

DOI: 10.19783/j.cnki.pspc.190379

基于知识图谱的低压配电网拓扑结构辨识方法

高泽璞¹, 赵云², 余伊兰¹, 罗永建¹, 徐紫薇¹, 张莲梅¹

(1. 武汉大学电气与自动化学院, 湖北 武汉 430072; 2. 南方电网科学研究院, 广东 广州 510080)

摘要: 低压配电网中正确的拓扑关系是至关重要的。低压配电网实际拓扑结构因运维需要变动频繁且巨大, 配电网信息系统中的数据更新不及时、流通性低以及质量差等不能正确反映低压配电网的实际拓扑结构, 有必要进行拓扑辨识。知识图谱技术可以清晰地反映出数据间的现有关系, 推理挖掘出隐藏知识, 适用于低压配电网的拓扑辨识。因此应用知识图谱技术于拓扑辨识中, 首先分析知识图谱构建方法, 运用知识图谱技术, 对多个低压配电网信息系统中的数据进行整合, 推理出缺失数据, 挖掘出各数据之间的关系。然后构建出低压配电网拓扑结构的知识图谱。最后结合《低压配电网基建工程典型设计规范》以及语义分词技术, 对低压配电网信息系统中的户变关系进行辨识。算例实验效果良好, 表明所提出的辨识方法具有理论价值和实际应用价值。

关键词: 知识图谱; 语义分词技术; 知识推理; 户变关系; 拓扑辨识

Low-voltage distribution network topology identification method based on knowledge graph

GAO Zepu¹, ZHAO Yun², YU Yilan¹, LUO Yongjian¹, XU Ziwei¹, ZHANG Lianmei¹

(1. School of Electrical Engineering and Automation, Wuhan University, Wuhan 430072, China;

2. Electric Power Research Institute, CSG, Guangzhou 510080, China)

Abstract: The correct topological relationship in the low-voltage distribution network is crucial. The low-voltage distribution network structure changes frequently and tremendously because of the need of operation and maintenance. It is necessary to identify the topology of the incorrect updating of data, low circulation and poor quality, which can not correctly reflect the actual topological structure of low-voltage distribution network. Knowledge graph technology can clearly reflect the existing relationship between data, reason mining hidden knowledge, and apply to low-voltage distribution network topology identification. Therefore, this paper applies knowledge graph technology in topology identification. Firstly, it analyses the construction method of knowledge graph, uses knowledge graph technology to integrate data in multiple low-voltage distribution network information systems, deduces missing data, excavates the relationship between data, and then constructs the knowledge graph of low-voltage distribution network topology structure. Finally, combined with "Typical Design Specification of Low Voltage Distribution Network Infrastructure Project" and semantic word segmentation technology, the household transformer relationship in low voltage distribution network information system is verified and identified. The experimental results of the example are great, which show that the identification method proposed in this paper has theoretical value and practical application value.

This work is supported by Special Project of China Southern Power Grid (No. ZBKJXM20170078) and National Natural Science Foundation of China (No. 51277134) "Mechanism and Method Research for Mass Electricity Physical Quantity Light-weight Transmission in Smart Grid".

Key words: knowledge graph; semantic word segmentation technology; knowledge reasoning; household relationship; topology identification

0 引言

低压配电网的信息系统正确记录配电网拓扑结

构是电网精细化管理和电网安全运行的前提^[1]。在拓扑结构中, 最重要的要素是户变关系。目前, 信息系统中的低压配电网拓扑数据由人工录入, 信息的正确性得不到保证。而随着电网规模的不断扩大, 低压配电网的结构错综复杂, 营销、计量、GIS 等信息系统的数据量急剧增加, 各信息系统独立运行, 数据之间流通性差, 对拓扑数据辨识造成困难, 亟

基金项目: 南方电网公司专项课题资助(ZBKJXM20170078); 国家自然科学基金项目资助(51277134)“智能电网海量电物理量数据轻型传输机理与方法研究”

需进行拓扑辨识^[2]。

拓扑辨识是对信息系统中的低压配电网拓扑数据的正确性进行验证。当前拓扑辨识方法分为在线方法和离线方法。离线方法使用硬件设备进行现场检测辨识, 成本高且效率低; 在线方法包括户变工频过零相关性分析法、历史停电事件相关性判别法和电压皮尔逊相关系数法等^[3-9]。这些方法是对计量系统中的相关信息进行分析而达到拓扑辨识的目的, 其具有成本低、实时性高的特点, 逐渐成为当前拓扑辨识的热点方法; 但在线方法对数据量和数据质量的要求高, 依赖采集数据, 占用通信信道。每种拓扑辨识方法均有其一定的优点、缺点、应用场景和应用效果, 现在尚未有一种方法能够完全解决实际复杂运行场景中的低压拓扑结构辨识难题, 所以, 现实中必须综合采用多种辨识方法^[10]。

知识图谱技术可以清楚地分析出实体与实体之间的关系, 通过实体间关系的挖掘和推理, 可以对缺失信息进行增补、判断现有信息的正确性以及挖掘出系统中隐藏的内部联系^[11]。知识图谱的特点完全符合拓扑辨识的要求, 并且减小了拓扑辨识对数据的依赖性^[12]。因此本文首次提出一种基于知识图谱的拓扑辨识方法, 该方法可以克服在线辨识方法需要海量高质量运行数据和占用通信信道的缺点, 针对特定场景, 具有极好的辨识效果。

1 知识图谱

1.1 知识图谱的定义与现状

随着计算机科学与互联网技术的飞速发展和数据规模的爆炸式增长, 知识图谱(Knowledge Graph)应运而生^[13]。其是显示知识发展进程与结构关系的一系列各种不同的图形, 用可视化技术描述知识资源及其载体, 挖掘、分析、构建、绘制和显示知识及它们之间的相互联系^[14]。其可以动态、直观地展示知识内部结构以及知识之间的联系。

目前, 知识图谱多在互联网领域得到推广, 而在电力行业中知识图谱应用很少, 寥寥的应用实例仅限于与电网相关的边缘功能, 例如电网公司客户服务知识图谱等。对于电网中较为核心的营销、GIS、计量等功能, 还未有知识图谱的相关应用^[15]。

1.2 知识图谱的类别

根据应用领域的不同, 知识图谱可分为: 通用知识图谱和行业知识图谱^[16]。

通用知识图谱不限定特定领域, 知识覆盖面广, 知识之间的关系网复杂, 通常以自底向上的方式来构建。最具代表性的是谷歌知识图谱, 它包含 5 亿实体对象和 350 亿条实体间关系信息。国内的典型

代表有百度知心、搜狗知立方等。

行业知识图谱针对特定领域, 知识广度低, 深度深, 有固定数据来源, 具有针对性^[17]。行业知识图谱通常采用自顶向下和自底向上相结合的构建方式。行业知识图谱的例子有: 包含全球最全地理知识知识图谱 Geonames、全球最大的电影知识图谱 Linked Movie Database 等^[18]。

1.3 知识图谱的构建方法

知识图谱的构建方法可分为: 依靠人工构建的知识图谱、基于群体智能的知识图谱、基于互联网链接数据构建的知识图谱和基于机器学习和信息抽取构建的知识图谱^[19]。

知识图谱的构建过程主要包括:

(1) 知识抽取。从源数据中抽取实体、关系、属性等知识要素, 并转化为机器可理解的知识。

(2) 知识融合。从不同知识源获取的知识可能存在矛盾, 需进行消歧整合。

(3) 知识存储。融合后的知识之间产生联系, 以三元组的形式存储到图数据库中。

(4) 知识推理。在知识图谱基础上, 进行链接预测、知识问答等智能化应用, 是对现有知识图谱的完善与拓展。

2 低压配电网拓扑结构的知识图谱构建

2.1 低压配电网拓扑结构的知识图谱构建过程

低压配电网拓扑结构的知识图谱属于行业知识图谱, 其包含大量物理设备, 例如变压器、馈线等, 这些物理设备及其属性的数据来源为电网公司现有的配电网信息系统(营销、计量、GIS 系统等)数据库, 记录形式包括了自然语言、数字及地理坐标等^[20]。经广泛调研, 当前国内配电相关的信息系统所使用的数据库一般为结构化数据库(例如: Oracle 数据库)。结构化数据库存储数据的方式具有固定性、严密性。因此, 针对构建低压配电网拓扑结构的知识图谱, 在通常构建方法的基础上做出如下改进:

1) 在低压配电网的拓扑结构中, 参考通用信息模型(Common Information Model, CIM), 从物理设备出发, 首先建立物理设备之间的关系, 再考虑物理设备自身的属性, 故采用“自顶向下”的方式构建知识图谱。

2) 配电网信息系统采用结构化数据库, 结构化数据知识抽取较为简单, 抽取数据库数据知识的同时, 还需要抽取数据库中的字段。

3) 针对配电网各信息系统之间存在系统壁垒, 从多个信息系统数据库中抽取的知识需要进行知识融合, 使不同信息系统中的同一实体属性得到完善,

且区分出同名异体或异名同一实体。

4) 对低压配电网拓扑结构的知识图谱所涉及到的物理设备进行本体建模, 本体为该类设备的抽象概念, 并包含其所有属性。本体集合尽量小, 本体与本体之间的关系可人为定义。

5) 基于构建好的本体, 将从数据库中抽取并融合后的知识与本体对应, 形成实体。由于结构化数据的固定性, 实体与属性之间的关系可通过样本数据训练进行构建。

6) 将构建的实体、属性及关系组成(实体, 关系, 实体)或(实体, 关系, 属性)形式的三元组, 存入图数据库中, 构建完成低压配电网拓扑结构的知识图谱。

7) 采用可视化工具展示低压配电网拓扑结构的知识图谱, 分析知识间的关联关系, 推理出低压配电网拓扑结构中的错误。

图1为低压配电网拓扑结构的知识图谱的构建改进方法。根据数据源特点, 分析低压配电网拓扑结构的特点, 改进了常规知识图谱的构建方法, 此种构建方法具有针对性。构建后的低压配电网拓扑结构的知识图谱更具专业性与实用性, 各步骤的具体实施方法将在下文阐述。

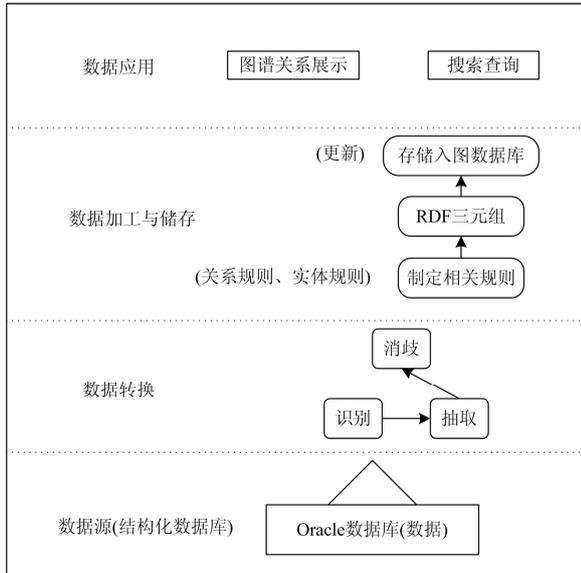


图1 低压配电网拓扑结构的知识图谱的构建改进方法
Fig. 1 Improved method for constructing knowledge graph of low-voltage distribution network topology

2.2 本体构建 O_i

本体是概念、概念的结构以及概念之间的关系的正式化表达, 作为构建实体时的通用模板^[21]。知识图谱中使用本体库来存储本体。本体构建结构如图2所示。

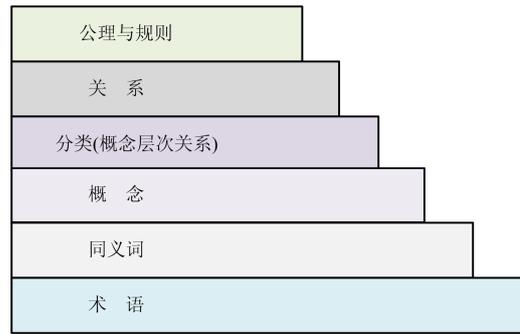


图2 本体构建结构
Fig. 2 Ontology building structure

通过分析低压配电网拓扑结构中所涉及到的物理设备及其信息, 本文所构建的低压配电网拓扑结构的知识图谱的本体为: 变压器、台区(台区总表)、馈线、低压用户表4大类。所需构建的本体从低压配电网信息系统的数据库中抽取。抽取对象为各数据类型字段, 将抽取出的数据类型与4大类本体之间相对应, 对于数据库中没有的本体属性, 可人为进行增补。将属于各本体的数据类型与本体连接组合, 即构成了该类本体。

本体构建表达式如式(1)。

$$O_i = (P_i, R_i) \quad (1)$$

式中: i 为针对户变关系辨识问题所建立的本体序号, 本文中 $i = \{1, 2, 3, 4\}$, O_1 表示变压器本体, O_2 表示台区本体, O_3 表示馈线本体, O_4 表示低压用户表本体; P_i 表示本体 O_i 的属性集合; R_i 表示本体 O_i 的关系属性集合。

每个本体 O_i 存在属性 $P_{ix} = ID$, 用于唯一标识实体, 以及属性 $P_{iy} = RID$, 其为关系标识。

本文利用 Protégé 软件辅助构建本体。构建包括: (1) 变压器、台区、馈线、低压用户表共4类物理设备及其通用属性与私有属性; (2) 各设备之间、设备属性之间以及属性与属性之间的通用关系。所构建的本体库具有通用性, 可节省存储空间。低压配电网拓扑结构中本体之间的关系如表1所示。本体构建结果如图3所示。从图3中可看到所构建的4大类本体及其相互间的关系, 各本体的属性类型隐藏在本体中, 可查看某类本体的属性。

2.3 实体构建 E_{ij}

从营销、计量、GIS 等信息系统的结构化数据库中抽取知识。结构化数据库中的字段表示数据库表每一列的数据类型, 数据库表的每一条记录(行)是对某一实体的描述, 每一条记录被每一列划分成多个实体属性(值)。对抽取出的知识进行分类, 根

表 1 低压配电网拓扑结构的本体间关系
Table 1 Ontological relations of low voltage distribution network topology

| 本体 | 关系 | 本体 |
|-------|------|-------|
| 变压器 | 分出线 | 馈线 |
| 变压器 | 供电范围 | 台区 |
| 馈线 | 连接 | 变压器 |
| 馈线 | 输送电能 | 台区 |
| 台区 | 所属 | 变压器 |
| 台区 | 包含 | 低压用户表 |
| 低压用户表 | 属于 | 台区 |

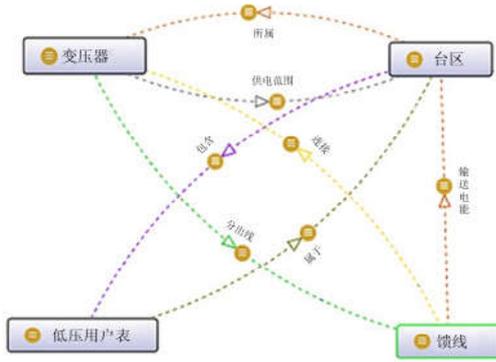


图 3 本体构建实例

Fig. 3 An example of ontology construction

据已构建好的本体, 将知识与本体进行关联, 形成低压配电网拓扑结构中的实体。例如: 1#馈线(长度、起点、终点、所属台区等)、1#台区(用户数量、名称、地理位置等)、1#台区(线损值、负荷值)。

实体构建表达式如式(2)所示。

$$E_{ij} = (P_{ij}, PV_{ij}) \quad (2)$$

式中: i 表示本体含义同式(1); E_{ij} 表示本体 O_i 的第 j 个实体; P_{ij} 表示本体 O_i 的第 j 个实体的属性集合; PV_{ij} 表示本体 O_i 的第 j 个实体的属性值集合。 P_{ij} 和 PV_{ij} 有一一对应关系。

每个实体 E_{ij} 都存在一个自身标志属性 $P_{ijx} = ID$ 和一个连接标志属性 $P_{ijx} = RID$, PV_{ijx} 在该类实体 E_{ij} 是唯一的。例如: 低压用户表本体 O_4 下的所有低压用户表实体都包含一个变压器连接标志属性 $P_{ijx} = TransformerID$, 表明其所属的 O_1 实体; 低压用户表本体 O_4 包含属性 $P_{ix} = Address$, 表示该低压用户表的用电地址。

上文提到各信息系统数据库之间是相互独立的, 对于同一实体的不同属性可能来自多个独立的

数据库, 因此对各配电网信息系统分别单独进行实体构建, 之后对比各实体间的自身标志属性, 把自身标志属性 $P_{ijx} = ID$ 相同的实体融合为一。此步的优点在于: 减少工作量, 同时完成了实体消歧和数据融合。

2.4 关系构建 $R_{ij \times mn}$

关系是知识图谱中的核心, 也是以知识图谱为工具处理问题的关键。关系构建包括: 实体与实体之间的关系; 实体与属性之间的关系。

实体与实体之间的关系可直接继承所对应本体之间的关系, 例如: “变压器-供电范围-台区”中的“供电范围”为“变压器”本体和“台区”本体之间的关系, 继承后实体之间关系为“变压器 A-供电范围-台区 1”。

实体与属性之间的关系可以根据知识抽取的字段加上相关修饰语来构成, 例如: “低压用户表 a-电压值为-220 V”, 其中“电压值为”即是实体“低压用户表 a”与其属性值“220 V”之间的关系, 此关系可通过知识抽取数据库字段后附加相应修饰语获得。

关系构建的表达式如式(3)。

$$R_{ij \times mn} = (E_{ij}, E_{mn}) \quad (3)$$

式中: i 表示含义同式(1); m 表示含义同 i ; n 和 j 表示数据在所属类中的位置; E_{ij} 、 E_{mn} 表示实体或属性; $R_{ij \times mn}$ 表示其之间的关系。2.2 节中人为构建的本体与本体之间的关系属性集合 R_i 也包含在 $R_{ij \times mn}$ 中。

2.5 完成知识图谱构建

本体、实体以及关系构建完成后, 实体与实体、实体与属性通过关系连接起来, 组合成三元组并存入图数据库中, 形成低压配电网拓扑结构的知识库。三元组形式如式(4)所示。

$$G = (E, R, S) \quad (4)$$

式中: G 表示三元组集合; E 表示实体的集合, 共包含 $|E|$ 种不同的实体, 每个实体有一个唯一确定的 ID; R 表示关系集合, 共包含 $|R|$ 种关系; S 表示属性集合。低压用户表和变压器之间的户变关系可由变压器和台区的关系与台区和低压用户表的关系推理出来, 表述为: 低压用户表实体 E_{4m} “属于”变压器实体 E_{1n} 的供电区域。

知识库中的三元组包含(实体 1, 关系, 实体 2)和(实体, 关系, 属性)两种形式。通过可视化的方式呈现出实体间错综复杂的关系图, 构成了低压配

电网拓扑结构的知识图谱。

3 低压配电网拓扑辨识方法

3.1 拓扑辨识方法流程

在低压配电网拓扑辨识中，最主要的目标是验证低压用户表与变压器之间的户变关系，将户变关系错误的低压用户表挑选出来，再由人工现场检验确认。目前在低压配电网的户变关系中，主要存在两类错误：

(1) 低压配电网信息系统中所记录的户变关系与实际户变关系不符；

(2) 低压配电网信息系统中遗漏部分低压用户表的户变关系信息。

知识图谱通过分析低压配电网拓扑结构中各实体之间的关系，可以辨识出配电网信息系统中的户变关系错误；通过知识推理与挖掘，可以增补完善配电网信息系统中所遗漏的低压用户表的户变关系信息。

上文中已构建完成低压配电网拓扑结构的知识图谱，各台区、变压器、馈线以及低压用户表都通过各自的关系联系在一起，属于同一台区的低压用户表以及本台区的变压器在图数据库中的存储距离较近，便于对数据的调取、查看与分析。基于已建立的知识图谱，根据现有数据，将同一台区下的低压用户表的用电地址按照地址标准规范化，然后进行语义分词处理。对照分析整个台区下所有低压用户表的用电地址的语义分词结果，把不符合《低压配电网建设工程典型设计规范》的低压用户表挑选出来，达到拓扑辨识的目的。其方法流程图如图 4 所示。

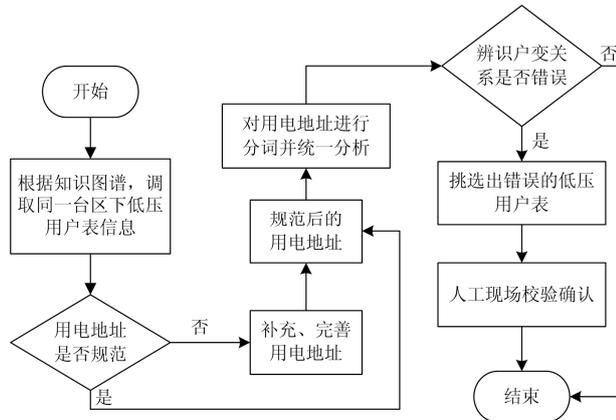


图 4 拓扑辨识方法流程图

Fig. 4 Flow chart of topology identification method

3.2 语义分词并增加属性

低压用户表的用电地址是由电网人员人工录入

的，严密性差、不规范。结合《国家地理信息标准地址描述规则》^[22]，本文将用电地址分为 10 层 (Level1—Level10)，从 Level1 到 Level10 各层地址辖区逐渐缩小。各层所代表含义见表 2。将所有低压用户表的用电地址按照层级规范完善，当某一层级地址不存在时，可以为空。

表 2 用电地址的各层代表含义

| 层级 | 代表含义 |
|---------|-----------|
| Level1 | 省、自治区等 |
| Level2 | 市等 |
| Level3 | 县、区等 |
| Level4 | 镇、乡等 |
| Level5 | 街道、路、大道等 |
| Level6 | 小区、花园、花苑等 |
| Level7 | 楼、栋等 |
| Level8 | 单元号等 |
| Level9 | 层号等 |
| Level10 | 号、室、房间等 |

例如：原用电地址为“东湖南路茶港小区 116 栋 2 单元 501 号”，规范后为“xx 省 xx 市 xx 区东湖南路茶港小区 116 栋 2 单元 501 号”。

在知识图谱中同一台区或小区的低压用户表之间具有关联关系，对于不规范的用电地址，可以与同小区或同台区中规范的用电地址对比，进行规范化处理。以图 5 为例说明。

图 5 中，低压用户表 A、B 属于同一小区或台区，低压用户表 B 用电地址不规范，只有 Level6—Level10 的地址层，通过同用户表 A 的用电地址做对比分析，将用户表 B 的用电地址补充完整。

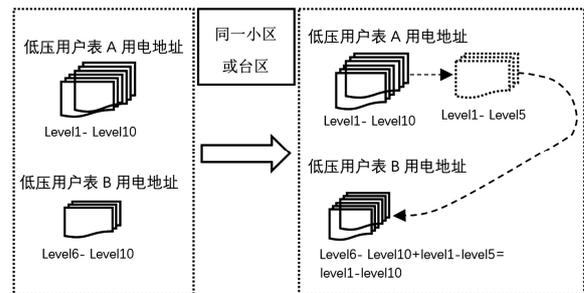


图 5 用电地址规范方法示例

Fig. 5 An example of the power address specification method

针对缺失用电地址的低压用户表，可调取出其所属变压器下的所有低压用户表的用电地址，推理出楼层数和户数，每层每户与低压用户表一一对应后，余留下的层户信息即为缺失的低压用户表的用电地址。

根据《地名地址数据规范》，对规范化后的用电

地址进行语法分词。与规范化地址时的层级相同, 地址分词也为 10 级。分词后的 Level1-Level10 地址级作为低压用户表的新增属性加入到低压配电网拓扑结构的知识图谱中。例如: {User1, Level1, Value1_1} 表示低压用户表 User1 的属性 Level1 的值为 Value1_1, 其他地址层级属性及其值依次类推。

3.3 拓扑辨识过程

基于知识图谱, 抽取某台区变压器 T_m 下的所有用户所涉及的楼栋集合 B_m , $B_m = \{B_{m1}, \dots, B_{mm}\}$, 查找 B_m 所包含的所有用户集合 U_m , $U_m = \{U_{m1}, \dots, U_{mk}\}$ 。Level1—Level6 属性值相同时, 即为同一个小区的低压用户表; Level1—Level7 属性值相同时, 即为同一栋楼的低压用户表; Level1—Level8 属性值相同时, 即为同一栋楼的同一单元的低压用户表; Level1—Level9 属性值相同时, 即为同一栋楼的同一单元的同一层的低压用户表。

当为同一栋楼的低压用户表时, 根据知识图谱中的关系, 可以统计供此栋楼的变压器数量 N_B , 如式(5)。

$$N_B = \begin{cases} 1 \\ 2 \\ \text{others} \end{cases} \quad (5)$$

根据变压器容量 S_{bN} 、变压器设计功率因数 $\cos\varphi$ 、用户表数 N_{user} 、《10 kV 及以下配网基建工程典型设计》给出的低压用户的典型负荷 P_{bN} , 可以初步判断符合设计规范的变压器台数 N_{BD} , 如式(6)所示。

$$N_{BD} = \left\lceil \frac{S_{bN} \cdot \cos\varphi}{N_{user} \cdot P_{bN}} \right\rceil \quad (6)$$

当 $N_B - N_{BD} \geq 1$ 时, 表示该栋楼户变关系存在错误, 比较同一栋楼的低压用户表的 Level8—Level10 的属性值, 根据《10 kV 及以下配网基建工程典型设计》^[23]: 中低层楼房用户数较少, 一台变压器即可满足用电需求, 则同一小区的一栋楼的低压用户表属于同一变压器供电区域; 高层楼房用户数较多, 一般 2~4 台变压器才能满足用电需求, 则同一小区同一栋楼同一层的低压用户表属于同一变压器。把不符合设计规则的低压用户表挑选出来, 进行修正: 将错误低压用户表所属变压器修改为同栋楼或同层户变关系正确数较多的 TransformerID。

同一变压器下的低压用户表地址为同一栋楼或同一层, 同一栋楼各层的用户表数量一般应相同, 各层用户表数相对比, 用户表数少的楼层即为配电网信息系统中户变关系信息缺失的低压电表。将

此低压用户表的户变关系信息修改为与同层其他户变关系正确的低压用户表相同的 TransformerID。

3.4 知识图谱拓扑辨识方法应用场景

本文所述的基于知识图谱的拓扑辨识方法适用于以下应用场景:

1) 城市居民低压用户及其台区。城市居民所组成的台区通常用电地址之间的关联关系较为密切, 有规律可循, 运用知识图谱推理后的结果准确性高。

2) 楼层较高的楼栋。随着社会的发展, 楼房的高度日益增加, 三十、四十层以上的楼房通常由 2~4 台变压器供电, 在供电变压器衔接的楼层十分容易造成户变关系错误。知识图谱拓扑辨识方法可以很好地找出错误信息并修改。

3) 城市小区。小区中各低压用户表的用电地址规范性高, 对不规范地址规范化较为容易。通过知识图谱拓扑辨识方法可以对整个小区每栋楼每户的户变关系进行辨识, 辨识效果良好。

3.5 本算法的优点

本文所述的知识图谱拓扑辨识方法辨识准确率高, 成本低, 所用数据量小, 不占用通信信道, 实用性强, 对数据量和数据质量的依赖性低, 可以对低压配电网信息系统中的数据进行增补。可以在城市范围内广泛推广, 并且可以基于已建立的知识图谱做其他方面的应用研究。

4 算例分析

4.1 算例基本情况

为检验基于知识图谱的拓扑辨识方法的效果, 本文分别以模拟数据与某供电公司的实际数据进行实验。

算例基本情况如下:

1) 模拟数据: 共模拟 3 个台区, 6 台变压器, 7 条馈线以及若干低压用户表, 实体数量为 400 左右。人为将其中部分低压用户表信息设置为错误信息, 用于检验知识图谱拓扑辨识方法的效果。

2) 电网公司数据: 使用某供电公司 GIS、营销以及计量等信息系统中的真实数据, 选取 3 个典型台区, 其中包含 432 个低压用户表。因为无法对现有电网数据的正确性做实地考察, 因此认为现有信息数据为正确数据。将现有数据中某台区下的用户人为调整到其他台区, 作为基于知识图谱的拓扑辨识方法的目标。

本文使用用电地址规范率以及拓扑辨识成功率两个指标来评价基于知识图谱的拓扑辨识方法的效果。两个指标的计算方法如式(7)、式(8)所示。

$$\text{用电地址规范率} = \frac{\text{规范化用电地址数}}{\text{用电地址总数}} \times 100\% \quad (7)$$

$$\text{拓扑辨识成功率} = \frac{\text{辨识出的错误信息数}}{\text{错误信息总数}} \times 100\% \quad (8)$$

4.2 模拟数据算例分析

模拟算例中, 从各本体中挑选出具有代表性的属性作为模拟模板, 具体属性见表 3。为验证本文所用方法的适用范围, 设置不同的楼层高度, 低层(6层)、中层(12层)和高层(24层)。

表 3 各本体属性代表

Table 3 Representation of ontology attributes

| 本体 | 变压器 | 台区 | 馈线 | 低压用户表 |
|----|--------------|---------|-------|---------|
| 属性 | 变压器名称 | 台区名称 | 线路名称 | 用户名称 |
| | 变压器类型(公变/专变) | 台区 ID | 线路起始端 | 户号 |
| | 一次侧额定电压/V | 所连接馈线名称 | 线路终点端 | 相别 |
| | 二次侧额定电压/V | 台区类型 | 所连变压器 | 用电地址 |
| | 额定容量/kVA | 台区地址 | 所连台区 | 所属台区 ID |
| | 设备型号 | 运行状态 | | 所属变压器 |
| | 配电变压器所属台区 | 所含变压器 | | 电量 |
| | | | | 电压 |

将模拟数据导入数据库中作为原始数据。按照上文所述的知识图谱构建方法, 构建出包含 387 实体节点、1 935 属性节点以及 2 814 条关系的低压配电网拓扑结构的知识图谱。然后对知识图谱中的低压用户表用电地址进行规范化, 规范化结果见表 4。将规范后的地址分级存入知识图谱中, 作为低压用户表的新增属性。然后对目标变压器或台区进行拓扑辨识分析, 辨识结果见图 6、表 5。

从图 6、表 5 中可以看出: 用电地址规范率和拓扑辨识成功率均为 100%。在规范用电地址只占

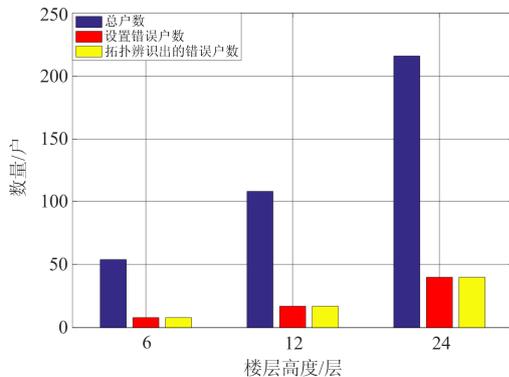


图 6 拓扑辨识结果图

Fig. 6 Topology identification result diagram

表 4 用电地址规范化结果(模拟)

Table 4 Results of standardization of electric address (simulation)

| 情况 | 规范地址数 | 不规范地址数 | 用电地址规范率 |
|--------|-------|--------|---------|
| 地址规范化前 | 36 | 342 | 9.52% |
| 地址规范化后 | 378 | 0 | 100% |

表 5 拓扑辨识结果(模拟)

Table 5 Topology identification results (simulation)

| 楼房高度 | 总户数 | 设置错误信息数 | 知识图谱拓扑辨识数 | 拓扑辨识成功率 |
|------|-----|---------|-----------|---------|
| 6层 | 54 | 8 | 8 | 100% |
| 12层 | 108 | 17 | 17 | 100% |
| 24层 | 216 | 40 | 40 | 100% |

比 9.5%的情况下, 通过对整个台区下低压用户的用电地址统一语义分析后, 用电地址的规范率可以提升到 100%。分别对低、中、高不同类型的楼层高度的台区进行拓扑辨识, 通过知识图谱的方法, 辨识出了不同情况下的全部错误低压用户表, 准确率高, 辨识效果好。

4.3 电网公司实际数据算例分析

4.2 节模拟数据的实验结果良好, 模拟数据的配电网特性突出、完整, 没有干扰因素的存在。而配电网中的实际情况更加复杂、多变, 因此使用电网实际数据进一步检验知识图谱拓扑辨识方法的实用性和适应性, 并与电压皮尔逊相关系数法进行对比分析。

根据实际配电网信息系统数据, 构建出低压配电网拓扑结构的知识图谱, 部分知识图谱展示如图 7 所示。

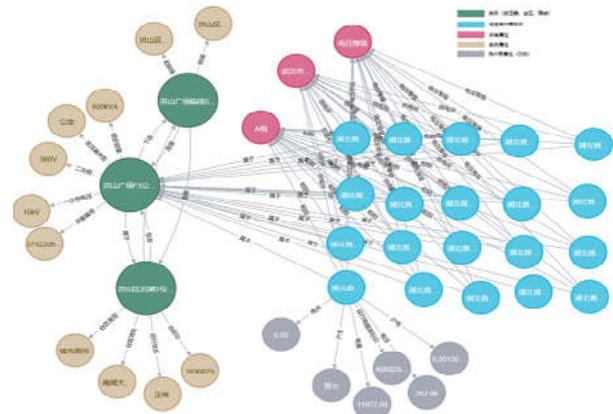


图 7 低压配电网拓扑结构的知识图谱示例

Fig. 7 An example of knowledge map of the topology of a low-voltage distribution networks

实际情况中, 用电地址输入随意、规范性更差, 用电地址规范化结果如表 6 所示。无效地址指低压

用户表不存在用电地址属性, 只能根据整栋楼用电地址之间的语义关系进行补充。从表 6 中可以看出, 实际中不规范地址可以使用知识图谱配合语义分析转化为规范地址, 而对于无效地址, 只能通过整栋楼、单元或小区各用户地址中户号所存在的缺失进行填补, 因此地址规范化无法达到 100%。

表 6 用电地址规范化结果(实际数据)

Table 6 Results of standardization of electric address (actual data)

| 情况 | 规范地址数 | 不规范地址数 | 无效地址数 | 用电地址规范率 |
|--------|-------|--------|-------|---------|
| 地址规范化前 | 21 | 404 | 7 | 4.86% |
| 地址规范化后 | 429 | 0 | 3 | 99.31% |

拓扑辨识结果如图 8、表 7 所示。共有 4 个实际台区。当台区下的低压用户表数量较少时, 辨识成功率较低, 其原因是由于数据量少, 关系之间的关联程度相对较小, 因而辨识成功率相对较低; 另外, 由于用电地址无法 100%规范化, 也是造成台区 A 和台区 C 拓扑辨识成功率没有达到 100%的原因。

表 7 拓扑辨识结果(实际数据)

Table 7 Topology identification results (actual data)

| 台区名称 | 总户数 | 设置错误信息数 | 知识图谱拓扑辨识数 | 拓扑辨识成功率 |
|------|-----|---------|-----------|---------|
| 台区 A | 27 | 4 | 2 | 50% |
| 台区 B | 101 | 14 | 14 | 100% |
| 台区 C | 128 | 17 | 16 | 94.12% |
| 台区 D | 176 | 21 | 21 | 100% |

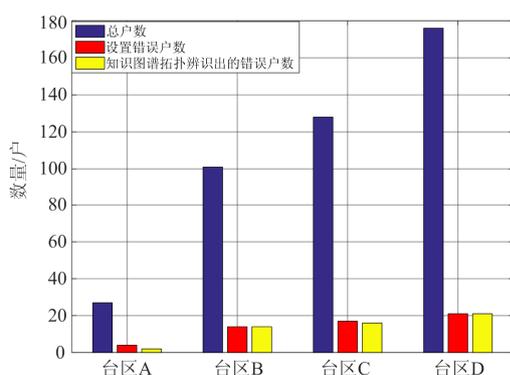


图 8 拓扑辨识结果图

Fig. 8 Topology identification result diagram

举例说明本文方法的拓扑辨识过程如图 9。台区 C 中 301 号变压器下的低压用户表的用电地址全部位于福宁园 14 栋, 14 栋中的 5 层 503 号和 9 层 902 号在配电信息系统中记录的所属变压器为 602 号变压器, 通过调取 602 号变压器下所有低压用户表的用电地址, 全为福宁园 3 栋用户, 由此可判断

出福宁园 14 栋 5 层 503 号和 9 层 902 号低压用户表的户变关系错误, 应修改为与福宁园 14 栋大部分低压用户表相同的 TransformerID=变压器 301。

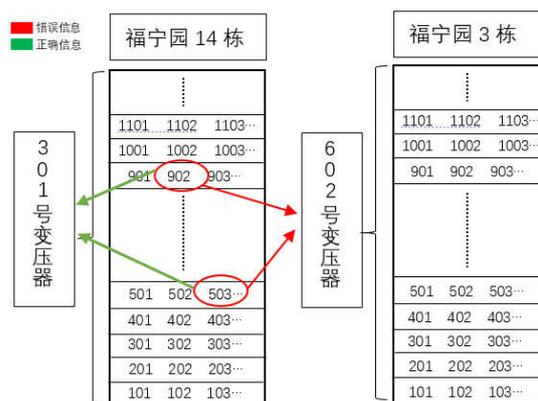


图 9 拓扑辨识前后结果示例

Fig. 9 An example of results before and after topology identification

使用文献[24]中的电压皮尔逊相关系数法, 采用实测电压数据(不停电, 每 15 min 采样一次, 数据时段 14 天), 针对特定场景和本文所适用场景进行试验, 辨识结果如表 8 所示。

表 8 电压皮尔逊相关系数法辨识结果

Table 8 Identification results of Pearson correlation coefficient method for voltage

| | 设置错误信息数 | 电压皮尔逊相关系数法辨识数 | 拓扑辨识成功率 |
|---------|---------|---------------|---------|
| 特定场景 | 12 | 11 | 91.67% |
| 本文所适用场景 | 12 | 1 | 8.33% |

电压皮尔逊相关系数法的拓扑辨识成功率极大依赖于特定的拓扑错误场景, 在小区、楼房等电气距离较近的场景中, 其拓扑辨识成功率大大降低。电压皮尔逊相关系数法所用的电压数据为所涉及到的每个用户 14 天每天 96 个计量的数据量, 所有数据需要通过通信信道传输至辨识中心, 而本文方法所使用的低压用户表的用电地址数据, 可由辨识中心直接从数据库中读取, 不需要占用通信信道。

现实中必须综合采用多种辨识方法, 本文所提出的辨识方法的辨识结果可以和电压皮尔逊相关系数法的辨识结果相互印证, 两种方法找出同一错误用户的可靠性优于单一方法找出的错误用户。

5 结论

本文将知识图谱技术引入到低压配电网中, 并首次将其应用在低压配电网拓扑辨识中。提出了基于知识图谱的拓扑辨识方法, 详细阐述了低压配电网

网拓扑结构的知识图谱的构建过程、用电地址规范化语义分词方法以及知识图谱拓扑辨识方法。通过模拟数据和实际数据的算例证明了此方法在用电地址规范化中规范率高,在拓扑辨识中成本低,所用数据量小,不占用通信信道,针对典型高楼场景时准确率高,可以改善低压配电网信息系统中的数据质量,可广泛应用在高层、小区等城市居民范围内。针对典型高楼场景时,与现有拓扑辨识方法相比,具有显著的优越性和实用性。

如果制定更多有关拓扑辨识的规则,研究除用电地址外的其他属性关系,能进一步提高知识图谱的完整性和拓扑辨识的准确率,并且在低压配电网的其他方面也可以运用知识图谱的方法解决问题,这也是后续研究的重点。

参考文献

- [1] 刘羽霄, 张宁, 康重庆. 数据驱动的电力网络分析与优化研究综述[J]. 电力系统自动化, 2018, 42(6): 157-167.
LIU Yuxiao, ZHANG Ning, KANG Chongqing. A review on data-driven analysis and optimization of power grid[J]. Automation of Electric Power Systems, 2018, 42(6): 157-167.
- [2] 郎燕生, 李静, 罗雅迪, 等. 基于图划分的大电网拓扑分析[J]. 电力系统保护与控制, 2017, 45(23): 108-115.
LANG Yansheng, LI Jing, LUO Yadi, et al. Large power grid topology analysis based on graph partitioning[J]. Power System Protection and Control, 2017, 45(23): 108-115.
- [3] 李遥, 许烈, 李永东. 模块化多电平矩阵变换器开关次数最优的拓扑连接策略[J]. 电工技术学报, 2018, 33(2): 342-352.
LI Yao, XU Lie, LI Yongdong. Topology connecting strategy of the modular multilevel matrix converter for optimal switching times[J]. Transactions of China Electrotechnical Society, 2018, 33(2): 342-352.
- [4] 曾磊磊, 曾鑫, 吕嘉. 考虑网络结构和运行状态的电网脆弱线路辨识[J]. 智慧电力, 2018, 46(8): 8-12, 51.
ZENG Leilei, ZENG Xin, LÜ Jia. Identification of vulnerable line of power grid in view of grid structure and running state[J]. Smart power, 2018, 46(8): 8-12, 51.
- [5] WANG C, ZHAO Y, ZHAO S, et al. Topology identification of general uncertain complex dynamic networks with time delay and noise perturbation based on generalized lag synchronization[C] // International Conference on Information Science & Control Engineering, IEEE Computer Society, 2017.
- [6] LUAN W, PENG J, MARAS M, et al. Distribution network topology error correction using smart meter data analytics[C] // 2013 IEEE Power and Energy Society General Meeting (PES), July 21-25, 2013, Vancouver, BC, Canada.
- [7] 应俊, 梅军, 王勇, 等. 配电主站馈线拓扑模型校核方法研究[J]. 电力系统保护与控制, 2018, 46(7): 83-89.
YING Jun, MEI Jun, WANG Yong, et al. Research on verification method of feeder topology model for distribution main station[J]. Power System Protection and Control, 2018, 46(7): 83-89.
- [8] LOURENCO E M, COELHO E P R, PAL B C. Topology error and bad data processing in generalized state estimation[J]. IEEE Transactions on Power Systems, 2015, 30(6): 3190-3200.
- [9] SAMUDRALA A N, AMINI M H, KAR S, et al. Optimal sensor placement for topology identification in smart power grids[C] // 2019 53rd Annual Conference on Information Sciences and Systems (CISS), 2019.
- [10] 严鋈, 王金全, 黄克峰, 等. 脉冲功率负载等效拓扑及潮流计算模型分析[J]. 电工技术学报, 2018, 33(23): 5523-5531.
YAN Qi, WANG Jinquan, HUANG Kefeng, et al. Analysis of equivalent topology and power-flow calculation model for pulsed power load[J]. Transactions of China Electrotechnical Society, 2018, 33(23): 5523-5531.
- [11] 谢晓帆, 刘秋林, 李斌, 等. 基于主成分分析法与对应分析法的县域配电网状况评估[J]. 智慧电力, 2018, 46(6): 68-73.
XIE Xiaofan, LIU Qiulin, LI Bin, et al. Evaluation of county distribution network status based on principal component analysis and correspondence analysis[J]. Smart Power, 2018, 46(6): 68-73.
- [12] 周冬旭, 张明, 朱红, 等. 新电改形势下智能配电网调度互动研究应用[J]. 电力工程技术, 2018, 37(2): 89-94.
ZHOU Dongxu, ZHANG Ming, ZHU Hong, et al. Research and application of intelligent distribution network dispatching interactive under the background of electric power system reformation[J]. Electric Power Engineering Technology, 2018, 37(2): 89-94.
- [13] 刘峤, 李杨, 段宏, 等. 知识图谱构建技术综述[J]. 计算机研究与发展, 2016, 53(3): 582-600.
LIU Qiao, LI Yang, DUAN Hong, et al. Knowledge graph construction techniques[J]. Journal of Computer Research and Development, 2016, 53(3): 582-600.
- [14] 刘梓权, 王慧芳. 基于知识图谱技术的电力设备缺陷记录检索方法[J]. 电力系统自动化, 2018, 42(14): 158-164.

- LIU Ziquan, WANG Huifang. Retrieval method for defect records of power equipment based on knowledge graph technology[J]. Automation of Electric Power Systems, 2018, 42(14): 158-164.
- [15] ZHU G, IGLESIAS C A. Sematch: semantic similarity framework for knowledge graphs[J]. Knowledge-Based Systems, 2017: S0950705117302447.
- [16] SINGH M. Protection coordination in distribution systems with and without distributed energy resources-a review[J]. Protection and Control of Modern Power Systems, 2017, 2(2): 294-310. DOI: 10.1186/s41601-017-0061-1.
- [17] UYAR A, ALIYU F M. Evaluating search features of Google knowledge graph and Bing Satori[J]. Online Information Review, 2015, 39(2): 197-213.
- [18] SHI B, WENINGER T. Discriminative predicate path mining for fact checking in knowledge graphs[J]. Knowledge-Based Systems, 2016, 104: 123-133.
- [19] 李新鹏, 徐建航, 郭子明, 等. 调度自动化系统知识图谱的构建与应用[J]. 中国电力, 2019, 52(2): 70-77, 157. LI Xinpeng, XU Jianhang, GUO Ziming, et al. Construction and application of knowledge graph of power dispatching automation system[J]. Electric Power, 2019, 52(2): 70-77, 157.
- [20] WANG C, MA X, CHEN J, et al. Information extraction and knowledge graph construction from geoscience literature[J]. Computers & Geosciences, 2018, 112: 112-120.
- [21] 石访, 张林林, 胡熊伟, 等. 基于多属性决策树的电网暂态稳定规则提取方法[J]. 电工技术学报, 2019, 34(11): 2364-2374. SHI Fang, ZHANG Linlin, HU Xiongwei, et al. Power system transient stability rules extraction based on multi-attribute decision tree[J]. Transactions of China Electrotechnical Society, 2019, 34(11): 2364-2374.
- [22] 数字城市地理信息公共平台地名/地址编码规则: GB/T23705—2009[S]. Coding rules for place names/addresses of digital city geographic information public platform: GB/T 23705—2009[S].
- [23] 广州供电局. 10 kV 及以下配网基建工程典型设计[Z]. 2017. Typical design of infrastructure project for distribution network of 10 kV and below in Guangzhou Power Supply Bureau[Z]. 2017.
- [24] 肖勇, 赵云, 涂志东, 等. 基于改进的皮尔逊相关系数的低压配电网拓扑结构校验方法[J]. 电力系统保护与控制, 2019, 47(11): 110-117. XIAO Yong, ZHAO Yun, TU Zhidong, et al. Topology checking method for low voltage distribution network based on improved Pearson correlation coefficient[J]. Power System Protection and Control, 2019, 47(11): 110-117.

收稿日期: 2019-04-08; 修回日期: 2019-07-01

作者简介:

高泽璞(1994—), 男, 硕士研究生, 主要研究方向为电力系统自动化, 大数据挖掘及可视化应用技术; E-mail: 137681373@qq.com

赵云(1985—), 男, 博士, 高级工程师, 主要研究方向为电网数据模型、大数据挖掘及可视化应用技术;

张莲梅(1970—), 女, 通信作者, 工学博士, 副教授, 研究领域为人工智能、电力系统规划、能源互联网. E-mail: lloottuuss@163.com

(编辑 葛艳娜)