

电力安全知识图谱研究综述

周梓俊¹ 喻召杰² 罗艾珂¹

1 广东电网有限责任公司东莞供电局, 广东东莞, 523129;

2 南网数字运营软件科技(广东)有限公司, 广东深圳, 518000;

摘要: 随着电力行业智能化转型加速, 电力安全管理的复杂性与重要性日益凸显。知识图谱作为结构化知识表示与语义推理的核心工具, 为电力安全领域提供了新的解决方案。本文系统梳理了电力安全知识图谱的研究现状, 聚焦其构建方法、关键技术及应用场景, 并提出未来发展方向。论文的研究可为电力安全智能化提供一定的技术支撑和参考借鉴。

关键词: 电力安全; 知识图谱; 故障诊断; 风险预警

DOI: 10.69979/3060-8767.26.01.001

引言

随着电力行业的快速发展和智能化转型, 电力安全管理的复杂性和重要性日益凸显。知识图谱作为结构化知识表示和语义推理的有效工具, 在电力安全领域展现出广阔的应用前景。本文系统梳理了电力安全知识图谱的研究现状, 重点分析其构建方法、关键技术及应用场景。

1 知识图谱基础与相关技术

1.1 知识图谱的核心理念与行业价值

知识图谱的本质是应对信息爆炸时代知识管理复杂性的解决方案。其核心在于将异构、碎片化的数据转化为具有明确语义关联的网络化知识体系, 从而为机器理解语义、进行深度推理奠定基础。在电力安全这一垂直领域, 知识图谱的价值尤为凸显, 它能够将分散在设备台账、操作日志、事故报告中的信息, 整合成一张清晰的“安全知识地图”, 直观展现“设备-隐患-操作-人员”之间的复杂关联, 为智能诊断和风险预警提供支撑。具体而言, 该知识体系通过“节点”表示实体(如变压器、巡检员、安全规程), “边”表示实体间的关系(如“位于”、“负责”、“违反”), 进而通过(头实体, 关系, 尾实体)构成的三元组作为基本表达单元。例如, 事实“巡检员张三负责巡视 500kV XX 变电站”可形式化表示为(张三, 负责巡视, 500kV XX 变电站)。这种结构化表示使得知识不再是孤立的数据点, 而是一个相互连接的、可追溯的动态网络。

1.2 面向电力安全的逻辑架构设计

知识图谱的架构设计需服务于其应用目标。针对电力安全领域对准确性和实时性的高要求, 其逻辑架构需在模式层和数据层实现紧密协同^[1]。

模式层(概念模型): 此层如同知识图谱的“宪法”, 由本体库构成。它定义了电力安全领域内所有概念(如“隐患”、“操作票”)的分类体系、属性规范以及概念间的约束关系(如“一种隐患”必然“关联于”“某个设备”)。一个设计良好的本体库是确保知识一致性和质量的关键, 例如, 它能有效区分“接地线”作为“安全工具”和“接地操作”作为“作业步骤”的不同语义。

数据层(实例数据): 此层是模式层的实例化, 存储着具体的事例知识, 通常依托于 Neo4j、Nebula Graph 等图数据库。在电力安全场景下, 数据层不仅存储静态知识(如设备参数), 更需支持动态数据的接入与更新(如实时告警、巡检记录), 从而形成一个鲜活演化的知识体系。

1.3 知识图谱的构建范式演进

知识图谱的构建并非单一僵化的流程, 而是存在不同的方法论范式, 以适应不同的应用场景和数据条件。

自顶向下范式: 该范式适用于业务规则明确、行业标准成熟的领域, 如电力安全规程。它强调先由领域专家主导, 构建起顶层本体模型, 明确核心概念与关系, 再基于此模型去填充和约束具体实例数据。这种方式能确保知识体系的规范性和逻辑严谨性, 但对专家知识依赖度高。

自底向上范式: 该范式则更适用于从海量、开放

域数据中发掘知识。它首先利用自然语言处理等技术从文本中自动化抽取实体和关系，形成大量三元组，然后通过聚类、归纳等方式，自底向上地抽象出本体模式。这种方式自动化程度高，但生成的本体在一致性上可能面临挑战。

混合迭代范式：对于电力安全这类兼具强规范性与数据多样性的领域，单纯的“自顶向下”或“自底向上”均存在局限。因此，一种“混合迭代”的构建范式更为可行，首先基于行业标准建立一个轻量级的核心本体框架（自顶向下），然后利用自动化工具从业务数据中抽取知识并不断丰富实例（自底向上），在此过程中持续校验、修正和扩展本体模型，形成一个螺旋式上升的构建闭环。

1.4 知识图谱的核心技术环节

知识图谱的构建并非一蹴而就，而是一个环环相扣的技术流程。其中，知识抽取与知识融合是决定知识库规模与质量的两个核心环节，它们共同将原始数据转化为纯净、可用的结构化知识。

（1）知识抽取：从非结构化文本到结构化知识

知识抽取承担了从海量、多源的电力安全文本（如巡检记录、工作票）中自动化提取事实知识的任务。其技术演进体现了从依赖规则到依靠模型的转变，主要包括三个子任务：

实体抽取：旨在从文本中定位并分类关键概念，如设备名称、隐患类型、作业人员等。早期的规则与词典方法虽直观但难以扩展。当前主流方案是基于深度学习的序列标注模型，通过预训练语言模型获取词汇的深度语义表示，再结合双向长短句记忆网络（BiLSTM）捕捉上下文特征，最后由条件随机场（CRF）层保证标签序列的全局最优性，从而有效应对电力领域专业术语的识别挑战。

关系抽取：在识别出实体后，需进一步判断实体间的语义关系（如“位于”、“引发”）。其方法从依赖句法分析的管道式抽取，发展为基于神经网络的端到端联合抽取模型。后者能同时抽取实体和关系，有效解决了传统管道式方法存在的错误累积、实体冗余等固有问题，尤其适合处理电力文本中常见的复杂关系模式。

属性抽取：负责充实实体的详细信息，例如从操作日志中提取某项作业的“开始时间”、“负责人”等属性。其技术思路常与关系抽取相似，可视为实体与属性

值之间的特定关系抽取，从而实现对实体特征的全面刻画。

（2）知识融合：构建统一知识视图的关键

从不同来源抽取的知识必然存在异构、冗余和冲突，知识融合技术正是为了构建统一、洁净的知识视图而生。它通过一系列精细化操作解决数据层面的不一致问题：

指代消解：解决同一段文本中代词或别称指向同一实体的问题（如“该断路器”…“它”），确保上下文提及的实体唯一。

实体链接：将文本中抽取的实体指称项（如“一号主变”）链接到知识库中的标准实体（如“#1 主变压器”），解决简称、别称等别名的归一化问题。

实体消歧：区分不同语境下同名实体所指向的不同对象（如“李伟”可能指代不同变电站的站长），避免错误合并。

知识合并：将来自外部知识库（如通用知识图谱或专业数据库）的相关知识集成到当前图谱中，消除冗余，丰富知识体系。

知识抽取负责“从无到有”地发现知识，而知识融合则负责“去伪存真”地精炼知识。二者协同工作，共同确保了电力安全知识图谱的完备性、准确性和一致性，为上层智能应用提供可靠的数据基石。

2 电力安全知识图谱构建方法

2.1 电力安全数据的多模态特征与治理挑战

电力安全知识图谱的构建质量，高度依赖于底层数据的全面性与准确性。电力系统的复杂性决定了其数据具有显著的多模态与多源性特征，主要体现为三种形态：

（1）结构化数据：主要来源于各类业务系统数据库，如生产管理（MIS）、设备资产（EAM）等系统，其格式规整、字段明确，例如设备台账、标准化作业流程等。这类数据是构建知识图谱中确定性事实的核心来源。

（2）半结构化数据：如工作票、操作票、日志文件等，它们虽有一定格式但灵活性高，通常包含大量非标准化的文本信息。如何从这些文本中精准提取关键实体和关系是主要挑战。

（3）非结构化数据：在电力安全领域中占比最大，包括事故分析报告、现场巡检记录、应急预案文档等。这些数据以自由文本为主，蕴含着丰富的专家经验和隐性知识，是知识抽取的重点和难点。

面对这些多模态数据，首要任务是进行系统性的数据治理。这远不止于传统的数据清洗（去重、纠错、补全），更包括针对电力领域特性的语义标准化工作。例如，对不同系统中指代同一概念的术语（如“接地变”与“接地变压器”）进行统一，对同一操作的不同描述进行规范化，为后续的知识融合打下坚实基础。

2.2 电力安全本体库的构建策略

本体库是电力安全知识图谱的语义基石，其构建策略直接决定了图谱的逻辑严谨性和实用性。如2.3节所述，混合迭代范式是当前最适合电力安全领域的构建策略。在这一范式指导下，具体实施路径如下：

(1) 核心本体框架设计（自顶向下阶段）：此阶段由领域专家和知识工程师协同完成。重点是基于电力安全规程、行业标准（如《电力安全工作规程》）以及专家经验，抽象出核心概念类别（如电力设备、作业活动、安全隐患、防护措施）并定义其层次结构与核心关系，形成一个稳定且可扩展的顶层本体框架。

(2) 知识实例化与本体丰富（自底向上阶段）：在核心本体的指导下，利用知识抽取技术（见3.3节）从多模态数据中抽取实例和关系。例如，从事故报告中抽取“（某开关，发生，绝缘故障）”这一实例，并将其关联到顶层本体的“安全隐患”概念下。这个过程会不断暴露出本体框架未覆盖的细粒度概念和新的关系，从而驱动本体的迭代丰富。

(3) 迭代优化与验证：通过持续将新增实例与现有本体进行匹配和冲突检测，不断修正和扩展本体模型。例如，当从大量数据中发现“超声波检测”是一种普遍存在的“检测手段”时，可将其作为“防护措施”的一个子类加入本体。这种迭代过程确保了本体库既能保持顶层规范，又能紧密贴合实际业务数据的多样性和动态性。

通过这种策略构建的本体库，不仅具备良好的理论支撑，更拥有强大的工程实践适应性，能够伴随业务发展而持续演化。

2.3 知识抽取方法

面向电力安全文本的知识抽取，需有效应对其专业术语密集、语法结构灵活及语义关系复杂等核心挑战。为应对这些挑战，研究方法逐渐从依赖显式规则转向依赖隐式数据模型。

在数据稀缺的场景下，基于规则的方法曾是一种直接有效的解决方案。该方法的核心在于人工构建高质量的领域词典与句法规则模板，通过模式匹配来识别目标知识。然而，其性能严重依赖于规则库的完备性与精确性，在面对电力现场产生的、表述多样化的非结构化文本时，其召回率往往不尽人意^[2,3]。

当前，基于深度学习的端到端模型因其强大的表示学习与泛化能力，成为研究的主流。这些模型通常遵循“表示-编码-推理”的计算框架：首先在表示层将文本映射为稠密向量；随后在编码层利用深度神经网络（如BiLSTM）捕获上下文特征；最后在推理层（如CRF）进行结构化预测。针对电力安全的具体需求，该框架衍生出诸多变体：例如，为处理实体重叠问题，^[4]提出了三元组分类模型；为捕捉时序依赖性，^[5]引入了基于GRU的序列建模；为实现更高效的信息抽取，^[6]设计了实体与关系的联合学习模型。这些进展显著提升了对复杂电力安全知识的自动化获取水平。

3 电力安全知识图谱应用场景

3.1 智能诊断

在电网运行中，设备故障的快速诊断是保障电力供应的关键。传统方法依赖人工经验判断，效率低下且易遗漏潜在隐患。基于知识图谱的智能诊断系统通过构建电网设备、运行参数、历史故障等实体及其关系的知识网络，实现故障的快速定位和根因分析。

3.2 作业安全

电力作业具有高风险性，传统培训方式存在参与度低、效果评估难等问题。结合VR技术和知识图谱的培训系统通过构建三维虚拟场景，将电力安全规程、操作流程、应急处置等知识以可视化方式呈现。

3.3 风险预警

电网运行中存在大量潜在风险，传统预警系统存在误报率高、预警滞后等问题。基于知识图谱的隐患识别系统通过整合气象数据、设备状态、历史故障等多源信息，构建包含风险因素、影响范围、处置措施等要素的知识网络。

4 未来研究方向

4.1 研究小样本学习降低知识标注成本

电力领域专业性强，标注高质量知识图谱需要大量

专家参与,成本高昂。未来研究可探索小样本学习技术,通过迁移学习、元学习等方法,利用少量标注数据训练模型,实现知识的自动抽取和标注。

4.2 研究语义对齐进行多源数据融合

电力数据来源广泛,格式各异,语义不一致问题突出。未来需加强语义对齐研究,构建统一的电力领域本体,实现不同数据源的知识融合。

4.3 融合大语言模型增强知识推理能力

大语言模型具有强大的语义理解和生成能力,未来可将其与知识图谱结合,构建“知识图谱+大语言模型”的混合系统,实现更复杂的推理和问答功能。

4.4 融合数字孪生探索新型应用模式

数字孪生技术可实时映射物理电网状态,未来可探索“知识图谱+数字孪生”的融合应用模式,构建虚实互动的智能电网系统。

参考文献

- [1] 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(03):582-600.
- [2] 潘道成,邓卫民,蒋祝巍,等. 基于电网安全隐患知识图谱的智能诊断技术研究及应用[J]. 河北电力技术,2022,41(04):20-24.
- [3] 郭素芹,郑建宁,陈坤,等. 基于知识图谱的变电站安全隐患动态分析方法[J]. 电力系统及其自动化学报,2021,33(12):125-133. DOI: 10.19635/j.cnki.csu-epsa.000791.
- [4] 张燎原,李英娜. 基于三元组分类器的电力安全领域知识抽取[J/OL]. 电力科学与工程:1-8[2024-05-29]. <http://kns.cnki.net/kcms/det1/13.1328.tk.20240412.1511.002.html>
- [5] 李洪海. 基于大数据分析的电力行业安全生产隐患识别研究[J]. 现代职业安全,2025,(03):14-17.
- [6] 李泽科,陈书里,陈斌,等. 基于知识图谱的电力调度自动化作业工作票生成方法[J/OL]. 福州大学学报(自然科学版),1-9[2025-04-13].

作者简介: 周梓俊(1994.07—),男,汉族,工程师。主要研究方向为安全应急管理。

喻召杰(1985.12—),男,汉族,高级工程师。主要研究方向为安全生产信息化。

罗艾珂(1995.02—),女,汉族,工程师。主要研究方向为安全管理。

基金项目:南方电网科技基金资助项目(031900KC23040016(GDKJXM20230399)、(031900KC23040017(GDKJXM20230401))