

DOI: 10.19783/j.cnki.pspc.211193

# 基于知识图谱的二次设备测试自动配置方法

叶远波<sup>1</sup>, 李端超<sup>1</sup>, 谢民<sup>1</sup>, 王志华<sup>2</sup>, 王骏东<sup>2</sup>

(1. 国网安徽省电力有限公司, 安徽 合肥 230022; 2. 武汉凯默电气有限公司, 湖北 武汉 430223)

**摘要:** 随着智能变电站的发展, 二次设备更新换代频繁且智能化程度不断提升, 以往采用手动测试配置的方法已难以满足工作要求。提出一种基于知识图谱的二次设备自动配置方法。首先, 模拟常规测试核心流程, 构建基于测试任务的信息点、基于设备厂家的信息点检索路径以及基于测试配置信息类型的命名实体知识三类知识图谱。然后, 根据测试设备及测试功能, 利用所构建的知识图谱检索需配置信息点, 确定设备厂家并基于子图匹配信息路径。采用改进的 BERT 模型计算配置信息点与描述文本之间的相似性, 从而完成映射配置。此外, 还加入了端口地址配置数据字符串校验。实验仿真结果表明, 基于知识图谱的二次设备测试自动配置方法合理有效, 并具有极高的准确性。

**关键词:** 二次设备测试; 知识图谱; BERT 模型; 文本向量; 路径推理

## Automatic configuration method of secondary equipment test based on a knowledge graph

YE Yuanbo<sup>1</sup>, LI Duanchao<sup>1</sup>, XIE Min<sup>1</sup>, WANG Zhihua<sup>2</sup>, WANG Jundong<sup>2</sup>

(1. State Grid Anhui Electric Power Co., Ltd., Hefei 230022, China; 2. Wuhan Kemov Electric Co., Ltd., Wuhan 430223, China)

**Abstract:** With the development of smart substations, the replacement of secondary equipment is frequent and intelligence is increasingly improved. The former manual test configuration method finds it difficult to meet the work needs. An automatic configuration method for secondary equipment based on a knowledge graph is proposed. First, the core process of conventional testing is simulated to build a knowledge graph based on three types of data, test task information points, information point retrieval paths of equipment manufacturers, and named entity knowledge of test configuration information types. Then, from the test equipment and testing capabilities, the information points needed to be configured are retrieved by the constructed knowledge graph, the device manufacturer is determined and the information path is matched based on the subgraph. The improved BERT model is used to calculate the similarity between the configuration information points and the description text to complete the mapping configuration. In addition, a string verification of port address configuration data is added. The simulation results show that the method based on the knowledge graph is reasonable and effective and has high accuracy.

This work is supported by the Science and Technology of State Grid Corporation of China (No. 5212002000AR).

**Key words:** secondary equipment test; knowledge graph; BERT model; text vector; path reasoning

## 0 引言

随着智能变电站快速推进, 变电站二次设备呈现出种类数量急剧增加、自动化程度明显提升、安全要求不断加强的特点<sup>[1-5]</sup>, 为变电站二次设备测试的发展带来了机遇与挑战。智能电子设备 IED(Intelligent Electronic Devices, IED)基于统一的数字化标准规约 IEC61850 实现通信<sup>[6]</sup>。同样, 二次设备测试也基于 IEC61850 规约解析 IED 文件实现测试配置<sup>[7]</sup>。

目前在变电站二次设备测试时, 检测人员需要根据应用功能种类与测试策略配置对应事件、信号、数据等信息点, 基于不同被测装置及应用业务的测试方案故障参数已经预制完成, 通过模拟故障设置测试参数并监测保护装置动作情况, 从而验证保护效果及装置参数设置准确与否<sup>[8-9]</sup>。

信息点配置是整个检测过程的关键环节, 即在导入全站系统配置 SCD(Substation Configuration Description)文件后, 根据出厂配置信息文件(ICD 文件), 基于 IED 输出端口地址配置数据及接口文本描述等信息特征, 实现待检测智能电子设备 IED 的模

基金项目: 国家电网公司科技项目资助(5212002000AR)

拟数据集与实例信息虚链接一一映射, 但因为 ICD 虚端子的 dU 或 dAI 描述没有统一规范, 所以大多采用手动配置的方法。变电站设备繁多, 人工配置测试导致效率低下、配置耗时耗力、测试周期长、人员经验及工作状态差异明显, 易出现缺项漏项, 出错后纠错的工作量也很大。

针对这一问题, 已有相关学者对于信息自动配置方案进行了探讨。文献[10]通过子模版匹配 SCD 文件的方式实现继电保护测试配置的实例化, 为二次设备测试提供了新思路, 但是该方案的可扩展性及泛化性有待提升。文献[11]提出基于 SCD 文件的专家系统完成检修二次信息的匹配, 实现智能变电站测试中安措票的配置, 但是分类较为简单且耗时较长。文献[9]及文献[12]均提出了基于利用自然语言处理技术实现二次设备自配置, 文献[6]采用 TextCNN 对智能录波器配置文件进行文本分类实现端口地址映射, 这种方法对于长文本效果较好, 但对于短文本具有局限性, 且未考虑其他端口地址数据; 文献[12]利用词向量表示技术计算虚端子语义相似度实现链接匹配, 能够得到较好的匹配效果, 但采用分词表示的方法在词向量表示上未考虑文本的全局语义信息, 仍有提升空间。通过以上讨论可以发现, 二次设备自动配置方案主要基于两种思路: 一是利用模板等先验知识进行匹配, 准确性高但泛化性低; 二是利用配置信息相似性进行分类或相似度计算, 适应性强但均为单设备配置。

知识图谱作为一种模拟人类思维方式的多元异构实体及关系表征大数据技术<sup>[13]</sup>, 可解释性及数据存储结构性能优越<sup>[14-15]</sup>, 已在数据检索、推荐系统、知识推理等方面得到广泛运用<sup>[16]</sup>, 同时在电力领域的数据运维、电力文本表征及电网动态辅助决策等方面做出了初步研究探索<sup>[17-20]</sup>。大量投运变电站的 SCD 可作为知识数据蒸馏样本基础, 且知识图谱技术能够很好地表征 IED 文件中的文本、字符等多源异构数据。因此, 本文提出基于知识图谱的二次设备测试自动配置方法, 将先验知识匹配与配置信息相似性计算两种配置方式的优势相结合, 利用知识图谱多跳路径匹配进行配置信息查询及抓取, 并采用改进 BERT 模型计算配置信息相似度, 从而实现准确高效的多设备多任务二次设备测试自动配置。

## 1 二次设备测试配置文件信息解析

智能变电站二次设备测试配置需根据测试任务映射对应数据集参数信息点, 不同测试任务的信息需求具有差异性, 在确认目标信息需求的基础上, 通过 IED 输出端口地址配置数据与文本描述信息映

射对应信息点, 即可实现测试信息配置。SCD 文件解析示意图如图 1 所示, 配置所需端口地址配置数据与文本描述信息均存储于 SCD 文件中, 该文件描述了智能变电站中各 IED 的数据模型、配置信息及相互间逻辑链接关系, 所有装置在 SCD 文件中均表现为一级 IED 节点及其节点内嵌套的各层子节点信息, 遵循 IEC61850 规约<sup>[21]</sup>, IED 结构主要包含访问点、逻辑设备、逻辑节点、数据属性、数据信息等。IED 之间通过虚端子链接, 主要包含采样值(SV)与通用对象变电站事件(GOOSE)信息, 解析配置文件即明确 IED 的 SV 与 GOOSE 的信息关系, 由于 SCD 数据嵌套复杂性及 SV 与 GOOSE 信息关系的隐晦性, 故通过匹配端口地址配置数据与文本描述完成信息关系映射与信息配置。

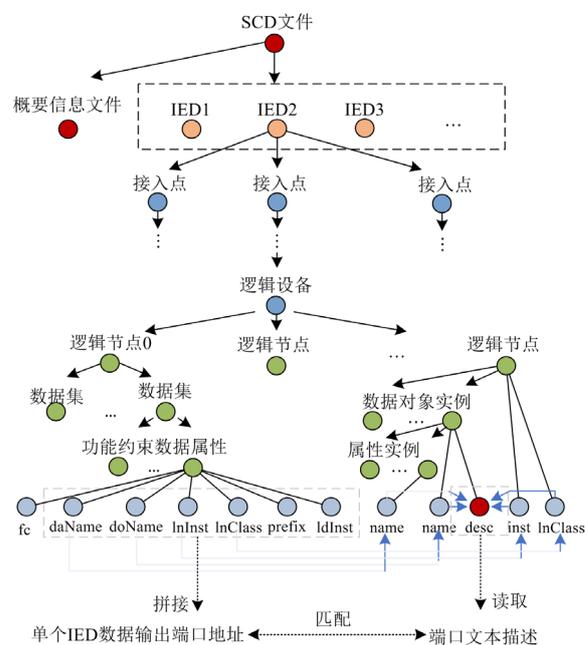


图 1 SCD 文件解析示意图

Fig. 1 Schematic diagram of SCD file parsing

本文采用元素树方式遍历 SCD 文件的嵌套数据, 以各设备 IED 节点为初始节点, 按照 IED 层次结构深度优先为原则访问其各层子节点的属性及元素, 然后通过字典列表的形式存储端口地址配置数据、文本描述及对应嵌套路径, 如此既获得了匹配需要的关键信息, 又保留了信息获取路径, 为构建二次设备测试自动配置知识图谱奠定数据基础。

## 2 知识图谱构建

智能变电站二次设备测试配置需要依次明晰三个内容, 首先是确定完成某种任务功能测试所需要配置的全部信息点, 随后是根据不同设备厂家的

IED 数据格式清楚对应信息的层次路径及描述特征，最终基于信息特征判断配置信息点的映射关系正确性。基于以上配置逻辑与流程，本文设计的二次设备测试自动配置知识图谱数据模型及架构共包含三类，如图 2 所示，分别是基于测试任务的信息点知识图谱、基于设备厂家的信息点检索路径知识图谱、基于测试配置信息类型的实体命名知识图谱。接下来针对不同部分知识图谱构建进行介绍。

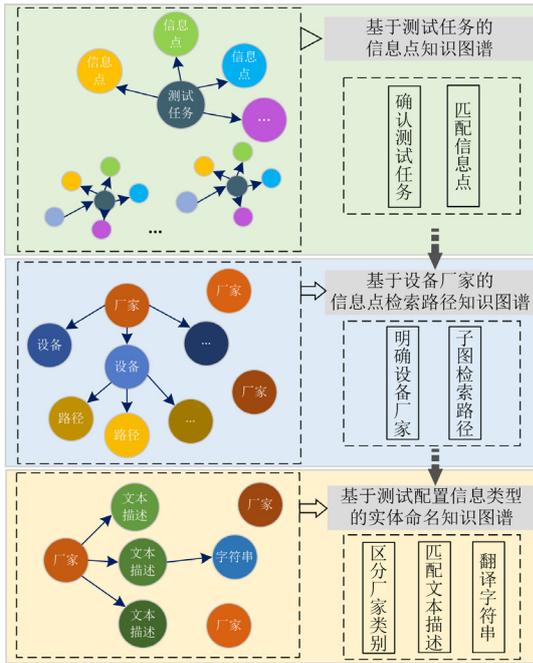


图 2 二次设备测试自动配置知识图谱数据模型及架构  
Fig. 2 Data model and architecture of knowledge graph for automatic configuration of secondary equipment test

### 2.1 测试任务信息点

智能变电站中需要测试的二次设备包括继电保护装置、合并单元、智能终端、智能录波器等多种设备，不同二次设备相对应的测试内容截然不同，且同种设备在不同状态下的测试内容也具有一定的差异性。例如继电保护装置测试具有变压器、线路、母线等多种位置场景，测试内容包含定值设置、采样精度、动作信息、开入开出量等，所需配置的信息点包含 SV 输入、GOOSE 输入及输出端口地址配置数据、保护功能压板、保护动作元件、保护动作相别及保护动作时间等。

针对该部分内容，在二次设备测试规定文件及厂家测试配置信息中已做出了对应说明。本文采用模板匹配的方式对文件进行抽取、匹配、融合并进行人工校验，构建的二次设备继电保护测试的部分信息点知识图谱数据示例如图 3 所示。

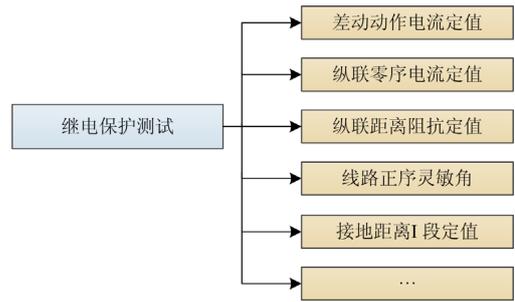


图 3 部分信息点知识图谱数据示例  
Fig. 3 Some information points knowledge graph data examples

### 2.2 设备厂家信息点检索路径

二次设备测试信息配置工作具有重复性及经验性，对应信息点在 SCD 文件中的位置在一定程度上相对统一，故而可通过寻找相似信息路径的方式来配置信息。

本文以 2.1 节图谱中的信息点为核心及出发点，在解析的字典中以深度优先的检索方式进行匹配，查找到对应信息则保存由 IED 节点至该信息点的检索路径，对同类信息的路径进行统计、筛查，保留路径相似度较高的路径作为检索路径。

但这种方法也存在一定的漏洞，例如二次设备厂家主要有南瑞、许继、四方等，各厂家间 IED 层次嵌套具有一定区别，不同版本设备信息表征方式同样可能具有差异性，且存在设备配置端口地址配置数据与描述信息规范性不足的问题。

针对以上问题提出两种解决思路：一是对不同厂家设备构建不同的信息点检索路径以解决厂家不同的问题；二是信息点检索采用多路径检索，因为通过观察解析保存的字典信息发现，一般前几层的嵌套信息路径基本统一，在逻辑节点后的从属层信息路径及内容才会出现差异，多为版本差异造成，因此通过保留多个相似度较高路径作为备选路径。

### 2.3 测试配置信息命名实体

当完成信息点确认及厂家信息点路径检索后，便到达二次设备配置的最终环节，即确认最终的测试点信息点与数据集的映射关系，也是整个配置环节中最关键的步骤。端口地址配置数据与文本描述信息虽然在格式上存在较大区别，但同一厂家设备信息逻辑上具有强关联性，相同实体命名及表述基本统一，因此挖掘其实体命名规律对于信息自动配置具有重要作用。

基于处理 SCD 文件获得的字典列表数据，抽取厂家、端口地址配置数据及文本描述三类信息，链接实体以 {厂家, 端口地址配置数据, 文本描述} 形

成三元组, 初始抽取结果存在大量冗余数据, 一类是完全重复的数据, 另一类是由于区分同类数据不同实体而给端口地址配置数据增加的无实意编号。

本文将三元组中端口地址配置数据信息中的数字量信息删除, 仅保留字母字符串量信息。然后以厂家与文本描述的融合信息构建索引, 统计每类索引中端口地址配置数据的种类及个数, 按照统计结果将三元组进行转化, 仅保留同类索引中相同端口地址配置数据的一个三元组。此外, 将相同三元组的统计个数以属性值性质嵌入, 从而实现将大量冗余实体消歧并融合, 最终形成厂家、文本描述、端口地址配置数据表述的简化子图, 基于测试配置信息类型的部分命名实体知识图谱数据示例如图 4 所示。

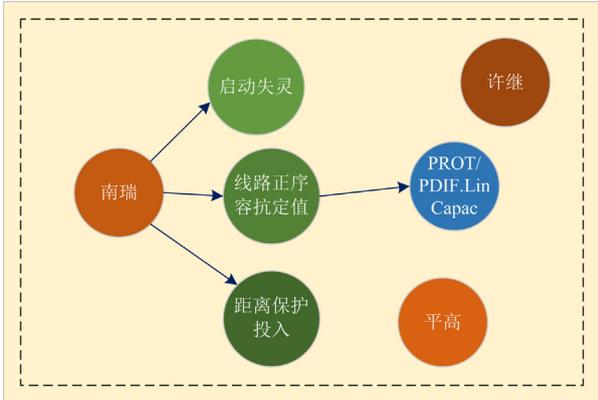


图 4 部分命名实体知识图谱数据示例

Fig. 4 Examples of partial named entity knowledge graph data

### 3 基于知识图谱的配置自动推理方法

#### 3.1 基于改进 BERT 模型的文本相似度匹配算法

短文本相似性不是直接计算两个原始文本, 需将源文本转换为计算机可以理解和处理的数据, 而计算机更善于处理向量模式数据, 所以一般将文本聚焦于向量空间中进行度量, 常见的模型有词袋模型和词嵌入。词袋模型通常与词频-逆文本频率联合使用, 常见的有向量空间模型、浅语义分析、隐含狄利克雷分布等, 词袋模型较为简单但具有一定的限制性, 没有考虑词序、稀疏性、复杂性问题, 对于处理本文较短的电力描述信息难以挖掘足够的语义信息。词嵌入常见的如 Word2Vec、TextCNN、BERT 模型等<sup>[22]</sup>, 在泛化性及语义信息理解上有了极大的提升。本文选取目前应用最为广泛的 BERT 模型为基础模型, 考虑文本较短的情况, 故而采用基于字向量的 BERT 模型对文本进行分布式嵌入。BERT 模型基础结构如图 5 所示。

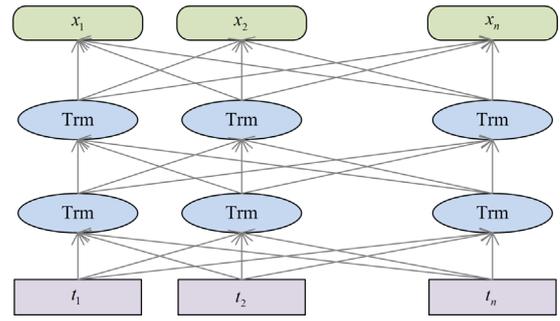


图 5 BERT 模型结构

Fig. 5 BERT model structure

BERT 模型利用了多头注意力机制的双 Transformer 编码器(Trm)结构<sup>[23]</sup>, 每个编码器内含有 6 个 Encoder block, 将按照字粒度切分后的文本表示为  $T = (t_1, t_2, \dots, t_n)$ , 其中  $t_i$  为文本  $T$  的第  $i$  个字。采用 word2vec 训练好的字粒度嵌入向量作为初始字向量, 同时加入片段向量及位置向量。三种特征向量加和得到综合嵌入特征向量  $X = (x_1, x_2, \dots, x_n)$ ,  $x_i$  为文本  $T$  的第  $i$  个字的综合字向量, 将向量  $X$  切割为  $a$  份, 将每部分切割后的矩阵  $X_j (j=1, 2, \dots, a)$  输入 Transformer 中, 按照式(1)~式(3)做线性化映射得到初始目标字向量  $Q$ 、上下文向量  $K$  以及价值向量  $V$ 。其中,  $W^Q$ 、 $W^K$ 、 $W^V$  分别为  $X_j$  映射到  $Q$ 、 $K$  以及  $V$  的线性变换矩阵。

$$Q = Linear_1(X_j) = XW^Q \quad (1)$$

$$K = Linear_1(X_j) = XW^K \quad (2)$$

$$V = Linear_1(X_j) = XW^V \quad (3)$$

将初始向量采用注意力机制求解, 采用 Softmax 函数归一化为概率分布, 如式(4)所示。

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

式中,  $d_k$  为  $K$  向量维度大小,  $\sqrt{d_k}$  为尺度标度以防止  $Q \cdot K$  结果过大。此时得到的字向量考虑了上下文信息。将  $a$  个  $Z_j$  拼接得到多头注意力矩阵, 训练机制如式(5)、式(6)所示。

$$\vec{Z} = MultiHead(Z_i) = Concat(Z_1, \dots, Z_n) \cdot W^0 \quad (5)$$

$$\text{where } Z_i = head_i = Attention(Q_i, K_i, V_i) \quad (6)$$

式中:  $W^0$  为训练初始矩阵;  $Concat(\cdot)$  表示拼接函数。将多头自注意力机制向量  $\vec{Z}$  与综合嵌入特征向量  $X$  进行标准化处理并输入前馈神经网络(Feedforward Neural Network, FNN), 如式(7)、式(8)。

$$\vec{Z} \leftarrow LayerNorm(\vec{Z} + X) \quad (7)$$

$$FFN(\vec{Z}) = g(\vec{Z} \cdot \mathbf{W}_1 + b_1)\mathbf{W}_2 + b_2 \quad (8)$$

式中： $\mathbf{W}_1$ 、 $\mathbf{W}_2$ 、 $b_1$ 及 $b_2$ 为训练需要的参数； $g(\cdot)$ 为激活函数。此时便得到了基于字粒度的短文本向量表示模型。

考虑基于字向量的文本嵌入模型对于电力领域词汇特征提取有一定的局限性，缺乏一定的先验知识与背景信息，对于专有词汇的表征可能会出现一定偏差。因此本文提出在字向量基础上融合二次测试专有词汇的词嵌入向量，从而丰富描述短文本的特征表征信息。

首先利用自定义的二次测试专有词汇词典抽取语料库短文本中的关键词，得到关键词序列  $\mathbf{KD} = (kd_1, \dots, kd_b)$ ，其中  $b$  为每个语料所含关键词个数，每个关键词依照出现顺序进行排列。将  $\mathbf{KD}$  作为文本序列输入 BERT 模型中进行训练，即每个  $kd$  相当于一个  $t_i$ ，所存在不同为本次训练无需分词，且以词粒度文本进行训练，其他训练过程与基于字粒度的 BERT 模型一致，最终可得到基于关键词的词粒度 BERT 模型语义向量表示模型。假设文本基于字粒度 BERT 模型所得表示向量为  $\mathbf{X}_d$ ，基于关键词的词粒度 BERT 模型所得表示向量为  $\mathbf{X}_{kd}$ ，将两个表示向量进行拼接，即得到最终由字向量及关键词向量结合的二次设备配置短文本分布式表示向量  $\mathbf{X}_g$ 。

$$\mathbf{X}_g = \text{Concat}(\mathbf{X}_d, \mathbf{X}_{kd}) \quad (9)$$

基于获得的文本向量，采用应用最为广泛的余弦相似度度量文本相似度，如式(10)。

$$\text{sim}(\mathbf{T}_1, \mathbf{T}_2) = \frac{\mathbf{X}_g^{\mathbf{T}_1} \cdot \mathbf{X}_g^{\mathbf{T}_2}}{\|\mathbf{X}_g^{\mathbf{T}_1}\| \times \|\mathbf{X}_g^{\mathbf{T}_2}\|} \quad (10)$$

求得的距离与相似度成反比，距离越小则相似度越大，与给定配置信息点距离最小的短文本描述即为所要匹配映射的配置信息描述。最终改进 BERT 模型处理流程如图 6 所示。

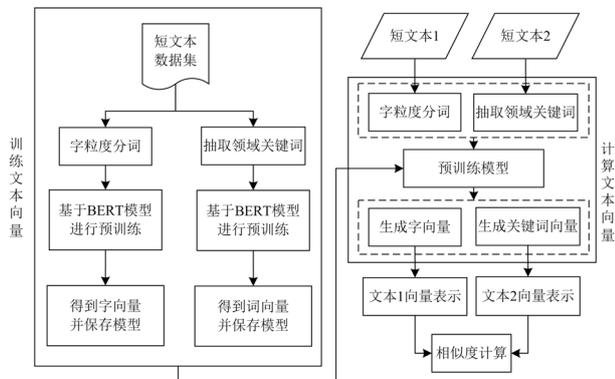


图 6 改进 BERT 模型处理流程

Fig. 6 Improved BERT model processing process

### 3.2 多图分级推理的信息配置模型

二次设备测试配置本质上是一项信息检索匹配任务，任务具有很强的程式性与逻辑性，模型需要协调信息配置顺序，强调配置结果准确率，同时兼顾配置效率，在考虑以上因素的基础上本文提出多图分级推理的信息配置模型来实现二次设备测试的自动配置，具体流程如图 7 所示。

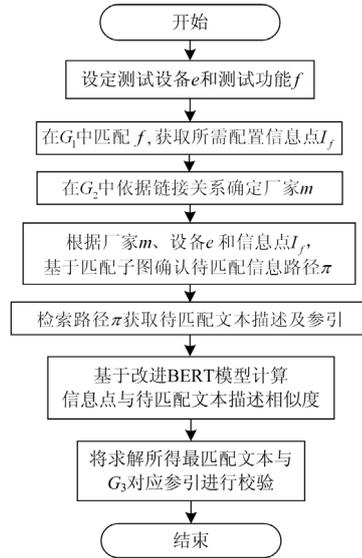


图 7 配置具体流程图

Fig. 7 Configuration flowchart

首先，给定测试设备  $e$  以及待测试功能  $f$ ，在基于测试任务的信息点知识图谱  $G_1$  中，依据关系链式 (11) 进行匹配，获取测试功能  $f$  需配置的所有信息点  $I_f$ 。

$$R_{(f,I)} : f \rightarrow \text{需配置} \rightarrow I_f \quad (11)$$

随后，在基于设备厂家的信息点检索路径知识图谱  $G_2$  中，依据关系链(12)确定设备  $e$  的厂家类别  $m$ 。

$$R_{(e,m)} : e \rightarrow \text{厂家} \rightarrow m \quad (12)$$

接下来，将  $G_1$  的信息点实体与  $G_2$  中厂家实体建立链接，通过厂家、设备及信息点构造子图，通过子图检索确认待匹配信息路径，所要匹配子图构造如图 8 所示。

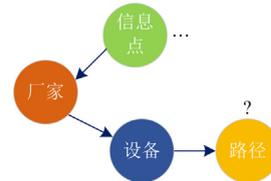


图 8 检索路径匹配子图模板构造

Fig. 8 Construction diagram of retrieve path matching the subgraph template

在获得实体待匹配信息路径  $\pi$  后, 根据路径即可找到全部  $l$  个待匹配本文描述数据集合  $\{T_1, T_2, \dots, T_l\}$ ,  $l \in N^+$ , 计算文本集合中每个短文本与信息点  $I_f$  中对应文本描述的距离值, 距离值最小的文本为所得映射的配置信息描述  $T_0$ 。

通过实验发现计算文本相似度值可以准确地将所要映射的文本检测出来, 但有时匹配结果不止一个, 这是因为配置时部分描述文本构建不完整或重复错误映射配置, 此时要想区分数据映射只能通过端口地址配置数据信息, 每个数据的端口地址配置数据是唯一的。因此, 本文提出基于测试配置信息类型的实体命名知识图谱  $G_3$ , 通过端口地址配置数据校核以解决文本描述无法确定映射数据的情况, 利用端口地址配置数据能够辨别映射关系是因为端口地址配置数据的命名具有一定规律性, 大部分字符串具有实际意义, 且在  $G_3$  中已将字符串与对应描述建立索引形成先验知识。

字符串与描述具有一对多的相互映射关系, 因此将  $I_f$  的文本进行字符串翻译, 若按厂家类别能匹配上  $G_3$  中描述则将对应位置文本转换为对应字符串, 若出现未翻译文字描述则将其删除, 但保留字母及数字描述, 按信息点描述信息顺序重组得到翻译字符串  $S^0$ 。基于 Levenshtein 编辑距离计算相似度<sup>[24]</sup>, 编辑距离以字符串转换操作次数作为衡量指标计算, 如式(13)。

$$D(S^0, S') = D_{ij} = \begin{cases} 0, & i=0 \text{ and } j=0 \\ \min \begin{cases} D_{i-1, j-1} + c_a & 0 \leq i \leq L_{S^0} \\ D_{i-1, j} + c_b, \text{ otherwise} & 0 \leq j \leq L_{S'} \\ D_{i, j-1} + c_c \end{cases} \end{cases} \quad (13)$$

式中:  $S^0$  为翻译重组的目标字符串;  $S'$  为待匹配的源字符串;  $c_a$ 、 $c_b$ 、 $c_c$  分别表示删除、插入、替换三种操作权重, 本文设置  $c_a = 1$ 、 $c_b = 1$ 、 $c_c = 2$ ;  $L_{S^0}$ 、 $L_{S'}$  分别表示字符串  $S^0$ 、 $S'$  长度。

基于编辑距离计算字符串相似度公式为

$$\text{sim}(S^0, S') = 1 - \frac{D(S^0, S')}{L_{S^0} + L_{S'}} \quad (14)$$

$\text{sim}(S^0, S')$  值越大, 则字符串相似度越高, 选取相似度最高的字符串作为最终匹配的映射配置信息, 考虑到二次设备测试配置错误易造成设备损坏或测试结果无效, 设置相似度最低阈值  $\varepsilon$ , 低于阈值则匹配失败转由人工校核。

## 4 算例结果与分析

### 4.1 实验数据

为了验证本文所提模型及方法在二次设备测试配置任务中的有效性, 本文选取许继、南瑞、平高三个厂家的设备进行测试, 重点研究保护、测控、智能终端及合并单元典型测试任务<sup>[25]</sup>, 抽取各厂家已配置成功的测试设备映射关系 IED 数据 5 000 条作为配置样本。样本按厂家随机分成 5 份: 选择其中 4 份作为训练集, 构建知识图谱并训练文本向量; 1 份作为测试集, 且测试集不分厂家随机打乱, 利用测试集衡量算法及模型效果。部分数据样本如表 1 所示, 可以看出文本描述与配置信息点相似度很大, 但在描述上也具有一定区别。端口地址配置数据大部分字符串具有一定解释性, 例如“PIGO”表示跳闸, “PROT”表示保护, 但同样两个闭锁重合闸配置信息点端口地址配置数据却完全不同, 是因为厂家命名区别, 且存在例如“Amp1”“Ind90”等释义难以确定字符串, 但其在描述上具有高度相似性, 这也是本文选择主要以文本匹配信息点的原因之一。

表 1 部分样本示例

Table 1 Examples of partial samples

配置信息点	端口地址配置数据	文本描述
额定延迟时间	MUSV/LLN0.DelayTRtg	额定延时
保护 1 电流 A 相 1	MUSV/PATCTR1.Amp1	A 相保护电流 1
保护 1 电流 B 相 1	MUSV/PBTCTR1.Amp1	B 相保护电流 1
接地距离III段定值	PROT/PDIS9.GndStr	接地距离III段定值
闭锁重合闸	RPIT/YxGGIO1.Ind90	闭锁重合闸-1
闭锁重合闸	PIGO/PTRC3.BlkRecST	闭锁重合闸
永跳	PIGO/PTRC3.BlkRecST	永跳

### 4.2 评价指标

本文测试数据均为配置完成数据, 因此检索结果具有唯一性, 即通过最终基于本文所提方法的配置数据判断检索结果。匹配成功则为正确, 匹配失败则为错误, 匹配失败则继续匹配, 至多匹配 5 次。将一个信息点匹配作为一次匹配任务, 通过建立准确度(ACC)和平均倒数排名(MRR)作为匹配结果的评价指标<sup>[26]</sup>, 以评估任务路径匹配与文本相似度匹配效果, 如式(15)、式(16)所示。

$$ACC = \frac{n_a}{N_A} \quad (15)$$

$$MRR = \frac{1}{N_A} \sum_{i=1}^{N_A} \frac{1}{rank_i} \quad (16)$$

式中:  $n_a$  为匹配任务正确次数;  $N_A$  为总的任务匹

配次数;  $rank_i$  为一次任务匹配出正确结果的次序, 第一次匹配成功即为 1, 第二次即为 2, 以此类推,  $rank_i \leq 5$ 。

### 4.3 实验环境及参数

本次实验环境为 Intel(R) Xeon(R) CPU E5-2620 v4 @ 2.10 GHz 八核, NVIDIA Corporation GP102 [GeForce GTX 1080 Ti], Ubuntu 16.04 操作系统, 采用 python 语言编程, 改进 BERT 模型基于 TensorFlow 深度学习框架构建。模型中各参数设置如表 2 所示。

表 2 模型重要参数设置

Table 2 Setting of important parameters of the model

模型参数	数值
输入文本最大长度	64
批处理大小	32
文本向量维度	300
Transformer 层数	12
隐层维度大小	768
自注意力头数目	12
学习率	0.000 5
训练轮数	3
字符串相似度阈值	0.5

### 4.4 实验设置及结果分析

#### 4.4.1 路径检索方法验证

本文提出的二次设备测试配置方法首先要找到所要匹配的配置信息在 IED 文件中的嵌套位置, 即检索位置路径。通过在  $G_1$  图谱中匹配配置信息点可极大减少依靠经验的重复性工作, 实验中发现构建每类任务的配置信息点虽然具有一定的工作量, 但是构建完成后可一直使用, 能够提升配置准确率, 减少操作失误, 单次检索时间小于 0.001 s。此外, 基于  $G_2$  图谱关系抽取, 可自动确定待测试设备厂家, 实现对设备厂家快速分类并为后续子图匹配及信息匹配奠定基础。以上两项任务在实验中均未发生匹配错误的情况, 验证了其逻辑合理并处理高效。

基于图谱  $G_2$  的子图匹配方式与关系型数据库两种方式检索路径的时间复杂度如表 3 所示, 可以看出子图匹配的方式能够极大地减少时间支出, 且检索逻辑清晰明了, 相比较而言通过关系型数据库查找相对复杂一些, 对于配置人员来说难度极大。

表 3 不同数据库检索路径时间复杂度对比

Table 3 Comparison of time complexity of different database retrieval paths

数据库类型	时间复杂度
关系型	$O(n^3)$
图数据模型	$O(n \log(n))$

为了验证路径检索结果性能, 本文通过设置单次任务 1~4 条最相似路径分别在测试集进行路径匹配, 抓取路径下所有待匹配文本描述与端口地址配置数据, 所获得数据中含有映射正确的待匹配信息点数据则为正确, 计算 ACC 指标结果如表 4 所示。

表 4 不同数量相似路径下检索结果对比

Table 4 Comparison of retrieval results under similar paths with different quantities

匹配相似路径数量/条	1	2	3	4
ACC	0.954	0.986	0.989	0.990

从表 4 中可以看出, 基于路径检索获取待匹配信息完全能够满足任务需求, 只匹配最相似路径便可获得 95% 以上的准确率, 并且在匹配两条最相似路径时准确率又得到了 0.32% 的提升。通过观察结果发现, 设备版本变化带来的 IED 文件嵌套格式变化波动不大, 随着后续匹配路径的数量增加, 准确率上升趋势减缓, 这也说明了信息匹配路径具有极大相似性, 同时验证了本文所提路径检索方法有效。

#### 4.4.2 改进 BERT 模型性能评估

为了评估本文提出的基于字向量并融入关键词向量的改进 BERT 文本嵌入式表示模型, 选取 VSM、Word2Vec 以及考虑细粒度、多头注意力机制及关键词的 BERT 模型进行比较, 待匹配文本数据基于路径检索正确情况下所获得的数据, 结果对比如表 5 所示。

表 5 不同文本匹配模型结果对比

Table 5 Comparison of results of different text matching models

实验序号	模型	ACC	MRR
1	VSM	0.822	0.839
2	Word2Vec	0.903	0.921
3	BERT(基线模型)	0.936	0.942
4	BERT(多头注意力机制+词粒度)	0.946	0.955
5	BERT(多头注意力机制+字粒度)	0.957	0.967
6	BERT(多头注意力机制+字粒度+关键词)	<b>0.982</b>	<b>0.989</b>

从上述各方法匹配的结果来看, 采用融合多头注意力机制、字粒度、关键词的 BERT 模型取得了最高的准确率及平均倒数排名, 实验 1 采用传统向量空间模型 VSM, 准确率仅达到 82.2%, 而实验 2~6 准确率均达到了 90% 以上, 原因在于 VSM 对于数据依赖性较强, 过于机械化, 而实验 2~6 属于深度学习模型, 生成的表示向量具有一定的语义特征, 更能够反映真实文本含义。同时, 在相同的 BERT 模型下, 实验 5 相较于实验 4 在指标上分别获得了 1.1% 准确率提升和 1.2% 的平均倒数排名提升, 说明

针对二次设备测试配置短文本采用字粒度能够获得更好的词义表示向量。此外, 本文融合基于关键词的词粒度 BERT 模型语义表征向量在评估结果中取得了最优效果, 相对于未考虑关键词的实验 5 提升了 2.5% 的准确率以及 2.2% 的平均倒数排名, 两项指标均达到了 98% 以上, 甚至前 5 平均倒数排名达到了 98.9% 的极优性能, 说明考虑领域关键词相当于对深度学习 BERT 模型增加了一部分先验知识, 起到对描述文本向量表征的补充作用。

#### 4.4.3 自动配置方法评价

本文所提出的二次设备测试自动配置方法主要由两个关键环节组成。一是图谱匹配获取信息点路径抓取待匹配输出端口地址配置数据及接口文本描述数据; 二是在抓取的数据中基于改进 BERT 进行相似度计算得到匹配数据并校验。下面将各环节结合为整体进行评估, 考虑不同的最相似路径抓取数量、文本相似度匹配模型及是否加入字符串校验因素, 以最终匹配准确率及用时作为评价指标, 实验设置 10 轮, 每轮实验设置配置 100 个测试任务, 耗时为每轮任务平均耗时, 实验结果如表 6 所示。

表 6 不同文本匹配模型结果对比

Table 6 Comparison of results of different text matching models

实验序号	路径抓取数量	相似度计算模型	是否校验	ACC	耗时/s
1	1	BERT	否	0.887	23.1
2	2	BERT	否	0.923	40.6
3	3	BERT	否	0.924	61.9
4	4	BERT	否	0.924	82.4
5	1	改进 BERT	否	0.939	25.7
6	2	改进 BERT	否	0.966	42.2
7	3	改进 BERT	否	0.968	66.1
8	4	改进 BERT	否	0.967	84.1
9	1	BERT	是	0.902	29.2
10	2	BERT	是	0.932	45.0
11	3	BERT	是	0.934	67.9
12	4	BERT	是	0.934	87.5
13	1	改进 BERT	是	0.950	30.1
14	2	改进 BERT	是	0.982	46.9
15	3	改进 BERT	是	0.982	68.7
16	4	改进 BERT	是	0.983	90.2

观察表 6 可发现, 16 个实验设置互为对照, 从实验结果可以对路径抓取、模型改进、校验成效三个方面实验结果进行综合讨论。首先, 实验 1~4、实验 5~8、实验 9~12 以及实验 13~16 四组实验在模型及是否校验统一的情况下, 抓取路径数量与耗时

成正比增加, 不难理解, 抓取数量增加则意味着所要进行的相似度计算规模也相应扩大, 从而造成时间成本的增加。实验 4 比实验 1 耗时增加了近 4 倍, 虽然牺牲了时间但换取了准确率的提升, 当抓取数量从 1 升至 2 时, 四组对比实验准确率均上升了 3% 左右。在抓取数量从 2 升至 3 及从 3 升至 4 时, 准确率提升较小, 甚至出现了准确率下降的情况。依据配置错误实验结果发现, 抓取路径增多直接带来待匹配数据量成倍增加, 由此加大了匹配难度, 带来相似度计算性能下降, 又因为版本变化带来路径变化的情况较少发生, 出现三种或四种路径的情况较少, 所以出现了路径数量为 2 时准确率提升较大, 后续增加路径数量提升较少的情况。

实验 1~4 与实验 5~8 互为对照, 同样实验 9~12 与实验 13~16 互为对照, 这两组对照的变量为模型, 通过观察实验结果不难发现, 改进 BERT 模型比基础 BERT 模型在准确率上有较大提升, 两组 8 个实验对比准确率平均提升了 4.7%, 再一次说明了本文所提改进模型的有效性, 同时, 改进模型带来的时间成本增加在 2 s 左右, 在可接受范围内, 在实际应用时, 可根据任务环境选择模型。

实验 1~8 与实验 9~16 在是否进行字符串校验的变量因素上形成互为对照, 由于增加了编辑距离计算过程, 所以造成了一定的时间成本增加, 每轮平均耗时增加 4.9 s。但从准确率上来说, 增加字符串校验对于准确率的提升具有一定积极作用, 平均提升 1.3% 的准确率。

通过对上述对比试验讨论, 证明了路径检索方法、改进 BERT 模型及字符串校验在二次设备测试配置中具有显著作用。此外, 从整体上看, 实验 16 准确率达到 98.3%, 再次直观验证了本文所提二次设备自动配置方法切实有效。

## 5 结论

本文考虑到二次设备测试人工配置效率低且易出错的问题, 提出了一种基于知识图谱的二次设备自动配置方法, 结论如下:

1) 通过构建二次设备测试知识图谱, 将 IED 文件与测试任务、测试设备、配置信息点、信息路径建立知识关联, 在此基础上实现了自动抓取待匹配信息点的端口地址配置数据及文本描述, 从而为信息映射匹配奠定了基础。实验证明, 路径检索准确率至少可达到 95% 以上。

2) 以 BERT 模型为基础, 在字向量 BERT 模型

基础上融合二次测试专有词汇的词嵌入向量,有效增加了文本相似度识别过程中的先验知识,准确率平均提升了4.7%,最高可达到96.7%。增加了端口地址配置数据字符串校验环节以解决文本描述无法确定映射数据的情况,进一步将准确率提升约1.3%,有效地实现信息点与文本描述的映射匹配。

3) 通过设置多变量对照实验,配置准确率最高达到98.3%。并对路径抓取、模型改进、校验成效进行综合评估,实验发现,增加校验环节及采用改进BERT模型对于准确率提升均有较大作用,且耗时上不会有较大影响。增加最相似路径抓取数量对于提升待匹配数据中配置映射准确率有一定效果,但在整个配置环节中准确率并非随着路径数量增加而线性增加,因为路径数量增加带来的待匹配描述文本数量成倍增加,由此带来的耗时增加及可能的相似度计算准确率下降仍需纳入工作考量。

### 参考文献

- [1] HE Yuqing, CHEN Yuehui, YANG Zhiqiang, et al. A review on the influence of intelligent power consumption technologies on the utilization rate of distribution network equipment[J]. Protection and Control of Modern Power Systems, 2018, 3(2): 183-193.
- [2] 张巧霞, 王广民, 李江林, 等. 变电站远程运维平台设计与实现[J]. 电力系统保护与控制, 2019, 47(10): 164-172.  
ZHANG Qiaoxia, WANG Guangmin, LI Jianglin, et al. Design and implementation of substation remote operation and maintenance platform[J]. Power System Protection and Control, 2019, 47(10): 164-172.
- [3] 李宝潭, 赵丹, 李宝伟, 等. 智能变电站在线防误系统关键技术的研究[J]. 智慧电力, 2020, 48(6): 73-78, 111.  
LI Baotan, ZHAO Dan, LI Baowei, et al. Key technologies of online error prevention system in smart substation[J]. Smart Power, 2020, 48(6): 73-78, 111.
- [4] 钱平, 杨松伟, 张永, 等. 面向智能变电站信息流可靠性分析方法研究[J]. 电测与仪表, 2021, 58(10): 106-111.  
QIAN Ping, YANG Songwei, ZHANG Yong, et al. Research on reliability analysis method of information flow in intelligent substation[J]. Electrical Measurement & Instrumentation, 2021, 58(10): 106-111.
- [5] 范卫东, 董金星, 冯晓伟, 等. 智能变电站二次操作校核方法研究[J]. 电测与仪表, 2020, 57(13): 40-46, 54.  
FAN Weidong, DONG Jinxing, FENG Xiaowei, et al. Research on checking method of secondary operation in smart substation[J]. Electrical Measurement & Instrumentation, 2020, 57(13): 40-46, 54.
- [6] 吴在军, 胡敏强. 基于IEC 61850标准的变电站自动化系统研究[J]. 电网技术, 2003, 27(10): 61-65.  
WU Zaijun, HU Minqiang. Research on a substation automation system based on IEC61850[J]. Power System Technology, 2003, 27(10): 61-65.
- [7] 陈德辉, 须雷, 赵希才. 智能变电站自动化与保护核心 IEC 标准的进展与分析[J]. 电力系统自动化, 2019, 43(21): 229-239.  
CHEN Dehui, XU Lei, ZHAO Xicai. Progress and analysis of intelligent substation automation and protection core IEC standards[J]. Automation of Electric Power Systems, 2019, 43(21): 229-239.
- [8] 诸源胜. 智能变电站二次设备一体化测试系统的研制及应用[D]. 广州: 华南理工大学, 2020.  
CHU Yuansheng. Development and application of integrated test system for secondary equipment of the intelligent substation[D]. Guangzhou: South China University of Technology, 2020.
- [9] 陈旭, 张弛, 刘千宽, 等. 基于深度语义学习的智能录波器自配置方法[J]. 电力系统保护与控制, 2021, 49(2): 179-187.  
CHEN Xu, ZHANG Chi, LIU Qiankuan, et al. Automatic configuration method of intelligent recorder based on deep semantic learning[J]. Power System Protection and Control, 2021, 49(2): 179-187.
- [10] 张晓莉, 刘慧海, 李俊庆, 等. 智能变电站继电保护自动测试平台[J]. 电力系统自动化, 2015, 39(18): 91-96.  
ZHANG Xiaoli, LIU Huihai, LI Junqing, et al. Automatic test platform in smart substation for relay protection[J]. Automation of Electric Power Systems, 2015, 39(18): 91-96.
- [11] FAN Wen, LIAO Yuan. Wide area measurements based fault detection and location method for transmission lines[J]. Protection and Control of Modern Power Systems, 2019, 4(1): 53-64.
- [12] 范卫东, 冯晓伟, 董金星, 等. 基于历史数据语义相似度的智能变电站虚端子自动连接[J]. 电力系统保护与控制, 2020, 48(17): 179-186.  
FAN Weidong, FENG Xiaowei, DONG Jinxing, et al. Automatic matching method of a virtual terminal in intelligent substation based on semantic similarity of historical data[J]. Power System Protection and Control, 2020, 48(17): 179-186.
- [13] SINGHAL A. Introducing the knowledge graph: things,

- not strings[J]. Official Google Blog, 2012.
- [14] WANG Xiang, WANG Dingxian, XU Canran, et al. Explainable reasoning over knowledge graphs for recommendation[C] // Proceedings of the Thirty-Third AAAI Conference on Artificial Intelligence, AAAI, 2019, Honolulu, USA: 5329-5336.
- [15] JI Shaoxiong, PAN Shirui, CAMBRIA E, et al. A survey on knowledge graphs: representation, acquisition and applications[J]. arXiv preprint arXiv: 2002.00388, 2020.
- [16] 饶子昀, 张毅, 刘俊涛, 等. 应用知识图谱的推荐方法与系统[J]. 自动化学报, 2021, 9(3): 2061-2077.
- RAO Ziyun, ZHANG Yi, LIU Juntao, et al. Recommendation methods and systems using knowledge graph[J]. Acta Automatica Sinica, 2021, 9(3): 2061-2077.
- [17] DAI Renchang, LIU Guangyi, WANG Zhiwei, et al. A novel graph-based energy management system[J]. IEEE Transactions on Smart Grid, 2019, 11(3): 1845-1853.
- [18] 刘广一, 王继业, 李洋, 等. “电网一张图”时空信息管理系统[J]. 电力信息与通信技术, 2020, 18(1): 7-17.
- LIU Guangyi, WANG Jiye, LI Yang, et al. “One graph of power grid” spatio-temporal information management system[J]. Electric Power Information and Communication Technology, 2020, 18(1): 7-17.
- [19] 乔骥, 王新迎, 闵睿, 等. 面向电网调度故障处理的知识图谱框架与关键技术初探[J]. 中国电机工程学报, 2020, 40(18): 5837-5848.
- QIAO Ji, WANG Xinying, MIN Rui, et al. Framework and key technologies of knowledge-graph-based fault handling system in power grid[J]. Proceedings of the CSEE, 2020, 40(18): 5837-5848.
- [20] 余建明, 王小海, 张越, 等. 面向智能调控领域的知识图谱构建与应用[J]. 电力系统保护与控制, 2020, 48(3): 29-35.
- YU Jianming, WANG Xiaohai, ZHANG Yue, et al. Construction and application of knowledge graph for intelligent dispatching and control[J]. Power System Protection and Control, 2020, 48(3): 29-35.
- [21] APOSTOLOV A. Efficient maintenance testing in digital substations based on IEC 61850 edition 2[J]. Protection and Control of Modern Power Systems, 2017, 2(4): 407-420.
- [22] GUTIÉRREZ L, KEITH B. A systematic literature review on word embeddings[C] // International Conference on Software Process Improvement, Springer, Cham, 2018: 132-141.
- [23] DEVLIN J, CHANG M W, LEE K, et al. Bert: pre-training of deep bidirectional transformers for language understanding[J]. arXiv preprint arXiv: 1810.04805, 2018.
- [24] LI Yujian, LIU Bo. A normalized Levenshtein distance metric[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2007, 29(6): 1091-1095.
- [25] 陈桂芳, 董秀成, 郑永康, 等. 基于长短期记忆网络的继电保护测试故障诊断研究[J/OL]. 电力系统保护与控制: 1-9[2021-12-20]. <https://doi.org/10.19783/j.cnki.pspc.210624>.
- CHEN Guifang, DONG Xiucheng, ZHENG Yongkang, et al. Fault diagnosis of a relay protection test based on a long short-term memory network[J]. Power System Protection and Control: 1-9[2021-12-20]. <https://doi.org/10.19783/j.cnki.pspc.210624>.
- [26] YU Y, LI Y, SHEN J, et al. STEAM: Self-supervised taxonomy expansion with mini-paths[C] // Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, 2020: 1026-1035.

收稿日期: 2021-08-30; 修回日期: 2021-12-21

作者简介:

叶远波(1973—), 男, 硕士, 教授级高工, 研究方向为电力系统继电保护; E-mail: 13905608638@139.com

王骏东(1996—), 男, 通信作者, 硕士研究生, 研究方向为电力知识图谱。E-mail: 1825893405@qq.com

(编辑 魏小丽)