

电力企业领域知识图谱构建与应用研究*

杜一玮¹ 柴丽萍² 李题印¹ 刘宗泽¹ 卢家驹¹

(1. 国网浙江省电力有限公司杭州供电公司 浙江杭州 311100; 2. 国网浙江省电力有限公司 浙江杭州 311100)

摘要: 电力企业领域知识的来源可能有企业文档数据、经营信息、档案材料、数据库数据和互联网公开数据等形式。海量的数据知识给电力企业日常经营管理带来了一定的挑战。开展电力企业领域知识图谱构建与应用研究,有助于推荐电力企业智慧化办公、智慧化管理,提高电力企业的管理效率。

关键词: 电力企业; 领域知识图谱; 知识管理; 数据管理

Abstract: The sources of knowledge in the field of power enterprises may include enterprise document data, business information, archival materials, database data, and publicly available internet data. The massive amount of data knowledge has brought certain challenges to the daily operation and management of power enterprises. Conducting research on the construction and application of knowledge graphs in the field of power enterprises can help recommend intelligent office and management for power enterprises, and improve their management efficiency.

Keywords: electric power enterprises; Domain Knowledge Graph; Knowledge management; Data management

一、引言

随着数字智能化时代的带来,数据管理问题成为社会各行各业的一项新的挑战。电力企业作为作为关系国家安全和国民经济命脉的特大型国有重点骨干企业,业务复杂,数据繁多。在电网业务的经营与管理过程中,往往会产生海量数据。如何对海量数据进行采集、整理与分析,从中挖掘出正确有用的信息,并辅助电力企业工作人员进行管理决策,一直是电力企业数智化探索的一项重要任务。知识图谱技术与方法已经成为企业信息管理领域的一大主流方法与技术,其为电力企业海量大数据的数据关系的构建与表达提供一种科学可行的解决方案。

知识图谱可依据其所涵盖的知识形式进行分类,可以根据侧重点在广度或深度分为两大类:通用知识图谱和领域知识图谱。通用知识图谱面向普遍性通用针对通用的环境和场景,覆盖了众多领域的知识,拥有大量的知识单元。领域知识图谱专注于特定的行业领域,其知识内容依赖于特定行业的数据,对于数据深度、质量以及准确性的要求比通用知识图谱更高。本文重点围绕电力企业构建领域知识图谱,旨在满足电力企业管理层的需求以及技术人员的专业知识需求。该知识图谱属于领域知识图谱的范畴,被称为电力企业领域知识图谱。因此,本文围绕电力企业领域知识图谱的构建与应用进行思考与实践。

二、知识图谱及其在电力企业领域的应用

知识图谱是美国科技公司谷歌于2012年提出的概念,是一种以符号形式表示真实世界中的各类概念与关系的结构化语义知识库^[1]。谷歌公司发展知识图谱技术的初衷是以知识图谱为核心基础,构建人工智能驱动的全新一代搜索引擎,提高互联网用户信息检索效应与体验。知识图谱概念中的关键技术主要是从海量信息中抽取概念实体并构建实体之间的关联关系。

自从谷歌公司将知识图谱运用到其搜索引擎之后,知识图谱技术就引发了工业和学术界的广泛关注。伴随着计算机视觉(CV)以及自然语言处理(NLP)等人工智能领域的重大突破,知识图谱已在众多行业如网络搜索、医疗信息、电商交易、证券金融等领域得到广泛运用。它在智能推荐、问题解答、数据分析及决策支持系统等方面具有关键性的影响力。知识图谱的基本构成单位是由“实体-关系-实体”组成的三元组。在这个架构中,可以为实体和关系增加详细的属性信息。实体之间的关系将各类实体关联在一起,以知识网络的结构,组成了一个连通的、完整的、巨大知识体系,通过图示化展现知识的结构关系。

知识图谱一种以语义网络为可视化方案的知识库。真实世界中的各类事物及事物之间的关联关系被抽象为知识实体与关系。因此,可以将知识图谱视作实体与关系的集合体。基于图论的数学视角看,知识图谱是由表示实体的知识节点,以及表示实体关系的边所构成的有向图。

知识图谱往往以三元组的形式进行表示,即:

$$KG = (E, R, S)$$

其中 $E = \{e_1, e_2, e_3, \dots, e_n\}$ 表示的是知识实体节点的集合,共包含 n 个知识实体节点; $R = \{r_1, r_2, r_3, \dots, r_m\}$ 表示的是知识实体的关系的集合,共包含 m 种不同知识关系; $S \subseteq E \times R \times E$, 表示的是该知识图谱中的“知识实体-

知识关系-知识实体”的知识三元组集合。

从知识图谱的知识构成层次看,一个知识图谱由模式层与数据层两个层次构成。目前,构建知识图谱方式主要有“自上而下”和“自下而上”两种技术思想^[2]。其中,“自上而下”指的是,在构建知识图谱时,首先构建知识图谱的模式层,定义抽象层面上的知识本体,再根据知识本体的具体规则将知识实体添加到知识图谱。与“自上而下”不同,“自下而上”构建方式指的是先基于知识抽取技术,从海量数据中抽取知识实体与知识关系,从中选取置信度较高的知识实体与知识关系加入知识图谱,再根据知识图谱对模式层的知识本体进行定义。

电力企业领域知识图谱是知识图谱在电力企业领域的实践应用,它是一种将电力企业产生、可运用的知识以及这些知识之间的相互关联以一种有向图的形式来表示和组织的工具。这种方式使得知识能够被计算机所理解、计算和推理,从而实现了对于知识的高效组织、管理和利用。在自适应学习系统中,知识图谱的使用对于构建领域模型具有重要意义。这不仅有助于知识的有序排列,而且还能为学习者提供个性化的学习资源和学习路径,提供根据使用者需求的自适应的学习服务。

随着电力企业的数据需求的提高与知识图谱技术的发展,电力企业领域知识图谱的构建研究开始引起学术界的关注。束嘉伟^[3]等提出面向电力知识图谱构建的重叠实体关系联合抽取方法,通过改进的序列标注方案进行联合抽取,构建了电力领域专属预训练模型,并增加对抗训练,提高了模型抽取电力知识的准确度和对陌生信息的预测能力。简家骏^[4]等从电力运检的功能需求出发提出引入 MMKG 解决综合智能问答系统和故障处置问题,重点介绍了针对电力运检数据的 MMKG 构建技术,总结了 MMKG 能够发挥作用的电力运检功能场景,并对未来发展方向做出展望,最后深入分析了发展 MMKG 会面临的挑战,为电力运检智能化发展与建设提供参考。王逸帆^[5]指出随着电力系统规模的扩大,产生的信息大量增加,为了对信息进行快速响应和处理,提升对电力领域知识管理的能力,需要在电力调度领域中开展知识图谱的研究,而电力调度规程是电力生产和管理的规章和依据,对其知识图谱进行研究具有很强的实用意义。刘炎昌指出随着国家电网公司信息化建设的快速发展,公司通信设备数量不断增加,对于电力通信设备故障的处置难度也在不断提高。若能将积累的历史故障文本数据加以组织和利用,将电力通信设备当中发生的故障信息进行平台化整合,建立电力通信设备故障知识图谱,可以较好的解决电力通信设备故障的处置问题^[6]。

梳理和分析现有电力企业领域知识图谱构建的研究发现,国内研究者对于电力企业技术层面知识图谱模型构建的研究关注较多,但对电力企业管理层面领域知识图谱构建的关注较少。即使有少数涉及电力企业管理知识图谱的研究,也多是只涉及单个管理场景下知识图谱应用的研究,缺乏以企业为主体的系统性的专业性电力企业知识图谱模式构建研究,同时存在管理视角下描述视角单一、不面对实际问题等问题。

三、电力企业领域知识图谱的构建

电力企业领域知识图谱系统地整合了电力公司管理的各项需求,它根据电力公司的管理目标,运用知识图谱技术来描绘和组织电力公司

运营所必需的各种知识、技巧以及它们之间的关联,并将被广泛应用于电力企业日常经营管理工作。

通过利用现有的电力企业的业务成果和内部资料,以及互联网上收集的数据,我们成功地提取出了一套计算机可以解读的业务知识。以支持电力企业业务行为与管理决策辅助为目标,本文整合电力企业特点,给出了电力企业领域知识图谱构建技术框架。如图1所示。

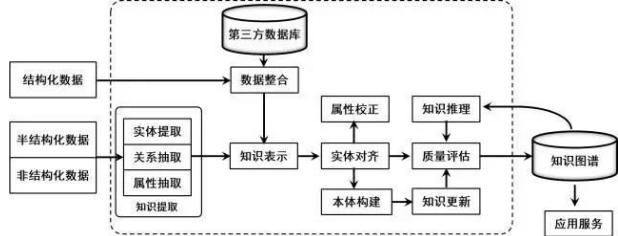


图1 电力企业领域知识图谱构建技术框架

图1中虚线框内的部分为知识图谱的构建过程,主要依赖于自下而上的方式,从高质量结构化数据中抽取专业程度高的实体及其关联,然后将这些实体和关系与顶层本体相匹配并构建知识图谱。随后,从其他数据源抽取实体、关系以及属性,并将它们融合到现有的实体和关系中。在抽取过程中,顶层本体与知识图谱是不断优化和迭代的。通过运用特定的技术工具,我们能够从广泛收集的数据源中筛选出有价值的资源模式。然后,我们进一步挑选出这些模式中的高置信度内容,并经过审核后纳入知识库。

在构建电力企业领域知识图谱的过程中,首先需要从原始数据入手。这个过程涉及到利用各种自动或半自动技术来挖掘出隐藏在原始数据中的知识实体,并将它们存储在知识库的两个主要层次——数据层和模式层。这一过程是一个持续更新和优化的过程,其核心是基于知识获取的逻辑。以下是每一步的关键技术要点:

(1) 知识提取

在构建电力企业领域知识图谱的过程中,知识提取是首要且至关重要的一步。这个过程旨在自动或半自动地从各种形式的非结构化和无结构数据中提取出实体的结构化信息,如实体、关系和属性等。这一过程面临的主要挑战在于如何从多元化的数据源中获取关键信息,并形成候选知识单元。这其中涉及到一些复杂的问题,例如知识的评估和重复处理等。这一过程的主要任务包括实体提取、关系提取和属性提取等。

(2) 数据整合

绝大部分数据资源的价值总量虽高,但由于其数据量的浩繁,往往其价值密度较低,这是数据资源独有的特点。虽然企业档案数据所承载的企业数据价格密度较大、可信度和可靠性较高,但是由于企业档案数据类型多样、结构复杂且多数情况下信息较为分散。电力企业领域知识的来源可能包括企业文档数据、经营信息、档案材料、数据库数据和互联网公开数据。其中,企业文档数据是电力企业领域知识图谱的核心数据源,其特点是由于高度结构化与规范性所保证的规范性。知识提取后可能包含大量的重复和冗余信息,数据之间的关系也是扁平化的。因此,需要对其进行清洗和整合。通过数据整合,可以消除概念的歧义,剔除冗余和错误概念,保证知识的质量。在电力企业领域知识图谱的构建过程中,数据整合的关键内容包括:实体对齐和数据整合等。

(3) 知识表示

知识表示是将现实世界的各种知识转化为计算机可以储存和处理的数据结构的过程。为了实现知识智能表示,机器必须具备丰富的知识储备,尤其是基本的常识性知识。自人工智能的发展历程起,知识表示的研究便已存在。在电力行业领域,知识图谱的知识表示方式通过结构化的手段来描绘客观世界中的概念、实体以及它们之间的联系,从而将网络上的信息转换为一种更贴近人类认知世界的模式,为理解电力企业的知识奠定了基础。

(4) 知识加工

通过知识抽取,可以从企业文档数据、经营信息、档案材料、第三方数据库以及互联网公开数据中提取出实体、关系与属性等知识要素。

再经过数据整合,可以消除实体指称项与实体对象之间的歧义,得到一系列基本的事实表达。然而,事实本身并不等于知识,在数字化转型的大环境以及服务智慧化的大趋势下,业务场景发生了诸多变化。所以,必须以用户需求为导向,结合场景完成知识加工的过程,才能收获到企业需要的知识。在电力企业领域知识图谱的构建过程中,知识加工的关键内容包括:本体构建、知识推理和质量评估等。

(5) 知识推理

电力企业领域知识图谱以图的方式展现了实体展示了实体、实例以及它们之间的关联。这种图形化的知识存储和检索方式旨在设计出能够有效处理大规模图形数据的存储策略,从而提高对知识图谱内知识的高效检索效率。在电力企业领域知识图谱的基础上,我们能够通过推理技术揭示新的实体及其相互关联。这种推理方式主要包括两种类型:一种是以符号为基础的推理,另一种则是以统计为基础的推理。在人工智能研究领域,前者通常依赖于逻辑规则或者其变体。这种类型的推理能从现有的知识图谱中挖掘出新的实体联系,这对于构建新知识或者检查知识图谱中的逻辑矛盾都非常有帮助。后者通常采用关系机器学习的方式,也就是通过统计规律来从知识图谱中提取新的实体关系。在知识计算过程中,知识推理扮演着关键角色,例如知识分类、知识验证、知识链接预测以及知识补充等任务。

四、电力企业领域知识图谱的构建

电力企业领域知识来源包括企业文档数据、经营信息、档案材料、第三方数据库以及互联网公开数据。海量的数据知识给电力企业日常经营管理带来了一定的挑战。当前社会已经进入数字智能化时代,各类型企业都面临数智化转型问题。随着信息技术的持续以及无纸化办公的盛行,大量公开数据被积累下来,对这些历史数据和文档进行开发和利用,能够就为今后工作的开展提供极大便利。开展电力企业领域知识图谱构建与应用研究,有助于推荐电力企业智慧化办公、智慧化管理,提高电力企业的管理效率。本文探讨了电力企业领域知识图谱的构建过程,介绍了电力企业领域知识图谱技术框架,对电力企业领域知识图谱的应用及实践成果进行了初步的探索、分析与思考,后续将重点围绕电力企业领域知识图谱的数据层构建问题展开研究,通过知识抽取、知识融合、知识加工等步骤实现电力企业领域知识图谱在数据层的构建。同时,注重知识图谱在电力企业中的应用模式,探索具体的管理应用场景。

参考文献:

[1]张吉祥,张祥森,武长旭等.知识图谱构建技术综述[J].计算机工程,2022,48(03):23-37.
 [2]杨玉基,许斌,胡家威等.一种准确而高效的领域知识图谱构建方法[J].软件学报,2018,29(10):2931-2947.
 [3]束嘉伟,杨挺,耿毅男等.面向电力知识图谱构建的重叠实体关系联合抽取方法[J/OL].高电压技术:1-11.
 [4]蔺家骏,闫玮丹,胡俊华等.多模态知识图谱在电力运检中的应用与展望[J/OL].综合智慧能源:1-10.
 [5]王逸帆.安徽省电力系统调度规程的知识图谱模式层构建研究[D].安徽财经大学,2023.
 [6]刘炎昌.基于知识图谱的电力通信设备故障诊断方法研究及应用[D].东北电力大学,2023.

* 基金项目: 国网浙江省电力有限公司科技项目(B311HZ220004)

作者简介: 杜一玮(1982-),男,硕士,高级工程师,主要从事电气工程、能源电力政策研究;
 柴丽萍(1983-),女,硕士,政工师,主要从事能源电力政策研究;
 李题印(1980-),男,博士,高级工程师,主要从事电气工程、能源电力政策研究;
 刘宗泽(1994-),男,学士,助理工程师,主要从事电气工程及其自动化研究;
 卢家驹(1994-),男,学士,助理工程师,主要从事电气工程及其自动化研究。