文章编号：1000-3673（2025）11-4876-12　　中图分类号：TM 721　　文献标志码：A　　学科代码：470·40

融合 PERT 与高效全局指针网络的电力变压器缺陷文本实体识别方法

林蔚青[1]，郑垂锭[2]，陈静[1]，江灏[1]，肖洒[3]，王铭海[1]，缪希仁[1]

（1．福州大学电气工程与自动化学院，福建省 福州市 350108；
2．国网福建省电力有限公司闽侯县供电公司，福建省 福州市 350116；
3．国网福建省电力有限公司超高压分公司，福建省 福州市 350013）

Text Entity Recognition Method for Power Transformer Defects Combining PERT and Efficient Global Pointer

LIN Weiqing[1], ZHENG Chuiding[2], CHEN Jing[1], JIANG Hao[1], XIAO Sa[3], WANG Minghai[1], MIAO Xiren[1]

(1. College of Electrical Engineering and Automation, Fuzhou University, Fuzhou 350108, Fujian Province, China;
2. Minhou County Electric Power Supply Branch of State Grid Fujian Electric Power Co., Ltd., Fuzhou 350116, Fujian Province, China;
3. Extra High Voltage Branch Company of State Grid Fujian Electric Power Co., Ltd., Fuzhou 350013, Fujian Province, China)

ABSTRACT: Power transformer defect texts contain information about equipment reliability, aiding intelligent operation, and lifecycle management. According to the bidirectional encoder representation from transformers (BERT) model, this study proposes a method that combines pre-training BERT with the permuted language model (PERT) and the efficient global pointer (EGP) for entity recognition in defect texts. The PERT model is initially developed by pre-training a vast Chinese corpus with a permuted language model. Then, PERT, as the semantic encoding layer, uncovers the text's internal semantic dependencies and complexities. EGP, as the information decoding layer, focuses on precisely locating key information and extracting the distribution features of entities within defect texts, accurately identifying defect entities. Finally, the PERT-EGP model identifies five entity types in defect texts: defective equipment, defective components, defective locations, defective phenomena, and defective severity. Experiments demonstrate that the proposed approach significantly improves handling complex composite entities and lengthy texts, significantly reducing model training time and providing enhanced text recognition performance.

KEY WORDS: defective text; transformer; entity recognition; permuted language model; efficient global pointer network

摘要：电力变压器缺陷文本蕴含大量与设备可靠性密切相关的信息，可为变压器的智能化运维及寿命周期管理提供重要支撑。依托基于 Transformer 的双向编码器表示(bidirectional encoder representation from transformers，BERT)模型，文章提出一种融合乱序语言模型预训练 BERT(pre-training BERT with permuted language model，PERT)与高效全局指针(efficient global pointer，EGP)网络的电力变压器缺陷文本实体识别方法。首先，在大规模中文语料库上利用乱序语言模型进行预训练以形成 PERT 模型。其次，PERT 作为语义编码层，以深入挖掘实体内部的语义依赖关系，并捕捉复杂文本中的语言特征；EGP 作为信息解码层，专注于精确定位关键信息并提取实体在缺陷文本中的分布特征，进而准确识别缺陷实体。最后，运用 PERT-EGP 模型识别缺陷文本中包含的缺陷设备、缺陷部件、缺陷部位、缺陷现象和缺陷程度 5 类实体。算例结果表明，相较于现有方法，该方法不仅在成分复杂的复合实体和长文本上效果提升显著，而且大幅缩短模型训练时间，具有更好的文本识别性能。

关键词：缺陷文本；变压器；实体识别；乱序语言模型；高效全局指针网络

DOI：10.13335/j.1000-3673.pst.2024.1527

0　引言

电力数据挖掘是智能电网发展的前沿领域，为了实现状态检修的目标，相关学者们利用数据驱动方法以挖掘在线监测的多源结构化数据，如预测、诊断及评估电力变压器及其部件的运行状态及故障风险[1-3]，展示了较好的应用前景。在电力变压器的全生命周期中，除了结构化数据之外，电网公司还积累了大量非结构化数据，如文本、图像、视频

基金项目：福建省高校产学合作项目(2023H6006)。

Project Supported by the Industry-academy Cooperation Project of Universities in Fujian Province (2023H6006).

等。其中，缺陷文本是运检人员在日常巡检过程中所记录的设备缺陷描述。准确高效地从中提取出缺陷部件、缺陷现象、缺陷程度等细粒度实体信息，可为缺陷分析和全寿命周期管理等研究提供额外的数据支撑[4]，有助于优化变压器运维管理策略和探究不同缺陷间的潜在规律[5]。

目前，缺陷文本的细粒度信息提取主要依靠人工处理，耗时且耗力。命名实体识别(named entity recognition，NER)技术可以有效提取自然语言文本中预定义的细粒度实体信息[6]，为自动识别缺陷文本中的关键信息提供了技术支撑。

面向各类中文电力文本数据，常采用基于规则的方法开展实体识别研究[7]。该类方法根据文本特点，利用人工制定的大量规则匹配文本中的特定实体。文献[8]针对电网缺陷文本的语法特征，定义了相应的语义框架与语义槽，通过槽匹配的方式提取实体信息。文献[9]出一种基于“左贪心”出栈规则和依存关系状态转移的依存句法树构建方法，通过树匹配方式实现电力设备缺陷文本中关键实体的精确识别。文献[10]提出一种两阶段实体抽取方法，基于依存句法匹配符合特定的实体，再利用相似度算法进一步精确实体范围，以精准识别配电线路跳闸填报文本中故障现象及故障原因等实体。

根据上述工作，在特定数据集中，基于规则的实体识别方法展现出良好的性能。但是，编写精确的规则需要大量的人力和丰富的专业知识，随着实体类型和语料库的增加，维护和更新规则库的成本会显著增加。此外，规则的制定极度依赖于特定数据集的语言特征，限制了该类方法的泛化能力。

近年来，随着深度学习技术的日臻成熟，并在实体识别方面取得了显著的性能提升[11]。基于深度学习的方法通常由字符表示、上下文编码器和标签解码器三层架构组成[12]，其能够通过非线性激活函数从文本数据中学习更加复杂的特征，逐渐成为电力领域实体识别的主流方案。其中，双向长短时记忆(bidirectional long short-term memory，BiLSTM)网络是目前主流的上下文编码器，加入条件随机场(conditional random field，CRF)对错误的标签进行约束后，可使模型的识别效果获得一定提升[13]。文献[14]提出基于可选词向量的 BiLSTM-CRF 电力缺陷文本命名实体识别模型，在字符表示层添加文本中数字量的向量表示，使模型具备更好的特征抽象能力及更高的实体识别精度。针对 BiLSTM 模型易受局部语义信息干扰的问题，文献[15]面向非结构化的电网调控文本，在 BiLSTM-CRF 中嵌入注意力机制，以优化权重参数，使模型聚焦于关键信息提取，进而提升实体识别的准确性。

基于 Transformer 的双向编码器表示(bidirectional encoder representation from transformers，BERT)模型，采用堆叠的 Transformer 编码器进行文本语义特征提取，展现了相较于传统深度学习方法更加强大的特征学习能力，其结合大规模语料库预训练，使得实体识别性能达到新的高度。文献[16]将 BERT 与 BiLSTM-CRF 相结合，对缺陷文本中的实体加以识别，并验证了基于 BERT 预训练的模型能更好地识别存在错别字的实体信息。文献[17]基于电力设备运行规程、电力技术标准及检修记录等文本构建大规模电力专业语料库，结合多种掩码机制和动态加载策略训练出适用于电力领域的 PowerBERT 模型，在电力缺陷记录实体识别任务中具有较好的性能。

虽然基于 BERT 预训练的深度学习模型在电力文本实体识别任务上取得了一定进展，但现有研究普遍利用序列标注范式，通过预测字符标签的方式识别实体，只考虑相邻字符标签间的依赖关系，忽略了实体内部字符之间的相互依赖关系，对于缺陷文本中由多个细粒度词语组成的复合实体表现不佳。例如，“接地电流监测装置”只识别出“监测装置”，而遗漏另一重要部分“接地电流”。另外，该范式需采用 CRF 对序列标签进行约束，使得模型训练代价大、计算复杂度较高且耗时。

针对上述问题，本文以非结构化的电力变压器缺陷文本数据为典型对象，提出一种融合乱序语言模型预训练 BERT(pre-training BERT with permuted language model，PERT)与高效全局指针网络(efficient global pointer，EGP)的实体识别方法，采用片段抽取的新范式，将实体视为一个整体进行判别，从而提高实体识别的准确率和效率。本文的核心贡献在于：

1）将待识别文本序列表示为上三角矩阵，该标注策略使得模型可以直观地捕捉到实体内部字符之间的语义依赖关系；

2）基于乱序语言模型对通用的 BERT 模型加以预训练，构建 PERT 模型，从而减轻模型预训练与实体识别阶段存在的差异；

3）将 EGP 作为解码层，充分学习实体在缺陷文本中的分布特征，以精准识别变压器缺陷文本中的关键实体。该结构支持并行计算，显著降低训练时间，提高模型在实际工程中的实用性。

1 电力变压器缺陷文本特点

在日常人工巡视、自维护计划、检修试验等工作中，运检人员会将电力变压器缺陷事件情况进行详细记录，包含缺陷发生日期、缺陷内容的详细描述、缺陷性质分类等信息，形成设备缺陷记录。这些记录通常存储在电网的生产管理系统或资产管理系统中，从数据库中导出的变压器缺陷记录数据组成如图 1 所示，主要包括两大类数据。

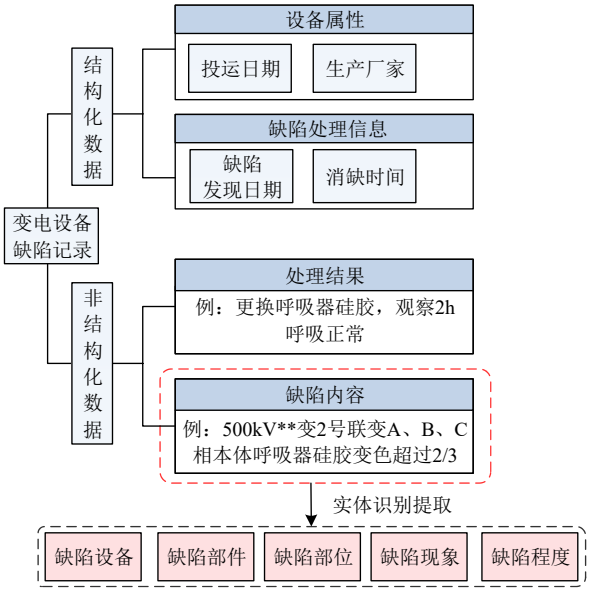

图 1 变压器缺陷记录数据组成

Fig. 1 Defect record data composition of transformer

第 1 类数据为设备属性以及缺陷处理流程中记录的结构化数据，该类数据可以直接被计算机处理，进行缺陷统计学分析等相关研究。

第 2 类数据为缺陷内容及处理结果这类以短文本形式记录的非结构化数据。其中，缺陷内容文本(以下简称缺陷文本)包含缺陷发生的部件、部位、现象等具体的缺陷细节信息，这些信息被转化为结构化数据后，可与结构化的设备属性信息、缺陷处理信息相结合，用于变电设备缺陷规律和模式的深入分析，从而为设备的生产、验收、运维检修等环节的优化提供有针对性的指导建议。

考虑到缺陷文本记录的过程因运检人员记录习惯而不同，没有固定的结构，导致文本中隐含的信息无法直接被应用。因此，需要利用实体识别技术提取其中的关键信息，实现非结构化文本的结构化转换，从而为电力变压器智能运维中各类更高级应用提供更加丰富的数据来源。

2 融合 PERT 和 EGP 的实体识别模型

本文所提 PERT-EGP 实体识别方法主要包括 PERT 预训练、实体识别模型训练以及模型预测 3 个阶段，其具体流程如图 2 所示。

1）PERT 预训练阶段：利用中文维基百科及其他百科、新闻、问答等大规模中文语料库，基于乱序语言模型对 PERT 进行预训练，从而使 PERT 初步学习中文文本的语义特征。同时生成字符索引表，以便后续阶段形成每个字符的初始向量表示。

2）实体识别模型训练阶段：将缺陷文本的训练集输入 PERT-EGP 模型，针对缺陷文本实体识别任务不断迭代模型参数，使其获得缺陷文本的语义理解能力。训练过程中以验证集 F_1 分数作为保存模型的依据，当验证集 F_1 分数连续迭代 10 次没有提升则停止训练。

3）模型预测阶段：将测试集中的缺陷文本依次输入训练完毕的 PERT-EGP 模型，输出每条文本中存在的特定实体。最后，利用评价指标对识别效果进行评价。

2.1 PERT 预训练机制

以 BERT 为代表的预训练语言模型，使用大规模文本语料库和无监督、自监督学习算法提取中文字、词、句子中的语义，获得了较强的自然语言表示能力，并在实体识别[18-19]、知识问答[20]、文本分类[21]等下游任务上取得了提升效果。但是 BERT 在预训练过程中使用的掩码语言模型任务和下句预测任务，分别以预测被遮盖的字符和句子的顺序为训练目标，与下游的实体识别任务仍存在较大差异。为了进一步提升实体识别效果，减轻模型预训练阶段与实体识别阶段存在的差异，本文采用基于乱序语言模型任务的 PERT 作为预训练模型，以词、短语为单位进行预训练，其结合了全词掩码和 N-gram 掩码策略选择候选的乱序标记，能够更好地提升在实体识别上的表现。

PERT 的主旨思想为：虽然句子中的某些字词顺序混乱，但仍可以捕获句子表达的核心意思。预训练过程中，PERT 将 BERT 中的掩码预测任务替换成词序预测任务[22]，选择文本中 15%的词或短语作为候选标记，这些候选标记词或短语有 90%的概率被随机打乱字符顺序，10%的概率保持不变，训练目标是预测标记的原始位置。

如图 3 所示，算法随机选取句子中的“冷却器”、“电源”2 个词进行乱序处理，其余词保持语序不变，然后预测句子中被乱序处理的 2 个词的真实顺序。PERT 通过学习上下文信息学习，判断出“冷”在原句中的排序是 1，“却”在原句中的排序是 2，“器”在原句中的排序是 3，“电”在原句中的排序是 6，“源”在原句中的排序是 7。

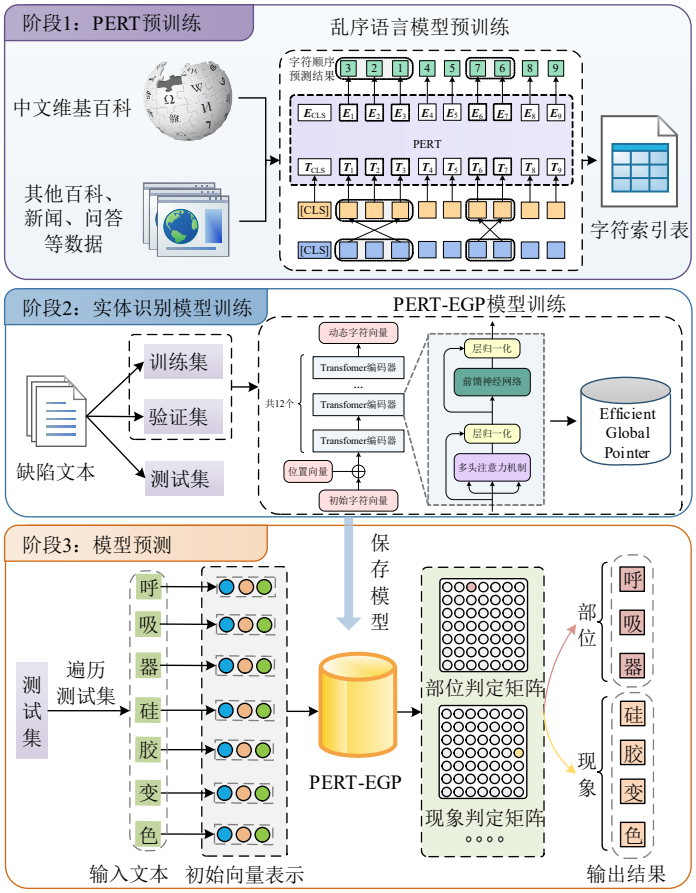

图 2　PERT-EGP 实体识别流程

Fig. 2　Entity recognition process of PERT-EGP

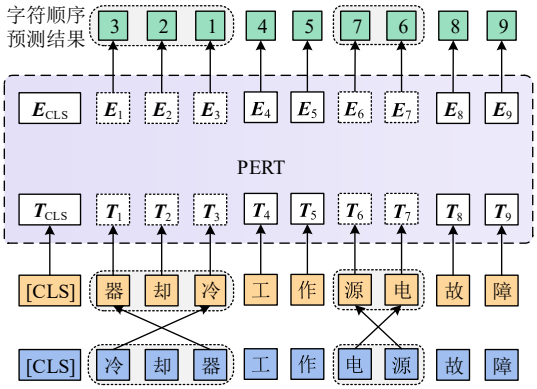

图 3　PERT 预训练原理

Fig. 3　Pre-training principles of PERT

2.2　PERT 字符编码流程

PERT 通过在大规模中文语料库进行预训练，获得了包含大量通用知识的初始字符向量。为了给每个字符向量融入中文变电设备缺陷文本的上下文语义信息，本文采用预训练后的 PERT 为每个字符进行动态字向量编码，该过程由 12 个首尾串联的 Transformer 编码器[23-24]实现。

PERT 动态字符向量编码原理如图 4 所示，PERT 动态编码层的输入由初始字符向量和位置向量组成，初始字符向量通过查询 PERT 预训练的字符索引表得到，位置向量由算法自动生成，二者叠加后输入编码层抽取文本特征。字符向量编码层通过 12 个堆叠的 Transformer 编码器来建模输入文本的语义信息，以生成每个字符的动态稠密向量。

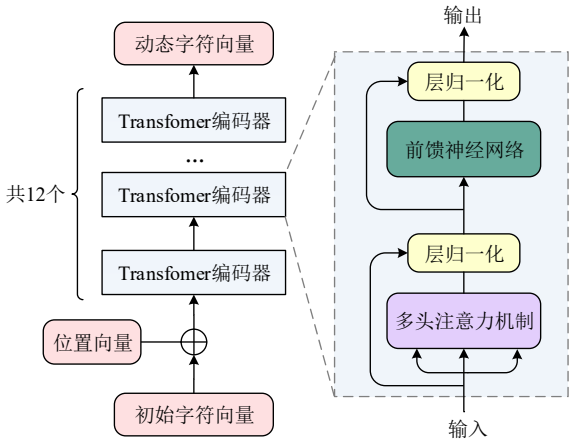

图 4　PERT 动态字符向量编码原理

Fig. 4　Dynamic character vector encoding principle of PERT

PERT 中的 Transformer 编码器结构在层归一化网络和前馈神经网络的基础上加入了多头注意力机制(multi-head attention，MHA)，该机制使得模型可以关注文本不同抽象级别上的特征，从而更好地处理变电设备缺陷本中存在的大量专业术语和复杂语言结构。

多头注意力机制模型结构如图 5 所示，多头注意力机制由自注意力层、拼接(Concat)层和线性变换(Linear)层组成[25]。其中，自注意力层包括点积(Matmul)模块和 Softmax 函数模块。

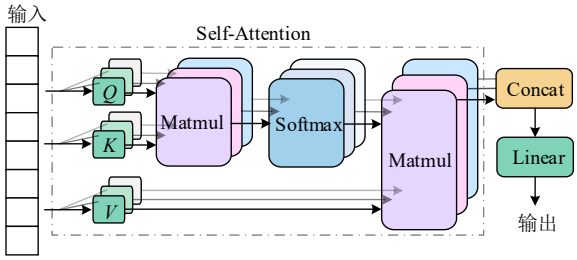

图 5　多头注意力机制模型结构

Fig. 5　Model structure of MHA

将长度为 T 的变电设备缺陷文本 $\boldsymbol{X}=(x_1,x_2,\ldots,x_T)$ 输入多个自注意力层中，计算出每层的注意力值矩阵。对第 m 个自注意力层初始化 3 个不同的线性投影矩阵 $\boldsymbol{W}_m^Q$、$\boldsymbol{W}_m^K$、$\boldsymbol{W}_m^V$，将输入文本序列 $\boldsymbol{X}$ 映射到问询 $\boldsymbol{Q}_m$、键 $\boldsymbol{K}_m$ 与值 $\boldsymbol{V}_m$ 矩阵，而后计算出每层的结果。计算过程如式(1)所示：

$$\begin{cases}\boldsymbol{Q}_m=\boldsymbol{X}\boldsymbol{W}_m^Q\\ \boldsymbol{K}_m=\boldsymbol{X}\boldsymbol{W}_m^K\\ \boldsymbol{V}_m=\boldsymbol{X}\boldsymbol{W}_m^V\\ \boldsymbol{H}_m=\mathrm{Softmax}\left(\dfrac{\boldsymbol{Q}_m\boldsymbol{K}_m^{\mathrm{T}}}{\sqrt{d}}\right)\boldsymbol{V}_m\end{cases}\tag{1}$$

式中：$\boldsymbol{H}_m$ 为第 m 个自注意力层的输出结果；d 为 3 个线性投影矩阵的维数；Softmax() 函数用于限制矩阵数值范围。

将多个自注意力层计算结果进行拼接，进而将拼接结果进行线性变化，输出注意力值矩阵：

$$\boldsymbol{H}=\mathrm{Concat}(\boldsymbol{H}_1,\boldsymbol{H}_2,\ldots,\boldsymbol{H}_m)\boldsymbol{W}^{\mathrm{o}}\tag{2}$$

式中：$\boldsymbol{W}^{\mathrm{o}}$ 为线性变换矩阵。

2.3　全局指针网络实体解码流程

经过 PERT 编码生成缺陷文本的动态字符向量表示后，本文采用全局指针网络精确定位关键信息，并学习各类型实体在上下文中的分布特征，从而识别出实体的首尾位置。其可提取出缺陷文本中所有可能为实体的文本片段，然后把这些片段输入各个类型的实体分类器进行判别，以确定文本片段所属的实体类型，从而完成文本中所有类型实体的识别。相较于 CRF，全局指针网络不需要根据上一状态计算下一状态，实体分类过程可以并行计算，以有效降低模型的时间复杂度。

全局指针网络[26]采用基于跨度的标注方式，相比于传统的序列标注，该标注策略不再单独标记每个字符，而是直接标记文本中的实体跨度，即实体的起始和结束位置，使模型可直观地捕捉到实体的整体性。这种标注方式将输入文本表示为上三角矩阵，矩阵中的每一个格子对应一个实体序列，将实体首尾作为一个整体考虑。如图 6 中的现象矩阵，第 1 行第 4 列数值为 1，表示输入文本“硅胶变色超过三分之二”中位置索引 1 至 4 的字符构成现象实体，故提取出的实体为“硅胶变色”；同理，在程度矩阵中抽取出的程度实体为“超过三分之二”。若矩阵中数值全为 0，代表文本中不含相应实体。

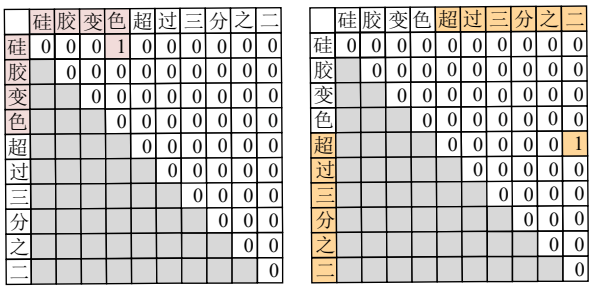

图 6　全局指针网络标注示例

Fig. 6　Annotation example of global pointer network

假设只有一种实体类型，$\boldsymbol{p}_i$ 和 $\boldsymbol{p}_j$ 为输入缺陷文本第 i 和 j 处的字符向量表示，经过起始索引前馈层和结束索引前馈层分别得到该片段的起始位置表征向量 $\boldsymbol{u}_i$ 和结束位置表征向量 $\boldsymbol{z}_j$，则缺陷文本序列第 i 位至第 j 位之间的片段被视为该类型实体的打分函数 $S(i,j)$，计算如下：

$$\boldsymbol{u}_i=\boldsymbol{\omega}_u\boldsymbol{p}_i+\boldsymbol{b}_u\tag{3}$$

$$\boldsymbol{z}_j=\boldsymbol{\omega}_z\boldsymbol{p}_j+\boldsymbol{b}_z\tag{4}$$

$$S(i,j)=\boldsymbol{u}_i^{\mathrm{T}}\boldsymbol{z}_j\tag{5}$$

式中：$\boldsymbol{\omega}_u$ 和 $\boldsymbol{b}_u$ 分别为起始索引前馈层的权重和偏置；$\boldsymbol{\omega}_z$ 和 $\boldsymbol{b}_z$ 分别为结束索引前馈层的权重和偏置。

全局指针网络在面对规模较小的语料库时，可能遭遇性能不佳的问题，这一局限性源自于网络结构中没有包含相对位置编码的机制。在没有位置信息指导下，网络的分类器有可能将实体的起始和终止位置进行随机匹配，导致识别出错误的实体组合。例如，对于图 6 中的文本，模型识别出实体起始位置在 1 和 5，实体终止位置为 4 和 10，在预测时可能会将 1 和 10 作为实体首尾位置，抽取出“硅胶变色超过三分之二”这一错误的现象实体。为了解决此问题，本文采用旋转式位置编码(rotary position embedding，RoPE)[27]将相对位置信息引入全局指针网络中，增强了模型对于实体长度的敏感性，以减少实体边界随机匹配的情况。

RoPE 通过将旋转式位置编码变换矩阵与起始位置表征向量 $\boldsymbol{u}_i$ 和结束位置表征向量 $\boldsymbol{z}_j$ 进行乘积操

作，可以使得打分函数变换为关于相对位置$(j-i)$的函数，从而融入实体的相对位置信息。

对于文本中的第 n 处字符，其对应的旋转式位置编码变换矩阵 $\boldsymbol{R}_n$ 如式(6)所示：

$$\boldsymbol{R}_n=\begin{pmatrix}\cos n\theta_1 & -\sin n\theta_1 & 0 & 0 & \cdots & 0 & 0\\ \sin n\theta_1 & \cos n\theta_1 & 0 & 0 & \cdots & 0 & 0\\ 0 & 0 & \cos n\theta_2 & -\sin n\theta_2 & \cdots & 0 & 0\\ 0 & 0 & \sin n\theta_2 & \cos n\theta_2 & \cdots & 0 & 0\\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots\\ 0 & 0 & 0 & 0 & \cdots & \cos n\theta_{\frac{d}{2}} & -\sin n\theta_{\frac{d}{2}}\\ 0 & 0 & 0 & 0 & \cdots & \sin n\theta_{\frac{d}{2}} & \cos n\theta_{\frac{d}{2}}\end{pmatrix} \tag{6}$$

式中：$\theta_t=10000^{-2(t-1)/d}, t\in[1,2,\ldots,d/2]$；$d$ 为该字符对应位置表征向量的维数，是一个偶数。

根据式(6)计算出实体起始位置和结束位置对应的旋转式位置编码矩阵 $\boldsymbol{R}_i$ 和 $\boldsymbol{R}_j$ 存在以下关系：

$$\boldsymbol{R}_i^{\mathrm{T}}\boldsymbol{R}_j=\boldsymbol{R}_{j-i} \tag{7}$$

引入相对位置编码后的打分函数 $S'(i,j)$ 如下：

$$S'(i,j)=(\boldsymbol{R}_i\boldsymbol{u}_i)^{\mathrm{T}}(\boldsymbol{R}_j\boldsymbol{z}_j)=\boldsymbol{u}_i^{\mathrm{T}}\boldsymbol{R}_{j-i}\boldsymbol{z}_j \tag{8}$$

全局指针网络采用的损失函数特别考虑了类别分布的均衡性，避免了将多类别的实体识别问题简化为若干独立的二分类问题所带来的类别不平衡问题。损失函数通过计算目标类别得分与非目标类别得分之间的差异，采取对比学习的方式，实现了各类别权重的自我调节。这种机制不仅平衡了各个类别之间的重要性，还提高了模型在处理不同数据分布时的泛化能力，使得全局指针网络在实体识别任务中，尤其存在丰富类别和不均衡数据集时，具有更好的适应性和鲁棒性。对于输入缺陷文本，假设需识别文本中存在的类别标签为 a 的实体，其损失函数 L 如式(9)所示：

$$L=\ln\left(1+\sum_{\chi_{(i,j)}\in G_a}e^{-s'_a(i,j)}\right)+\ln\left(1+\sum_{\chi_{(i,j)}\in U_a}e^{s'_a(i,j)}\right) \tag{9}$$

式中：$\chi_{(i,j)}$ 是以输入缺陷文本中第 i 处的字符作为首字符，第 j 处的字符作为尾字符，所组成的文本片段；$S'_a(i,j)$ 是文本片段 $\chi_{(i,j)}$ 相对于类别 a 的打分函数；G_a 是该缺陷文本中所有类别标签为 a 的实体集合；U_a 是该缺陷文本中所有非实体片段或类型非 a 的实体构成的集合。

2.4　参数共享优化的高效全局指针(EGP)

虽然全局指针网络在资源丰富的计算环境中，能通过内存或显存的充足供给，实现推理速度的优化，其性能提升足以弥补空间资源的额外消耗。然而，该网络在减少参数量以实现模型压缩方面尚存在提升空间。特别在标注层，过多的参数可能在训练数据量不足且类别众多的场景下，导致模型过拟合训练数据的特定特征，无法泛化到新的数据样本上。为缓解该问题，本文引入一种参数共享优化策略，使全局指针网络能够在更少的参数下执行。

具体而言，实体识别任务可视为 2 个子任务，包括提取候选实体的抽取和对候选实体的类别进行预测。候选实体抽取过程可被视为一个二分类问题，旨在判断给定的文本子序列是否构成一个实体。上述过程可为所有实体类别共用参数，从而实现参数共享。由于候选实体抽取阶段的参数在不同实体类别间共享，模型的参数量主要受实体类别预测阶段的影响。这意味着即便实体类别的数量增加，增加的参数量也集中在实体类别预测阶段。因此，这种划分子任务的策略有效降低了整体模型的参数量和计算复杂度，尤其在面对实体类别繁多的情况下，该方法能显著减轻模型负担，提高处理效率和灵活性。记优化后的全局指针网络为高效全局指针网络(Efficient Global Pointer)，其打分函数 $S''(i,j)$ 如下：

$$S''(i,j)=\boldsymbol{u}_i^{\mathrm{T}}\boldsymbol{z}_j+\boldsymbol{w}^{\mathrm{T}}\boldsymbol{\lambda}_{i,j} \tag{10}$$

式中：$\boldsymbol{w}$ 为针对实体类别的权重向量；$\boldsymbol{\lambda}_{i,j}$ 为该片段的表示向量。

2.5　评价指标

为了准确评估本文所提方法的实体识别效果，使用精确率(δ_{Pre})、召回率(δ_{Rec})和 F_1 分数(δ_{F1})作为评价指标衡量缺陷文本实体识别模型的性能。

其中，精确率(δ_{Pre})是模型正确识别的实体数量与识别出的总实体数量之比，能够反映模型正确识别各类实体的能力；召回率(δ_{Rec})是模型正确识别的实体数量与全部待识别实体总数量之比，能够反映模型找全各类实体的能力；F_1 分数(δ_{F1})是δ_{Pre}和δ_{Rec}的调和平均值，用于综合评估模型在准确性和完整性方面的表现。

3　算例分析

为验证本文 PERT-EGP 方法对变压器缺陷文本

实体识别的有效性，以国内某省份的电力变压器缺陷文本数据为例进行验证分析，变压器电压等级包括交流 500kV 及 1000kV。这些缺陷数据以非结构化文本的形式存在，具体是指变压器在运行、维护和检修过程中被记录下的缺陷或故障文本信息。缺陷文本记录时间为 2013-07-02 至 2022-07-23，共包括 1397 条数据，数据集按照 8:1:1 的比例分为训练集、验证集和测试集。

为方便与基于序列标注的实体识别方法进行比较，使用经典的 BIO(Begin, Inside, Outside)模式和片段标记格式对文本数据进行实体标注，共包括缺陷设备、缺陷部件、缺陷部位、缺陷现象、缺陷程度 5 类实体。为方便描述，后续简称为设备、部件、部位、现象、程度实体。所有实验均在配置为 Intel Xeon w5-2465X CPU、NVIDIA RTX4090 GPU、64GB DDR5 RAM 的服务器上完成。

3.1 实验数据集分布

数据集中的缺陷文本长度分布如图 7 所示，文本长度从 0 到 250 个字符，大多数长度在 100 字符以下，最高频的长度在 30 至 50 个字符。这表明大部分缺陷文本相对简短，在实际的缺陷记录过程中，运检人员倾向于录入关键信息。仅有小部分情况复杂的缺陷，运检人员会对其进行详细描述。

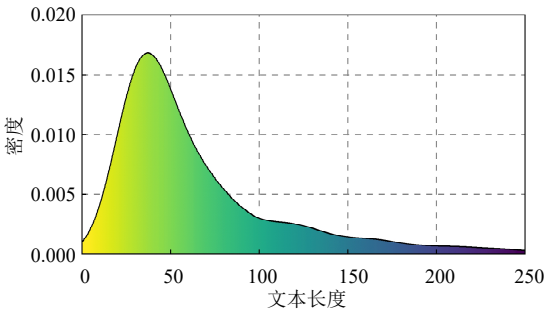

图 7 缺陷文本长度分布

Fig. 7 Length distribution of defect texts

标注完的数据集中，各类型实体长度分布如图 8 所示。由于电力领域对设备及其零部件的描述具体且精炼，设备、部件、部位实体的长度多集中于 2 至 6 个字符。而缺陷现象和缺陷程度涉及多种多样的状况，且受到运检人员记录习惯影响，对于同一个缺陷现象或缺陷程度记录的详略情况不一，因此这两类实体长度分布相对更广，集中在 2 至 14 个字符之间，甚至存在超过 15 个字符的描述。

数据集实体数量分布如表 1 所示。对于一条缺陷文本，运检人员均会记录缺陷发生的设备和现象，故这 2 类实体数量最多，分别达到 1396 和 1424。其中，现象实体大于缺陷文本数 1396，这是由于存

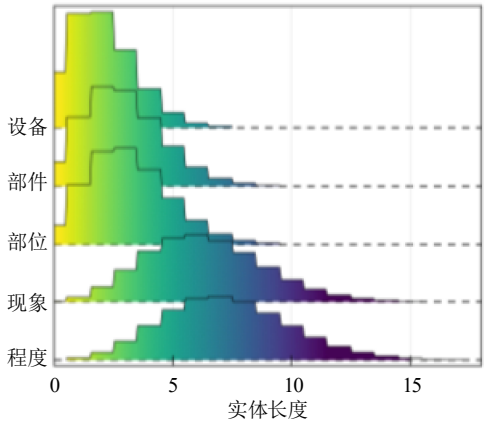

图 8 实体长度分布

Fig. 8 Length distribution of entities

表 1 数据集实体数量分布

Table 1 Distribution of the number of entities in the dataset

实体类型	训练集	验证集	测试集	合计
设备	1117	139	140	1396
部件	585	70	75	730
部位	802	95	92	989
现象	1138	144	142	1424
程度	245	37	35	317

在部分文本，一条文本记录了多个缺陷现象。正常情况下，部件和部位实体都应该被完整记录，但在实际记录中，由于运检人员描述习惯的差异，可能仅记录其中一部分，故这 2 类实体的数量偏少。另外，部位实体数量为 989，多于部件实体，这是由于缺陷部位是缺陷发生的具体位置，记录人主观上更倾向于描述自己观察的具体部位，易遗漏部件实体的记录。程度实体的数量最少，这是由于只有漏油、硅胶变色等部分特定缺陷现象，运检人员才会对其进行记录。

3.2 实体识别效果对比分析

为了充分验证本文方法在缺陷文本实体识别上的优越性，在测试集上将其与各类实体识别方法加以对比。包括传统机器学习模型隐马尔可夫模型(hidden Markov mode，HMM)、CRF；基于传统深度学习的模型膨胀卷积神经网络-条件随机场(iterated dilated convolutional neural network-conditional random field，IDCNN-CRF)、BiLSTM-CRF；基于 BERT 预训练的深度学习模型 BERT-IDCNN-CRF[28]、BERT-BiLSTM-CRF[29]；基于片段抽取范式的 PERT 编码机器阅读理解模型(machine reading comprehension based PERT，PERT-MRC)[30]、PERT 编码标签知识增强表示模型(label-knowledge enhanced representation based PERT，PERT-LEAR)[31]。所有模型的超参数均在验证集上加以优化，选择验

证集上最佳δ_{F1}的超参数为最终设置，用于测试集评估。本文所提 PERT-EGP 模型的最大文本输入长度为 384，批次大小为 16，PERT 层学习率为 0.0001，EGP 层学习率为 0.00001，慢热学习的比例和随机失活率设置为 0.2。

不同模型的总体识别效果对比如表 2 所示，由表可知，相比于其他实体识别模型，本文所提 PERT-EGP 模型的δ_{Pre}、δ_{Rec}和δ_{F1}分别为 91.21%、92.15%和 91.68%，在δ_{Rec}和δ_{F1}上均达到了最优，虽然δ_{Pre}低于 PERT-LEAR 模型 0.03%，但它的δ_{Rec}高出 3.93%，这意味着在保持相似的同等精度下，PERT-EGP 能够识别出更多的正确实体。从δ_{F1}上看，PERT-EGP 高出基于序列标注的模型(HMM、CRF 、IDCNN-CRF 、BiLSTM-CRF 、BERT-IDCNN-CRF 、BERT-BiLSTM-CRF)3.06%~25.12%，与基于片段抽取范式的模型(PERT-MRC、PERT-LEAR)对比也能高出约 2%，证明 PERT-EGP 模型能更好地学习缺陷文本语义信息，更有效地识别文本中的特定实体。

表 2　不同模型的总体识别效果对比
Table 2　Comparison of overall recognition effects of different models

模型	δ_{Pre}/%	δ_{Rec}/%	δ_{F1}/%
HMM	55.77	81.82	66.33
CRF	81.52	82.02	81.77
IDCNN-CRF	86.40	81.40	83.83
BiLSTM-CRF	85.48	85.12	85.30
BERT-IDCNN-CRF	88.30	88.84	88.57
BERT-BiLSTM-CRF	86.44	90.91	88.62
PERT-MRC	89.22	90.70	89.96
PERT-LEAR	**91.24**	88.22	89.71
PERT-EGP(本文方法)	91.21	**92.15**	**91.68**

基于序列标注范式的模型中，HMM 表现不佳，其δ_{F1}为 66.33%。CRF 的δ_{Pre}、δ_{Rec}和δ_{F1}相比 HMM 分别提高了 25.75%、0.2%和 15.44%，这是因为其通过对序列标签加入约束，有效解决了准确率低的问题。IDCNN-CRF、BiLSTM-CRF 模型分别通过 IDCNN 和 BiLSTM 实现了更有效的序列特征提取，以及更好的缺陷文本上下文建模能力，使得实体识别性能进一步提升。基于 BERT 预训练的模型，相比 IDCNN-CRF、BiLSTM-CRF 在δ_{F1}上获得了显著进步，分别提升了 4.74%、3.32%，这是由于 BERT 在处理每个字符时，能够综合缺陷文本前后文信息理解字符在特定语境中的含义，故在同义词、多义词的区分以及复杂语言结构理解上更具优势，使其在基于序列标注范式的模型中达到最优性能。

3.3　模型提升效果分析

为充分验证本文所提 PERT-EGP 模型在缺陷文本实体识别上的提升效果，从传统机器学习模型、基于传统深度学习的模型、基于 BERT 预训练的深度学习模型、基于片段抽取范式的模型中各选取δ_{F1}最高的 CRF、BiLSTM-CRF、BERT-BiLSTM- CRF 和 PERT-MRC 模型，对比各模型在不同实体类型及不同缺陷文本长度下的识别性能。

各模型在 5 类实体类型上的识别效果如图 9 所示，本文所提 PERT-EGP 模型在各类型实体的识别上都达到了最佳性能。对于设备实体，各模型的δ_{F1}都接近 100%，因为其表达统一、标准化、长度短，能够被轻易识别；对于部件和程度实体，因其在文本中的出现模式较为固定，最优δ_{F1}也能够达到 90%以上；对于部位实体，PERT-EGP 得益于能够学习实体内部语义信息的优势，在一些复合实体的识别上效果更佳，相较于 BERT-BiLSTM-CRF 和 PERT-MRC，其δ_{F1}提升了约 4%；对于现象实体，各模型的识别性能偏低，这是由于对缺陷现象的描述一般长度较长且叙述形式多样，给模型识别带来了困难。但是，PERT-EGP 和 PERT-MRC 在此类型实体的表现要明显优于 BERT-BiLSTM-CRF。由此可见，基于片段抽取范式的模型对于长度较长的实体具有更好的语义理解能力。

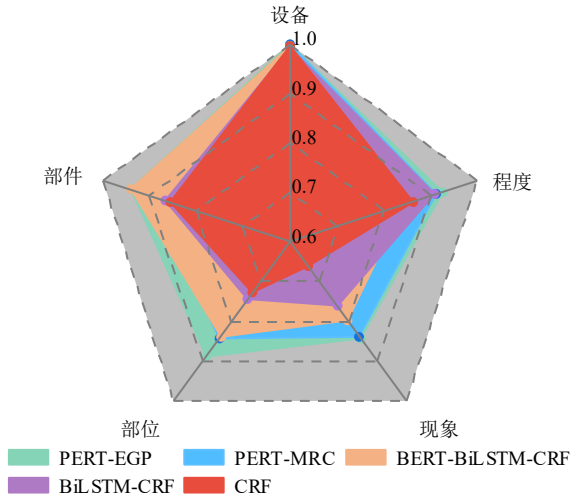

图 9　各模型在 5 类实体类型上的δ_{F1}
Fig. 9　δ_{F1} of each model on 5 types of entities

为了探究各模型在不同长度缺陷文本上的实体识别效果，采用四分位数将缺陷文本按长度分为 A(0~33 字)、B(33~48 字)、C(48~80 字)、D(80 字以上)4 类，各模型在 4 类不同长度文本上的识别效果如图 10 所示。总体来看，各模型在 0~33 字的短文本上识别效果最优，δ_{F1}都能够达到 90%以上。随着缺陷文本长度增加，各模型的识别效果逐渐递减，在 D 类文本上，基于序列标注范式的模型性能下降最为显著。本文所提 PERT-EGP 模型在 D 类长文本

上展现出明显的优势，较 BERT- BiLSTM-CRF 在δ_{F1}上高出 7.44%，较 PERT-MRC 高出 2.94%，证明本文方法在复杂、冗长的缺陷文本中具有更强的理解和识别实体的能力。

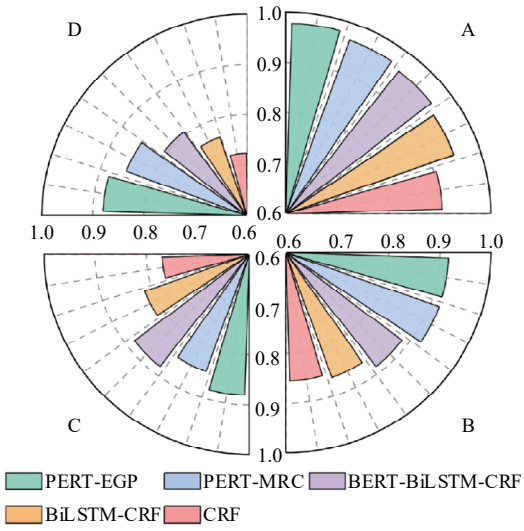

图 10 各模型在不同长度缺陷文本上的δ_{F1}

Fig. 10 δ_{F1} of each model on defect texts of different lengths

3.4 模型效率分析

本节通过模型的训练耗时以及测试耗时，对所提 PERT-EGP 模型的计算效率进行分析。

模型训练耗时以及相应的δ_{F1}如图 11 所示。可以看出，HMM 模型训练耗时不足 1min，在所有模型中所需时间最少，但其δ_{F1}仅为 66.33%。本文所提 PERT-EGP 模型次之，训练耗时为 13.41 min，同时在所有模型中δ_{F1}最高，表明其在缺陷文本实体识别任务中，能够兼顾效率与识别性能。对比 BERT-IDCNN-CRF 模型与 BERT-BiLSTM-CRF 模型，PERT-EGP 模型的训练时间降低了 49.42%、62.71%，这是由于高效全局指针网络不需要像 CRF 一样进行递归计算，可以并行处理实体分类，极大降低了模型的时间复杂度。PERT-MRC 的训练时间最长，其通过扩增样本的方式提高识别性能，导致训练时间大幅增加。

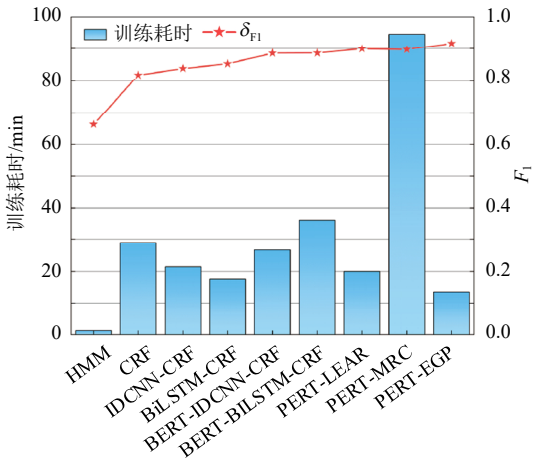

图 11 模型训练耗时分析

Fig. 11 Training time analysis of models

模型测试耗时以及相应的δ_{F1}如图 12 所示。HMM 模型测试耗时最长，达到 4.33s。其余未采用预训练的模型测试耗时在 1.35~1.98s。由于模型结构更加复杂，识别效果较好的预训练模型测试耗时稍长，在 2.52~3.62s。本文 PERT-EGP 模型测试耗时为 2.92s，进一步证明其在实际工程应用的适用性。

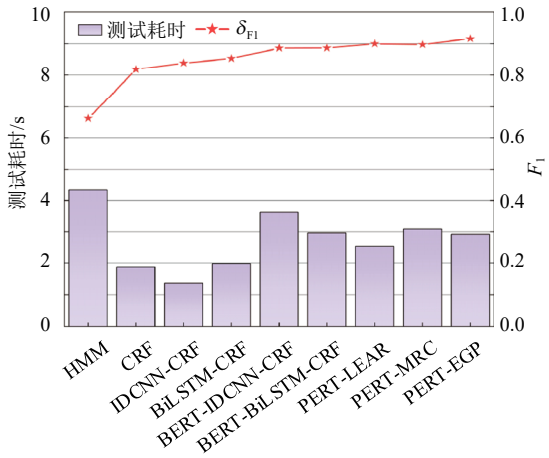

图 12 模型测试耗时分析

Fig. 12 Testing time analysis of models

3.5 PERT 有效性分析

为了验证本文方法融合 PERT 的编码方法在缺陷文本实体识别上的优越性，保持解码层不变，采用不同的预训练语言模型作为编码层进行实体识别。分别使用 BERT、引入全词掩码(whole word masking，WWM)策略的 BERT-WWM、引入全词掩码策略的鲁棒优化 BERT(a robustly optimized BERT pretraining approach with whole word masking，RoBERTa-WWM)与 PERT 进行比较，PERT 有效性分析如图 13 所示。

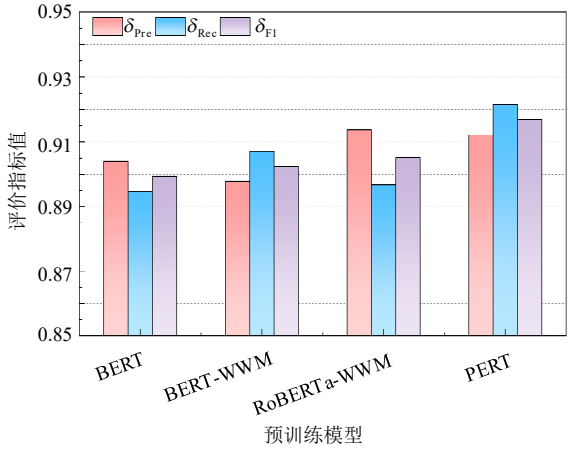

图 13 PERT 有效性分析

Fig. 13 Effectiveness analysis of PERT

由结果可知，基于 PERT 的编码方法综合效果最佳，相较 BERT、BERT-WWM、RoBERTa-WWM

方法，其在δ_{F1}上分别提升了 1.75%、1.44%和 1.17%。PERT 虽然在精确率上比 RoBERTa-WWM 编码方法低 0.16%，但 PERT 得益于乱序语言模型的预训练，其在预训练阶段的目标是预测被打乱字符的原始位置，该过程在无需引入掩码标记的情况下自监督地学习文本语义信息，有效减轻了预训练阶段与识别实体阶段之间的差异，因此能够识别出更多的实体，δ_{Rec}较 RoBERTa-WWM 提升了 2.48%，具备更优秀的综合性能。

3.6　EGP 有效性分析

为验证本文所用基于 EGP 的解码方法在提升缺陷文本实体识别性能上的有效性，以 PERT 作为编码层保持不变，采用不同的解码方法进行实体识别。分别使用 IDCNN-CRF、BiLSTM-CRF、MRC、LEAR 与 EGP 进行比较，对比结果如图 14 所示。

由图可知，相比其他解码方法，所提基于 EGP 的解码方法的精确率、召回率、F_1 分数分别为 91.21%、92.15%、91.68%，在召回率和 F_1 分数指标上达到了最优，精确率与最高的 LEAR 几乎持平。从 F_1 分数上看，基于 EGP 的解码方法要高出其他方法 1.72%~3.02%，证明其能够更好地捕捉实体在缺陷文本中的分布特征，从而更加有效地辨别出缺陷文本中的实体。

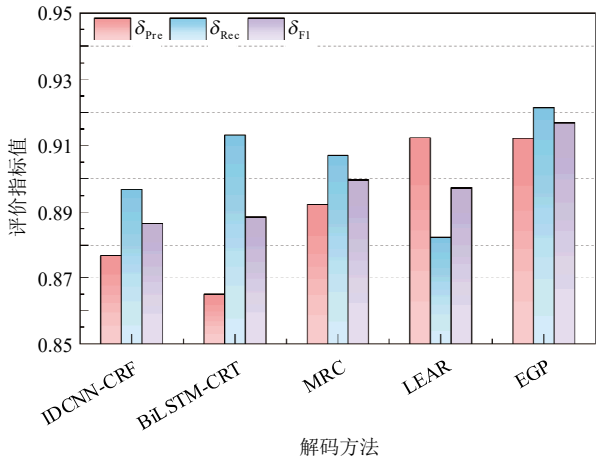

图 14　EGP 有效性分析

Fig. 14　Effectiveness analysis of EGP

3.7　典型样例分析

为了评估不同模型的实际识别效果，从测试集中选择出典型句子并对比分析不同模型的识别结果。本文方法 PERT-EGP 与 CRF、BiLSTM-CRF、BERT-BiLSTM-CRF、PERT-MRC 模型在样例文本中的识别结果如图 15 所示。

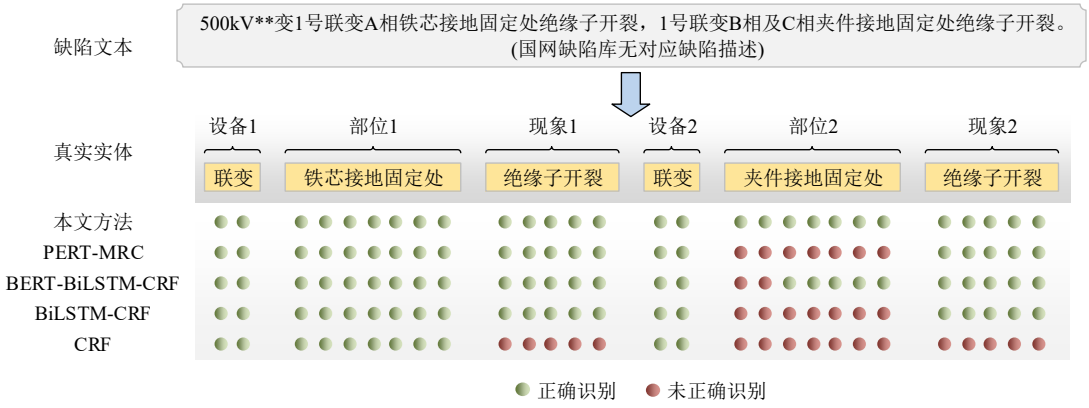

图 15　典型样例识别效果分析

Fig. 15　Analysis of typical sample recognition results

由图可知，本文方法正确识别出了该条缺陷文本中包含的各类实体。CRF 由于没有 BERT、BiLSTM 等上下文语义学习模块，将“绝缘子”误识别为部位实体，导致现象实体“绝缘子开裂”整体判断错误，同时未识别出“夹件接地固定处”这一部位实体。BiLSTM-CRF、BERT-BiLSTM-CRF、PERT-MRC 加强了对上下文语义的学习，成功识别出“绝缘子开裂”这一现象实体，但在对文本中位置偏后的部位实体“夹件接地固定处”的识别上都产生了错误。BiLSTM-CRF 和 PERT-MRC 遗漏了整个实体，BERT-BiLSTM-CRF 只识别出“接地固定处”而忽略了字符“夹件”，这是对实体组成特征学习不充分所导致。本文 PERT-EGP 方法通过高效全局指针加强了对实体内部语义特征的学习，对于由字符“夹件”和字符“接地固定处”组成的复合实体“夹件接地固定处”仍能够正确识别。

4　结论

针对现有缺陷文本实体识别方法对复杂实体内部依赖关系考虑不足以及计算效率低的问题，本文提出一种融合 PERT 与 EGP 的实体识别方法(PERT-EGP)。PERT 和 EGP 分别作为语义编码层和信息解码层，充分提取实体内部的语义依赖关系及其在缺陷文本中的分布特征。本文采用片段抽取的

新范式，优化实体识别准确率与模型复杂度，以提升实际工程中的实用性。结论如下：

1）相比于 BiLSTM-CRF、BERT-BiLSTM-CRF 和 PERT-MRC 等多类模型，本文模型采用矩阵标记文本中的实体跨度，使其可以直观地捕捉到实体的整体性，从而获得更佳的识别性能，在组成成分复杂的复合实体以及长度 80 字以上的冗长缺陷文本中，实体识别优势显著。

2）本文方法通过高效全局指针网络实体分类过程的并行计算以及参数共享策略，能够大幅缩短模型的训练时间。

3）相比于 BERT、BERT-WWM 等编码方法，基于 PERT 的编码方法能够有效减轻预训练阶段与识别实体阶段之间的差异，有助于提升模型实体识别效果。

参考文献

[1] 刘云鹏，许自强，李刚，等．人工智能驱动的数据分析技术在电力变压器状态检修中的应用综述[J]．高电压技术，2019，45(2)：337-348.
LIU Yunpeng，XU Ziqiang，LI Gang，et al．Review on applications of artificial intelligence driven data analysis technology in condition based maintenance of power transformers[J]．High Voltage Engineering，2019，45(2)：337-348(in Chinese).

[2] 缪希仁，林蔚青，肖洒，等．基于条件互信息与 LSTNET 的特高压变压器顶层油温预测方法[J]. 电网技术，2022，46(7)：2601-2609.
MIAO Xiren，LIN Weiqing，XIAO Sa，et al．Forecasting method for top oil temperature in ultra-high voltage transformers based on conditional mutual information and LSTNET[J]．Power System Technology，2022，46(7)：2601-2609(in Chinese).

[3] 林蔚青，缪希仁，肖洒，等．基于时空特征挖掘的特高压变压器热状态参量预测方法[J]．中国电机工程学报，2024，44(4)：1649-1661.
LIN Weiqing，MIAO Xiren，XIAO Sa，et al．Forecasting method for thermal state parameters in ultra-high voltage transformers based on spatial-temporal features mining[J]. Proceedings of the CSEE，2024，44(4)：1649-1661(in Chinese).

[4] TIAN Mingwei，ZHANG Lie，GUO Peng，et al．Data dependence analysis for defects data of relay protection devices based on Apriori algorithm[J]．IEEE Access，2020，8：120647-120653.

[5] ZHOU Bin，LI Jie，LI Xinyu，et al．Leveraging on causal knowledge for enhancing the root cause analysis of equipment spot inspection failures[J]．Advanced Engineering Informatics，2022，54：101799.

[6] CHEN Yanping，HU Ying，LI Yijing，et al．A boundary assembling method for nested biomedical named entity recognition[J]．IEEE Access，2020，8：214141-214152.

[7] WANG Huifang，CAO Jing，LIN Dongyang．Deep analysis of power equipment defects based on semantic framework text mining technology[J]．CSEE Journal of Power and Energy Systems，2022，8(4)：1157-1164.

[8] 曹靖，陈陆燊，邱剑，等．基于语义框架的电网缺陷文本挖掘技术及其应用[J]．电网技术，2017，41(2)：637-643.
CAO Jing，CHEN Lushen，QIU Jian，et al．Semantic framework-based defect text mining technique and application in power grid[J]．Power System Technology，2017，41(2)：637-643(in Chinese).

[9] 邵冠宇，王慧芳，吴向宏，等．基于依存句法分析的电力设备缺陷文本信息精确辨识方法[J]．电力系统自动化，2020，44(12)：178-185.
SHAO Guanyu，WANG Huifang，WU Xianghong，et al．Precise information identification method of power equipment defect text based on dependency parsing[J]．Automation of Electric Power Systems，2020，44(12)：178-185(in Chinese).

[10] 刘蓓，尚银辉，刘绚，等．配电线路跳闸填报文本智能挖掘方法[J]．高电压技术，2021，47(2)：445-453.
LIU Bei，SHANG Yinhui，LIU Xuan，et al．Inteligent mining method for line trip filling texts in distribution systems[J]．High Voltage Engineering，2021，47(2)：445-453(in Chinese).

[11] SANTISO S，PÉREZ A，CASILLAS A．Adverse drug reaction extraction：tolerance to entity recognition errors and sub-domain variants[J]. Computer Methods and Programs in Biomedicine，2021，199：105891.

[12] LI Jing，SUN Aixin，HAN Jianglei，et al．A survey on deep learning for named entity recognition[J]. IEEE Transactions on Knowledge and Data Engineering，2022，34(1)：50-70.

[13] AN Ying，XIA Xianyun，CHEN Xianlai，et al. Chinese clinical named entity recognition via multi-head self-attention based BiLSTM-CRF [J]．Artificial Intelligence in Medicine，2022，127：102282.

[14] LI Jianbin，FANG Suwan，REN Yuqi，et al. SWVBiL-CRF：selectable word vectors-based BiLSTM-CRF power defect text named entity recognition[C]//Proceedings of 2020 IEEE International Conference on Big Data (Big Data)．Atlanta：IEEE，2020：2502-2507.

[15] 丁禹，尚学伟，米为民．基于深度学习的电网调控文本知识抽取方法[J]．电力系统自动化，2020，44(24)：161-168.
DING Yu，SHANG Xuewei，MI Weimin．Deep learning based knowledge extraction method for text of power grid dispatch and control[J]．Automation of Electric Power Systems，2020，44(24)：161-168(in Chinese).

[16] 陈鹏，邰彬，石英，等．融合 BERT、双向长短记忆网络和条件随机场的电力设备缺陷文本实体抽取[J]．电网技术，2023，47(10)：4367-4375.
CHEN Peng，TAI Bin，SHI Ying，et al. Text entity extraction of power equipment defects based on BERT-BI-LSTM-CRF algorithm[J]．Power System Technology，2023，47(10)：4367-4375(in Chinese).

[17] 贾骏，杨强，付慧，等．基于电力设备大数据的预训练语言模型构建和文本语义分析[J]．中国电机工程学报，2023，43(3)：1027-1036.
JIA Jun，YANG Qiang，FU Hui，et al．Research on pre-training language model construction and text semantic analysis based on power equipment big data[J]. Proceedings of the CSEE，2023，43(3)：1027-1036(in Chinese).

[18] LI Ren，MO Tianjin，YANG Jianxi，et al．Bridge inspection named entity recognition via BERT and lexicon augmented machine reading comprehension neural model[J]．Advanced Engineering Informatics，2021，50：101416.

[19] ZHU Peng，CHENG Dawei，YANG Fangzhou，et al．Improving Chinese named entity recognition by large-scale syntactic dependency graph[J]．IEEE/ACM Transactions on Audio，Speech，and Language Processing，2022，30：979-991.

[20] 乔凯，陈可佳，陈景强．基于知识图谱与关键词注意机制的中文医疗问答匹配方法[J]. 模式识别与人工智能，2021，34(8)：733-741.
QIAO Kai，CHEN Kejia，CHEN Jingqiang．Chinese medical question answering matching method based on knowledge graph and keyword attention mechanism[J]．Pattern Recognition and Artificial Intelligence，2021，34(8)：733-741(in Chinese).

[21] 虞佳淼，王慧芳，张亦翔，等．基于 BERT 的电网现场作业风险自动评级方法[J]．电网技术，2023，47(11)：4746-4754．
YU Jiamiao，WANG Huifang，ZHANG Yixiang，et al．Automatic risk rating method for power grid field operation based on BERT[J]．Power System Technology，2023，47(11)：4746-4754(in Chinese)．

[22] WANG Xiaodi，LIU Jiayong．A novel feature integration and entity boundary detection for named entity recognition in cybersecurity[J]．Knowledge-Based Systems，2023，260：110114．

[23] 董家富，万雄，王岩，等．基于 XGB-Transformer 模型的短期电力负荷预测[J]．电力信息与通信技术，2023，21(1)：9-18．
DONG Jiafu, WAN Xiong , WANG Yan, et al．Short-term power load forecasting based on XGB-Transformer model[J]．Electric Power Information and Communication Technology，2023，21(1)：9-18(in Chinese)．

[24] 李妍，孟洁，何金，等．面向电力业务数据的命名实体识别[J]．电力信息与通信技术，2022，20(4)：24-31．
LI Yan，MENG Jie，HE Jin，et al．Named entity recognition for electric power business data[J]．Electric Power Information and Communication Technology，2022，20(4)：24-31(in Chinese)．

[25] 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：Curran Associates Inc．，2017：6000-6010．

[26] LI Hongjun，CHENG Mingzhe，YANG Zelin，et al．Named entity recognition for Chinese based on global pointer and adversarial training[J]．Scientific Reports，2023，13(1)：3242．

[27] SU Jianlin，AHMED M，LU Yu，et al．RoFormer：enhanced transformer with rotary position embedding[J]．Neurocomputing，2024，568：127063．

[28] 李妮，关焕梅，杨飘，等．基于 BERT-IDCNN-CRF 的中文命名实体识别方法[J]．山东大学学报(理学版)，2020，55(1)：102-109．
LI Ni，GUAN Huanmei，YANG Piao，et al．BERT-IDCNN-CRF for named entity recognition in Chinese[J]．Journal of Shandong University (Natural Science)，2020，55(1)：102-109(in Chinese)．

[29] 郭知鑫，邓小龙．基于 BERT-BiLSTM-CRF 的法律案件实体智能识别方法[J]．北京邮电大学学报，2021，44(4)：129-134．
GUO Zhixin，DENG Xiaolong．Intelligent identification method of legal case entity based on BERT-BiLSTM-CRF[J]．Journal of Beijing University of Posts and Telecommunications，2021，44(4)：129-134(in Chinese)．

[30] PENG Cheng，YANG Xi，YU Zehao，et al．Clinical concept and relation extraction using prompt-based machine reading comprehension[J]．Journal of the American Medical Informatics Association，2023，30(9)：1486-1493．

[31] YANG Pan，CONG Xin，SUN Zhenyu，et al．Enhanced language representation with label knowledge for span extraction[C]//Proceedings of 2021 Conference on Empirical Methods in Natural Language Processing．Association for Computational Linguistics，2021：4623-4635．

林蔚青

在线出版日期：2025-02-14。
收稿日期：2024-08-16。

作者简介：
林蔚青(1997)，男，博士研究生，研究方向为电气设备在线监测与故障诊断，E-mail：weiqing3306@qq.com；
郑垂锭(1998)，男，硕士，研究方向电力大数据分析及应用，E-mail：zhengchuiding @163.com；
陈静(1988)，女，通信作者，副教授，硕士生导师，研究方向为人工智能在电力系统中的应用、电力故障检测识别等，E-mail：chenj@ fzu.edu.cn。

（责任编辑　张京娜）