

[故障处理]

发电设备故障知识图谱可视化构建与智能化应用研究

文/李丰均¹ 徐云柯¹ 彭元苏¹ 顾宝² 顾俊锴³

1. 浙江浙能绍兴滨海热电有限责任公司; 2. 浙江浙能数字科技有限公司; 3. 杭州电子科技大学

为提高发电设备的维护与运行效率，研究基于发电设备故障数据，探究故障知识抽取的方法以构建知识图谱，并实现设备故障智能查询、故障解决方案生成等应用，同时支持发电设备故障知识图谱的动态更新，从而有效提高发电设备故障处理的效率，增强电厂的产能，提高其智能化程度。

随着中国电力行业的迅速发展，电厂规模逐渐扩大，电力设备规模不断增大，其结构也呈现出复杂化的趋势。尤其是火力发电厂作为电力供应的主要来源，在保障稳定且高质量供电方面具有重要地位。然而，在这一领域，设备故障的快速处理对于维持高效运行至关重要。在实际处理故障时，常常需要检修员依赖于结合实际经验与专业知识的方式，根据故障现象迅速锁定问题设备，并进行设备故障排查，从而使得设备能够快速恢复正常运行。然而，上述方法的有效性性与检修员的专业素养密切相关，包括其专业知识储备、逻辑分析能力以及信息处理能力，这些因素直接影响着设备维护的成果。检修员需要集中管理和处理各种不同类型的故障数据以及维修规程日志，进而将这些数据整合成一个故障-设备关系数据库，用于辅助故障定位。这个数据库能够帮助检修员快速搜索相关信息，提供可能的故障原因和解决方案。但在生产过程中，设备产生的故障数据和现有的维修规程数据往往种类繁多、数量庞大，从而使得对于这些多样化且异构的故障数据进行集中管理和快速搜索成为一项挑战。

知识图谱在应对电力行业的发展需求方面具有

突出优越性，尤其在整理和归纳海量数据方面表现出显著优势，从而在信息管理、数据分析和智能应用领域得到了广泛的关注和应用。知识图谱作为一种结构化的知识表示方式，能够将领域内的知识、概念和关系进行有效的组织和连接，为问题求解和决策制定提供有力支持。在电力领域，构建电厂设备故障知识图谱能够将设备、故障和解决方案等关键信息进行关联，为电厂故障管理和运维提供全面而深入的支持。通过构建故障数据知识库，知识图谱能够显著提升故障数据的检索效能，使之更为精准。此外，对于知识管理和语义搜索等诸多领域的需求，知识图谱能够与发电设备的文本语言解析相关技术有机融合，将发电设备的非结构化数据，如维修规程和工作手册等，有序整合至知识图谱中，从而推动面向电力设备维护领域的故障智能搜索等应用。然而，在当前的背景下，电厂领域迫切需要解决多源异构故障数据集中管理和高效检索的难题。为达成这一目标，亟需运用先进技术手段，如知识图谱和数据挖掘技术，建构一个综合性系统，以融合、加工和分析故障相关数据，从而极大提升故障处理效率与准确性。

相关工作

1. 电厂领域知识图谱构建

在电厂领域，构建知识图谱的动机源自于处理和应用海量复杂数据的需求，而传统的数据管理和分析方法已无法满足高效处理和智能应用的要求。因此，创建电厂领域的知识图谱成为一项创新性的解决方法。在电厂复杂的背景下，知识图谱的构建首先在关键设备故障查询方面得到应用。肖发龙等^[1]采用知识图谱技术，提出关联电厂设备多模态数据的故障处理方案，取得在电网日常工作、修护和管理方面的显著成果。熊奥等^[2]基于知识图谱技术，整合核电设备故障知识，解决电厂设备维修数据孤岛的问题。知识图谱的构建不仅可以帮助专家用户查询故障设备和快速定位故障源，还能协助管理者迅速响应故障设备，并构建电力设备故障知识库，实现对知识的有效管理和应用^[3]。张虹等^[4]提出一种基于反向等量配对调整法的断面越限处置策略计算方法，利用知识图谱实现交直流大电网断面越限处置策略的快速生成流程。同时，吴勤浩等^[5]利用知识图谱、语义库和相似性分析等新技术，通过对物资属性和使用历史进行相似性分析，建立管理规范，显著提高重码的识别率和可信度。然而，针对实际电厂领域的研究需要超越仅仅针对故障识别或后续管理。面对电厂数据的多模态、复杂性以及实时更新需求，如何实现知识图谱的动态更新，并关注后续功能应用，仍然是一个亟需解决的问题。

2. 电厂领域高级应用

在不断变化的电力市场和技术环境中，为满足能源需求、环境保护和运营效率提升的挑战，电厂必须采用前沿技术和方法，以实现能源的高效利用与可持续发展，从而推动电厂领域的高级应用不断提升。当前，电厂应用领域不仅借助自动化检测技术实现生产数据的实时共享^[6]及故障检测^[7]，还着眼于智能电厂构建，其中蕴含着数字孪生^[8]、一体化大数据平台^[9]等重要内容。电厂实际需求更趋向于针对关键设备的智能监测，以确保电力生产过程的安全性和效率。为此，本文基于电厂关键设备故障知识图谱构建，实现了故障设备的智能检索、故障解决方案的智能生成以及专家知识的有效存储等高级应用，以交互式分析的方式，有效地应对发电故障并保持知识图谱的动态更新。

需求分析

1. 数据描述

在日常的生产中，发电设备总是会发生各种各样的故障。在检修员解决故障之后，会将故障数据以图像、表格和文本等格式来保存，并且大部分发电设备的维修规程也是一种PDF格式文件，它也是一套以故障数据组成的数据集。上述两种情况是故障数据的两大来源。故障知识的抽取主要目的是从原始数据中抽取故障设备、故障现象、故障原因和解决方法四个类别的数据，同时确定上述数据间的关系，从而为构建知识图谱的节点和关系提供数据基础。

2. 需求分析

在与电力生产领域专家进行多次交流之后，了解到面向电厂设备故障复杂多样的问题，主要面临日志数据多源异构、设备关系复杂、故障类型错综复杂、故障事件难以识别、自动标注准确低以及手工标注费时费力等关键性问题，进而导致在构建电厂设备故障知识图谱时会出现实体抽取效率低、关系构建不明确和知识推理准确率低等挑战。为此，本文将开展基于深度学习的语料标注研究，构建多类型语料数据半自动化标注工具，实现实体、关系的快速准确抽取，构建知识图谱，进而推进面向电厂领域的高级应用，并提炼出如下需求。

1) 多类型日志数据的文字自动识别。由于电厂故障数据过于庞大，且人为读写过于复杂繁琐，为此运用OCR技术将设备故障图像信息转化为文本，提供给维修师傅进行故障检测分析，以大大提升其修理效率。

2) 面向多类型数据的聚类与可视化。目的在于获取具有代表性的文本数据，降低重复标注相似语料数据的可能性，减少手工标注数据耗费的人力物力，提高实体抽取和关系抽取模型的泛化能力。

3) 实体识别与关系识别。实体抽取与关系抽取是构建发电设备故障知识图谱的关键环节，通过知识抽取将原始的非结构化文本数据转换为结构化数据，其准确率决定了图谱构建的质量。

4) 知识图谱构建及高级应用。经过知识融合与知识加工后，构建知识图谱网络，并进一步推进基于故障知识图谱的智能检索、故障解决方案生成和专家知识入库等高级应用，满足对知识管理、语义

检索和数据分析等方面的需求。

文本数据结构化处理

1. 文字提取

故障知识数据呈现多样且复杂的特征，同时涵盖多种数据格式。为了实现对这些数据的有效处理，本研究采用了程序辅助提取和光学字符识别（OCR）文本识别，并结合人工校验的策略，将故障知识数据统一转化为非结构化文本数据类型，如图1所示。对于EXCEL表格文件，采用程序直接读取的方式将其导入系统数据库。而对于PDF和图像等难以被程序直接解读的数据格式，主要依赖光学字符OCR进行文本识别，将其转化为文本数据，并在后续进行人工校验，以确保文本数据的准确性。文本检测阶段采用诸如二值化、连通域分析和投影分析等技术，结合机器统计学习方法，以提升识别准确度。

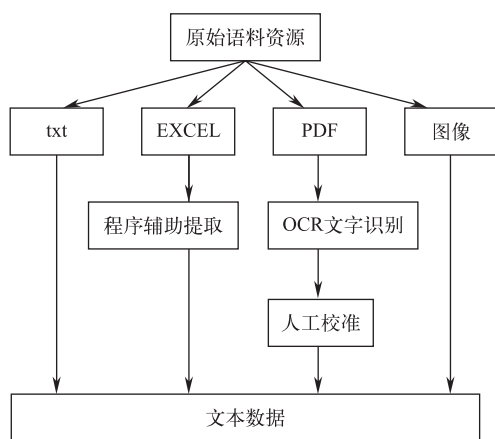


图1 文本的预处理过程

由于相机和扫描仪捕捉的图片无法直接应用于计算辅助翻译（CAT）工具，而手动输入方式又繁琐，因此在面向智能电厂业务场景中，构建适用的文本转换工具具备重要意义。对于PDF和图像数据，本研究依托OCR为基础的文本识别转换工具，并结合人工校验，旨在实现非文本格式数据的快速转换。这一方法不仅在转化速度上具备便捷性，更通过人工校验进一步提升了文本数据的准确性。

2. 语料预处理

由于文本数据的种类和数量繁多，且其分类存在不足、描述方式异质、关系混杂，导致语料标注的难度巨大。为应对这一问题，本研究首先运用深

度学习中的Doc2Vec算法来计算文档集内所有文本之间的语义相似度。通过这种方式，将引入的语料数据映射到向量空间中，并可将其用于采样人工标注的语料。这种向量化的语料数据有助于更便捷地利用程序进行建模和分析。

为确保模型训练的可靠性以及实体输出的精确性，所输入模型的文本数据需要经过人工标注。本研究采用人机交互式的语料标注方式，以提高标注的效率和准确性。经过标注的实体语料将被储存在语料库中，为训练结果的准确性校验提供了标准支持。

在对语料进行向量化后，采用K-means算法，这是业界认可度较高的聚类算法之一，因其简单原理和强可解释性而得到广泛应用。通过K-means算法，将语料数据进行聚类，以满足在采样时的均匀性要求。这种方式有利于提高模型训练的效率，同时减少了对无效类别的重复训练，提高实体抽取和关系抽取模型的泛化能力。

3. 知识抽取

由于从文字中提取出的文本数据缺乏明显的结构性，无法直接用于构建电厂设备故障知识图谱，因此本研究需要借助电厂相关的知识抽取技术，将这些文本数据转化为结构化形式，以便为知识图谱的构建做好准备。

研究实体抽取和关系发现技术可以追溯到20世纪60年代。其核心任务是从包含预定义语义类型的语料文本中自动识别命名实体，并为其确定相应的类别，如设备名称、部件名称、故障名称和故障原因等。这个过程涵盖了两个主要方面，即命名实体的识别和分类。在本研究中，通过人机交互式语料标注的辅助方式，将非结构化文本数据转化为结构化形式，从而实现了知识的抽取，如图2所示。这一过程为构建知识图谱提供了重要的基础。

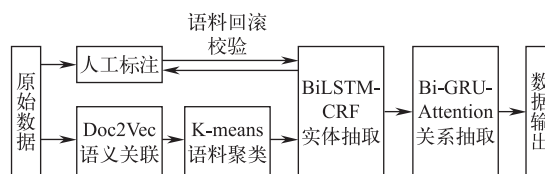


图2 知识抽取流程图

（1）基于条件随机场算法的双向长短期记忆网络模型训练（BiLSTM-CRF）

BiLSTM算法可以从双向来捕获上下文的信息，还能从文本中挖掘隐藏特征，所以用于知识抽取有绝佳的效果反馈。将标注好的数据作为训练集输入BiLSTM-CRF深度学习模型训练。首先使用双向LSTM循环神经网络进行标注语料接收和处理，然后通过CRF输出契合标注转移约束条件的、相似度最高的预测标注序列。在完成一次训练迭代之后将训练结果输入到语料库，进行精确度匹配。由此不断循环迭代，训练出精确度满足需求的模型。

最终基于该模型来预判待标注的语料标签，并抽样部分进行人工语料标记检测，保证精确性后获取实体命名标记。

(2) 基于注意力机制的双向门控循环神经网络结构模型 (Bi-GRU-Attention)

Bi-GRU-Attention模型在Bi-GRU的基础上增加关系词注意力层。首先，该模型中的GRU结构利用Dord2Vec模型对输入的实体数据语料进行向量化，获得词向量表示。然后，利用Bi-GRU层学习和计算词向量信息，得到包含词向量上下文语义信息和位置向量的新词向量信息。接着，利用注意力机制，通过计算特征出现的概率，对关系特征进行加权增强，进一步提高测试模型的有效程度。最后，通过全连接层得到句子的特征向量，最终输入到Softmax分类器获取关系分类结果。

最终基于该模型来预测实体之间的关系标签，并抽样部分进行人工语料标记检测，保证精确性后获取关系命名标记。

4. 设备树构建

本项目的目标为发电厂，其包含的设备复杂，层级较多，但又关系紧密。设备树 (Device Tree) 在知识图谱的构建过程中起到根基的作用，通过设备的关联逻辑，可以将“故障-原因-解决方案”三元组数据挂载到设备树上，更加符合用户的使用习惯，也使得图谱的逻辑性、合理性和布局美观性更强。

设备树的构建涉及设备专有名词的整理以及相应的关系数据，进而将这些数据导入图形标记中，并赋予特定编号以确定层次关系。以DA02为根编号，通过在其基础上添加01形成二级编号，层级关系逐步展开，继续递增编号以处理剩余数据。

最终，构建好的设备树以EXCEL格式呈现。通过开发设备树，可以更加直观地管理和展示设备信

息，同时也能便捷地关联故障、属性等相关信息。这样建立的设备树为知识图谱的管理与应用提供了必要的基础条件。

5. 图谱构建

(1) 图谱数据关联

在经过知识抽取之后，获得大量的结构化数据，这些数据相对于之前的原始数据，规则化程度较高。发电厂的设备一般处于一个相对封闭且稳定的环境中，设备的工作目标相对明确，并且设备之间相互紧密关联，由此产出的设备树条理十分清晰。为此，将结构化数据依附于设备树之上，由此可以构建逻辑性更强，也更加清晰的图谱。数据关联的过程可以是人工标注，或者是关键词匹配，但是效率最高，精确度最强的还是面向关系数据库的模式匹配方法。这种匹配方法基于关系型数据库，具有高效的数据处理能力，并且具有高聚合、低耦合的普适性，封装性更强，而且极大降低了电厂用人成本。

(2) 图谱可视化

由上述设备树和三元组数据构建出来的知识图谱的设备层级复杂，分支众多，所以本研究采用力导向图 (FDG) 来将图谱图像化展现。力导向模型 (FDM) 把图像定义为一个物理空间，空间内所有节点都会受到其他节点的影响，吸引或者排斥，当整体平衡时，就会展现出最优布局规划。这些节点的位置并不是人为设置的，而是根据力的作用排布。此外，根据不同的实体类型，通过赋予节点颜色，用以区分具体的实体类型。

电厂设备故障知识图谱高级应用。

1. 智能检索

在知识图谱被构建出来之后故障事件搜索与处置建议推送模块支持依托应用系统提供的搜索服务、知识推送服务，在知识库中一站式搜索和查阅设备或故障的全面信息，如图3所示。左侧列表展示设备的层次关系，以某电厂为根节点的设备列表为例，通过点击左侧设备树列表中的8机组，图谱更新，页面中展示以8机组为根节点的设备知识图谱，以此实现图谱下钻探索功能。其中，绿色节点为设备节点（深绿色节点为原有设备节点，浅绿色节点为根据缺陷日志数据提取出的新设备节点），蓝色节点为故障节点。

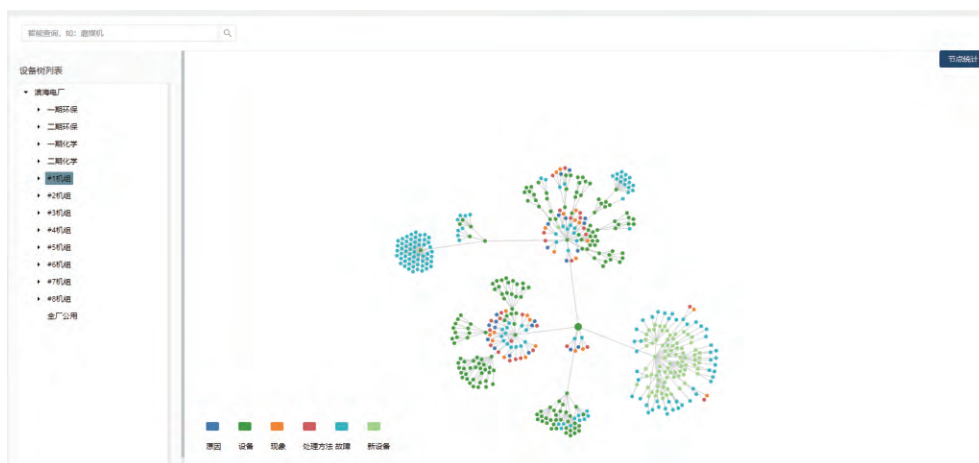


图3 设备知识图谱展示

此外，除了支持知识图谱的展示之外，还支持设备的智能检索，如图4所示。

在搜索栏中，输入“冷却塔”，点击搜索可以实现图谱更新，更新后将展示与冷却塔相关的设备

知识图谱，并展示与冷却塔相关联的现象、故障，以及和故障相关联的原因和解决方案。其中，现象是橙色节点，故障原因是深蓝色节点，故障解决方案是红色节点。

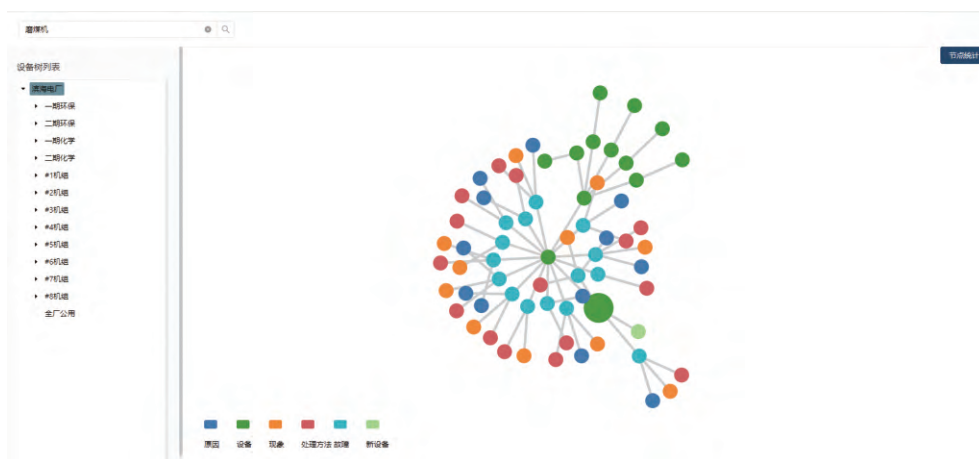


图4 设备信息智能检索

2. 设备故障解决方案

基于知识图谱的智能检索，除了能够提供设备信息检索之外，同样支持故障信息智能检索。通过在搜索框中输入故障信息，知识图谱将更新为与该故障相关的设备信息及故障原因、解决方案知识图谱。

如图5所示，在左侧故障信息列表上点击某个故障，可以在知识图谱上高亮显示从根节点（电厂信息节点）到末端叶节点（故障处理方法节点）的节点链接图。在右侧中，也将展示与该故障相关的故障名称、故障现象、故障原因和故障解决方案，以此方便用户快速定位故障位置，并生成故障解决方

案，提升电厂生产过程中的故障识别、定位以及处理效率。

3. 专家知识入库

针对已有的设备及故障信息，已构建的知识图谱能够有效实现信息的智能检索以及故障解决方案的智能生成，然而当全新的设备故障信息发生时，现有的知识图谱无法满足相关故障信息检索和方案生成。为此，需要领域专家参与图谱更新，将新的故障信息以及相关解决方案进行人工录入。如图6所示，在设备“DA02010101”上，出现“通风量不足”的问题，这是一个全新的故障信息，为此需要专家进行方案录入。录入故障现象“风机转

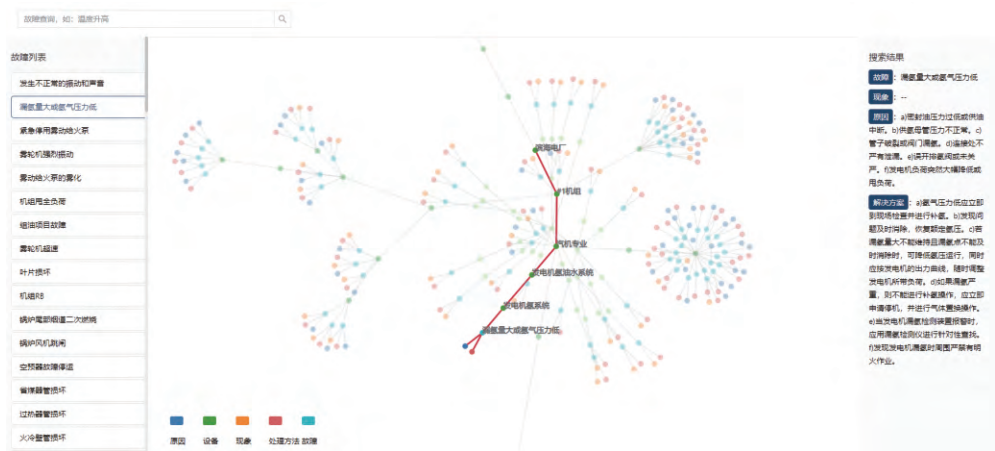


图5 设备故障解决方案生成

速低”，故障原因“转动皮带松弛，轴承润滑不良”，以及处理方法“调整地脚螺栓位置”。基于上述实现知识图谱的实时更新。

图6 故障信息专家知识库

结束语

本文面向电厂设备故障复杂多样的问题，主要面临日志数据多源异构、设备关系复杂、故障类型错综复杂、故障事件难以识别、自动标注准确低以及手工标注费时费力等关键性问题，进而导致在构建电厂设备故障知识图谱时会出现实体抽取效率低、关系构建不明确和知识推理准确率低等挑战。本工作将开展基于深度学习的语料标注研究，构建多类型语料数据半自动化标注工具，实现实体、关系的快速准确抽取，构建知识图谱，进而推进面向电厂领域的关键设备故障智能检索、故障解决方案生产以及专家知识库等高级应用。**EA**

(收稿日期: 2023-12-05)

参考文献

[1] 肖发龙, 吴岳忠, 沈雪豪, 等. 基于深度学习和知识图谱的变电站设备故障智能诊断[J]. 电力建设, 2022, 43(3): 66-74.

[2] 熊奥, 高畅, 赵明辉, 等. 基于知识图谱的核电设备健康管理知识建模与分析[J]. 科技促进发展, 2021, 17(4): 640-649.

[3] 巩宇, 李碧薇, 李德华, 等. 基于知识图谱的电力设备故障知识库构建方法[J]. 电子产品可靠性与环境试验, 2021, 39(4): 72-77.

[4] 张虹, 景欣, 阮梦宇, 等. 基于知识图谱的交直流大电网断面越限处置策略快速生成方法[J]. 现代电力, 2021, 38(4): 455-464.

[5] 吴勤浩, 沈炫辰, 陈雨, 等. 基于知识图谱和语义库的物资相似性分析[J]. 中国核电, 2022, 15(5): 654-659.

[6] 郝欣, 牟长信, 郑伟. 应用实时数据库实现电厂生产过程信息共享[J]. 东北电力技术, 2006(1): 17-19.

[7] 何佳. 电气自动化监控技术在电厂中的应用发展[J]. 科技视界, 2014(25): 187+347.

[8] 陶飞, 刘蔚然, 张萌, 等. 数字孪生五维模型及十大领域应用[J]. 计算机集成制造系统, 2019, 25(1): 1-18.

[9] 张帆. 智慧电厂一体化大数据平台关键技术及应用分析[J]. 华电技术, 2017, 39(2): 1-3+7+76.