

DOI: 10.19783/j.cnki.pspc.200608

基于知识图谱技术的风电数据管理与应用研究

吕梦平¹, 段斌¹, 蒋海辉¹, 邓栋²

(1. 湘潭大学自动化与电子信息学院, 湖南 湘潭 411100; 2. 湘电风能有限公司, 湖南 湘潭 411100)

摘要: 随着我国风电领域的发展, 行业内产生的相关数据在来源和数量上都达到了一个新的高度。这些数据被分散在各个业务系统中, 若能对其进行统一管理并应用, 则对行业具有积极作用。提出了一种用知识图谱管理风电数据的方法。首先, 从多源风电数据中抽取出实体、关系、属性等知识图谱元素。在此过程中, 分别给出了针对结构化与非结构化数据的抽取方法。其次, 将得到的元素在图数据库 Neo4j 中构建出了风电数据全景知识图谱, 实现了不同类型、业务数据间的贯穿统一。最后, 通过故障推理、业务查询等典型案例与传统的基于关系型数据库的数据管理模式作对比, 结果显示提出的方法具有较好的运用效果。

关键词: 风电大数据; 知识图谱; Neo4j 图数据库; 故障推理; 业务查询

Research on management and application of wind power data based on knowledge graph technology

LÜ Mengping¹, DUAN Bin¹, JIANG Haihui¹, DENG Dong²

(1. School of Automation and Electronic Information, Xiangtan University, Xiangtan 411100, China;

2. Xiangdian Wind Energy Co., Ltd., Xiangtan 411100, China)

Abstract: With the development of the wind power industry in China, data from the field has reached a new stage in terms of sources and magnitude. These data are scattered over each business system, and if they can be managed intensively and applied, they will be helpful to industry. A wind power data management method based on a knowledge graph is proposed. First, knowledge graph elements such as entity, relation and attribute are extracted from multi-source wind power data, and the extraction methods for structured and unstructured data are given. Then those elements are used to construct a panoramic knowledge graph of wind power data in the graph database Neo4j to achieve the connection between different types and business data. Finally, typical cases such as fault reasoning, business query, etc. are compared with the traditional relational database-based data management mode. The results show that the proposed method is good in application.

This work is supported by the National Natural Science Foundation of China (No. 61379063).

Key words: wind power data; knowledge graph; Neo4j graphical database; fault reasoning; business query

0 引言

截至 2019 年, 我国并网风机装机容量达 198 GW, 与此同时, 系统中积累的数据规模越来越大、类型越来越丰富, 并且呈指数级增长, 风电大数据时代已经到来。面对如此庞大且丰富的数据, 如果能对其高效利用, 可以为风电数据挖掘和分析提供基础, 从而促进风电领域的发展^[1]。

目前, 风电企业产生的数据主要包括风机运行数据、设备信息、运维工单和运营数据等^[2], 它们

因格式、用途等原因被竖井式的、孤立地存储在各个业务部门中, 最终形成了信息孤岛^[3]。所以这些数据包含的信息往往具有隐蔽性、分散性和异构性等特征^[4], 传统的基于关系型数据库的数据管理平台无法对这些多源异构数据进行分析。文献[5-7]提出用基于 Hadoop 框架的 HDFS 和 MapReduce 技术对风电数据进行分析, 虽然这种方法一定程度上提高了数据计算能力, 但它仍然是对各类数据进行分开存储管理的, 在数据分析时局限于单源数据, 没有考虑多源数据间的隐含关系。而在大数据环境下, 要想真正发挥数据的潜能, 需要将不同来源、不同结构的数据有机地统一结合起来, 做到跨

基金项目: 国家自然科学基金项目资助(61379063)

领域、跨时间、跨类型的全景数据分析，最终找出数据中隐藏的价值信息^[8]。

为充分挖掘出风电数据中的价值，更好地服务风电企业，本文提出了一种基于知识图谱技术的风电数据管理方法。知识图谱作为一门新兴技术，具有模型简单高效、便于构建和维护、对数据的描述能力强等特点。它已经在电网客服系统^[9]、智能调控^[10]和低压配电网^[11]等领域广泛使用。本文根据风电数据的结构特点，针对结构化与非结构化数据分别给出了实体、关系、属性等知识图谱元素的抽取方法。然后基于图形数据库 Neo4j 构建出了风电数据全景知识图谱，并通过实际案例与传统的关系型数据库存储方式作了对比。

1 风电数据的应用案例

随着风电技术的发展，风电数据发挥着越来越重要的作用，各国都积极开展风电大数据的研究。丹麦维斯塔斯风力技术集团通过分析大量的气象报告、潮汐相位、地理空间、卫星图像等数据，优化了风力涡轮机的布局，提高了风力发电效率^[12]；IBM 基于成熟的数值天气模型，加入大气动力过程和物理过程，提出了精确的风电场微观选址方案，可以有效解决风资源最大化利用问题和风机维护成本最小化问题^[13]；文献[14]在海量风电数据环境下，结合 SCADA 系统的实时监控数据和风场实时风速建立了风电机组齿轮箱故障诊断与预警模型，实验结果表明，该模型在精度和加速比方面比传统的方法具有很好的效果；文献[15]基于风速、风电出力、风电机组地理位置信息等数据，考虑风电场间时空相关性的影响因素，采用非序贯蒙特卡洛法对风电机组容量可信度进行计算，其结果更加接近实际情况。

根据上面分析可知，正如引言所述，风电数据的运用是结合多种数据将分散在其中的信息连贯起来，找到其中规律实现目标优化，所以建立一个统一的数据管理平台是十分必要的。而知识图谱得益于本身模型结构的优势，可以把多源异构数据灵活地组织到一起^[16]。

2 风电数据知识图谱自动构建

2.1 风电数据知识图谱结构

知识图谱本质上是一个知识库，由节点、关系、属性等元素组成。在知识图谱中，节点表示实体，关系作为有向边将实体连接起来，最终形成“实体-关系-实体”的图模型结构。其中，实体和关系都可以具有属性，作为对实体和关系的具体描述^[17]。风电数据知识图谱结构如图 1 所示，图中以风机的齿轮箱为例，不仅记录了齿轮箱包含的部件以及部件厂家信息，还有对应的检修工单、试验单、发生过的故障和维修人员等全方位的信息。这种结构模式的特点是可以有效地表示风电数据间的各种联系，方便贯穿企业内的多源数据以构成一个完整的全景数据库，利于管理。

2.2 风电数据知识图谱自动构建过程

风电数据可以分为两类：结构化数据与非结构化数据。结构化数据主要指风机 SCADA 系统中的设备运行数据和故障数据，以表格为主，是系统中的主要数据源。这类数据的知识图谱构建比较简单，只需将表格的行作为知识图谱的实体节点名称，列作为关系，单元格中的值作为属性值。非结构化数据主要指一些文本类的数据，如运维工单、检修记录等。对于这类数据的处理相对复杂，主要分为以下步骤：(a) 实体/属性抽取；(b) 共指消解；(c) 关系抽取；(d) 数据融合。具体构建过程如图 2 所示。

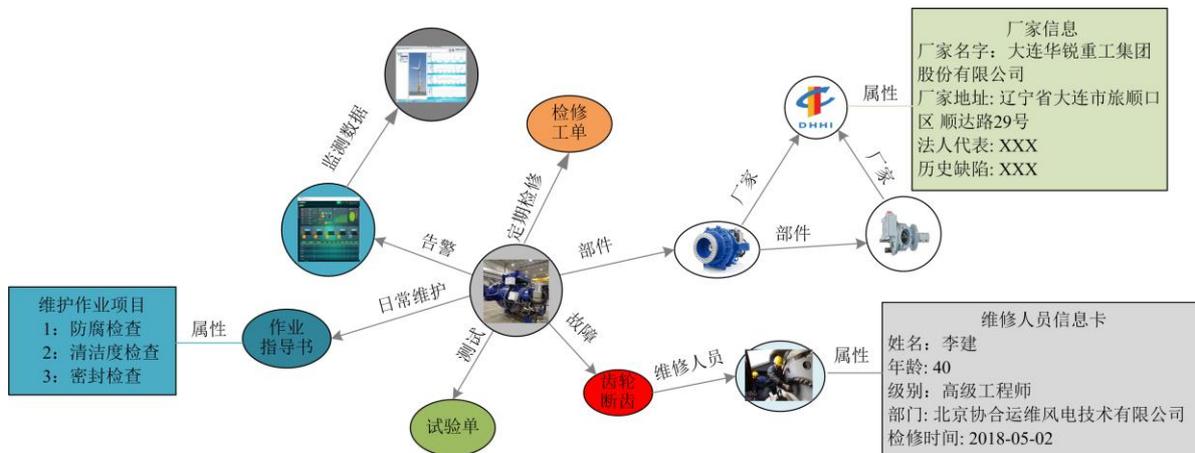


图 1 风电数据知识图谱结构

Fig. 1 Structure of wind power data knowledge graph

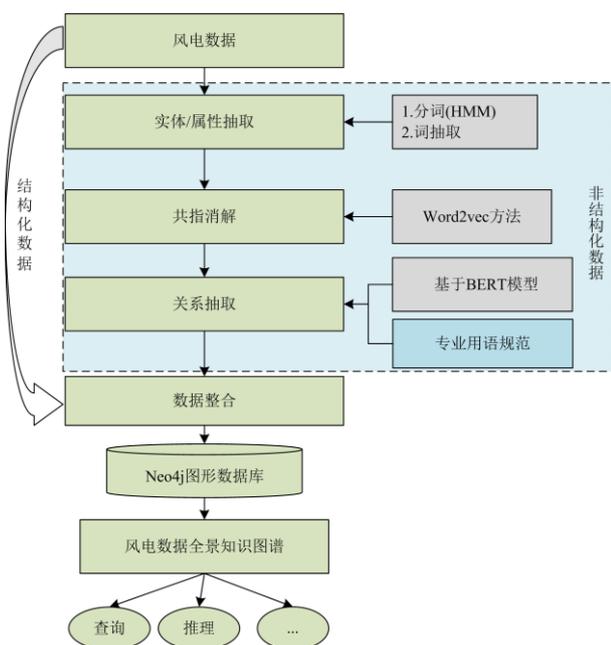


图2 风电数据知识图谱构建过程

Fig. 2 Construction process of wind power data knowledge graph

2.3 实体/属性抽取

实体/属性抽取是从非结构化风电数据中抽取出来表示实体/属性的词^[18], 比如现在有一条风机设备故障语录文本: “风机 P1#的齿轮箱油温过高, 温度达到 102°”。其中“风机 P1#”、“齿轮箱”和“油温过高”是实体, “温度”是属性名, “102°”则是属性值, 需要从文本中把这些词抽取出来, 具体步骤如下:

1) 分词: 将风电数据文本切割为单独的词, 如上述的“风机 P1#”、“齿轮箱”等。采用隐马尔可夫模型(HMM)^[19]对其分词。根据 HMM 原理, 在分词过程中可以认为每个字(观测值)有一个对应的状态(状态值), 状态集合用 $G=\{B, M, E, S\}$ 表示。其中, B: 一个词的开始; M: 一个词的中间; E: 一个词的结束; S: 单字成词。例如故障文本:

风机 P1# 的 齿 轮 箱 油 温 过 高 , 温 度 达 到 102°
B E S S B M E B E B E S B E B E S

可以根据这句话的状态序列将句子切分为: 风机/P1#的/齿轮箱/油温/过高/, /温度/达到/102°。但实际情况下“风机 P1#”和“油温过高”应该属于单个词, 不应被分开, 说明 HMM 是基于常用词语切分, 不具有领域知识, 所以导入风电领域专业术语集进行辅助分词可以提高分词准确率。最终句子被切分为: 风机 P1#的/齿轮箱/油温过高/, /温度/达到/102°。

2) 词抽取: 将上一步中切分好的词逐一在领域专业术语集中进行检索, 若能检索到匹配项, 则抽

取该词作为知识图谱的实体/属性节点^[18]。

2.4 共指消解

共指消解^[20]是将风电数据文本中具有相同含义但表述不同的实体/属性进行合并。风电数据文本以自然语言的形式记录, 由于个人习惯、工作经验和圈内流行用语等因素, 运维人员在记录时不会严格按照规范书写, 所以针对同一实体/属性经常会出现不同的名称, 如“齿轮箱油冷系统失效”和“齿轮箱冷却系统失效”。这其中“油冷系统”和“冷却系统”属于同一实体, 需要在知识图谱中合并为一个节点。共指消解的步骤如下。

(1) 根据词性分类: 在风电数据文本中, 如果两个词属于同一个实体/属性, 那么它们的词性一定相同。因此可以将文本中表示实体/属性的词按词性分类成不同的词性集合(对于词性的标注在上节分词步骤中用 Python 中的 jieba.posseg 函数可以顺带完成), 然后对每个词性集合进行共指消解。

(2) 余弦相似度计算: 首先使用 word2vec 算法^[21]将每个词性集合中的词向量化, 向量维度选择 100 维, 得到词向量 $V_{word} = (v_1, v_2, v_3, \dots, v_{100})$, 然后通过

$$\cos \theta = \frac{x_1 \cdot y_1 + x_2 \cdot y_2 + x_3 \cdot y_3 + \dots + x_{100} \cdot y_{100}}{\sqrt{x_1^2 + x_2^2 + x_3^2 + \dots + x_{100}^2} \cdot \sqrt{y_1^2 + y_2^2 + y_3^2 + \dots + y_{100}^2}}$$

分别计算词向量之间的余弦值。根据余弦定理可知, 余弦值越大, 两个词向量间的夹角越小, 从而判定两个词越相似, 当余弦值达到某个阈值时, 就认为两个词相同。在本文中, 经过大量数据验证发现当 $\cos \theta \geq 0.86$ 时, 基本可以认为两个词属于同一实体/属性。

(3) 形成同义词集^[18]: 根据上一步的余弦相似度计算可以找出表示同一实体/属性的词, 将这些词整理成集合, 从而形成若干个同义词集。然后, 基于风电领域专业术语集在每个同义词集中选择标准化的词作为知识图谱的实体/属性。

2.5 关系抽取

关系抽取^[22]是从风电数据文本中判别实体之间是否存在关系以及关系类型。由于风电领域属于特定领域, 关系类型数量有限, 文本格式也有一定的规范, 所以采用基于 BERT 模型抽取。BERT 是一种经过预训练的语言表征模型, 与传统的单向语言模型不同, 它嵌入了新的掩码语言模型 (MLM), 能够实现融合左右上下文信息的深层双向语言表征^[23]。在新的训练任务下, 只需少量的标注样本模型参数就可以快速调优。在本例中, 通过 5 轮的训练后模型便有了很强的泛化能力, 基于 BERT 模型的关系抽取流程如图 3 所示。

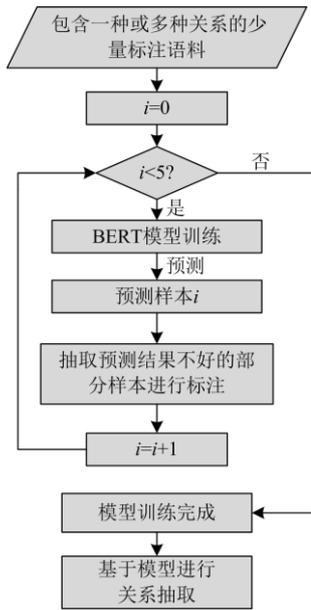


图3 基于BERT模型的关系抽取流程

Fig. 3 Relationship extraction process based on BERT model

3 案例分析

使用了某风电公司的 0.6TB 数据，其中 0.2TB 的结构化数据来自风机 SCADA 系统、定检记录等。0.4TB 的非结构化数据来自风机故障记录、维修工单和故障处理规程等。这些数据通过图 2 的步骤处理后，在图形数据库 Neo4j 中构建出的风电数据知识图谱如图 4 所示。由于图谱过大，这里只展示部分。

3.1 故障推理

基于风电数据知识图谱可以实现风机部分故障的辅助推理。风机故障主要分为单一性故障和综合性故障两种，单一性故障是指故障原因是确定的，对于这类故障知识图谱可以直接给出故障原因及处理措施。以主变频器 MCB 故障为例，通过检索知识图谱得出结果如图 5 所示。

知识图谱的优势真正体现在对综合性故障的诊断上，综合性故障的原因复杂多样，通常需要利用多维度的数据进行分析，是一个典型的知识、信息

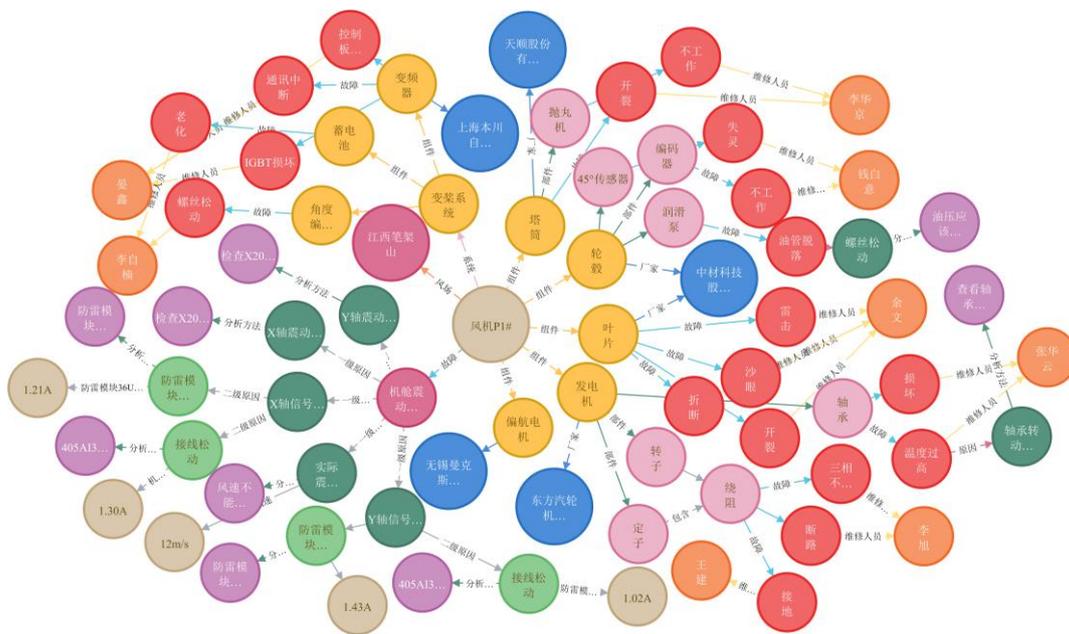


图4 风电数据知识图谱(部分)

Fig. 4 Knowledge graph of wind power data (part)



图5 主变频器 MCB 故障树

Fig. 5 Fault tree for main transformer MCB

所以图谱中包含大量有关故障名称、故障频率、故障原因和分析方法等信息，积累了丰富的专家经验，是一棵天然的故障树。以机舱震动偏移故障为例，基于知识图谱检索返回的故障树如图 6 所示。

图 6 中的故障树中还包含故障发生时传感器传回的相关具体参数，结合故障树给出的分析方法，工程师可以快速确定故障原因。基于知识图谱的机舱震动偏移故障推理流程如图 7 所示。

密集型任务。上文提到，风电数据知识图谱整合了历史故障记录、维修工单和故障处理手册等数据，

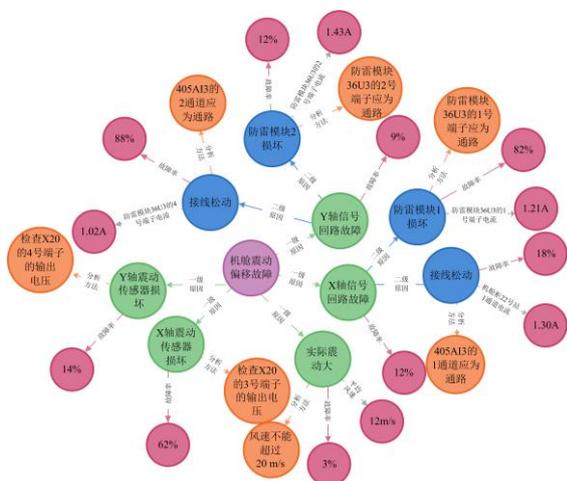


图 6 机舱震动偏移故障树

Fig. 6 Fault tree for engine room vibration offset

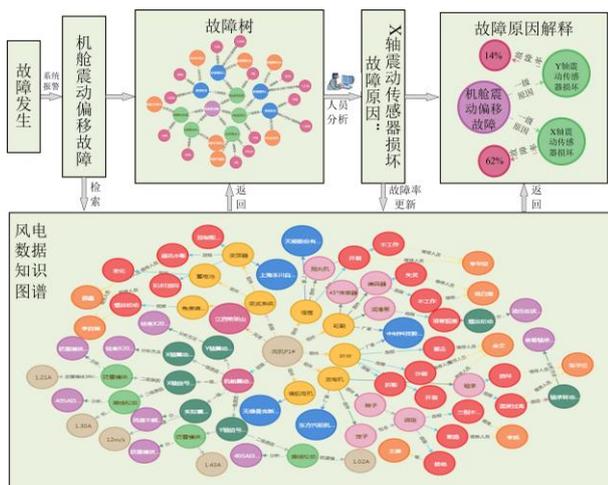


图 7 基于知识图谱的机舱震动偏移故障推理流程

Fig. 7 Fault reasoning process of engine room vibration offset based on knowledge graph

需要说明的是, 有些故障原因是没有传感器数据提供判断的(如图 6 中的 Y 轴震动传感器损坏、X 轴震动传感器损坏), 这时故障率的作用就体现出来了, 如果部分原因已通过传感器数据排除, 剩下的原因可以参考历史故障率来判断。如图 7 中基于故障树分析得出的故障原因是 X 轴震动传感器损坏, 原因解释是根据传感器数据可以排除 X、Y 轴信号检查回路故障和实际震动过大故障, 对于 X 轴或 Y 轴震动传感器损坏没有直接数据判断, 但是参考历史记录统计的故障率发现 X 轴震动传感器损坏的故障率(62%)远大于 X 轴震动传感器损坏的故障率(14%), 所以基本可以判定是 X 轴震动传感器损坏。另外, 此次故障发生之后故障率得到更新。

由上述可知, 利用风电数据知识图谱对故障进

行分析的优点在于一方面可以很大程度上摆脱对人员知识、经验的依赖, 提高故障诊断的准确率, 另一方面可以缩短运维周期, 减少风机停机时间, 提高发电量。

然而, 整个过程如果基于关系型数据库来分析故障是很难完成的, 原因有两点: (1) 关系型数据库的数据展示没有知识图谱那么直观, 不便于分析。(2) 关系型数据库中各类型的数据是分散存储的, 多源数据间没有关联, 更没有故障树提供参考, 工程师进行故障分析时完全靠经验去调取相关数据, 很容易忽略潜在信息(如故障率)导致分析结果不准确。

3.2 业务查询

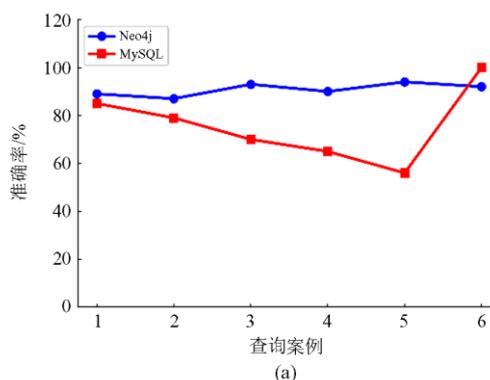
设计了风电业务中 6 个典型的案例模板, 分别基于知识图谱的方法和基于 MySQL 关系型数据库的方法对其查询, 通过准确率、召回率和查询时间等方面对两种方法进行比较。其中, 准确率和召回率的值越大, 说明查询结果越好。案例模板如表 1 所示, 每个案例通过修改变量查询 20 次, 查询结果取平均值。基于知识图谱的查询使用 Neo4j 自带的 cypher 语言, 而传统的关系数据库使用基于关键词的 SQL 语句进行查询。

表 1 搜索模板

Table 1 Search template

序号	搜索条件
1	风机 XXX 发生故障的次数
2	风机 XXX 在 2019 年 5 月到 6 月内发生偏航压力低故障的具体时间段
3	风机数量超过 XXX 个且海拔超过 2 km 的风场
4	风机 XXX 发电机的定子制造商还生产哪些风机部件在 XXX 年到 XXX 年内, 故障率最高的风机部件以及相应的维修工程师
5	
6	XXX 风场年龄较老的风机

通过对表 1 中的案例查询结果如图 8(a)、图 8(b)、图 8(c)所示。



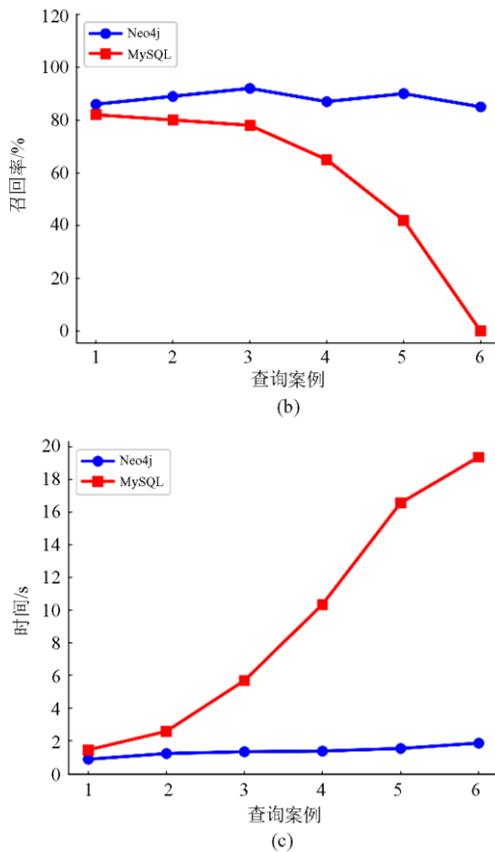


图 8 查询结果指标对比图

Fig. 8 Comparison chart of query result indicators

由图 8 可看出，在准确率、召回率和查询时间等方面，基于知识图谱技术的查询结果均比基于关键词的查询结果优越，且基于知识图谱技术的查询结果波动不大，而基于关系数据库的查询结果却越来越差，主要是因为 6 个案例中的约束条件越来越多(查询深度增加)，在关系型数据库中查询时需要大量的表连接(JOIN 操作)，多表之间关系断连的概率大大增加，所以准确率和召回率下降，而且此过程急剧消耗硬件资源，导致查询时间变长。而基于知识图谱技术的查询是通过子图匹配的方式查询，即通过将查询案例描述成子图谱去图数据库中匹配，它不受查询深度和数据规模的影响，所以准确率和召回率较高且查询时间短。对于案例 6，基于关系型数据库查询准确率为 100%，召回率为 0，原因是 SQL 语句无法对其进行合理的描述，关系型数据库无法理解“年龄较老”的意思，导致返回结果为空。而基于知识图谱技术的查询结果依旧较好，说明知识图谱在语义理解方面具有明显优势。

4 结论

文章提出了一种以知识图谱形式存储风电数据

的新思路，结合风电数据特点给出了知识图谱的具体构建方法，并通过实例对比验证了本文提出的风电数据管理方法在运用层面具有显著的优势，能够更好地满足企业业务需求。

然而，本文的工作还处于初步阶段，目前本文的风电知识图谱存在的问题有：(1) 在共指消解、关系抽取方面需要更多的标注样本来验证和训练。(2) 知识图谱规模不够大，很多故障的故障树并不完整，后期需要加入更多更广的业务数据来改善。(3) 高质量的知识图谱对其实际应用具有巨大影响，目前没有考虑图谱的质量控制。后期需要采用先进算法对知识图谱质量进行评估，以更好地为应用提供保障。

参考文献

- [1] 范强, 文贤旭, 林呈辉, 等. 大数据技术在风电领域应用研究[J]. 电力大数据, 2017, 20(9): 55-58.
FAN Qiang, WEN Xiankui, LIN Chenghui, et al. Application and research of big data technology in wind power field[J]. Power System and Big Data, 2017, 20(9): 55-58.
- [2] VALIZADEHHAGHI H, LOTFIFARD S, QU Z. A multivariate predictive model of wind power data for robust optimization of energy storage[J]. IEEE Transactions on Industrial Informatics, 2016, 12(4): 1350-1360.
- [3] 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.
MA Yisong, WU Zhigang. Modeling and analysis of big data for power grid based on Neo4j[J]. Advanced Technology of Electrical Engineering and Energy, 2016, 35(2): 24-30.
- [4] 张兴. 风电运营大数据关键技术[J]. 电子技术与软件工程, 2019(10): 154-156.
ZHANG Xing. Key technologies of big data in wind power operation[J]. Electronic Technology & Software Engineering, 2019(10): 154-156.
- [5] 王寅生, 朱志成, 王其台, 等. 风电大数据系统搭建策略的研究[J]. 风力发电, 2017(1): 8-11.
WANG Yinsheng, ZHU Zhicheng, WANG Qitai, et al. Wind big data system based on cloud platform to build strategy research[J]. Wind Power, 2017(1): 8-11.
- [6] WANG H, ZHAO S, ZHAO H, et al. Research on data processing for condition monitoring of wind turbine based on Hadoop platform[C] // 2017 IEEE International Conference on Mechatronics and Automation (ICMA), August 6-9, 2017, Takamatsu, Japan: 322-326.
- [7] 肖凯, 刘鹏, 车建峰, 等. 基于电力大数据平台的海量风电数据处理架构与应用研究[J]. 电力信息化, 2017, 15(7): 13-19.

- XIAO Kai, LIU Peng, CHE Jianfeng, et al. Research on the architecture and application of massive wind power data processing based on electric power big data platform[J]. *Electric Power Information and Communication Technology*, 2017, 15(7): 13-19.
- [8] SABITHA A S, PUNHANI R. Identification of potential regions for wind power development using data mining[C] // 2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon), February 14-16, 2019, Faridabad, India: 306-313.
- [9] 田晓, 刘勇超, 王婧, 等. 电网公司客户服务知识图谱构建的应用价值[J]. *山东电力技术*, 2015, 42(12): 65-67, 80.
- TIAN Xiao, LIU Yongchao, WANG Jing, et al. Application value of building knowledge graph system for hotline customer service in State Grid Corporation[J]. *Shandong Electric Power*, 2015, 42(12): 65-67, 80.
- [10] 余建明, 王小海, 张越, 等. 面向智能调控领域的知识图谱构建与应用[J]. *电力系统保护与控制*, 2020, 48(3): 29-35.
- YU Jianming, WANG Xiaohai, ZHANG Yue, et al. Construction and application of knowledge graph for intelligent dispatching and control[J]. *Power System Protection and Control*, 2020, 48(3): 29-35.
- [11] 高泽璞, 赵云, 余伊兰, 等. 基于知识图谱的低压配电网拓扑结构辨识方法[J]. *电力系统保护与控制*, 2020, 48(2): 34-43.
- GAO Zepu, ZHAO Yun, YU Yilan, et al. Low-voltage distribution network topology identification method based on knowledge graph[J]. *Power System Protection and Control*, 2020, 48(2): 34-43.
- [12] 陈华军. 大数据变革—南方电网大数据应用探索[J]. *中国电力企业管理*, 2015(17): 25-28.
- CHEN Huajun. Big data change—big data application of southern power grid[J]. *China Power Enterprise Management*, 2015(17): 25-28.
- [13] 彭小圣, 邓迪元, 程时杰, 等. 面向智能电网应用的电力大数据关键技术[J]. *中国电机工程学报*, 2015, 35(3): 503-511.
- PENG Xiaosheng, DENG Diyuan, CHENG Shijie, et al. Key technologies of electric power big data and its application prospects in smart grid[J]. *Proceedings of the CSEE*, 2015, 35(3): 503-511.
- [14] 张少敏, 毛冬, 王保义. 大数据处理技术在风电机组齿轮箱故障诊断与预警中的应用[J]. *电力系统自动化*, 2016, 40(14): 129-134.
- ZHANG Shaomin, MAO Dong, WANG Baoyi. Application of big data processing technology in fault diagnosis and early warning of wind turbine gearbox[J]. *Automation of Electric Power Systems*, 2016, 40(14): 129-134.
- [15] 候卫萍, 蔺红. 基于大数据技术风电机组容量可信度计算[J]. *电测与仪表*, 2020, 57(14): 39-44.
- HOU Weiping, LIN Hong. Wind turbine capacity reliability calculation based on big data technology[J]. *Electrical Measurement & Instrumentation*, 2020, 57(14): 39-44.
- [16] 李雪. 基于智能交互的物体识别增量学习研究[D]. 青岛: 山东科技大学, 2017.
- LI Xue. Study on hybrid incremental learning based on intelligent human-computer interaction systems[D]. Qingdao: Shandong University of Science and Technology, 2017.
- [17] FU L, BAI Y, ZHONG Z. Constructing a vertical knowledge graph for non-traditional machining industry[C] // 2018 IEEE 15th International Conference on Networking, Sensing and Control (ICNSC), March 27-29, 2018, Zhuhai, China.
- [18] 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.
- [19] HUANG W, ZHANG N, KANG C, et al. From demand response to integrated demand response: review and prospect of research and application[J]. *Protection and Control of Modern Power Systems*, 2019, 4(2): 148-150. DOI: 10.1186/s41601-019-0126-4.
- [20] 刘岍, 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.
- [21] JATNIKA D, BIJAKSANA M A, SURYANI A A. Word2Vec model analysis for semantic similarities in English words[J]. *Procedia Computer Science*, 2019, 157: 160-167.
- [22] YU H, LI H, MAO D, et al. A relationship extraction method for domain knowledge graph construction[J]. *World Wide Web: Internet and Web Information Systems*, 2020, 23(2): 735-753.
- [23] YIN X, HUANG Y, ZHOU B, et al. Deep entity linking via eliminating semantic ambiguity with BERT[J]. *IEEE Access*, 2019, 7: 169434-169445.

收稿日期: 2020-05-31; 修回日期: 2020-07-27

作者简介:

吕梦平(1996—), 男, 通信作者, 硕士研究生, 研究方向为知识图谱技术应用; E-mail: 2543654969@qq.com

段斌(1966—), 男, 博士, 研究方向为电力系统自动化, 工程教育。

(编辑 葛艳娜)