

DOI:10.13296/j.1001-1609.hva.2026.02.006

基于关系导向的电力设备故障缺陷文本实体及关系联合抽取方法

李艾青, 宋辉, 田嘉鹏, 盛戈峰, 江秀臣
(上海交通大学电气工程系, 上海 200240)

摘要: 电力设备故障缺陷知识图谱能够有效提升设备运维的智能化、自动化水平,而实体及关系的抽取对图谱的构建至关重要。然而故障缺陷文本中的实体关系三元组往往互相重叠或嵌套,使得传统方法难以处理,并伴随着误差传递、冗余实体推断等问题。针对这些问题,文中提出了一种面向电力设备故障缺陷领域的实体及关系联合抽取方法。该方法将三元组抽取任务建模为不同关系类型下头实体映射到尾实体的过程,通过首先抽取出头实体,再为已识别头实体针对每一种关系分别标记其对应的尾实体,从而有效缓解了三元组重叠嵌套及冗余推断等问题。实验表明,所提出的方法相较于基线模型在三元组出现不同程度重叠或嵌套时表现地更加鲁棒,其 F_1 值提升了8.57%~25.19%,验证了所提模型的有效性与可行性。

关键词: 电力设备; 故障缺陷文本; 知识图谱; 知识抽取; 深度学习

Relation-oriented Method for Jointly Extracing Text and Relations of Power Equipment Fault Defect Based on Relationships

LI Aiqing, SONG Hui, TIAN Jiapeng, SHENG Gehao, JIANG Xiuchen
(Department of Electrical Engineering, Shanghai Jiao Tong University, Shanghai 200240, China)

Abstract: Fault or defect knowledge graphs of power equipment could effectively improve the intelligent and automatic level of operation and maintenance of the equipment and, however, the extractions of text and relations is crucial to the construction of the graph. However, in fault or defect text, the entity relationship triples often overlap or embedded each other, resulting in not only difficult treatment by traditional method, but also accompanying such issues as error transmission and redundant entity inference. As for these issues, in this paper a joint extraction method based on relation-oriented with Bi-LSTM is proposed. In this method, the triplet extraction task is modeled as the process of mapping the head entity to the tail entity under different relationship types. First, the head entity is extracted and then the corresponding tail of each type of relationship is marked for the identified entity, thus effectively mitigating the issues related to overlapping and nesting triples as well as redundant entity-relationship inference. The experiments show that the proposed method is more robust than the baseline model when triples overlap or nest occurring at different degrees, the F_1 value is improved by 8.57% to 25.19%, which verifies both effectiveness and feasibility of the proposed model.

Key words: power equipment; fault or defect text; knowledge graph; knowledge extraction; deep learning

0 引言

电力设备作为电力系统的重要组成部分,其安全稳定运行是确保电网持续向用户提供安全、优

质、经济电能的先决条件,近年来电力设备的智能运维方法逐渐受到重视^[1-2]。在设备日常运维过程中,积累了大量记录设备故障缺陷情况、产生原因与对应运维措施的文本数据,这类故障缺陷文本数

收稿日期:2025-07-11; 修回日期:2025-10-18

基金项目:国家重点研发计划项目(2020YFB1709701)。

Project Supported by National Key Research and Development Program of China(2020YFB1709701).

据由于非结构化、多源异构性高等特点,难以直接从中提取出有效知识并加以利用,往往被闲置或需要人工反复查询。知识图谱(knowledge graph)是一种能够聚集文本知识的结构化知识表示方法,具有较高的可视化程度,能够有效完成海量数据组织管理、知识挖掘总结与智能问答等任务,在电力领域的故障预测^[3]、故障处置决策^[4-6]、缺陷检索^[7]、电力调度^[8-9]、设备健康管理^[10]、辅助故障定位^[11]等方面具有良好的应用效果。借助于知识图谱的高效数据调取与复杂关联表示^[4],对设备故障缺陷文本进行解析、录入与推理,可有效学习到有关电力设备的先验知识,进一步为输变电设备状态评估、决策生成提供知识支持。作为知识图谱构建最基础的部分,知识抽取是指从自然语言文本中抽取出头、尾实体及其相互关系,构成图谱的基础单元。在电力设备故障缺陷文本中,存在着大量实体间互相重叠或嵌套的情况,严重影响知识抽取的准确度。

设备故障缺陷文本通常由运维人员人工记录,用语凝练简洁,实体密度较高,并由于电力领域信息密集的特点以及自然语言表述的特性,使得其中几乎所有句子都具有重叠或嵌套实体。此外,由于不同运维人员的记录习惯存在差异,导致设备缺陷文本中嵌套实体结构复杂多变,其嵌套层数与颗粒度缺乏规律。上述嵌套特点极大地增加了实体识别的难度,并严重制约了知识抽取的准确性。而现有研究工作大多侧重于对非嵌套实体的识别,并采用流水线工作模式进行知识抽取^[4, 8, 12-17]。其中,通常采用序列标注^[12-17]方式完成实体识别,通过为每个字符赋予标签来识别该字符所属实体的类型与其在实体中的位置。此类方法在缓解实体边界模糊问题上具有一定优势,但受限于标记次数或标签数量,不能灵活处理嵌套实体或重叠关系,无法直接应用在故障缺陷文本上。此外,由于流水线式知识抽取方法将实体抽取(name entity recognition, NER)与关系抽取(relation classification, RC)割裂为两个先后独立的任务分别处理,往往会导致误差传递、冗余实体推断与子任务缺少交互问题^[18],限制了知识抽取的效率与准确性。

文中针对设备故障缺陷文本中存在三元组互相重叠或嵌套的特点,提出了一种基于关系导向的故障缺陷文本的实体关系联合抽取方法。该方法通过对知识三元组进行联合建模^[19-23]的方式完成知识抽取,以减小误差累积;并将三元组抽取任务视为在关系导向下两实体的映射过程,以避免冗余的实体关系推断。同时通过以关系为导向对输入句

子进行分层标记实体,突破了传统方法中字符的单标签限制,保证了对重叠或嵌套三元组的处理能力。实验证明,该方法能有效地抽取故障缺陷文本的实体和关系三元组,在不同程度的重叠或嵌套情况下表现得更加鲁棒。

1 设备故障缺陷文本的特点与抽取难点

设备故障缺陷文本其存在以下几个特点。

1)实体嵌套。设备缺陷文本中实体分布密度较高,存在大量“一个实体嵌套另一个实体”的情况,即同一个词可能同时属于多个实体,见图1。而目前广泛使用的基于序列标注的实体识别方法,通过使用“Type+B/I/E”单标签来区分实体的开头、内部与结尾以及实体的类型,其中每个字符只能被识别为一个标签,无法将“35 kV 母线压变间隔柜”中的“3”同时标记为电压等级与缺陷设备的开头,因而不能有效解决实体嵌套问题。

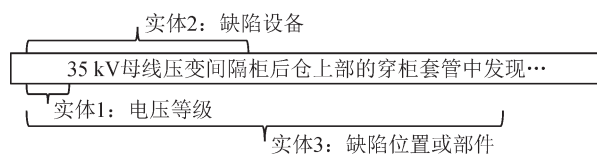


图1 嵌套实体示例

Fig. 1 Example of nesting entities

2)关系重叠。故障缺陷文本实体间关系复杂,大多数句子中一个实体同时与其他多个实体存在关系,使得不同三元组中存在某一相同的实体(single entity overlap, SEO^[20]);此外,还存在着少量不同关系类型的三元组共享同一实体对的现象(entity pair overlap, EPO^[20]),见表1。而传统流水线式抽取方法通过先抽取出实体,再对两两实体对进行关系分类,不仅存在误差累积与冗余关系推断,同时也难以从单一的实体元组向量中识别出多种关系,因此无法有效处理该类关系重叠问题。

表1 设备故障缺陷文本关系重叠样例

Table 1 Sample relation overlapping in fault or defect text

缺陷文本	SEO	EPO
开关管穿柜	{开关管, 缺陷位置为穿柜套管内部}	{穿柜套管, 位于开关管}
套管内部出现悬浮放电	{开关管, 缺陷类型为悬浮放电}	{开关管, 薄弱环节, 穿柜套管}
	{开关管, 包含部件穿柜套管}	

2 框架设计

针对上述特点,文中采用头实体触发,以关系为映射导向,对尾实体进行分层标记来进行故障缺

陷文本的三元组抽取任务,整体框架见图2。

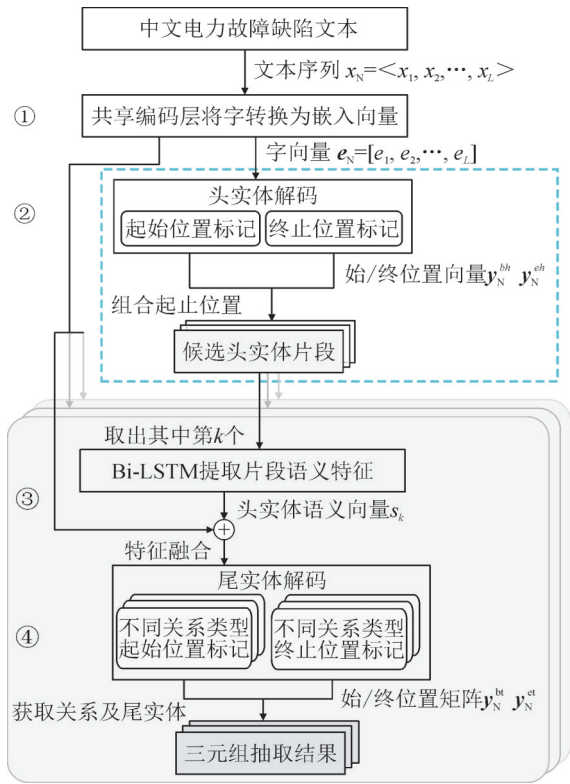


图2 模型框架图

Fig. 2 Framework of the model

对于任意三元组 \$(h, r, t)\$, 其识别过程可建模为

$$P[(h, r, t)|x] = P(h|x)P[(r, t)|h, x] = P(h|x) \prod_{r \in R} P_r(t|h, x) \quad (1)$$

由此三元组识别过程被转化为两个子任务:一是从句子 \$x\$ 中抽取出头实体 \$h\$;二是在不同关系类型 \$r\$ 下,根据头实体 \$h\$ 的先验信息,与句子 \$x\$ 共同解码得出对应的尾实体 \$t\$。这样建模具有以下优势:

1)通过以关系为映射导向,而非将其作为被动分配给实体对的标签,使得相同的实体对之间可通过不同的映射方式连接,从而有效处理重叠关系。

2)由于关系种类数量远小于实体个数,可使原先对两两实体进行关系抽取的 \$O(n^2)\$ 复杂度降低至 \$O(n)\$,避免了冗余的实体关系推断。

3)通过在不同关系类型下对输入句子进行分层标记相应实体,使得同一字符能够重复、多次出现在不同的实体中,突破了传统方法中字符的单标签限制,从而保证了对嵌套实体的处理能力。两种方法对重叠嵌套三元组的标记方式见图3。

3 模型结构

文中提出的 JRB(joint extraction model based on relation-oriented with Bi-LSTM)模型结构见图4。首

设备缺陷文本	开	关	管	穿	柜	套	管	内	部	出	现	悬	浮	放	电
传统序列标注	E	E	E	C	C	C	C	P	P	O	O	T	T	T	T
	B	I	E	B	I	I	E	B	E	O	O	B	I	I	E
头实体标记	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
“缺陷位置”	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
尾实体标记	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
“包含部件”	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
尾实体标记	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
“缺陷类型”	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
尾实体标记	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
“薄弱环节”	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
尾实体标记	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0

图3 与传统方法的标记方式对比

Fig. 3 Tagging comparison with traditional approach

先,通过按字分词避免不同分词工具的误差累积;接下来,将分词后的文本序列使用语言模型 BERT (bidirectional encoder representation from transformers) 进行编码,并输入到头实体解码器中。头实体解码器标记头实体的边界,输出两组位置向量,分别表示所有头实体起始和终止位置的分布情况。之后,提取被识别为头实体的向量序列,送入语义提取器中解析其语义特征,获取头实体的语义向量。最后,将头实体语义向量和句子的 BERT 编码共同输入到尾实体解码器中,解码得到两组位置矩阵。该矩阵由多个位置向量组成,其中每个位置向量都分别表示与给定头实体存在特定关系类型的尾实体位置分布情况。

3.1 BERT 编码层

BERT 模型基于 Transformer 结构,利用多头注意力层计算词间关联度来克服长句信息遗失问题,同时加入残差网络与层归一化结构来避免梯度消失并加快网络的收敛速度。通过对大规模、未标注的语料进行自监督训练,BERT 能够有效学习到文本的通用语言特征,并获得深层次的文本词向量表示,可作为后续两个子任务的共享编码,表达式为

$$e_N = \text{BERT}(x_N) \quad (2)$$

式(2)中: \$x_N = \langle x_1, x_2, \dots, x_L \rangle\$ 为输入句子,长度为 \$L\$; \$e_N = [e_1, e_2, \dots, e_L]\$ 为 BERT 模型输出的编码序列,其中 \$e_i\$ 为 \$x_i\$ 对应位置的嵌入向量。

3.2 头实体解码器

基于指针标注^[24]思想,头实体解码器采用两个独立的全连接分类器来分别标记各个头实体的起始与终止位置,通过为输入序列的每个位置赋予标签“0”或“1”来判定其是否为某一实体的边界,进而

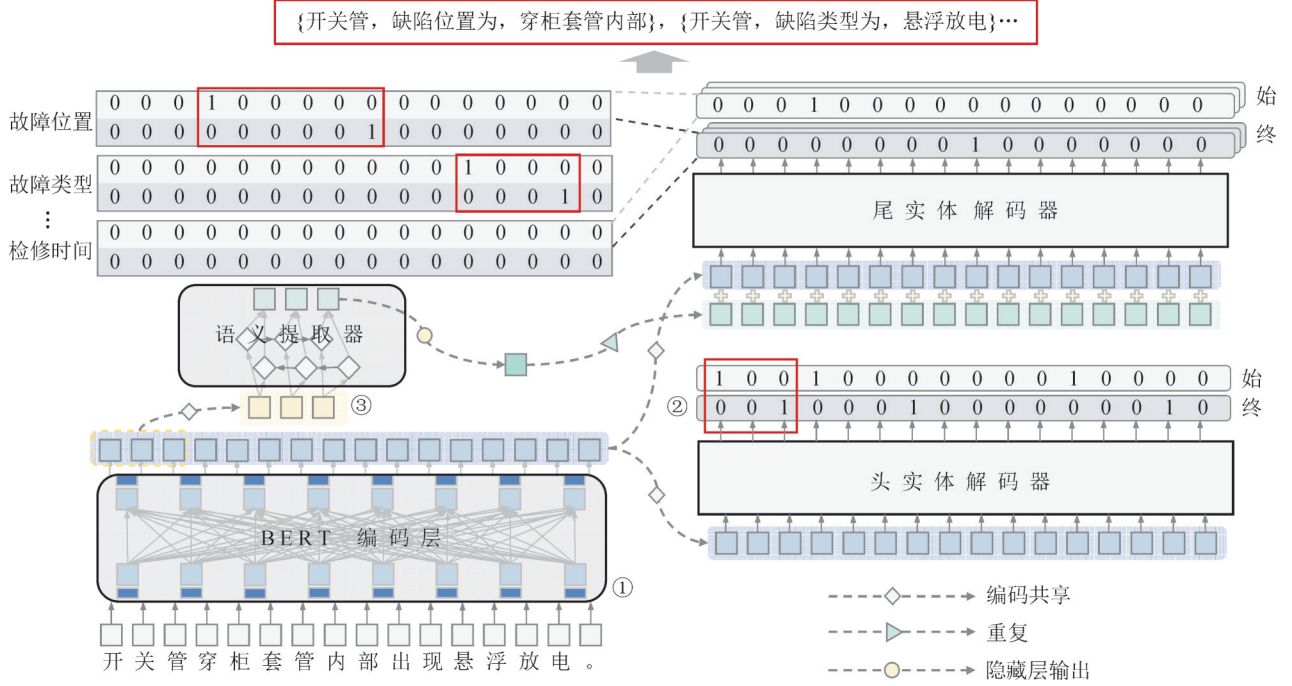


图4 实体关系三元组抽取模型结构

Fig. 4 Structure diagram of entity-relationship triples extraction model

识别出句子中所有可能的头实体,表达式为:

$$P_i^{bh} = \sigma(W_{bh}e_i + b_{bh}) \quad (3)$$

$$P_i^{eh} = \sigma(W_{eh}e_i + b_{eh}) \quad (4)$$

$$s(x) = \begin{cases} 0 & x < \delta \\ 1 & x > \delta \end{cases} \quad (5)$$

$$y_i^t = s(P_i^t) \quad t \in \{bh, eh\}, i \in \{1, \dots, L\} \quad (6)$$

式(3)-(6)中: P_i^{bh} 与 P_i^{eh} 分别为输入序列中的第 i 个位置作为头实体的开始或结束位置的概率,将概率值经过一个平移缩放的符号函数 $s(\cdot)$ 获取该位置的输出; δ 为设定阈值; $W(\cdot)$ 为权重; $b(\cdot)$ 为偏置分量; $\sigma(\cdot)$ 为 sigmoid 激活函数; 头实体解码器输出的位置向量记为 $y_N^{bh}, y_N^{eh} \in R^L$ 。

3.3 语义提取器

为了从识别出的头实体片段序列中提取头实体语义信息,采用双向长短期记忆网络 (bidirectional long short-term memory, Bi-LSTM) 对输入的向量序列进行特征解析。LSTM 的核心在于利用记忆细胞保存长期的历史信息,并利用门控结构 (遗忘门、输入门与输出门) 管理信息量。通过控制记忆单元中信息的保存、更新、衰减、输入、输出等动作, LSTM 模型提供了多层次的特征选择方法,能有效获取序列的语义特征。

为了更好地利用上下文信息,充分提取头实体片段包含的特征,可采用 Bi-LSTM 来捕捉双向的语义依赖。通过为输入序列分别训练一个向前和向后的 LSTM 网络来获取各位置的双向语义特征,其

表达式为:

$$\vec{h}_n, \vec{c}_n = \overrightarrow{\text{LSTM}}(x_n, \vec{h}_{n-1}, \vec{c}_{n-1}) \quad (7)$$

$$\overleftarrow{h}_n, \overleftarrow{c}_n = \overleftarrow{\text{LSTM}}(x_n, \overleftarrow{h}_{n+1}, \overleftarrow{c}_{n+1}) \quad (8)$$

$$h_n = [\vec{h}_n; \overleftarrow{h}_n] \quad (9)$$

式(7)-(9)中: x_n 为序列中第 n 个位置的输入; $\vec{h}_n, \vec{c}_n \in R^d$ 与 $\overleftarrow{h}_n, \overleftarrow{c}_n \in R^d$ 分别为前向与反向的 LSTM 在第 n 个位置对应的输出与细胞状态, d 为隐含层大小; $[\cdot; \cdot]$ 为向量拼接; h_n 为 Bi-LSTM 在第 n 个位置的输出。

抽取出第 k 个头实体对应区间内所有位置的 BERT 编码,记为 $E_k = [e_{kb}, e_{kb+1}, \dots, e_{ke-1}, e_{ke}]$ 。将 E_k 输入到 Bi-LSTM 层中获取头实体特征信息,取其中最后一个位置 e_{ke} 的输出作为表征该头实体特征的语义向量 S_k 为

$$S_k = \text{BiLSTM}(E_k) = h_{ke} \quad (10)$$

式(10)中,为了使 S_k 的维度与 e_N 相同,需将 Bi-LSTM 的隐藏层大小设置为嵌入维度的 1/2。

3.4 尾实体解码器

尾实体的解码过程需要同时考虑句子本身的输入特征与头实体的语义信息。对应地,尾实体解码器的输入可表示为

$$e^k = S_k + e_N \quad (11)$$

在尾实体解码器中,通过对每个头实体,检查在所有关系类型下,是否存在对应的尾实体,并标

记出位置,从而同时确定头、尾实体与关系三元组。其中,尾实体解码器同样采用两个独立的全连接分类器进行二分类,通过对每个位置在不同关系类型下分层标记0/1,来判断该位置是否为某个头实体在特定关系下的尾实体的开始或终止位置,从而获取该头实体对应的所有尾实体及其之间的关系类型,具体为:

$$P_i^{bt} = \sigma(W_{bt} e_i^k + b_{bt}) \quad (12)$$

$$P_i^{et} = \sigma(W_{et} e_i^k + b_{et}) \quad (13)$$

$$y_i^t = s(P_i^t) \quad t \in \{bt, et\}, i \in \{1, \dots, L\} \quad (14)$$

式(12)-(14)中, e_i^k 为与第 k 个头实体对应的尾实体解码器输入序列中第 i 个位置的向量。尾实体解码器输出的位置矩阵记为 $\mathbf{y}_N^{bt}, \mathbf{y}_N^{et} \in R^{R \times L}$, 其中 R 为预先设定的关系类别总数,可认为是由每个关系类型 $r \in R$ 各自的尾实体位置向量组成。

3.5 起止位置组合

对于实体解码器输出的两组位置向量(矩阵),采用最近匹配策略组合各个实体的边界,其中每个结束位置匹配其左边最接近的起始位置,每个起始位置匹配其右边最接近的结束位置,共同表示实体所在区间。由于头实体解码器采用位置向量来表示所有头实体的分布情况,当其中部分头实体互相嵌套时可能会引起位置向量的组合混乱。而尾实体解码器得益于在关系导向下的分层标记方式,对应于不同关系类型的尾实体位置向量被彼此分隔开,避免了起止位置组合混乱的情况。

3种嵌套头实体类型见图5。经分析发现,只有当两个头实体处于中间嵌套时,位置向量的组合会出错。因此,将此类实体中跨度较大者作为尾实体处理,相应地改写三元组关系,以此避免由头实体嵌套引起的位置组合混乱问题。

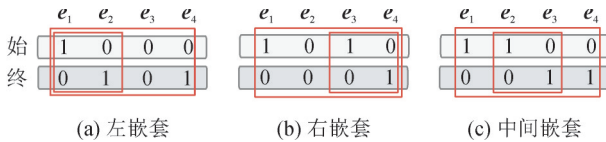


图5 头实体嵌套类型

Fig. 5 The types of head entities nesting

3.6 损失函数

文中选择二分类交叉熵损失函数对头、尾实体解码过程进行训练,整个模型的联合损失可认为是两个子任务的损失相加,表达式为:

$$L_h = \sum_{t \in \{bh, eh\}} -\frac{1}{L} \sum_{i=1}^L [z_i^t \lg P_i^t + (1 - z_i^t) \lg (1 - P_i^t)] \quad (15)$$

$$L_t = \sum_{t \in \{bt, et\}} -\frac{1}{L} \sum_{i=1}^L [z_i^t \lg P_i^t + (1 - z_i^t) \lg (1 - P_i^t)] \quad (16)$$

$$L_{total} = \sum_{h \in T} L_h + \sum_{t \in T|h} L_t \quad (17)$$

式(15)-(17)中: z 为样本标签,若样本属于正例,则取值为1,否则为0; T 为输入句子的三元组集合; $T|h$ 为句子中头实体为 h 的所有三元组集合。

通过共享编码并最小化联合损失来训练模型,可以使得头实体、尾实体与关系的抽取相互影响,从而使得子任务产生的错误被另一个约束,以此增加子任务间的交互性,并减少错误传播。

4 实验验证

4.1 数据集

文中采用的电力设备故障缺陷文本是某电力公司历经10年时间采集的真实现场记录。在对原始样本进行分析整理后,拆分出2160条故障缺陷数据,其中包含设备缺陷、运行状态、故障过程与解体维护经过等信息,并以6:1:2的比例划分训练集、验证集与测试集。考虑到故障缺陷文本存在“实体间互相重叠或嵌套”的特点,设计出4种不同重叠程度的数据集,用以探究模型对重叠或嵌套三元组的抽取效果,见表2。其中“重叠占比”的含义为:出现与其他三元组所含实体重叠或嵌套的实体所属三元组在总三元组数中的比重,表达式为

$$\text{重叠占比} = \frac{\text{含重叠或嵌套实体的三元组数量}}{\text{三元组总数}} \quad (18)$$

表2 数据集三元组个数与重叠占比

Table 2 Number of triples and overlap ratio of dataset

数据集名称	三元组个数			重叠占比/%
	训练集	验证集	测试集	
FT ₁	5 908	967	2 006	0.00
FT ₂	10 933	1 788	3 610	14.10
FT ₃	6 458	1 050	2 122	23.90
FT ₄	4 482	738	1 464	34.44

4.2 参数设置

文中使用BERT嵌入模型作为共享编码层,输出编码维度为768;在头实体解码器与尾实体解码器中,对起始与终止位置的判定阈值设置为0.5;在语义提取器中,将Bi-LSTM的隐藏层大小设置为384;学习率设置为 1×10^{-4} ;使用小批量梯度下降方法更细模型参数,将batch size设置为16;训练期间使用Adam优化器,在验证集上获取 F_1 分数最高的模型,用于测试集进行性能测试。实验基于Python3.7,采用tensorflow2.5框架,并在一块RTX 3090 GPU上完成训练与测试。

4.3 实验设置

为了比较流水线式抽取方法与联合抽取方法对设备故障缺陷领域中实体与关系抽取任务的性能优劣,文中选用命名实体识别领域中普遍使用的基于条件随机场(conditional random fields, CRF)的双向长短时记忆网络结构和基于BERT的关系预测方法^[25]组合作为流水线式基线模型,与JRB模型进行对比实验。首先分别对4种文本数据集进行抽取,以探究三元组嵌套重叠程度对模型的影响。其次,再针对不同的关系类型进行抽取效果分析。最后,根据在不同数据集上的收敛性能分析其效率。文中选用实体关系抽取领域的3个典型评价指标:精确度(*precision*)、召回率(*recall*)与 F_1 值来评价故障缺陷文本中实体和关系的抽取效果。

4.4 实验分析

4.4.1 三元组重叠嵌套程度的影响分析

JRB模型与基线模型对4种故障缺陷文本数据集进行实体及关系抽取的结果见表3。从表3中可以看出,随着数据集三元组重叠程度的增加,基线模型的召回率与 F_1 值明显下降,而JRB联合抽取方法的各项指标始终维持在较高的水平。

表3 流水线式与联合抽取方法的性能比较

Table 3 Performance comparison between pipeline extraction method and joint extraction method %

数据集	模型	<i>P</i>	<i>R</i>	F_1
FT ₁	流水线式	98.35	98.35	98.35
	联合式	96.75	96.41	96.58
FT ₂	流水线式	94.41	82.85	88.26
	联合式	95.48	98.23	96.83
FT ₃	流水线式	91.59	72.90	81.19
	联合式	96.49	97.17	96.83
FT ₄	流水线式	89.18	57.99	70.28
	联合式	95.12	95.83	95.47

作为经典的流水线式实体及关系抽取模型,BERT+Bi-LSTM+CRF基线方法可以完成三元组抽取任务,但在数据集三元组重叠度与分布复杂度高时效果差强人意,这是因为命名实体识别任务本质上是个序列标注问题,其中每个字符只能被识别为一个标签,从而导致该字符只能属于特定实体的固定位置,使其无法处理嵌套或重叠实体。而有别于传统的流水线式抽取方法将三元组识别过程视为头、尾实体组合映射到关系的过程 $f(s,o) \rightarrow r$;以关系为导向的联合抽取方法则将三元组抽取任务建模为不同关系类型下头实体映射到尾实体的过程 $f_r(s) \rightarrow o$ 。通过给定头实体,对输入序列进行不同关系类型下的分层标记,天然地具有处理三元组重

叠或嵌套问题的能力,因此在数据集三元组重叠度高或分布复杂时仍能保持较高的召回率与 F_1 值。

4.4.2 关系类型对模型的影响分析

为进一步分析不同关系类型对模型抽取效果的影响,文中分别统计了数据集中7种关系类型的三元组抽取结果,见表4。

表4 不同关系类型的识别效果比较

Table 4 Performance comparison between different

关系类型	模型	relationship types %		
		<i>P</i>	<i>R</i>	F_1
所属站点(场所)	流水线式	96.51	97.74	97.12
	联合式	99.29	99.58	99.44
故障(缺陷)特征	流水线式	98.05	97.51	97.78
	联合式	100.00	100.00	100.00
故障(缺陷)检测时间	流水线式	98.32	97.38	97.85
	联合式	99.76	98.57	99.16
故障(缺陷)位置	流水线式	48.97	4.71	8.59
	联合式	92.33	95.69	93.98
故障(缺陷)现象	流水线式	91.13	90.60	90.87
	联合式	94.89	97.49	96.17
故障(缺陷)类型	流水线式	97.20	97.72	97.46
	联合式	99.32	99.32	99.32
停电检修时间	流水线式	84.94	80.30	82.55
	联合式	86.39	88.48	87.43

从实验结果可看出,联合抽取方法的整体效果较为均衡,其 F_1 值均保持在87%以上。而流水线方法对“故障(缺陷)现象”、“停电检修时间”与“故障(缺陷)位置”关系的预测结果明显低于联合抽取方法的对应结果。其中“故障(缺陷)位置”关系的召回率严重偏低,是拉低流水线模型效果的主要因素。通过对这3种关系对应的语料文本和最终生成的三元组预测结果进行分析,发现文本中存在“故障(缺陷)现象”关系的实体对之间,有的距离较远,通常难以通过流水线方法识别出来;同时“故障(缺陷)检测时间”与“停电检修时间”关系的尾实体都在命名实体识别中被标记为“时间”类型,导致在关系分类时容易被混淆;此外,在抽取历史案例的位置信息时,需详细、明确到具体设备的具体部件上。这导致描述故障或缺陷位置的实体片段中往往包含了设备实体,使得关系类型为“故障(缺陷)位置”的三元组所含重叠或嵌套实体的占比偏高。在应对数据集中此类实体重叠或嵌套的文本时,流水线方法表现相较于对其他关系类型的三元组时下降的更为明显。

4.4.3 算法效率分析

通过对JRB联合抽取模型在不同数据集上的收敛性能对其效率进行分析,见图6。从模型训练时的收敛迭代次数来看,JRB模型在迭代3~5次后

即可收敛,并将 F_1 值突破至 90% 以上。得益于模型中基于指针标注的实体抽取设计,可将序列标注的多分类问题转变为位置标签的二分类问题,从而在实体识别时使用 sigmoid 函数代替 softmax 函数,使模型整体更易于训练。随着迭代次数的增加, JRB 模型在各数据集上的 F_1 值逐渐升高,分别在第 8、27、13、23 个训练周期时达到最高值,并根据 early stopping 机制在 6 个迭代周期后停止训练。

4.5 实体及关系抽取效果展示与图谱构建

文中提出的联合抽取方案能准确有效地识别出故障缺陷文本的实体以及实体之间的关系类型。文中模型对案例报告中关键信息的抽取效果

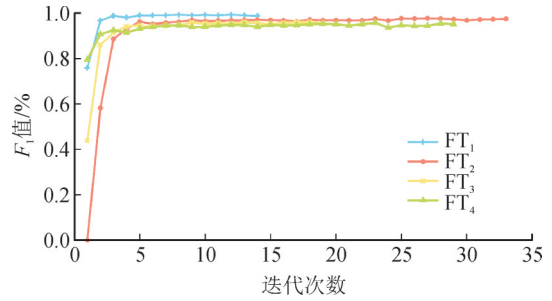


图6 JRB模型的收敛性能

Fig. 6 Convergence performance of JRB model

以及相应的案例知识图谱见图7。场站中积累的大量故障或缺陷案例、检修记录、运维预案等文本在抽取转化、融合汇集后形成规模庞大的设备运维图谱,可用以为电力设备的智能运维提供知识支撑。

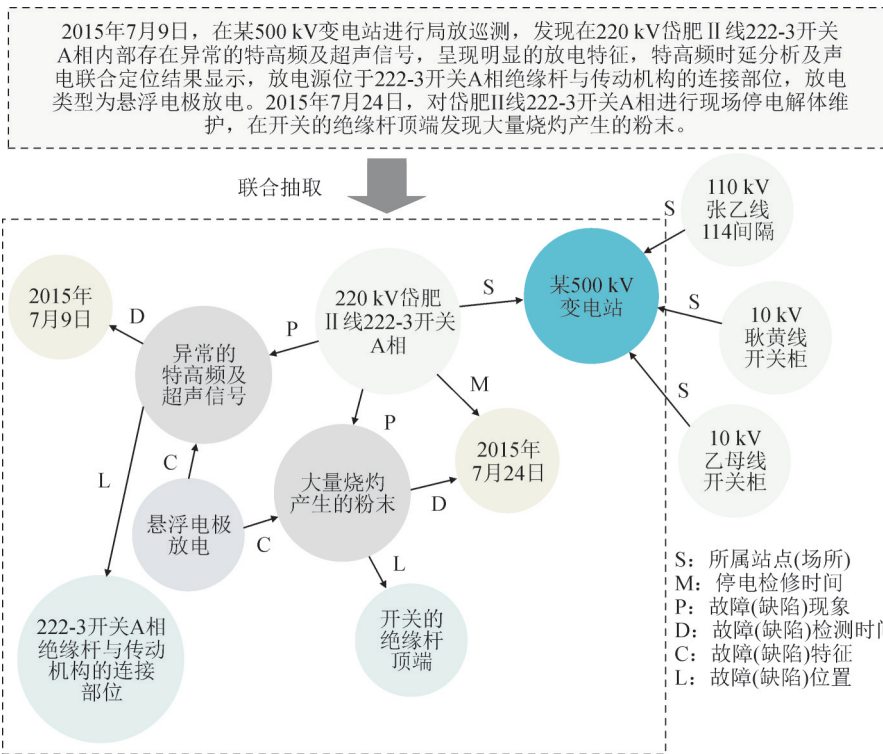


图7 JRB构建的某案例知识图谱

Fig. 7 A case knowledge graph constructed by JRB

当设备发生故障或检测出缺陷时,通过输出包含设备及现象的信息,利用文中模型完成对其中设备与现象描述的三元组抽取,再结合案例知识图谱,通过 Cypher 语言或子图匹配等进行快速检索,返回与当前事件相似的历史案例记录,获取其中该类设备发生相似现象时的故障或缺陷位置、原因等信息,完成智能诊断,丰富诊断方式;同时,结合设备的历史检修信息、运维规范以及故障预案等,生成处置决策以辅助运维人员工作。并将当前事件中抽取获得的实体及关系信息加入到图谱中实现动态更新。在使用知识图谱进行信息检索时,可

通过关系搜索、多跳搜索获取更多隐藏的知识与关联信息,有助于运维人员了解故障机理与缺陷起源、理解处置方案。

5 结论

针对设备故障缺陷文本中存在大量三元组重叠或嵌套的特点,文中提出了一种基于关系导向的联合抽取模型。该模型将三元组抽取过程建模为两个实体之间的映射联系,通过关系导向下的分层标记方式有效保证了对重叠或嵌套三元组的处理能力,并同时避免了冗余实体推断与误差累积。通

过实验验证得到了以下结论:

1)所提出的JRB模型在电力公司实际采集的样本上达到了95.47%以上的 F_1 值,能准确、有效地识别出样本中各类实体及关系信息,并在面临实体嵌套或重叠时仍能保持良好的抽取效果,克服了流水线模型中传统序列标注方法无法应对三元组嵌套与重叠问题,将 F_1 值提升了8.57%~25.19%。

2)所提出的JRB模型采用共享编码以及联合优化方法增强了子任务之间的交互联系;同时通过加入Bi-LSTM模块增加了模型对头实体语义特征的解析与提取能力;此外,在不同的故障缺陷文本数据集上,JRB模型仅需3~5次迭代便可收敛,并将 F_1 值提升至90%以上,保证了模型的训练效率。

参考文献:

- [1] 马富齐,穆睿昕,贾 嵘,等.基于声光融合成像特征解析的电力设备局部放电精细识别方法研究[J]. 电力系统保护与控制, 2025,53(11):51-62.
MA Fuqi, MU Ruixin, JIA Rong, et al. Refined identification method for partial discharge in power equipment based on acoustic-optical fusion imaging feature analysis[J]. Power System Protection and Control, 2025, 53(11): 51-62.
- [2] 赵洪山,王惠东,刘婧萱,等.考虑局部纹理特征和全局温度分布的电力设备红外图像超分辨率重建方法[J]. 电力系统保护与控制, 2025, 53(2): 89-99.
ZHAO Hongshan, WANG Huidong, LIU Jingxuan, et al. Super-resolution reconstruction method for infrared images of power equipment considering local texture features and global temperature distribution[J]. Power System Protection and Control, 2025, 53(2): 89-99.
- [3] 董理科,白 鹭,武 娜,等.基于知识图谱的电力变压器故障预测方法研究[J]. 高压电器, 2022, 8(11): 151-159.
DONG Like, BAI Lu, WU Na, et al. Research on fault prediction method of power transformer based on knowledge map[J]. High Voltage Apparatus, 2022, 8(11): 151-159.
- [4] 郭 榕,杨 群,刘绍翰,等.电网故障处置知识图谱构建研究与应用[J]. 电网技术, 2021, 45(6): 2092-2100.
GUO Rong, YANG Qun, LIU Shaohan, et al. Construction and application of power grid fault handling knowledge graph[J]. Power Grid Technology, 2021, 45(6): 2092-2100.
- [5] 叶欣智,尚 磊,董旭柱,等.面向配电网故障处置的知识图谱研究与应用[J]. 电网技术, 2022, 46(10): 3739-3749.
YE Xinzhi, SHANG Lei, DONG Xuzhu, et al. Knowledge graph for distribution network fault handling[J]. Power Grid Technology, 2022, 46(10): 3739-3749.
- [6] 王骏东,杨 军,裴洋舟,等.基于知识图谱的配电网故障辅助决策研究[J]. 电网技术, 2021, 45(6): 2101-2112.
WANG Jundong, YANG Jun, PEI Yangzhou, et al. Distribution network fault assistant decision-making based on knowledge graph [J]. Power Grid Technology, 2021, 45(6): 2101-2112.
- [7] 张焕龙,周珂燕,王延峰,等.基于掩码记忆的无人机电力设备分割跟踪方法研究[J]. 电力系统保护与控制, 2024, 52(21): 140-150.
ZHANG Huanlong, ZHOU Keyan, WANG Yanfeng, et al. A segmentation and tracking method for UAV power equipment based on mask memory[J]. Power System Protection and Control, 2024, 52(21): 140-150.
- [8] 饶 桐,龚泽威一,于 虹,等.知识图谱在输变电设备运维中的应用[J]. 电网与清洁能源, 2024, 40(6): 91-96.
RAO Tong, GONG Zeweiyi, YU Hong, et al. Application of knowledge graph in operation and maintenance of power transmission equipment[J]. Power System and Clean Energy, 2024, 40(6): 91-96.
- [9] 郑伟彦,杨 勇,卢家驹,等.面向配电网知识图谱的配电调度文本实体链接方法[J]. 电力系统保护与控制, 2021, 49(4): 111-117.
ZHENG Weiyang, YANG Yong, LU Jiaju, et al. Entity linking method of distribution dispatching texts for a distribution network knowledge graph[J]. Power System Protection and Control, 2021, 49(4): 111-117.
- [10] 李 刚,李银强,王洪涛,等.电力设备健康管理知识图谱:基本概念、关键技术及研究进展[J]. 电力系统自动化, 2022, 46(3): 1-13.
LI Gang, LI Yinqiang, WANG Hongtao, et al. Knowledge graph of power equipment health management: Basic concepts, key technologies and research progress[J]. Automation of Electric Power Systems, 2022, 46(3): 1-13.
- [11] LYU Ke, GAO Caixia, SI Jikai, et al. Fault coil location of interturn short-circuit for direct-drive permanent magnet synchronous motor using knowledge graph[J]. IET Electric Power Applications, 2020, 14(9): 1712-1721.
- [12] 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. Atlanta, USA: IEEE, 2020: 2502-2507.
- [13] YANG Q Y, JIANG J, FENG X Y, et al. Named entity recognition of power substation knowledge based on transformer-BiLSTM-CRF network[C]//Proceedings of 2020 International Conference on Smart Grids and Energy Systems(SGES). Perth, Australia: IEEE, 2020: 952-956.
- [14] ZHENG Kaihong, SUN Lingyun, WANG Xin, et al. Named entity recognition in electric power metering domain based on attention mechanism[J]. IEEE Access, 2021(9): 152564-152573.
- [15] 朱欣娟,童小凯,王西汉,等.一种融合协同因子的知识图谱传播推荐模型[J]. 西安工程大学学报, 2022, 36(2): 79-86.
ZHU Xinjuan, TONG Xiaokai, WANG Xihan, et al. A graph recommender model for knowledge graph propagation with collaborative factor[J]. Journal of Xi'an Polytechnic University, 2022, 36(2): 79-86.
- [16] 田嘉鹏,宋 辉,陈立帆,等.面向知识图谱构建的设备故障文本实体识别方法[J]. 电网技术, 2022, 46(10): 3913-3922.
TIAN Jiapeng, SONG Hui, CHEN Lifan, et al. Entity recognition approach of equipment failure text for knowledge graph construction[J]. Power Grid Technology, 2022, 46(10): 3913-3922.
- [17] 陈志忠,熊泽森,姚 东,等.基于知识图谱的电力杆塔主要构件识别方法研究[J]. 浙江电力, 2024, 43(5): 100-108.
CHEN Zhizhong, XIONG Zesen, YAO Dong, et al. Research on a recognition method of main components of electric power towers us-

(下转第70页)

- 程学报, 2019, 39(13): 3998-4007.
- LIU Yunpeng, XU Ziqiang, DONG Wangying, et al. Concentration prediction of dissolved gases in transformer oil based on empirical mode decomposition and long short-term memory neural networks [J]. Proceedings of the CSEE, 2019, 39(13): 3998-4007.
- [24] 舒畅, 金潇, 李自品, 等. 基于CEEMDAN的配电变压器放电故障噪声诊断方法[J]. 高电压技术, 2018, 44(8): 2603-2611.
- SHU Chang, JIN Xiao, LI Zipin, et al. Noise diagnosis method of distribution transformer discharge fault based on CEEMDAN[J]. High Voltage Engineering, 2018, 44(8): 2603-2611.
- [25] 唐竹, 肖宇航, 郭淳, 等. 基于CEEMDAN模态分解和TCN-BiGRU的短期电力负荷预测[J]. 智慧电力, 2024, 52(12): 59-64.
- TANG Zhu, XIAO Yuhang, GUO Chun, et al. Short-term electricity load forecasting based on CEEMDAN decomposition and TCN-BiGRU model[J]. Smart Power, 2024, 52(12): 59-64.
- [26] 李佳, 邓科, 侯玉莲, 等. 基于GRA-CEEMDAN-BiLSTM的变压器油中溶解气体浓度预测[J]. 变压器, 2022, 59(6): 42-47.
- LI Jia, DENG Ke, HOU Yulian, et al. Prediction of dissolved gas concentration in transformer oil based on GRA - CEEMDAN - BiLSTM[J]. Transformer, 2022, 59(6): 42-47.
- [27] 马秉伟, 陈晓国, 郑宇, 等. 电力变压器环保绝缘油研究进展与趋势[J]. 南方电网技术, 2024, 18(5): 12-21.
- MA Bingwei, CHEN Xiaoguo, ZHENG Yu, et al. Research progress and trends of eco-friendly insulating oil for power transformers[J]. Southern Power System Technology, 2024, 18(5): 12-21.
- [28] TORRES M E, COLOMINAS M A, SCHLOTTHAUER G, et al. A complete ensemble empirical mode decomposition with adaptive noise[C]//2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). CZ: IEEE, 2011: 4144-4147.
- [29] 陈铁, 陈一夫, 李咸善, 等. 基于SDS-SSA-LSTM的变压器油中溶解气体浓度预测[J]. 电子测量技术, 2022, 45(12): 6-11.
- CHEN Tie, CHEN Yifu, LI Xianshan, et al. Prediction of dissolved gas concentration in transformer oil based on SDS-SSA-LSTM[J]. Electronic Measurement Technology, 2022, 45(12): 6-11.
- [30] 王延年, 王栋, 廉继红, 等. 局部遮荫下基于IP & O-SSA的MPPT控制研究[J]. 西安工程大学学报, 2023, 37(4): 110-117.
- WANG Yannian, WANG Dong, LIAN Jihong, et al. Research on MPPT control based on IP & O-SSA under partial shading[J]. Journal of Xi'an Polytechnic University, 2023, 37(4): 110-117.
- [31] HEIDARI A A, MIRJALILI S, FARIS H, et al. Harris hawks optimization: Algorithm and applications[J]. Future Generation Computer Systems, 2019(97): 849-872.
- 傅雨晨(1999—), 女, 硕士研究生, 主要研究方向为人工智能应用(E-mail: 202108580021063@ctgu.edu.cn).
- 陈星(1988—), 女, 硕士, 讲师, 主要研究方向为电力设备状态监测与故障诊断(通信作者)(E-mail: chenxing20230511@163.com).
- 付文龙(1988—), 男, 博士, 副教授, 博导, 主要研究方向为人工智能应用、电力设备状态监测与故障诊断(E-mail: ctgu_fuwenlong@126.com).
- (上接第49页)
- ing knowledge graph[J]. Zhejiang Electric Power, 2024, 43(5): 100-108.
- [18] 王艳, 连洪铤, 王寅初, 等. 基于改进边缘注意力生成对抗网络的电力设备热成像超分辨率重建[J]. 电力系统保护与控制, 2024, 52(3): 119-127.
- WANG Yan, LIAN Hongbo, WANG Yinchu, et al. Super-resolution reconstruction of thermal imaging of power equipment based on improved edge-attentive generative adversarial networks[J]. Power System Protection and Control, 2024, 52(3): 119-127.
- [19] ZHENG Suncong, WANG Feng, BAO Hongyun, et al. Joint extraction of entities and relations based on a novel tagging scheme[C]// Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics. Vancouver, Canada: ACL, 2017: 1227-1236.
- [20] ZENG Xiangrong, ZENG Daojian, HE Shizhu, et al. Extracting relational facts by an end-to-end neural model with copy mechanism[C]// Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics. Brussels, Belgium: ACL, 2018: 506-514.
- [21] WEI Zhepei, SU Jianlin, WANG Yue, et al. A novel cascade binary tagging framework for relational triple extraction[C]// Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. Online: ACL, 2020: 1476-1488.
- [22] YU Bowen, ZHANG Zhenyu, SHU Xiaobo, et al. Joint extraction of entities and relations based on a novel decomposition strategy[C]// Proceedings of the 24th European Conference on Artificial Intelligence. Santiago de Compostela, Spain: ECAI, 2020.
- [23] WANG Yucheng, YU Bowen, ZHANG Yueyang, et al. TPLinker: Single-stage joint extraction of entities and relations through token pair linking[C]// Proceedings of the 28th International Conference on Computational Linguistics. Barcelona, Spain(Online): ICCL, 2020: 1572-1582.
- [24] LI Xiaoya, FENG Jingrong, MENG Yuxian, et al. A unified mrc framework for named entity recognition[C]// Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. Online: ACL, 2020: 5849-5859.
- [25] WU Shanchan, HE Yifan. Enriching pre-trained language model with entity information for relation classification[C]// Proceedings of the 28th ACM international conference on information and knowledge management. Beijing, China: ACM, 2019: 2361-2364.
- 李艾青(1999—), 女, 硕士研究生, 主要研究方向为输变电设备状态监测与智能化方面的研究工作(E-mail: li_aiqing@sjtu.edu.cn).
- 宋辉(1987—), 男, 博士, 副研究员, 主要从事电力设备智能运维工作(通信作者)(E-mail: songeos@163.com).