$

$$h_t = o_t \odot \tanh(c_t) \quad (13)$$

通过上一时刻的状态和遗忘门的点积，实现对当前时刻遗忘信息的确定；通过当前临时状态和记忆门的点积，实现对当前时刻记忆信息的确定，并最终将当前时刻的状态信息（记忆信息和以往信息）传递到下一时刻。

在双向 BiLSTM 中，模型同时从历史方向和未来方向对当前时刻进行表征，并将前向和后向通过 Concat 函数进行拼接：

$$\vec{h}_t = \text{LSTM}_{\text{fwd}}(x_1, x_2, \dots, x_t) \quad (14)$$

$$\tilde{h}_t = \text{LSTM}_{\text{bwd}}(x_T, x_{T+1}, \dots, x_t) \quad (15)$$

$$h_t^{bi} = \text{concat}(\tilde{h}_t, \tilde{h}_t) \quad (16)$$

通过双向门控记忆在每个位置融合过去与未来信息，为 NER 任务提供强顺序感知的逐位置表示。

2.4 CRF

混合注意力机制为模型提供了全方位的特征学习能力，BiLSTM 为模型提供了时序特征学习能力，但是 NER 任务如果仅依赖 BiLSTM 做最终的实体类别预测，会导致出现标签语义不一致的情况，例如 B-DEVICE 后面是 I-ROLE 或 I-DEVICE 出现在 B-DEVICE 之前。为解决这一问题，本文 BiLSTM 的基础上，引入了条件随机场（conditional random field, CRF），提出了 BiLSTM-CRF 模型。用公式表示为：

$$z_t = W_o h_t^{bi} + b_o \quad (17)$$

$$P = [z_1, z_2, \dots, z_L] \quad (18)$$

式中： $z_t \in \mathbb{R}^C$ 是标签打分向量，其中 C 是标签个数； $P \in \mathbb{R}^{L \times C}$ 是整个输入经过 BiLSTM 和线性投影后的结果，是 CRF 的输入。并结合转移评分矩阵 $A \in \mathbb{R}^{C \times C}$ 计算整个输入序列标签的分数。

设候选标签序列为 $y = [y_1, y_2, \dots, y_L]$ ，CRF 的全局打分函数为：

$$\text{score}_{\text{crf}}(x, y) = \sum_{i=1}^L (A_{y_{i-1}, y_i} + P_{t, y_i}) \quad (19)$$

式中： P_{t, y_i} 为第 t 个位置的发射函数； A_{y_{i-1}, y_i} 为从标签 y_{i-1} 转移到标签 y_i 的转移分数。条件概率定义为：

$$p(y|x) = \frac{\exp(\text{score}_{\text{crf}}(x, y))}{\sum_{y' \in y^L} \exp(\text{score}_{\text{crf}}(x, y'))} \quad (20)$$

通过全局打分函数，CRF 对正确标签和错误标签进行标记，实现了在学习输入与输出映射关系的同时，考虑相邻时刻输出的关系。

3 实验分析

3.1 评价指标及参数设置

为综合评估所提模型的实体抽取性能，本文选择精确率（Precision）、召回率（Recall）、 F_1 -score 作为评价指标，具体公式为：

$$P = \frac{TP}{TP + FP} \quad (21)$$

$$R = \frac{TP}{TP + FN} \quad (22)$$

$$F_1 = \frac{2 \cdot P \cdot R}{P + R} \quad (23)$$

式中：TP 为正确预测的实体个数；FP 为错误预测的实体个

数；FN 为未预测到的实体个数。

本文实验环境为 Python3.10.0，开发框架为 Torch2.8.0、Transformers4.55.2。优化器采用 Adam，GPU 为 NVIDIA GeForce RTX 4090。本文所提模型训练参数如表 2 所示，使用的预训练模型为 bert-base-chinese。

表 2 模型训练参数

参数	参数值
输入文本最大长度	100
Epoch	50
Batch_size	4
Dropout	0.5
Bert learn_rate	1e-5
CRF learn_rate	1e-4

3.2 对比实验及分析

为验证本文所提混合注意力机制在电力行业实体识别中有效性，基于 BiLSTM-CRF 模型，在所建立的数据集基础上，与目前常用的注意力机制进行了对比实验，其结果如表 3 所示。

表 3 注意力机制对比结果

单位：%			
注意力机制	Precision	Recall	F_1 -score
缩放点积注意力机制	89.54	77.68	79.21
交叉注意力机制	88.76	75.53	76.25
稀疏注意力机制	90.34	79.13	82.58
混合注意力机制	95.63	84.72	89.93

从表 3 可以看出，在 BiLSTM-CRF 上，混合注意力机制凭借全局趋势和局部趋势的共同学习，相较于其他注意力机制在电力行业实体数据集上具有更强的识别性能。尽管全局特征和局部特征的权重分配对于全局趋势和局部趋势的学习十分重要，但是经过门控机制的自适应学习，使得混合注意力在权重分配上根据字段位置的不同，学习不同的权重。

本文同时与现有的分词模型进行了对比。为验证本文所提模型在电力行业实体识别任务中的有效性，在所建立的电力行业实体数据集上，选择多个现有研究中较为常用的模型进行了对比实验：膨胀卷积神经网络 - 条件随机场（IDCNN-CRF）、双向 Transformer 编码器 - 双向长短期记忆网络 - 条件随机场（BERT-BiLSTM-CRF）、双向 Transformer 编码器 - 膨胀卷积神经网络 - 条件随机场（BERT-IDCNN-CRF）以及传统的神经网络模型：卷积神经网络、双向神经网络。其对

比结果如表 4 所示。从表中可以看出，所提模型相较于其他现有的模型具有更强的电力行业实体识别能力。混合注意力机制的全局特征和局部特征学习能力使得模型能够精准识别电力行业实体边界，准确提取电力行业实体。

表 4 不同模型之间的对比结果

模型	单位：%		
	Precision	Recall	F_1 -score
BiLSTM	84.43	74.36	78.62
CNN	85.35	76.81	78.59
IDCNN-CRF	88.12	79.58	80.34
BERT-BiLSTM-CRF	93.64	82.79	88.52
BERT-IDCNN-CRF	90.76	80.31	87.60
所提模型	95.63	84.72	89.93

为分析各模块对于所提模型的影响，本文在所建立的电力行业数据集上进行了消融实验。所提模型通过 BERT-based-chinese 进行分词，并通过人工进行实体标注，通过 BERT 的编码器进行编码，然后计算词向量的缩放点积注意力和局部窗口注意力，并通过门控机制进行自适应权重学习，将注意力得分进行拼接，输入 BiLSTM，学习数据的时序特征，并将其输出输入 CRF 中，得到最终的结果，消融实验结果如表 5 所示。

表 5 消融实验结果

	人工实 体标注	自适应权 重学习	混合 注意力	单位：%		
				Precision	Recall	F_1 -score
模型 1				79.83	70.62	73.24
模型 2	√			82.31	76.92	79.64
模型 3			√	85.64	79.39	80.73
模型 4		√	√	88.74	72.53	85.62
模型 5	√	√	√	95.63	84.72	89.93

其中，模型 1 为仅使用 BERT-basd-chinese 进行分词，并 按 照 [TIME、LOCATION、ORG、VALUE、DEVICE、EVENT、ROLE、RULE、AREA] 进 行 标 记；模 型 2 为 在 BERT-based-chinese 分词和标注的基础上，使用人工对标注结果进行纠错和标记；模型 3 为在 BERT-based-chinese 分词和标注的基础上，使用混合注意力机制，但权重均设定为 0.5；模型 4 为在 BERT-based-chinese 分词和标注的基础上，使用混合注意力机制和自适应权重；模型 5 为所提模型。

从表 5 可以看出，现有大模型由于缺少电力行业专业知

识，相关数据集较少，对于电力行业实体的识别性能并不理想，使得仅通过 BERT 标注的电力行业实体识别模型性能较差；当通过人工进行纠错和标注后，数据集质量提高，模型的性能有所提升，但仍受限于数据集规模较小，难以学习到正确的电力行业实体边界；当仅使用混合注意力机制时，尽管模型学习到数据的全局特征和局部特征，然而权重的设定使得模型对于全局特征和局部特征的重视程度相同，导致在部分语境中的识别效果较差；当使用门控机制对混合注意力机制的权重进行自适应学习时，模型能够根据上下文信息，自适应调整全局特征和局部特征的权重，模型的性能有了提高，但是受限于数据集的质量，其识别性能仍有提升的空间。上述消融实验证明了所提模型的模块在电力行业实体识别任务中的必要性以及所提方法的有效性。

4 结论

针对电力行业文本中命名实体识别任务存在的全局依赖和局部依赖并存、实体边界复杂、数据集较少的问题，提出了一种基于混合注意力机制和 RoBERTa-BiLSTM-CRF 的实体抽取框架。主要贡献为：

（1）针对传统语言模型存在的局部特征学习能力不足的问题，提出了基于局部窗口多头注意力机制，通过窗口限制注意力机制感受野，提高局部特征的学习能力；同时结合缩放点积注意力机制，通过门控机制进行混合注意力机制的融合，实现全局特征和局部特征的兼顾；

（2）针对电力行业数据集较少的问题，基于公开的国家标准、行业标准、行业报告等公开文件，通过“大模型 + 人工标注”的方式，建立了电力行业实体抽取数据集，为电力行业命名实体抽取任务提供高质量的数据集，并基于该数据集对所提模型进行了验证，证明了所提模型在电力行业命名实体识别任务中的有效性。

尽管如此，所提模型仍存在局限性：算力要求较高、长文本处理效率有限，且尽管建立了新的数据集，但其规模相较于大模型所需规模仍存在较大的差距。未来研究将探索轻量化预训练模型与知识蒸馏方法，引入电力知识图谱增强专业实体识别，并进一步研究数据增强方法，以支持更全面的电力知识建模与智能应用。

参考文献：

[1] 舒印彪,陈国平,贺静波,等. 构建以新能源为主体的新型电力系统框架研究[J]. 中国工程科学,2021,23(6):61-69.
[2] 肖先勇,郑子莹.“双碳”目标下新能源为主体的新型电力系统：贡献、关键技术与挑战[J]. 工程科学与技术,

- 2022, 54(1): 47-59.
- [3] 王小君, 窦嘉铭, 刘翌, 等. 可解释人工智能在电力系统中的应用综述与展望[J]. 电力系统自动化, 2024, 48(4): 169-191.
- [4] 韩泽雷, 鞠平, 秦川, 等. 面向新型电力系统的频率安全研究综述与展望[J]. 电力自动化设备, 2023, 43(9): 112-124.
- [5] 陈晓红, 傅文润, 刘朝明, 等. 人工智能大模型在电力设备运维场景中的应用探讨[J]. 中国工程科学, 2025, 27(1): 180-192.
- [6] JIA M S, CUI Z Y, HUG G. Enhancing LLMs for power system simulations: a feedback-driven multi-agent framework[J]. IEEE transactions on smart grid, 2025, 16(6): 5556 - 5572.
- [7] LIU Y X, JIA M S, ZHANG Y X, et al. RePower: an LLM-driven autonomous platform for power system data-guided research[J]. Patterns, 2025, 6(4): 101211.
- [8] 田红鹏, 吴璟玮. RIB-NER: 基于跨度的中文命名实体识别模型[J]. 计算机工程与科学, 2024, 46(7): 1311-1320.
- [9] 孙哲涛, 于红, 宋奇书, 等. 基于规则匹配与深度学习 Ab-Transformer 的渔业标准表格信息抽取方法[J]. 大连海洋大学学报, 2023, 38(1): 140-148.
- [10] 张晗, 张亚洲, 徐秉智, 等. 一种混合提示学习与规则的领域命名实体识别方法[J]. 郑州大学学报(理学版), 2025, 57(5): 31-38.
- [11] ZHENG K H, SUN L Y, WANG X, et al. Named entity recognition in electric power metering domain based on attention mechanism[J]. IEEE access, 2021, 9: 152564-152573.
- [12] KAUR N, SAHA A, SWAMI M, et al. BERT-Ner: a transformer-based approach for named entity recognition[C]//2024 15th international conference on computing communication and networking technologies (ICCCNT). Piscataway: IEEE, 2024: 1-7.
- [13] WU L T, LIN J R, LENG S, et al. Rule-based information extraction for mechanical-electrical-plumbing-specific semantic web[J]. Automation in construction, 2022, 135: 104108.
- [14] KONG Z, YUE C X, SHI Y, et al. Entity extraction of electrical equipment malfunction text by a hybrid natural language processing algorithm[J]. IEEE access, 2021, 9: 40216-40226.
- [15] DEVLIN J, CHANG M W, LEE K, et al. BERT: pre-training of deep bidirectional transformers for language understanding[C]//Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers). Brussels: ACL, 2019: 4171-4186.
- [16] CHANDRA C, OJIMA Y, BENDARKAR M V, et al. Aviation-BERT-NER: named entity recognition for aviation safety reports[J]. Aerospace, 2024, 11(11): 890.
- [17] AHMED A, HUANG D G, ARAFAT S Y, et al. Enriching urdu NER with BERT embedding, data augmentation, and hybrid encoder-CNN architecture[J]. ACM transactions on asian and low-resource language information processing, 2024, 23(4): 1-38.
- [18] 许力, 李建华. 基于 BERT 和 BiLSTM-CRF 的生物医学命名实体识别[J]. 计算机工程与科学, 2021, 43(10): 1873-1879.
- [19] HU X S, ZHANG H J, SUN Y J. Chinese medical short text matching model based on fine-tuning BERT-attention-BiLSTM[C]//2023 IEEE/ACIS 23rd International Conference on Computer and Information Science (ICIS). Piscataway: IEEE, 2023: 91-96.
- [20] LI D Y, YAN L, YANG J Z, et al. Dependency syntax guided BERT-BiLSTM-GAM-CRF for Chinese NER[J]. Expert systems with applications, 2022, 196: 116682.
- [21] LIU N, HU Q, XU H Y, et al. Med-BERT: a pretraining framework for medical records named entity recognition[J]. IEEE transactions on industrial informatics, 2021, 18(8): 5600-5608.

【作者简介】

高朋(1986—), 男, 江苏盐城人, 硕士, 助理工程师, 研究方向: 新能源发电场站智能辅控技术、计算机软件架构及强化学习等。

鲁斌(1984—), 男, 浙江衢州人, 硕士, 高级工程师, 研究方向: 风力发电中的电动变桨距系统及其相关技术。

王敦聪(1994—), 男, 江苏徐州人, 硕士, 助理工程师, 研究方向: 变电站智慧巡检、电力大模型。

别同(1997—), 男, 山东潍坊人, 硕士, 助理工程师, 研究方向: 新能源发电场站智能辅控技术、计算机软件架构及强化学习等。

胡恒华(1999—), 男, 黑龙江望奎人, 本科, 助理工程师, 研究方向: 新能源发电场站智能辅控技术、智慧工地等。

(收稿日期: 2025-10-14 修回日期: 2025-12-09)