

基于改进CasRel模型的风力发电机故障领域 实体关系抽取研究

张友星^{1*}, 王兴芬^{2#}

¹北京信息科技大学计算机学院, 北京

²北京信息科技大学管理科学与工程学院, 北京

收稿日期: 2025年4月18日; 录用日期: 2025年5月20日; 发布日期: 2025年5月30日

摘要

针对风力发电机故障领域关系抽取任务存在大量的专业术语、单实体重叠和多实体重叠、单句子关系多以及句子结构复杂的问题, 本文提出一种改进的CasRel二元级联标记框架CA-CasRel。由于ERNIE模型在预训练过程中引入了实体信息和知识图谱信息, 本研究首先用ERNIE模型代替BERT模型进行句子编码嵌入, 深层次地捕捉文本序列的实体关系信息。另外, 本研究在预测头实体之后, 对头实体向量用余弦注意力机制进行编码, 余弦注意力具有方向敏感性和语义纯度强化, 可高效提取语义的本质特征和对稀疏特征的无偏捕捉。最终, 所提架构和基线CasRel模型相比, 其精确率、召回率和F1值分别提升了1.71%、6.09%和4.05%, 达到了领域内的SOTA效果。

关键词

风力发电机故障, CasRel, 余弦注意力, SOTA

Research on Entity Relationship Extraction in Wind Turbine Fault Domain Based on Improved CasRel Model

Youxing Zhang^{1*}, Xingfen Wang^{2#}

¹Computer School, Beijing Information Science and Technology University, Beijing

²School of Management Science and Engineering, Beijing Information Science and Technology University, Beijing

Received: Apr. 18th, 2025; accepted: May 20th, 2025; published: May 30th, 2025

*第一作者。

#通讯作者。

文章引用: 张友星, 王兴芬. 基于改进 CasRel 模型的风力发电机故障领域实体关系抽取研究[J]. 人工智能与机器人研究, 2025, 14(3): 670-683. DOI: 10.12677/airr.2025.143066

Abstract

In the task of relation extraction within the wind turbine fault domain, there are numerous challenges, such as the presence of domain-specific terminology, overlapping of single and multiple entities, multiple relations in a single sentence, and complex sentence structures. This paper proposes an improved CasRel-based binary tagging framework, CA-CasRel. Since the ERNIE model incorporates entity information and knowledge graph data during pre-training, this study replaces the BERT model with ERNIE for sentence encoding and embedding, enabling a deeper capture of entity-relation information in text sequences. Furthermore, after predicting the head entity, the head entity vector is encoded using a cosine attention mechanism. Cosine attention, with its direction sensitivity and enhanced semantic purity, efficiently captures essential semantic features and unbiasedly detects sparse features. Compared to the original CasRel model, the proposed architecture improves precision, recall, and F1 score by 1.71%, 6.09%, and 4.05%, respectively, achieving state-of-the-art performance in the domain.

Keywords

Wind Turbine Faults, CasRel, Cosine Attention, SOTA

Copyright © 2025 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

近几年来, 人工智能技术得到飞速的发展, 无人值守智慧化风电厂开始崭露头角, 我国作为风力发电的大国, 风力发电机作为风力发电的主要部件, 容易在运行过程中出现风机叶片磨损、断裂、腐蚀等故障。在对风电机进行维修和处理时存在海量的风力发电机维修手册、故障手册以及技术说明书等。这些文本数据当中存在大量有价值的信息, 研究如何从这些文本数据中提取结构化知识的实体关系抽取方法, 对风力发电机故障智能诊断和预警具有重要意义。

目前, 风力发电机故障语料存在大量的重复表达、逻辑不够清楚以及三元组重叠的问题。此外, 该语料存在大量的专业术语, 单句子关系多, 句子关系复杂。这些问题容易导致模型在识别深层次语义信息时发生错误, 从而直接影响模型的关系提取性能。例如, 在句子“检查油位计是否损坏或者辅助点接触不实”, 部件名称实体“油位计”和其他实体存在不同类型的关系, 一个部件既需要进行“故障分类”, 同时还需要“检修和操作”。另外, 句子中存在多对多关系, 并且句子中有超过 4 类的关系。例如, 在句子“检查齿轮, 如有断齿、崩齿时, 如磨损严重时进行更换”, 一个部件存在多种异常状态, 如“齿轮”部件, 存在“断齿”、“崩齿”和“磨损”三种异常状态, 在这个句子中同时还存在其他两种关系。上述情况下容易导致模型在识别过程中漏识别或错误识别关系分类。

针对上述问题, 本文以风力发电机故障领域为背景, 构建了该领域内数据集, 同时以 CasRel 模型为基准进行改进, 本研究具体贡献如下: ① 构建风力发电机故障领域实体关系抽取数据, 共定义 7 类实体和 10 类关系, 人工标注大约 1000 条数据。② 在词嵌入编码层引入 ERNIE 模型替代 BERT 模型, ERNIE 模型融入知识图谱的信息, 能够更有效地把握实体与实体之间的关联。③ 本研究在预测头实体之后, 对头实体向量用余弦注意力机制进行编码, 由于余弦注意力具有方向敏感性和语义纯度强化, 可深层次地

去捕捉语义的细粒度信息,一定程度上提升了模型的预测效果。

2. 相关工作

实体关系抽取[1]是实体识别中的下游任务,旨在从无规则杂乱的文本中自动识别并提取实体间的关系。在风力发电机故障文本中,关系抽取通过分析文本语句,将实体之间的关联以结构化形式呈现出来。例如,从文本“测量回路的接线松动”可抽取出主体“测量回路”,客体“接线松动”,以及它们之间的关系“设备状态异常”,从而为风力发电机故障领域知识图谱构建、信息检索、推荐系统等提供数据基础。

早期的实体关系抽取方法通常可分为:有监督学习、半监督学习、无监督学习。Zhang 等[2]提出了一种将支持向量机和解析树上的卷积核相结合的方案进行实体关系抽取。Hasegawa 等[3]从无规则文本中抽取实体,随后分析该实体和上下文的语义相似度,然后通过计算出的相似度进行聚类。近年来,随着深度学习的发展,逐渐替代传统的方案,具体可以分为流水线和联合提取的方案。2015年,Xu 等[4]使用 LSTM 进行实体关系抽取。这个方法利用了句法依存分析树中的最短路径,充分将词性、句子向量、句法等进行多特征融合,提升了模型的泛化能力。在2021年,Zhong 等[5]提出了使用两个编码器组成的 pipeline 模型,采取两个独立的预训练模型进行编码,只使用实体模型为关系模型提供输入特征,两个编码器分别用于实体识别和关系抽取任务,实体建模使用 Span 级别,该模型分别在数据集 ACE04、ACE05 和 SciERC 取得了令人满意的成绩。Yang 等[6]研究出一种端到端生物学三元组提取流水线方法。该方法使用基于跨度的检测策略,通过枚举可能的候选跨度和实体对来检测重叠三元组,最终取得了不错的效果。

流水线方法在实体关系抽取任务中存在一定的问题。该方法将实体识别和关系抽取划分为独立的步骤,可能会导致错误在各阶段之间传播,存在误差累计。为了解决上述问题,目前很多学者将工作重心转移到联合实体关系抽取。联合实体关系抽取[7]是指在同一模型框架下,同时进行实体抽取和关系抽取的任务。Wei 等[8]提出一个全新的角度,构建了级联二元标记框架,之前,大多数实体关系抽取任务将关系类别视为离散的标签,作者将关系建模视为将主体映射到句子中的客体的函数,该方法可以有效地解决关系重叠问题。Zheng 等[9]将实体关系任务进行分解为三个子任务,即关系分类、实体识别以及主宾语对齐,提出了一个基于潜在关系和全局对应关系的联合关系三元组提取框架,解决了提取泛化性能差的原因。Wadhwa 等[10]将大模型运用到关系抽取任务,提出了思维链(Chain of Thought,简称 CoT)[11]方式来微调 and 增强关系抽取。在操作时,人为给一些提示样例,交给 GPT-3 [12]模型,补充其他数据的解释,作为数据增强,接着交给 Flan-T5 [13]小一点的模型去微调学习。

基于图结构的联合实体关系抽取方法同样是一种关键技术。Wang 等[14]在研究中提出了一种基于图架构的联合学习模型,该模型通过引入带有偏置权重的损失函数来增强头实体与尾实体之间的语义关联,从而有效缓解了关系重叠问题。实体关系抽取的远程监督方法同样是一种不错的方案。Zeng 等[15]提出了 PCNN 模型,将远程监督学习看作多实例学习问题,通过 CNN 自动提取文本语义特征,有效地解决了数据标注错误和错误传播等问题。

在工业领域相关研究中,戴志辉等[16]提出一种基于 MacBERT-BiLSTM-CRF 的模型架构,用于构建缺陷知识图谱,为继电保护缺陷管理提供了辅助决策支持。针对汽车故障领域的关系抽取难题,文雯[17]提出了一种结合多轮迭代对抗攻击和注入实体知识嵌入的端到端模型,以提升抽取精度。实验结果表明,该方法 F1 值相对于传统模型,有大幅度提升。

尽管目前在通用领域的实体关系抽取方法比较完善,但是对于风力发电机故障领域的研究相对较少,本文提出一种创新性的研究方法,在一定程度上解决了风力发电机故障领域实体关系抽取问题。

3. 模型架构设计

为了解决风力发电机故障语料中存在的关系重叠以及单句子关系多的问题, 本研究提出了一种改进的二元级联关系抽取架构 CA-CasRel, 模型结构在 CasRel 的基础上改变了嵌入模型和融入了余弦注意力机制, 其具体模型架构如图 1 所示。

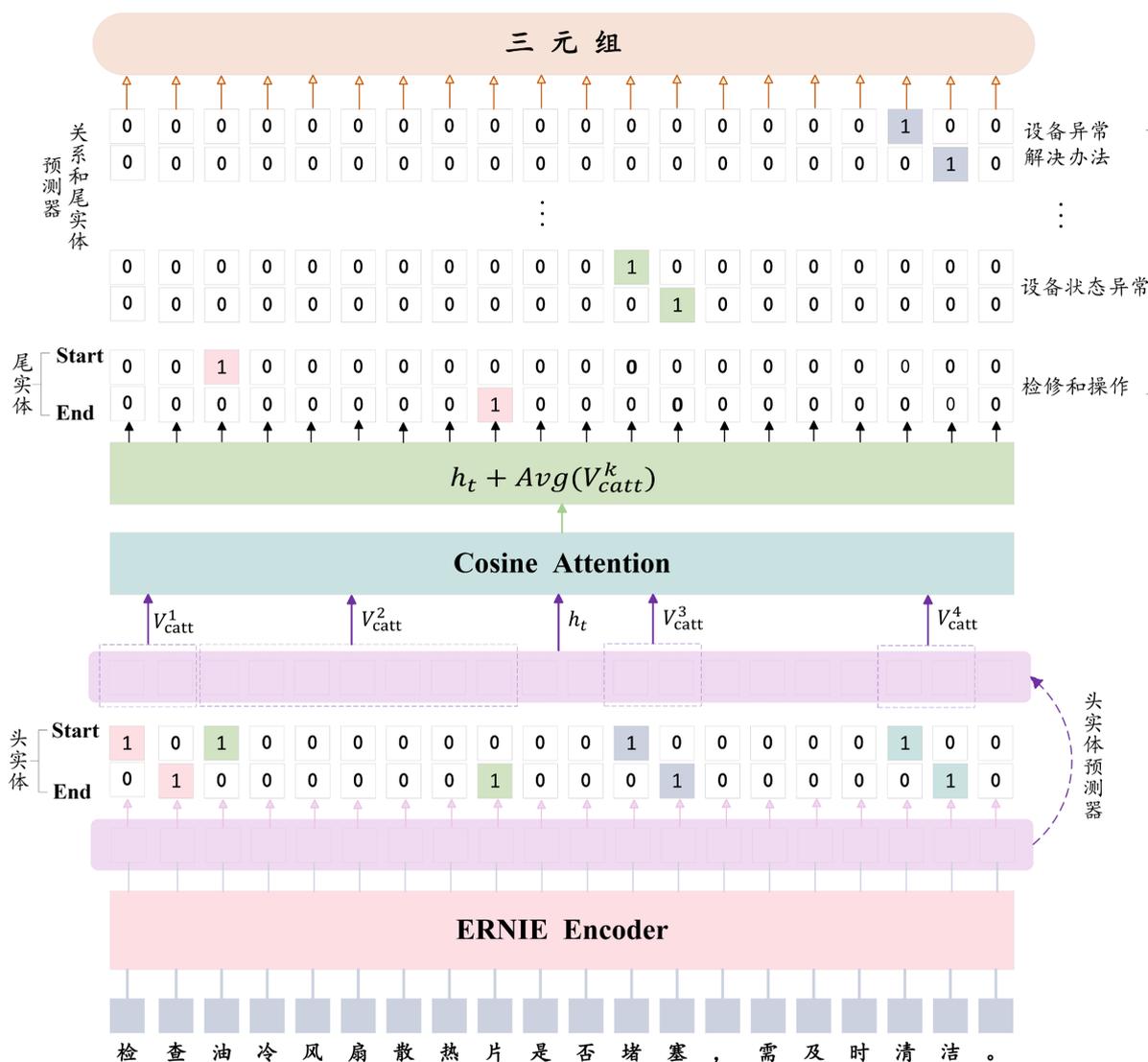


Figure 1. Overall architecture of CA-CasRel model

图 1. CA-CasRel 模型整体架构图

模型的整体架构主要分成四部分。首先, 通过 ERNIE 模型对输入文本进行编码处理, 提取文本的上下文相关特征, 同时生成每个词语的隐含向量表示。随后, 利用头实体预测模块对每个词语执行二分类任务, 判别其是否构成实体的起始或终止边界, 进而实现全局实体范围的识别定位。该方法采用双边界判定机制, 可有效避免实体重叠现象。然后, 在提取尾实体之前, 引入余弦注意力机制, 对头实体位置信息进行特征提取, 使用文本相似度对每个词向量计算相关性, 改变权重分布, 将头实体特征和全局上下文表示进行相加。最后, 使用关系和尾实体预测器预测出全部的关系分类以及对应的尾实体。最终经

过输出层, 得到所有的三元组信息。下面将依次介绍 ERNIE 嵌入层、头实体预测器、余弦注意力以及关系和尾实体预测器。

3.1. ERNIE 嵌入层

由于风力发电机故障语料领域内的特殊性, 并且数据集中存在大量的三元组重叠问题, 尤其是多对多关系较多, 传统的 BERT 模型在预训练过程中只学习了上下文的语义信息, 对实体信息、知识的显示捕捉不足。ERNIE 通过引入知识图谱和实体级别知识, 能够更好地理解实体及其上下文信息, 从而增强模型对实体的语义表示能力, 因此使用 ERNIE 模型代替 BERT 模型。对于关系重叠问题, 尤其是实体重叠和关系重叠类型的复杂情况, ERNIE 能通过知识增强机制更准确地捕捉实体和关系之间的语义关联。例如对于“电流传感器”来说, ERNIE 模型可以通过知识图谱去区分“电流”和“电流传感器”, 电流作为参数信息, 而“电流传感器”作为部件名称。另外, ERNIE 模型采用知识掩码方法, 可以同时捕捉字、词、短语和实体不同的细粒度的语义信息, 在解决实体重叠和关系嵌套问题上具有明显的效果。ERNIE 模型架构如图 2 所示。

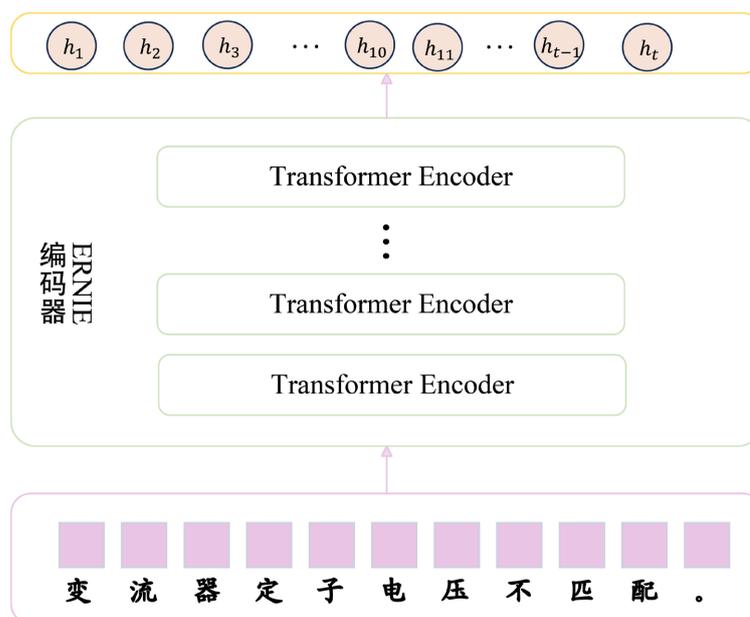


Figure 2. Overall architecture of the ERNIE model
图 2. ERNIE 模型整体架构

如图 2 所示, 模型的输入是一个句子, 首先对句子进行填充操作, 使其长度保持一致, 假设输入的文本序列表示为 $S = \{c_1, c_2, c_3, \dots, c_n\}$, 再经过 ERNIE 模型编码之后的语义向量 X :

$$X = \text{ERNIE}(S) \quad (1)$$

3.2. 头实体预测器

经过 ERNIE 模型对输入文本进行深度特征抽取后, 可获得蕴含上下文关联的语义表征序列 $H = \{h_1, h_2, h_3, \dots, h_t\}$, 随后, 将句子的上下文表示输入头实体预测器, 以识别句子中所有可能存在的实体。首先, 使用一个二分类层, 对每个词的位置信息, 分别识别出是否为头实体的起始位置和结束位置。然后, 使用激活函数, 对每个位置向量进行二分类, 输出两个概率分布矩阵。最后, 基于识别结果, 对概率

矩阵进行阈值判断, 将头实体起始位置设置为 1, 头实体结束位置设置为 0, 得到所有头实体的起始位置和结束位置。其具体的计算公式如下:

$$P_i^{\text{start},n} = \sigma(W_{\text{start}}m_i + b_{\text{start}}) \quad (2)$$

$$P_i^{\text{end},n} = \sigma(W_{\text{end}}m_i + b_{\text{end}}) \quad (3)$$

其中, W_{start} 和 W_{end} 分别表示训练权重, 主要作用是进行特征筛选, b_{start} 和 b_{end} 表示偏置值, m_i 代表第 i 个位置对应的向量。 $P_i^{\text{start},n}$ 和 $P_i^{\text{end},n}$ 分别表示第 i 个头实体位置的起始信息和结束信息, σ 表示 sigmoid 激活函数, 将输入映射到大于 0 和小于 1 的区间, 实现概率输出或非线性变换。

3.3. 余弦注意力

在 2017 年, Transformer 架构的横空出世, 注意力机制作为 Transformer 的核心组成部分, 起到至关重要的作用。注意力机制可以确保模型聚焦于关键实体, 比如“轴承”和“叶片”等, 其次还可以捕捉句子的上下文信息。基于注意力机制对文本单元进行动态加权, 引导模型聚焦于领域关键特征。然而, 注意力机制在学习的时候, 对幅度较大的向量赋予更高的权重, 忽略了语义相关性较强但幅度较小的特征。很难去识别到风力发电机故障领域占比较少的部件。因此, 本研究使用改进的注意力机制, 使用余弦相似度进行归一化得分, 使用余弦归一化替代自注意力机制中的点积计算。这样, 余弦注意力关注不同向量之间的方向相似程度, 衡量向量是否在相同方向, 对向量的幅度大小降低关注。这种特性使其对语义相关性更加敏感, 尤其在处理长文本时, 能够更好地捕捉全局语义信息。其架构如图 3 所示。

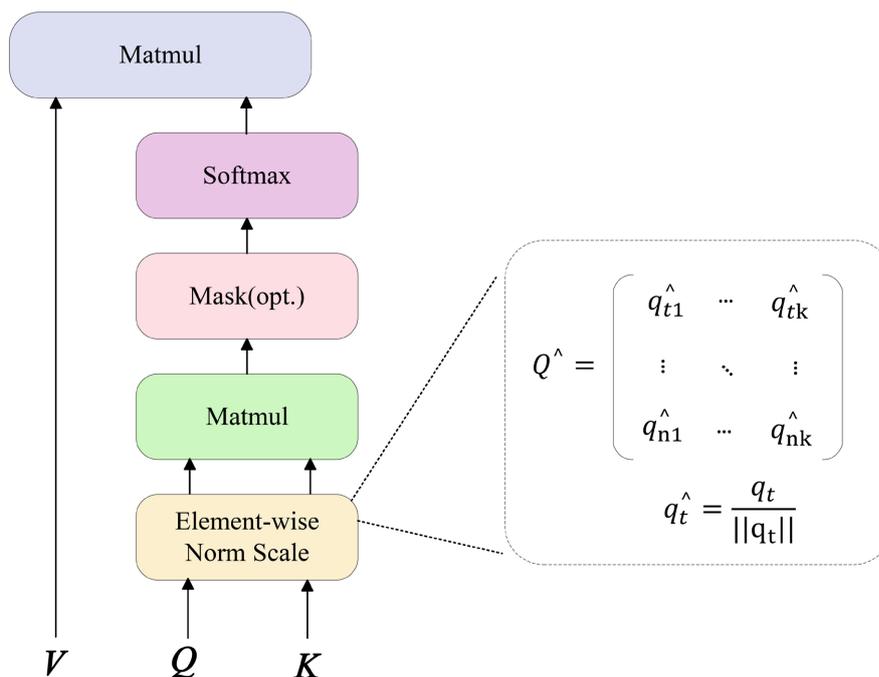


Figure 3. Cosine attention mechanism model structure diagram
图 3. 余弦注意力机制模型结构图

在自注意力机制中对键和查询向量执行矩阵乘法之前, 本文提出的方法首先将这两个向量除以它们各自的范数, 然后构造矩阵。 Q 和 K 是由除以各自范数的向量构成的查询矩阵和键矩阵。利用余弦注意力机制对头实体向量特征捕捉, 具体的计算如下:

$$\text{cosine}(\hat{m}, \hat{n}) = \frac{m}{\|m\|} \cdot \frac{n}{\|n\|} \quad (4)$$

$$\text{Attn}(q_i, k_i, v_i) = \sum_{i=1}^n \text{softmax}(\text{cosine}(\hat{q}_i, \hat{k}_i)) v_i \quad (5)$$

其中, 公式(4)为标准的余弦相似度公式。在公式(5)中, 首先对查询向量 q_i 和每个键向量 k_i 进行余弦相似度的计算, 然后对所有的键向量的余弦相似度进行归一化操作, 将相似度的值归一化概率分布, 作为注意力权重, 最后用注意力权重矩阵和值向量 v_i 进行动态加权求和, 将注意力 v_{catt}^k 作为最终结果, 与第一部分实体识别得到的实体向量进行相加输入到关系和尾实体预测器当中。

3.4. 关系和尾实体预测器

在 CA-CasRel 框架中, 关系和尾实体预测器的主要作用是根据识别出来的头实体, 预测其对应的关系以及尾实体。首先, 通过头实体的特征向量 h_i 和余弦注意力机制从上下文中聚合的关系特征, 取均值表示所有的关系类别的整体表示, 对所有的关系进行预测, 将所有与头实体相关的关系标记。然后, 根据已经得到的头实体和关系标签, 进一步预测尾实体的开始位置和结束位置。具体计算过程如下:

$$p_i^{\text{end}, n} = \sigma(W_{\text{end}}^r (m_i + \text{Avg}(v_{\text{catt}}^k))) + b_{\text{end}}^r \quad (6)$$

$$p_i^{\text{start}, n} = \sigma(W_{\text{start}}^r (m_i + \text{Avg}(v_{\text{catt}}^k))) + b_{\text{start}}^r \quad (7)$$

根据上述公式可以得到尾实体的起始位置和结束位置的概率分布, 是关系抽取任务中尾实体位置预测模块的核心公式。 $p_i^{\text{end}, n}$ 代表第 n 个关系中, 第 i 个位置是尾实体的结束索引的概率。 $p_i^{\text{start}, n}$ 表示第 n 个关系中, 第 i 个位置是尾实体起始索引的概率。 b_{end}^r 和 b_{start}^r 代表偏置值, v_{catt}^k 通过余弦注意力机制对上下文信息进行加权聚合后得到第 k 个头实体特征信息。这两个公式通过文本的局部和全局关系特征, 分别预测尾实体的起始位置和结束位置。这样的设计使模型能够更准确地处理复杂的关系抽取任务, 特别是在头实体与多个尾实体或关系重叠的场景中表现很好。

4. 实验

4.1. 数据集

本文数据来源于某智能风电厂提供的风力发电机维修手册、故障手册和运维报告等文本语料, 此外包括从《百度百科》和《维基百科》上查找的风电行业相关论文以及风电行业标准等文本。两者共同作为本文研究的数据基础。参考风电行业相关书籍、通用名称实体识别方法以及部分国际标准, 并且根据所提供的故障文本特点, 共定义了部件名称(PN)、故障类型(FT)、参数信息(PI)、维修方法(MRM)、正常状态(NS)、异常状态(AS)和标准和协议(SP) 7 个实体类。在 7 个实体类型基础上, 定义 10 类关系, 具体如下图 4 所示。

其中共标注 1035 条关系抽取数据集, 其中训练集、验证集和测试集的比例为 8:1:1, 具体关系数量分布和样例如表 1 所示。

Table 1. Number distribution and examples of relationship types

表 1. 关系类型数量分布和样例

关系类型	数量	样例
故障分类	232	{“主控系统安全链”, “故障分类”, “断开”}
属性信息	316	{“水冷系统”, “属性信息”, “进阀压力”}

续表

检修和操作	511	{“检查”, “检修和操作”, “变桨电机轴承”}
设备状态正常	158	{“控制柜”, “设备状态正常”, “完好”}
设备状态异常	538	{“齿轮”, “设备状态异常”, “断齿”}
符合和遵循	83	{“制动系统”, “符合和遵循”, “JBT10426.1-2004”}
参数信息状态	185	{“相电压”, “参数信息状态”, “跌落”}
参数调整	89	{“调节”, “参数调整”, “转速”}
故障类型解决方案	78	{“损坏”, “故障类型解决办法”, “更换”}
设备异常解决方案	109	{“过高”, “设备异常解决办法”, “检查”}

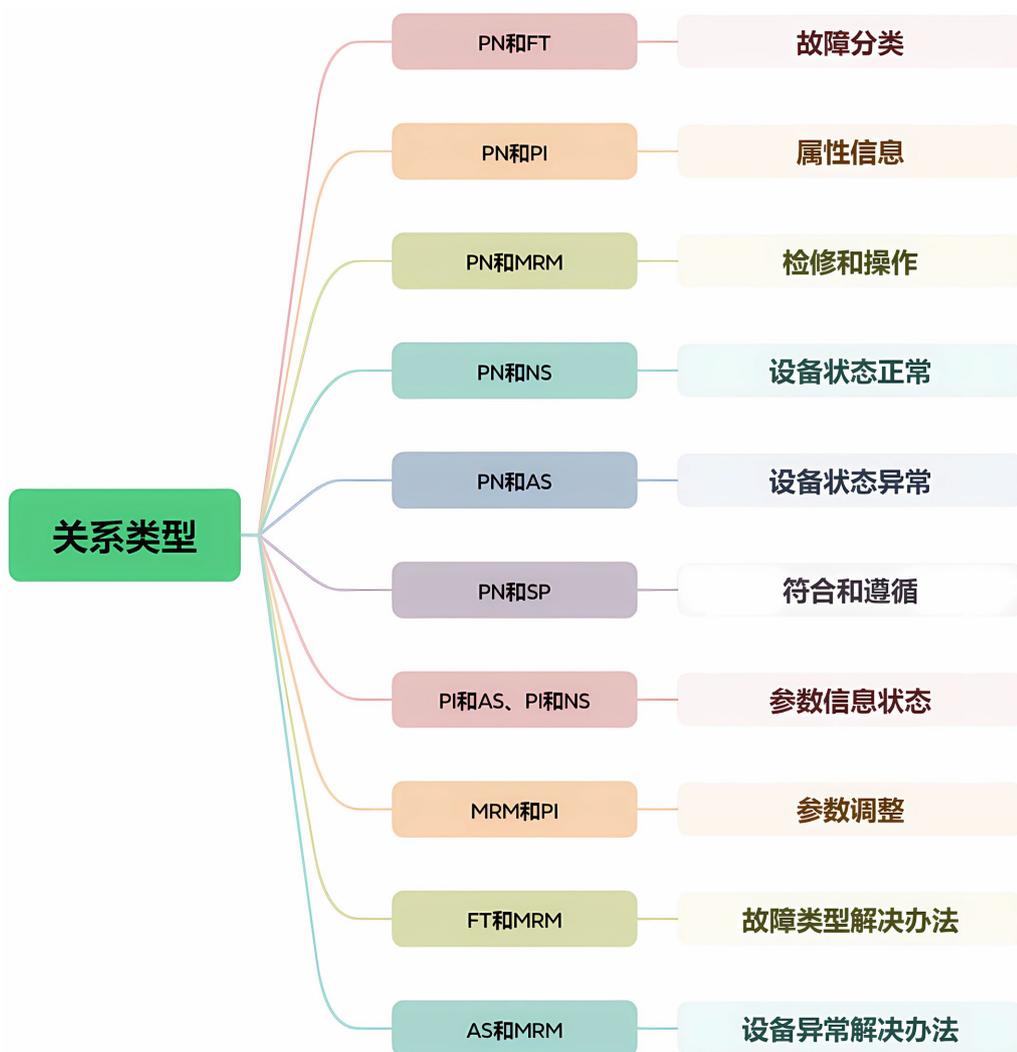


Figure 4. Relationship type distribution diagram

图 4. 关系类型分布图

4.2. 评价指标

本研究遵循命名实体识别任务的标准评估范式, 采用精确率(Precision)、召回率(Recall)及 F1 值构成

的三元指标体系, 具体计算公式如下:

$$P = \frac{TP}{TP + FP} \quad (8)$$

$$R = \frac{TP}{TP + FN} \quad (9)$$

$$F_1 = \frac{2 \times P \times R}{P + R} \quad (10)$$

4.3. 实验环境和模型参数

1) 实验环境

本文提出的模型均在以下环境搭建, 具体详情如表 2 所示。

Table 2. Experimental environment

表 2. 实验环境

实验环境	设置
操作系统	Ubuntu 20.0
GPU	Tesla V100
Python	3.7.16
PyTorch	1.12.1

2) 模型参数

通过实验可得, 本文所提模型的参数设置为表 3 的时候, 模型的表现效果最好, 所以, 模型的超参数设置为如下表 3 所示。

Table 3. Model parameters

表 3. 模型参数

超参数	数值
ernie_dim	768
max_seq_len	256
batch_size	16
att_head_num	4
lr	1e-5
epoch	300

4.4. 对比实验

为了证明所提模型结构的有效性, 本文在构建数据集上进行了一系列对比实验, 分别选取 CopyR [18]、GraphRel [19]、CasRel [8]和 PRGC [9]作为对比模型, 具体实验结果如下表 4 所示。

从表 4 得, 实验 1 和实验 2 分别采用单一联合解码器和应用多个分离解码器, F1 的分数分别达到 24.44%和 34.64%, 多个分离解码器比单一解码器表现出更好的性能, F1 分数上升 10.2%。实验 3 采用双阶段关系抽取, 先提取实体, 再提取关系, 用 GCN 模型进行关系加权, 相对于实验 2, F1 分数上升 5.55%。实验 4 将关系抽取任务进行任务分解, 设计了潜在关系预测组件、关系特定序列标注组件和全局

对应组件, 相对于实验 3, F1 的值达到 56.91%, 提升了 16.72%。实验 5 为本文的基线模型, 采用二元级联框架进行关系抽取器, 其 F1 值为 61.20%, 相对于实验 1 和 2, 该模型可以解决风力发电机故障领域存在的单实体重叠和多实体重叠问题。本文融入实体信息和知识图谱信息的 ERNIE 预训练模型代替传统的 BERT 模型, 可以充分捕捉领域内的特定名词实体和关系上下文信息, 通过引入基于余弦相似度的注意力计算模块, 能够高效捕捉风力发电机故障领域的本质特征, 对风力发电机稀疏关系具有更强的判别能力。最终, 与基线模型进行比较, 本文模型 CA-CasRel 模型的精确率、召回率和 F1 值分别提升了 1.71%、6.09%和 4.05%, 达到了领域内的 SOTA 效果, 其对应的柱状图信息如图 5 所示。

Table 4. Comparative experimental results

表 4. 对比实验结果

编号	模型	P	R	F_1
1	CopyROneDecoder	28.53%	21.37%	24.44%
2	CopyRMutilDecoder	39.61%	30.78%	34.64%
3	GraphRel	47.42%	34.88%	40.19%
4	PRGC	61.25%	53.14%	56.91%
5	CasRel	64.56%	58.17%	61.20%
6	CA-CasRel	66.27%	64.26%	65.25%

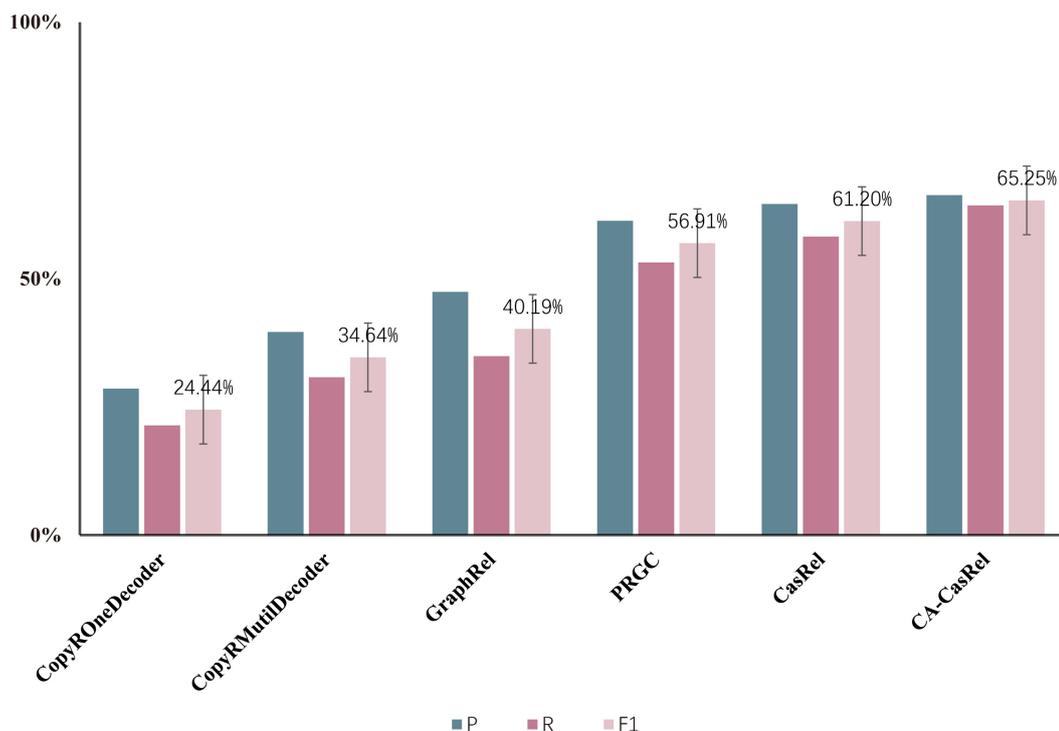


Figure 5. Histogram of comparative experimental results

图 5. 对比实验结果柱状图

4.5. 消融实验

为评估本研究模型各组成部分对关系抽取任务的性能影响, 本研究开展了多组模块化消融实验。

1) ERNIE 和 BERT 对比效果

为探究 ERNIE 模型在文本序列编码中的性能优势, 本研究设计了 BERT 与 ERNIE 两种编码方式的对比实验, 具体实验结果详见表 5。

Table 5. ERNIE and BERT coding ablation experimental results

表 5. ERNIE 和 BERT 编码消融实验结果

模型	<i>P</i>	<i>R</i>	<i>F</i> ₁
BERT 编码	64.56%	58.17%	61.20%
ERNIE 编码	68.64%	57.41%	62.53%

从表 5 可得, BERT 模型虽然在关系抽取任务中有一定效果, 但是对实体、关系语义识别不足。ERNIE 在预训练阶段融入了知识对齐任务, 如通过知识图谱对实体进行链接, 同时使用知识掩码的策略, 增加了实体级别嵌入和语义单元增强机制, 在风力发电机故障领域, 可以更准确地理解和区分不同设备和故障类型之间的关系, 避免由于术语相似或上下文复杂而导致的误判。例如“电压互感器”, 可以同时识别“电压”和“电压互感器”两个实体。与 BERT 嵌入相比, *F*₁ 分数值提升了 1.33%。可视化柱状图如图 6 所示。

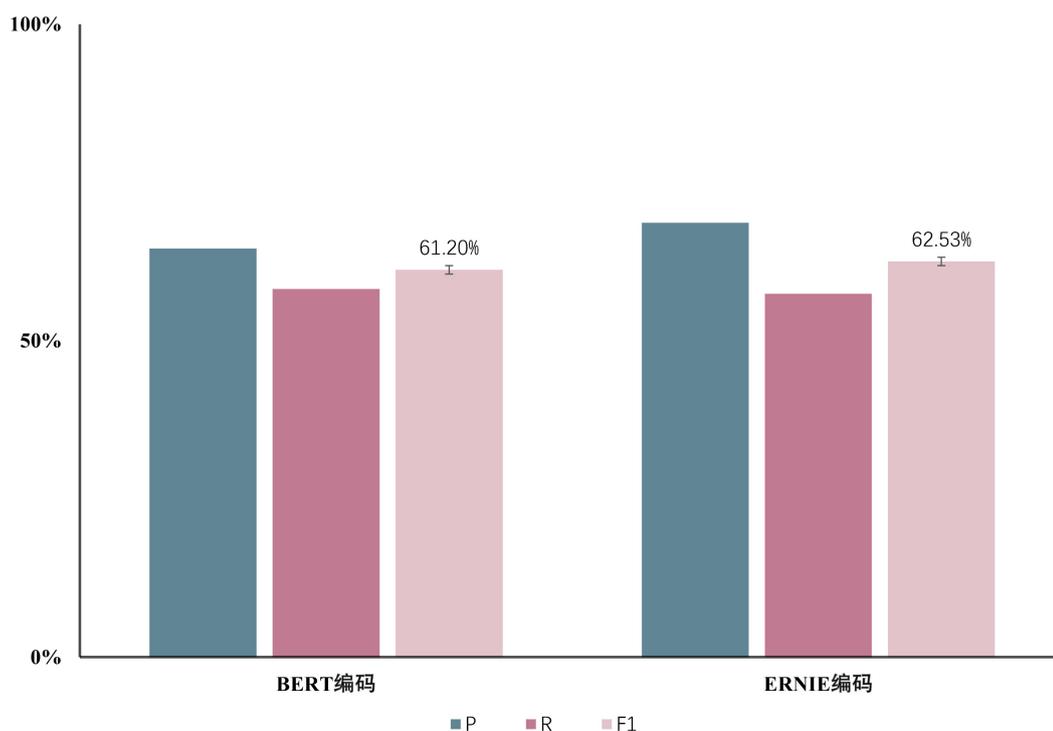


Figure 6. ERNIE and BERT ablation experiment results

图 6. ERNIE 和 BERT 消融实验结果

2) 余弦注意力和多头注意力对比效果

为评估不同注意力机制模块的性能差异, 本研究设计了三组对比实验方案: 基于无注意力结构、标准多头注意力结构以及提出的余弦注意力结构进行对比测试。如表 6 所示, 三种架构的量化评估结果展现出显著差异, 其中余弦注意力机制在多个指标上呈现最优表现。

Table 6. Ablation experiment results of cosine attention mechanism components
表 6. 余弦注意力机制组件消融实验结果

模型	<i>P</i>	<i>R</i>	<i>F</i> ₁
CasRel	64.56%	58.17%	61.20%
CasRel + MulATT	66.38%	58.56%	62.22%
CasRel + COSATT	66.95%	60.08%	63.33%

从表 6 可知, 在对头实体向量引入多头注意力机制之后, 与未引入相比, 其精确率、召回率和 *F*₁ 分数分别提升 1.82%、0.39% 和 1.02%。然而, 将多头注意力机制替换成余弦注意力机制以后, 与使用多头注意力机制相比较, 精确率、召回率和 *F*₁ 分数分别提升 0.57%、1.52% 和 1.11%。余弦注意力机制能够更好地捕捉输入之间的相似性, 尤其是文本数据中的相似词或实体。在风力发电机故障的关系抽取中, 设备部件、故障类型等信息往往具有较强的相似性, 余弦注意力机制能够更加准确地衡量不同实体之间的相似度。另外, 在计算注意力分布的时候, 余弦注意力通过计算向量之间的角度来判断它们的相似度, 而不直接依赖于加权的注意力分数。这使得在处理大量信息时, 计算效率更高。其可视化如图 7 所示。

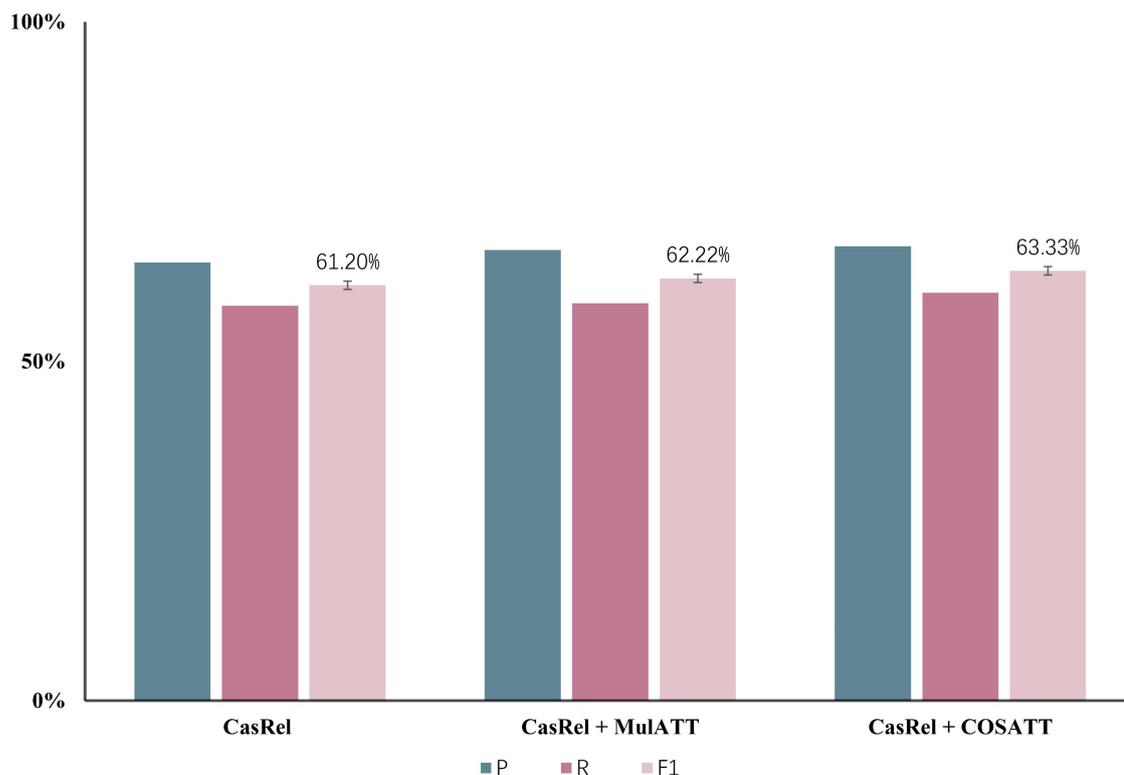


Figure 7. Ablation experiment results of cosine attention mechanism and multi-head attention mechanism
图 7. 余弦注意力机制和多头注意力机制消融实验结果

5. 结论

本文针对风力发电机故障领域实体关系抽取任务存在大量的专业术语、三元组重叠、单句子关系多等问题, 提出了一种改进的 CasRel 二元级联标记框架 CA-CasRel。在嵌入层, 使用 ERNIE 模型, 相对于 BERT 模型, 其对领域专有名词和实体关系更敏感, 能够有效地捕捉实体间细粒度关系。另外, 在头实体

预测阶段后, 采用余弦注意力机制对实体向量进行动态编码, 该方法通过方向感知的归一化计算范式, 实现双重优化: 一方面, 基于向量空间角度约束增强语义表征纯度, 聚焦核心语义关联而非强度干扰; 另一方面, 借助高维稀疏空间的方向对齐特性, 建立对长尾/低频特征的鲁棒表征机制, 有效克服传统注意力在稀疏分布下的特征遮蔽效应, 最终取得了领域内的 SOTA 效果。未来, 我们将进一步完善风力发电机故障语料库, 对实体关系类别作更加细致的划分, 提升风电机故障领域实体关系抽取的准确率。

参考文献

- [1] Detroja, K., Bhensdadia, C.K. and Bhatt, B.S. (2023) A Survey on Relation Extraction. *Intelligent Systems with Applications*, **19**, Article ID: 200244. <https://doi.org/10.1016/j.iswa.2023.200244>
- [2] Zhang, M., Zhou, G. and Aw, A. (2008) Exploring Syntactic Structured Features over Parse Trees for Relation Extraction Using Kernel Methods. *Information Processing & Management*, **44**, 687-701. <https://doi.org/10.1016/j.ipm.2007.07.013>
- [3] Hasegawa, T., Sekine, S. and Grishman, R. (2004) Discovering Relations among Named Entities from Large Corpora. *Proceedings of the 42nd Annual Meeting on Association for Computational Linguistics*, Barcelona, 21-26 July 2004, 415-422. <https://doi.org/10.3115/1218955.1219008>
- [4] Xu, Y., Mou, L., Li, G., Chen, Y., Peng, H. and Jin, Z. (2015) Classifying Relations via Long Short-Term Memory Networks along Shortest Dependency Paths. *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, Lisbon, September 2015, 1785-1794. <https://doi.org/10.18653/v1/d15-1206>
- [5] Zhong, Z. and Chen, D. (2021) A Frustratingly Easy Approach for Entity and Relation Extraction. *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, Online, June 2021, 50-61. <https://doi.org/10.18653/v1/2021.naacl-main.5>
- [6] Yang, C., Deng, J., Chen, X. and An, Y. (2023) SPBERE: Boosting Span-Based Pipeline Biomedical Entity and Relation Extraction via Entity Information. *Journal of Biomedical Informatics*, **145**, Article ID: 104456. <https://doi.org/10.1016/j.jbi.2023.104456>
- [7] Tuo, M. and Yang, W. (2023) Review of Entity Relation Extraction. *Journal of Intelligent & Fuzzy Systems*, **44**, 7391-7405. <https://doi.org/10.3233/jifs-223915>
- [8] Wei, Z., Su, J., Wang, Y., Tian, Y. and Chang, Y. (2020) A Novel Cascade Binary Tagging Framework for Relational Triple Extraction. *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, Online, July 2020, 1476-1488. <https://doi.org/10.18653/v1/2020.acl-main.136>
- [9] Zheng, H., Wen, R., Chen, X., Yang, Y., Zhang, Y., Zhang, Z., et al. (2021) PRGC: Potential Relation and Global Correspondence Based Joint Relational Triple Extraction. *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, Online, August 2021, 6225-6235. <https://doi.org/10.18653/v1/2021.acl-long.486>
- [10] Wadhwa, S., Amir, S. and Wallace, B. (2023) Revisiting Relation Extraction in the Era of Large Language Models. *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, Toronto, July 2023, 15566-15589. <https://doi.org/10.18653/v1/2023.acl-long.868>
- [11] Wei, J., Wang, X., Schuurmans, D., et al. (2022) Chain-of-Thought Prompting Elicits Reasoning in Large Language Models. *Advances in Neural Information Processing Systems*, **35**, 24824-24837.
- [12] Brown, T.B. (2020) Language Models Are Few-Shot Learners. arXiv: 2005.14165.
- [13] Chung, H.W., Hou, L., Longpre, S., et al. (2024) Scaling Instruction-Finetuned Language Models. *Journal of Machine Learning Research*, **25**, 1-53.
- [14] Wang, S., Zhang, Y., Che, W. and Liu, T. (2018) Joint Extraction of Entities and Relations Based on a Novel Graph Scheme. *Proceedings of the 27th International Joint Conference on Artificial Intelligence*, Stockholm, 13-19 July 2018, 4461-4467. <https://doi.org/10.24963/ijcai.2018/620>
- [15] Zeng, D., Liu, K., Chen, Y. and Zhao, J. (2015) Distant Supervision for Relation Extraction via Piecewise Convolutional Neural Networks. *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, Lisbon, September 2015, 1753-1762. <https://doi.org/10.18653/v1/d15-1203>
- [16] 戴志辉, 张富泽, 张近月, 等. 基于 MacBERT-BiLSTM-CRF 模型的继电保护装置缺陷知识图谱构建方法[J]. *电力系统保护与控制*, 2024, 52(20): 131-143.
- [17] 文雯. 面向汽车故障知识图谱构建的实体关系抽取算法研究[D]: [硕士学位论文]. 西安: 西安工业大学, 2024.
- [18] Zeng, X., Zeng, D., He, S., Liu, K. and Zhao, J. (2018) Extracting Relational Facts by an End-to-End Neural Model with Copy Mechanism. *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1:*

- Long Papers*), Melbourne, July 2018, 506-514. <https://doi.org/10.18653/v1/p18-1047>
- [19] Fu, T., Li, P. and Ma, W. (2019). GraphRel: Modeling Text as Relational Graphs for Joint Entity and Relation Extraction. *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, Florence, July 2019, 1409-1418. <https://doi.org/10.18653/v1/p19-1136>