

基于自然语言句法分析的电力变压器运行规范知识图谱构建

杨帆¹, 李靖², 王坚敏³, 丁克勤⁴, 谢元⁵, 密铁宾¹, 邱才明¹

(1. 大数据工程技术与人工智能中心(上海交通大学), 上海 200240; 2. 浙江华云电力实业集团有限公司, 浙江 杭州 310000; 3. 浙江华云信息科技有限公司, 浙江 杭州 310012; 4. 中国特种设备检测研究院, 北京 100029); 5. 密歇根州立大学, 密歇根州 东兰辛 48913

摘要: 为保证电力系统的安全可靠运行和实现供电系统的规范化、标准化和科学化管理, 电气设备需要严格按照安全运行管理规程进行维护操作。电力变压器是发电厂和变电所的主要设备之一, 其可靠运行对电能稳定传输有着重要意义。利用哈尔滨工业大学信息检索研究中心开发的语言技术平台, 对电力变压器运行规范进行句法分析和知识抽取。进一步基于图数据库Neo4j形成电力变压器运行规范知识图谱, 展示变压器各部件在不同情况下需要进行的操作, 为其运行维护提供了指导, 并通过实例验证了所提方法的有效性。

关键词: 电力变压器; 运行规范; 语言技术平台; Neo4j; 知识图谱

Construction of knowledge graph for power transformer operation management regulations based on natural language syntax analysis

YANG Fan¹, LI Jing², WANG Jianmin³, DING Keqin⁴, XIE Yuan⁵, MI Tiebin¹, QIU Caiming¹

(1. Research Center for Big Data and Artificial Intelligence Engineering and Technologies (Shanghai Jiao Tong University), Shanghai 200240, China; 2. Zhejiang Huayun Electric Power Industry Group Co., Ltd., Hangzhou 310000, China; 3. Zhejiang Huayun Information Technology Co., Ltd., Hangzhou 310012, China; 4. China Special Equipment Inspection and Research Institute, Beijing 10029, China; 5. Michigan State University, East Lansing 48193, USA)

Abstract: In order to ensure the safe and reliable operation of the power system and to realize the standardized and scientific management of the power supply system, the electrical equipment needs to be maintained in strict accordance with the safe operation management regulations. The power transformer is one of the main equipment for power plants and substations, and their reliable operation is of great significance for the stable transmission of electrical energy. The Language Technology Platform (LTP) developed by the information retrieval research center of Harbin Institute of Technology is used for syntactic analysis and knowledge extraction. Furthermore, a knowledge graph is formed for power transformer operation management regulations based on the graph database, i.e., Neo4j. The knowledge graph displays the operations required for each component in different situations and provides guidance for transformer maintenance, and the proposed method is verified by cases.

Key words: power transformer; operational management regulations; Language Technology Platform (LTP); Neo4j; knowledge graph

国家重点研发计划课题 (2018YFF0214704)

0 引言

电力变压器作为输变电系统的核心设备，其健康状态对于电力系统安全稳定运行至关重要。为保证设备管理的标准化，不同电压等级的变压器都必须严格遵守其运行规范进行验收、维护和异常处理等操作^[1]。然而，运行规范作为非结构化文本数据，其搜索系统都是基于关键词来实现的，存在不够直观、查准率和召回率低的问题，并且无法理解业务语言，也无法支持语义推理^[2]。因此，需要对该类数据进行自然语言句法分析及信息挖掘，形成可读性更高的图数据库，从而对设备验收、维护、巡视及缺陷管理提供更为高效的指导。

目前，自然语言处理技术的发展日趋完善，已经被广泛应用于医疗^[3-4]、互联网^[5]和情报学^[6]等各个领域，说明了该领域技术的通用性和可移植性，为电力文本挖掘技术的研究提供了思路。文献[7]提出了一种基于词库实现的分词算法，进而设计了一个适用于电力企业搜索引擎的分析器。文献[8]考虑电力设备缺陷文本的特点，基于卷积神经网络构建了电力设备缺陷文本分类模型，按缺陷严重程度分为“一般”“重要”和“紧急”三种类型。文献[9]基于词向量和循环卷积神经网络，挖掘电力变压器工作票信息，并将设备缺陷按等级进行分类。文献[10]针对电力设备缺陷记录语料，构建电力设备缺陷知识图谱，并基于所构建的知识图谱，提出了利用图谱搜索进行电力设备缺陷检索的方法。文献[11]基于Neo4j构建电网的全景数据库，根据电力网络拓扑建立设备映射表，从而进行网络参数的数据检索和信息提取。但是尚未开展针对设备运行规范进行信息提取，并形成支持推理的知识图谱，进而辅助电力设备运维管理的智能查询和策略推送的工作。

本文以66~500 kV油浸式变压器（电抗器）运行规范^[12]为例，基于语言技术平台（Language Technology Platform, LTP）对非结构化文本进行语法分析，编写规则进行知识抽

取，获取电力变压器运行规程中的主体及操作关系。进一步地，本文基于Neo4j形成电力变压器运行规程图数据库。数据库包含主要的节点、关系、属性和路径。可视化结果最终以关系图谱、关系表格的形式呈现，从而使得不同情况下需要对电力变压器进行的维护操作变得简明易懂，为变压器运行维护人员提供了便利。

1 自然语言句法分析及知识抽取

1.1 自然语言句法分析

实现自然语言句法分析，首先要正确地进行分词、词性标注和实体识别。目前这些技术已在众多自然语言处理平台中广泛使用，而现有的处理技术主要面向全领域，自带词库为常用词。电力行业由于包含较多的专业词汇，难以被正确切分及提取。因此本文根据文献[13]，首先对电力变压器建立了用户词典。用户词典包含词组及其词性，部分内容如下表所示。

表 变压器部分用户词典

Tab. Transformer user dictionary

词性	词组
名词	有载分接开关、绝缘油、气体继电器、主变压器、调度部门、排油管、分接开关、压力释放阀
动词	逸出、腐蚀、开裂、老化、变色
形容词	严重、轻微、正常、异常

基于预定义的用户词典和常用词库，电力变压器运行规范被依次进行分段、分句、分词、词性标注、命名实体识别、依存语义分析和语义角色标注等操作。该步骤借助哈尔滨工业大学信息检索研究中心开发的LTP实现。LTP是一个开放的中文语言处理平台，囊括了词法分析（包括分词、词性标注和命名实体识别）、句法分析（依存句法分析）和语义分析（词义消歧和语义角色标注）三方面六项语言处理基础技术，并且可以通过Python的PyLtp模块进行调用^[14]。以有载分接开关为例，“复装后注油前，应手摇操作，观察其动作切换情况是否正确，并测量变压器绕组直流电阻”，该句被解析如图1所示。

图1中，词性标注中n表示一般名词；v为动

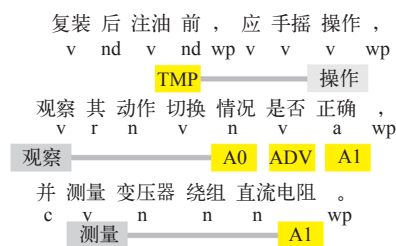


图1 自然语言句法分析

Fig. 1 Syntactic analysis of natural language

词；nd为方位名词；a为形容词；r为代词；c为连词；wp为标点符号；TMP为语义角色标注后的时间短语；A0通常表示动作的施事；A1通常表示动作的影响等；ADV为状语。句法分析是知识抽取的基础。

1.2 知识抽取

知识抽取主要任务是识别各实体/属性间是否存在关系及相应关系类型^[10]。具体而言，本节知识抽取的目的在于：抽取电力变压器运行规程中的各部件主体，如分接开关、套管、冷却装置和变压器油等，在安装过程中或者在不同异常情况下需要进行的操作。因此，抽取过程分为三步进行：

- 1) 抽取变压器所有部件名称，如套管、电流互感器、有载分接开关、气体继电器和压力释放阀等。
- 2) 对于每个部件，由语义角色标注找出时间短语（TMP），即该部件可能发生的不同情况。
- 3) 在时间短语（TMP）后找出动词（v）。同时，若句中含有动作施事（A0）或动作影响（A1），将动词及其施事、影响进行关联。

例如，编写关联规则后对图1例句进行知识抽取，即在复装后注油前，需做到：①手摇操作；②观察动作切换情况；③测量变压器绕组直流电阻。

2 知识图谱

2.1 知识图谱分类及构成

知识图谱的概念首先由Google进行实践并倡导，是下一代搜索引擎技术的核心，目的是加强智能搜索。一般地，知识图谱分为开放域和封闭域两类。开放域知识图谱知识覆盖范围较广，但存在实体歧义、关系难以穷举、专业性不强和噪声大等问题，主要用于搜索引擎；而封闭域知识图谱一般是行业知识图谱，只适用于特定行业，其知识专业性较强，噪声较小，实体、属性和关系也可根据需求进行限定和穷举，因此应用可以比较深入，更具有针对性^[15]。

本文所要构建的知识图谱属于封闭性知识图谱。作为结构化的语义知识库，其基本组成单元为“实体-关系-实体”或者“实体-关系-属性”三元组，实体间通过关系相互连接，构成网状的知识结构。相对于本体、传统的语义网络而言，知识图谱更强调对于实体描述的覆盖，而其语义关系也更加复杂而全面。

2.2 图数据库Neo4j

Neo4j是基于Java开发的高性能、高可靠性和可扩展性强的开源图数据库，完全兼容ACID，具备原子性（Atomicity）、一致性（Consistency）、隔离性（Isolation）和持久性（Durability）。本文知识图谱基于Neo4j搭建。

Neo4j的信息建模包括节点、边和属性三种构造单元如图2所示。其中构造单元的具体功能分述如下：

- 1) 点可以表示一个变压器的某个具体部件，如分接开关、温度计等。
- 2) 边可以表示设备（变压器）及其部件之间的关联，或者部件在不同情况需要进行的相关操作的关联。边可以带有方向。例如变压器含有绝缘油，表示为变压器=>绝缘油。
- 3) 属性针对点和边而言，不作任何具体约定且支持动态更新。

图数据库的创建和维护简单，节点和边的连接关系相当于传统关系数据库的一条记录，节点和边的属性则相当于记录中的字段，

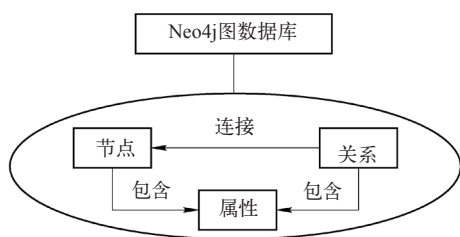


图2 Neo4j构造单元

Fig. 2 Tectonic unit

属性内容和个数可以动态变化，节点之间的边也可以自由删减并且不会影响已有数据结构的逻辑^[17]。因此，图数据库可以动态地进行修改、增加和删减，拥有灵活的表达能力和可扩展性。此外，Neo4j可以通过Python的Py2neo模块，导入图（Graph）、节点（Node）和关系（Relationship）等进行调用。图数据库创建完毕后，访问本地端口“http://localhost:7474/”，即可获得电力变压器运行规程知识图谱。

至此，电力变压器运行规程知识图谱构建完成，其流程如图3所示。

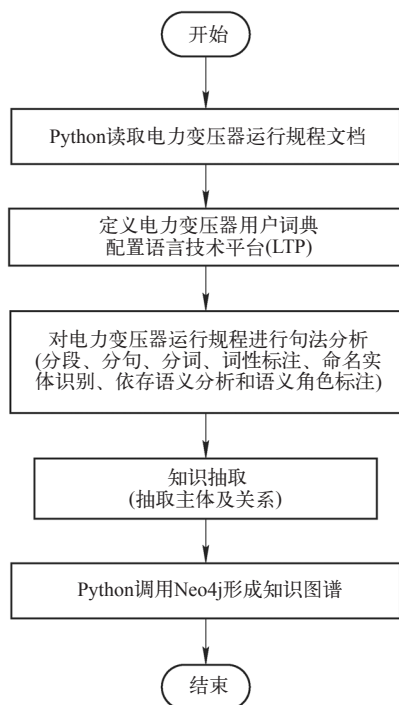


图3 电力变压器运行规程知识图谱构建流程图

Fig. 3 Construction flow chart of knowledge graph of power transformer operation rules

3 实例分析

电力变压器运行规程知识图谱给出变压器（主根节点）的不同部件（次根节点）在不同情况下，需要执行的操作。其基本架构如图4所示。

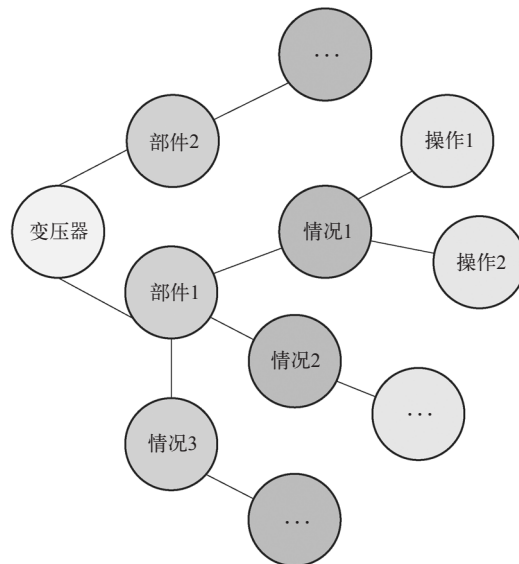


图4 电力变压器运行规程知识图谱基本架构

Fig. 4 Basic framework of knowledge graph of power transformer operation rules

3.1 单部件分析

现以66~500 kV油浸式变压器（电抗器）运行规范^[12]为例，对其分接开关维护注意事项进行解析并形成图谱，如图5所示。

3.2 多部件分析

节选66~500 kV油浸式变压器（电抗器）运行规范并进行解析。分析部件分别为吸湿器、阀门、净油器、二次回路和控制箱，形成图谱如图6所示。

4 结束语

本文基于LTP对电力变压器运行规范进行句法解析和知识挖掘，获取电力变压器各部件主体在安装过程中或者在不同异常情况下需要进行的操作。进一步地，通过图数据库Neo4j形成知识图谱，使得变压器运行规程简明易读，为操作人员提供了便利。

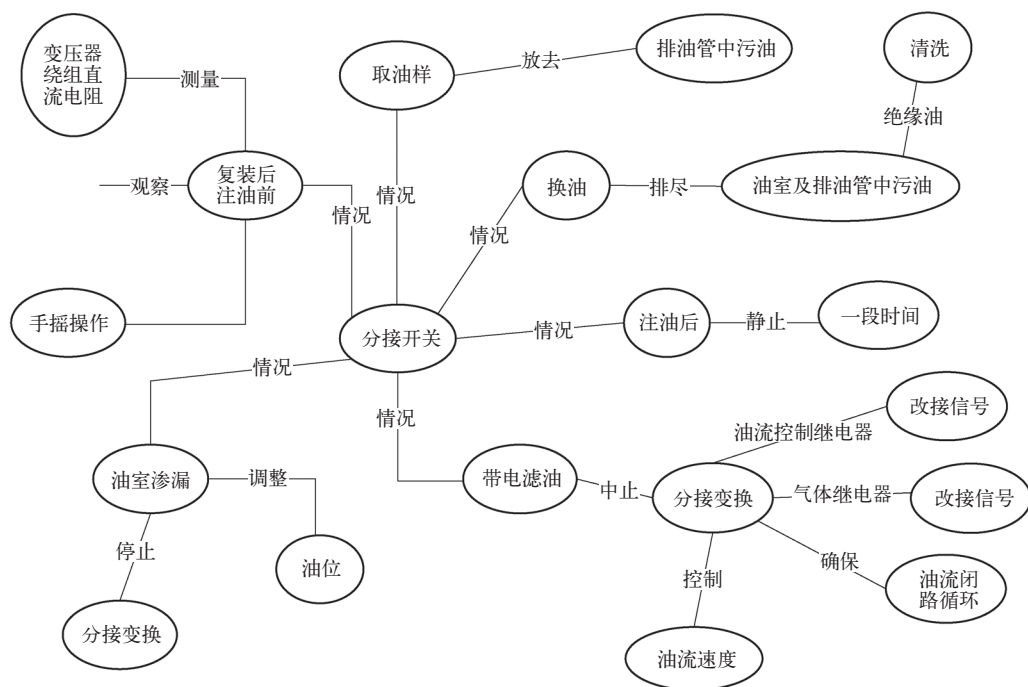


图5 分接开关维护注意事项

Fig. 5 Precautions for tap changer maintenance

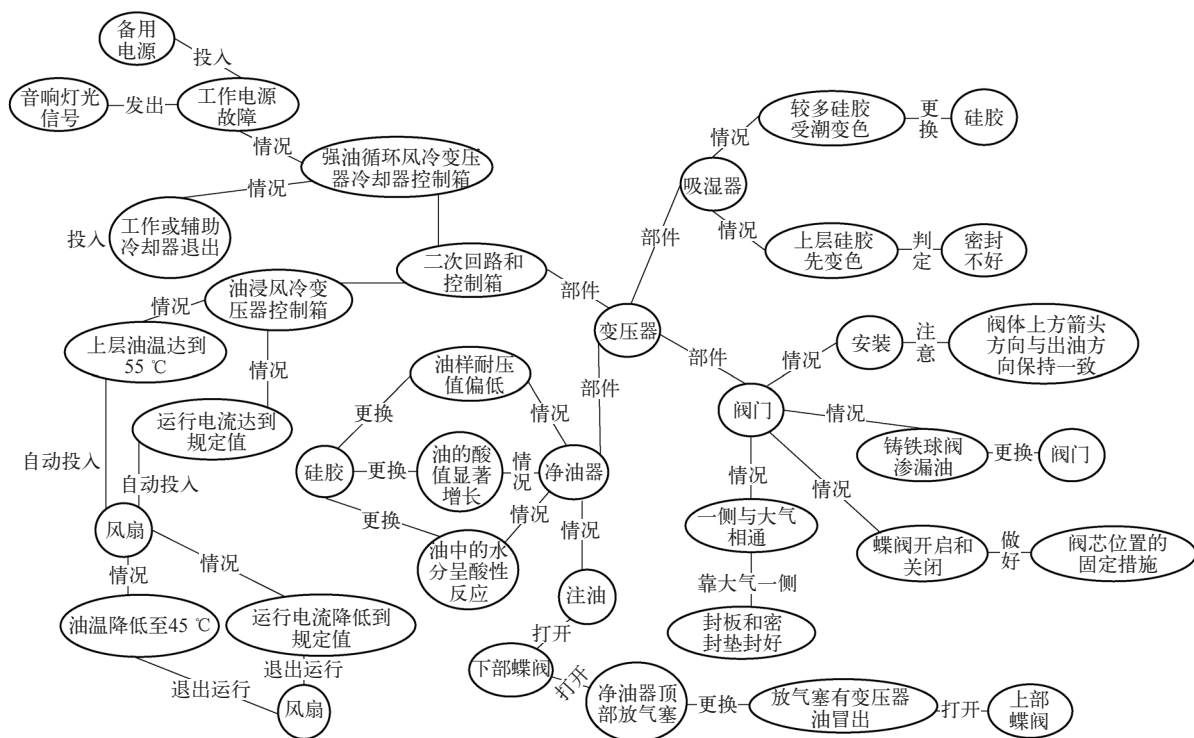


图6 多部件维护注意事项

Fig. 6 Precautions for multi component maintenance

参考文献

- [1] 电气装置安装工程 电力变压器、油浸电抗器、互感器施工及验收规范: GB 50148—2010[S].
- [2] 姬源, 谢冬, 周思明, 等. 电力领域语义搜索系统的构建方法[J]. 计算机系统应用, 2016, 25(4): 91-96.
- [3] 周晴宇. 基于医疗知识图谱的探索式搜索研究[D]. 湘潭: 湘潭大学, 2017.
- [4] 侯梦薇, 卫荣, 陆亮, 等. 知识图谱研究综述及其在医疗领域的应用[J]. 计算机研究与发展, 2018, 55(12): 2587-2599.
- [5] 郝云飞. 基于自然语言处理的互联网舆情高危信息处理模块的研究[D]. 呼和浩特: 内蒙古大学, 2019.
- [6] 刘旭东, 苏马婧, 朱广宇. 基于自然语言处理的多源情报分析系统的研究与设计[J]. 信息技术与网络安全, 2019, 38(5): 17-21.
- [7] 吴克河, 何霞, 李廷顺. 基于Lucene构建电力企业搜索引擎分析器[C]. 2008电力行业信息化年会.
- [8] 刘梓权, 王慧芳, 曹靖, 等. 基于卷积神经网络的电力设备缺陷文本分类模型研究[J]. 电网技术, 2018, 42(2): 644-651.
- [9] 蒋逸雯, 李黎, 李智威, 等. 基于深度语义学习的电力变压器运维文本信息挖掘方法[J]. 中国电机工程学报, 2019, 39(14): 4162-4172.
- [10] 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.
- [11] 马义松, 武志刚. 基于Neo4j的电力大数据建模及分析[J]. 电工电能新技术, 2016, 35(2): 24-30.
- [12] 国家电网公司. 10(66) kV~500 kV油浸式变压器(电抗器)运行规范[S/OL]. <https://wenku.baidu.com/view/b5ea4b87ec3a87c24028c4ec.html>, 2005.
- [13] 曹靖, 陈陆燊, 邱剑, 等. 基于语义框架的电网缺陷文本挖掘技术及其应用[J]. 电网技术, 2017, 41(2): 637-643.
- [14] CHE W, LI Z, LIU T. LTP: a Chinese language technology platform[J]. Journal of Chinese Information Processing, 2010, 2(6): 13-16.
- [15] 徐增林, 盛泳潘, 贺丽荣, 等. 知识图谱技术综述[J]. 电子科技大学学报, 2016, 45(4): 589-606.
- [16] 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.
- [17] 王余蓝. 图形数据库Neo4j与关系数据库的比较研究[J]. 现代电子技术, 2012, 35(20): 77-79.

作者简介



杨帆

杨帆(1995—), 女, 博士研究生, 主要研究方向为电力系统大数据。

(收稿日期: 2020-05-11) **EA**