

基于知识图谱的西藏能源多源异构大数据融合与推理研究

管奕

西藏民族大学，陕西咸阳，712082；

摘要：随着信息技术的不断进步和互联网的快速发展，大数据已经成为当今社会中的一项重要资源。对于西藏能源领域，多源异构数据的融合与推理研究，主要探讨的是基于知识图谱的数据融合与推理方法。通过对知识图谱的学习利用以及西藏能源数据的收集、预处理、知识抽取、融合以及推理机制的构建，旨在为西藏能源产业的智能化发展与可持续发展提供高效、精准的技术支持，具有重大的实践价值。

关键词：知识图谱；西藏能源；多源异构；大数据融合

DOI：10.69979/3041-0673.24.9.022

引言

位于中国西南边陲的省份—西藏，作为重要能源基地，拥有丰富的能源资源，包括太阳能、风能、水能、地热能等可再生能源以及传统的化石能源。党的二十大报告^[1]提出，“深入推进能源革命”“确保能源安全”，助推高原经济高质量发展，西藏能源开发与利用受到了广泛的关注。随着能源开发与利用的不断深入，涉及能源的数据呈现出多源异构的特性，这些数据分散于不同的部门、系统与监测设备中，数据格式、语义和存储方式都存在着差异。

1 西藏能源知识图谱的作用

知识图谱^[2]由谷歌 Google Knowledge Graph 提出而正式得名，它作为一个宏大的数据模型，能够构建庞大的“知识”网络，包含大量实体、属性以及关系等等，为人们提供一种快速便捷进行知识检索与推理的方法。在能源系统中，知识图谱能够以图结构的形式对海量异构数据进行处理与分析，实现能源领域数据的融合与挖掘；同时知识图谱也可以直观地展示系统中各组件之间的关系，找出数据间的内在联系，实现知识的有效组织与推理应用，从而制定更加合理的能源策略。在西藏这样特殊地理和气候的条件下，知识图谱技术可以提供更精确的数据支持，对西藏能源的合理规划、高效生产、精准配送与科学管理具有极为重要的意义。

2 西藏能源多源异构大数据融合与推理方法

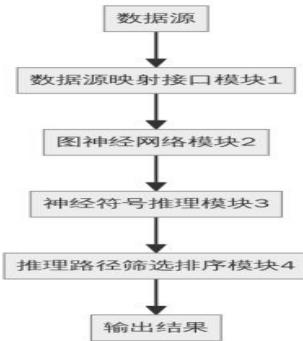


图 1：方法流程图

2.1 数据源

通过网络爬虫或 API 接口等方式获取多个数据源，每个数据源包含不同类型的实体，包括西藏能源^[3]企业生产运营数据（如发电量、开采量等）、能源基础设施监测数据（如电网运行参数、油气管道压力）、政府能源管理部门统计数据（能源存储数据、能源消费数据）、科研机构的能源研究数据（新能源开发技术参数、能源效能评估数据等）以及地理信息数据（能源资源分布地理位置、能源项目周边环境等），这一步骤的关键在于数据源的多样性和全面性。

对所述西藏能源数据源进行数据清洗和预处理，针对数据中的噪声、错误数据（如错误的能源计量数据）和缺失数据进行处理，可使用一些高效的数据清洗算法，如填充缺失值（如统计方法、机器学习方法）、纠正错误值（如哈希表、排序算法）、删除重复数据或者多重填补等方法。

2.2 数据源映射接口

构建多种数据源的映射接口,采集和存储数据源中的各实体,定义所述数据源和所述实体之间的映射关系规则,将不同数据源的数据映射到统一的知识图谱格式,构建知识图谱中节点,本次步骤是实现异构数据融合的关键。

在这一步骤中,我们可以设计灵活的映射规则,使用本体匹配算法^[4]来自动发现不同数据源之间的语义对应关系,能够有效地处理来自不同来源的异构数据,包括结构化数据库数据(如关系型数据库中的数据能源交易记录)、半结构化数据(如XML格式的能源设备配置文件)和非结构化数据(如能源项目报告文档、能源政策文本)等。

2.3 图神经网络

基于图神经网络^[5]的所述知识图谱,将每个节点用一个嵌入向量来表示,这个嵌入向量是在语义空间中的向量表示,它包含了节点在所述知识图谱中的信息。计算节点之间的相似度来表征节点的关联性,基于图神经网络的嵌入学习模型通过训练数据来学习节点的嵌入向量,并使得相似的结构或语义相近的节点具有相似或相近的嵌入向量。基于图神经网络学习各节点的嵌入向量,将图结构的所述知识图谱转化为语义向量。

通过所述嵌入向量和映射接口提取融合节点,对于具有相似节点的多个数据源,根据所述映射规则和所述嵌入向量,将节点进行融合,得到一个综合了多个数据源信息的融合节点,该融合节点中包含了来自多个数据源的信息,包括能源关系、能源地理位置和能源项目报告文档等。这一步骤的核心在于利用图神经网络的强大表示能力,将复杂的知识图谱结构转化为低维的向量表示。例如,我们可以使用GraphSAGE或GAT等先进的图神经网络模型,对包含数百万个节点的大规模能源知识图谱进行高效的表示学习。

2.4 神经符号推理

在神经网络中,将西藏地区能源融合节点的属性信息和文本描述输入到基于符号逻辑的神经网络中,转化为符号逻辑,符号逻辑中包括了节点本身的属性值、节点之间的关系以及约束条件。同时,通过注意力机制,对西藏能源融合节点的文本描述进行嵌入,得到文本特征向量。

把符号逻辑和文本特征向量结合起来进行知识补

全,输入到一个基于符号逻辑的神经网络中进行推理,该网络采用卷积神经网络或循环神经网络模型。然后,该推理网络会输出一个概率分布,表示所述符号逻辑中不同约束条件的可信度。最后,根据所述概率分布,对所述符号逻辑中的约束条件进行筛选,从而生成新的知识三元组。这一步骤结合了符号逻辑的可解释性和神经网络的学习能力,能够有效地推理出能源需求,有助于西藏地区能源企业更好地把握市场动态,制定合理的价格策略。

2.5 推理路径筛选排序

基于能源知识图谱和融合节点,采用深度优先搜索算法遍历能源知识图谱中的各节点和边,生成所有可能的推理路径,其中,所述推理路径是一系列知识三元组的集合。然后,将每一推理路径中的节点转化为嵌入向量,即转化为能源知识图谱的语义空间中的一系列向量,得到一条推理路径的嵌入向量集合。把能源推理路径的嵌入向量集合输入到注意力机制中,计算能源推理路径的相似度,即路径嵌入向量的余弦相似度。这一步骤通过系统性地探索知识图谱的结构,生成可能的推理路径。

基于注意力机制^[6]和路径相似性^[7]进行推理,将所有可能的推理路径根据相似度进行排序,其中相似度最高的推理路径,即路径嵌入向量的余弦相似度最高的推理路径被认为是最优推理路径。根据所述推理路径和推理目标,即查询语句中指定的关系,计算注意力权重,所述注意力权重是根据所述推理路径的不同节点与查询语句中指定的关系的距离、结构相关信息进行计算的,该注意力权重会赋予给不同推理路径,即根据路径不同节点与查询语句中指定的关系的距离、结构相关信息对相似度进行加权。最后,将所述注意力权重与所述相似度相结合,计算所述推理路径的分数,从而筛选并排序所述推理路径,得到最优推理路径。这一步骤通过引入注意力机制,能够更好地捕捉路径中的关键信息,提高推理的准确性。如从能源生产企业到能源消费区域的能源传输路径分析,挖掘能源配送过程中的关键环节和潜在瓶颈,为能源运输优化提供依据。

3 西藏能源多源异构大数据融合与推理方法的预测应用场景

3.1 能源产出与优化调度

通过构建知识图谱,整合西藏地区各类能源产出数

据（如水能、太阳能、风能等）、生产设备状态及生产效率信息。基于知识图谱的数据分析和挖掘技术，预测能源产出趋势，优化能源生产调度策略。能提高能源生产效率，确保能源供应稳定，降低生产成本。

3.2 能源运输与储存管理

利用知识图谱技术，建立涵盖运输路径、物流信息、储存设施等的综合知识图谱。通过图谱分析，预测能源运输需求、优化运输路线，同时监控储存设施状态，预测储存能力。能提升能源运输效率和安全性，优化储存布局，降低储存损耗。

3.3 能源销售与市场预测

整合销售数据、客户信息及市场趋势等多维信息，构建知识图谱。基于图谱分析，预测市场需求变化，制定针对性的销售策略。能提高市场占有率，优化客户关系管理，增强市场竞争力。

3.4 能源利用与能效分析

通过知识图谱技术，整合能源消耗数据、能效指标等信息。预测能源利用趋势，发现能效提升空间，制定节能优化方案。能提高能源利用效率，降低能源消耗成本，促进可持续发展。

3.5 智能电网与智慧城市

在智能电网和智慧城市建设中，利用知识图谱技术融合多源异构数据。预测电力需求、优化电力分配，同时监控城市能源使用情况，提升城市管理效率。实现能源资源的合理分配与高效利用，提升城市管理智能化水平。

4 总结与展望

本文主要探讨的是基于知识图谱的数据融合与推理方法，首先分析了中国西藏能源的现状，解释了知识图谱的相关概念，以及在能源领域所发挥的重要作用。在获取数据源时，采用的是网络爬虫或 API 等接口方式，再构建多种数据源的映射接口，根据所述映射规则和嵌入向量将节点进行融合，利用图神经网络将复杂的知识图谱结构转化为低维的向量表示，结合符号逻辑的可解

释性和神经网络的学习能力，采用深度优先搜索算法遍历能源知识图谱中的各节点和边，生成所有可能的推理路径，最后利用注意力机制和路径相似性进行推理，提高推理的准确性。该方法为西藏能源发展有着积极地推动作用。

然而，该方法仍面临一些挑战，如知识图谱的动态更新机制有待完善、推理效率和准确性在大规模复杂数据场景下仍需提升等。未来研究将重点关注这些问题，进一步优化知识图谱构建与推理技术，结合新兴技术（如区块链技术保障数据安全与可信共享），为西藏能源产业的智能化发展和可持续发展提供更强大的技术支撑。

参考文献

- [1] 以西藏能源高质量发展助推高原经济高质量发展 [J]. 新西藏(汉文版), 2023, (08): 21-24.
 - [2] 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. DOI: 10.19678/j.issn.1000-3428.0061803.
 - [3] 黄素文. 西藏: 联手大企业开发新能源 [N]. 中国能源报, 2009-08-24 (A05).
 - [4] 常万军, 张东方. 基于深度学习的海量本体匹配模型构建研究 [J]. 河南工学院学报, 2024, 32(02): 21-6.
 - [5] 刘雪洋, 李卫军, 刘世侠, 等. 基于图神经网络的知识推理方法研究综述 [J/OL]. 计算机工程与应用, 1-17 [2024-12-26]. <http://kns.cnki.net/kcms/detail/11.2127.TP.20241220.1535.009.html>.
 - [6] 张金柱, 孙雯雯, 仇蒙蒙. 融合异构网络表示学习与注意力机制的引文推荐研究 [J]. 数据分析与知识发现, 2024, 8(10): 14-27.
 - [7] 朱柏霖, 桂韬, 张奇. 基于模态相似性路径的统一多模态实体对齐 [J]. 中文信息学报, 2024, 38(06): 34-44.
- 作者简介: 管奕 (2003.06), 女, 汉族, 湖南衡阳人, 本科, 西藏民族大学, 研究方向数据科学与大数据技术。