



# 计算机科学

COMPUTER SCIENCE

## 知识驱动的机械设备故障诊断

董家祥, 翟纪宇, 马昕, 沈磊贤, 张力

引用本文

董家祥, 翟纪宇, 马昕, 沈磊贤, 张力 [知识驱动的机械设备故障诊断](#) [J]. 计算机科学, 2023, 50(5): 82-92.

DONG Jiexiang, ZHAI Jiyu, MA Xin, SHEN Leixian, ZHANG Li. [Mechanical Equipment Fault Diagnosis Driven by Knowledge](#) [J]. Computer Science, 2023, 50(5): 82-92.

---

## 相似文章推荐 (请使用火狐或 IE 浏览器查看文章)

**Similar articles recommended (Please use Firefox or IE to view the article)**

### [基于深度学习的智能设备故障诊断研究综述](#)

Review of Intelligent Device Fault Diagnosis Based on Deep Learning

计算机科学, 2023, 50(5): 93-102. <https://doi.org/10.11896/jsjcx.220500197>

### [基于FWA-PSO-MSVM的船舶区域配电电力系统故障诊断](#)

Fault Diagnosis of Shipboard Zonal Distribution Power System Based on FWA-PSO-MSVM

计算机科学, 2022, 49(11A): 210800209-5. <https://doi.org/10.11896/jsjcx.210800209>

### [基于通道拆分CLAHE和自适应阈值残差网络的变工况故障诊断](#)

Fault Diagnosis Based on Channel Splitting CLAHE and Adaptive Threshold Residual Network Under Variable Operating Conditions

计算机科学, 2022, 49(11A): 211100122-7. <https://doi.org/10.11896/jsjcx.211100122>

### [面向图像分类的小样本学习算法综述](#)

Survey on Few-shot Learning Algorithms for Image Classification

计算机科学, 2022, 49(5): 1-9. <https://doi.org/10.11896/jsjcx.210500128>

### [基于属性图模型的领域知识图谱构建方法](#)

Method of Domain Knowledge Graph Construction Based on Property Graph Model

计算机科学, 2022, 49(2): 174-181. <https://doi.org/10.11896/jsjcx.210500076>

# 知识驱动的机械设备故障诊断

董家祥 翟纪宇 马昕 沈磊贤 张力

清华大学软件学院 北京 100084

(dix20@mails.tsinghua.edu.cn)

**摘要** 随着社会经济的快速发展,现代工业逐渐呈现出研究对象复杂化、应用手段信息化和生产方式多元化的发展趋势。机械故障诊断作为现代工业重要的研究领域之一,由于故障机理研究不足和可参考知识匮乏等问题,仍然存在一系列技术瓶颈。为应对上述问题,文中提出了知识驱动的机械设备故障诊断方案,主要包括知识构建和诊断流程两个部分。在知识构建方面,提出了领域知识图谱构建方法;在诊断流程方面,设计了一个通用的机械设备故障诊断流程,该流程包括故障问诊、故障定位、起因定位和故障维修指导4个步骤。目前,该方案已经在国内某大型挖掘机维修商实际落地应用,并对其进行了有效性验证,实验结果表明该方案提高了挖掘机故障诊断领域的知识化程度和智能化水平,并表现出了较高的准确性和实用性,后续将在工业界持续推广使用。

**关键词:** 机械设备;故障诊断;知识驱动;领域知识图谱

**中图分类号** TP311

## Mechanical Equipment Fault Diagnosis Driven by Knowledge

DONG Jiexiang, ZHAI Jiyu, MA Xin, SHEN Leixian and ZHANG Li

School of Software, Tsinghua University, Beijing 100084, China

**Abstract** With the rapid development of social economy, modern industry now presents a trend featuring complex research objects, informationalized application methods and diversified production modes. Industrial fault diagnosis, as one of the most important research areas in modern industry, is still facing a series of technical bottlenecks due to the complexity of mechanical equipment and the lack of referential knowledge. In order to solve the above problems, this paper proposes a knowledge-driven fault diagnosis scheme for mechanical equipment, which mainly includes two parts—knowledge construction and diagnosis process. In terms of knowledge construction, this paper presents a domain knowledge graph construction method. In terms of diagnosis process, this paper designs a general mechanical equipment fault diagnosis process consisting of four steps, fault inquiry, fault location, fault cause location and fault maintenance guidance. To date, the scheme has been actually applied in a large excavator maintenance provider in China, and its effectiveness has been verified. Experimental results indicate the scheme improves the knowledge and intelligent level of excavator fault diagnosis and shows high accuracy and practicability. The application of the scheme will be further promoted in the industry.

**Keywords** Mechanical equipment, Fault diagnosis, Knowledge driven, Domain knowledge graph

## 1 引言

机械设备是工程建设中施工机械的总称<sup>[1]</sup>,被广泛应用于国防、水利、建筑和运输等工程领域。作为现代工业的重要生产力,机械设备的使用极大地加快了施工进度,提高了施工质量。机械故障诊断作为现代工业重要的研究领域之一,旨在了解和掌握机器运行状态,确定其整体或局部正常或异常,尽早发现故障及其产生原因,并能预报故障的发展趋势<sup>[2]</sup>。机械故障诊断能减少设备在发生故障后造成的经济损失,延长设备安全运转的时间周期,为企业创造更多的经济价值,也能

有效减少设备安全隐患,避免重大安全事故的发生,对工业生产具有重要意义。

目前,机械故障诊断的研究主要存在故障机理研究不足和可参考知识匮乏等挑战。故障机理研究指通过理论或大量实验数据分析,得出设备故障状态和设备系统参数之间的规律<sup>[3]</sup>,现有的机械故障诊断研究对机械故障机理研究的重视不足且缺乏合理的故障诊断流程和有效的知识表达方式,专业知识和维修经验难以凝练与应用,无法满足实际的工业应用需求且缺少普适性。

在实际的故障诊断过程中,由于经验丰富的专家资源

到稿日期:2022-11-18 返修日期:2023-02-24

基金项目:国家自然科学基金(71690231)

This work was supported by the National Natural Science Foundation of China(71690231).

通信作者:张力(lizhang@tsinghua.edu.cn)

匮乏,且对发生故障的机器进行实地勘察成本较高。为了充分利用专家资源,提升故障诊断效率,领域专家远程指导用户进行故障诊断成为了目前故障诊断的重要形式。在传统的远程故障诊断过程中,专家首先会对用户进行故障问诊,收集基本的故障信息,然后基于自身经验知识或查阅维修手册、故障案例和历史工单等手段,进行初步的故障定位。该过程主要存在以下问题:

(1)过分依赖专家经验。在远程故障诊断过程中,要求故障诊断专家具有丰富的经验,能够在获得用户提供的故障信息后尽快确定引起故障的原因。然而,故障诊断专家的培养成本较高,经验积累耗时,且专家经验受限于个人,难以进行知识的分享和积累,不利于知识的复用,当出现新问题时仍需要重复手动查阅各种资料,耗时费力。

(2)故障现象描述参差不齐。由于用户缺乏相关工业机械背景知识,在进行报修时往往会出现描述不准确、描述口语化等问题,导致专家无法获取准确、全面的故障现象,影响对故障起因及维修方案的判断。

(3)故障诊断过程繁琐。当遇到未曾出现过的或较难的问题时,故障诊断专家的诊断方式回退为原始的材料查找,需要手动翻阅典型故障案例、标准维修手册等进行确认,这个过程耗时费力,严重影响故障诊断的效率,并且无法实现知识共享,其他人员在遇到类似问题时仍需要手动查阅资料来解决。

为应对上述挑战,本文提出了知识驱动的机械设备故障诊断方案,主要包括领域知识图谱构建方法和通用机械设备故障诊断流程两个部分,并将其实际应用于挖掘机故障诊断领域。本文的主要贡献有以下3点:

(1)构建了故障诊断领域知识图谱,融合了多源异构数据,实现了知识自动化集成、知识积累和知识共享,降低了对专家诊断经验和知识储备的要求,辅助远程故障诊断应用。

(2)提出了描述标准化和补全推荐的故障问诊方法,在降低用户描述要求的同时,帮助故障诊断专家更好地理解用户描述。

(3)实现了基于文本相似性匹配的自动化故障诊断,能够基于用户提供的故障描述信息准确、高效地定位知识图谱中的标准故障模式,并生成合理的故障诊断方案辅助专家进行故障诊断,提升故障诊断的效率和准确率。

## 2 相关工作

### 2.1 工业知识图谱

知识图谱<sup>[4]</sup>是用图模型来描述知识和建模世界万物之间关联关系的网状结构,由节点和边共同构成。知识图谱中的节点是实体或是抽象的概念,边是实体的属性或是实体之间的关系<sup>[5]</sup>。知识图谱最早的定位是提升搜索引擎功能的知识库<sup>[6]</sup>,由本体层和实体层共同组成,本体层是知识图谱结构化的概念模板,实体层由一系列事实组成。构建知识图谱的本质是建立认知和理解世界的过程,知识是人工智能的基石,知识不断地积累与完善,机器才能逐渐具备认知能力。

知识图谱属于人工智能三大派系的符号主义学派,于2012年由谷歌首次提出,其设计思路受互联网创始人 Tim Berners-Lee 于1998年提出的语义网<sup>[7]</sup>启发,其初衷是希望

机器可以更好地获取、存储和使用知识。在此之前,知识数据还经历过命题逻辑、谓词逻辑、产生式表示、框架和语义网络等知识表示方式<sup>[8]</sup>。知识图谱按应用领域可以分为通用知识图谱和领域知识图谱两类,通用知识图谱主要面向全领域,以常识性知识为主,主要强调知识的广度,如历史上影响较广的 Cyc 知识库<sup>[9]</sup>、著名的词典知识库 WordNet<sup>[10]</sup>、MIT 媒体实验室的 ConceptNet<sup>[11]</sup>、维基百科结构化知识库 Freebase<sup>[12]</sup>、大规模多语言链接知识库 Wikidata<sup>[13]</sup>等。领域知识图谱主要面向特定行业,用于描述特定领域的垂类知识,对知识的准确性要求较高,该类知识图谱通常用于指导和辅助特定行业中复杂的分析应用与决策支持。目前,知识图谱已经成为学术界和工业界的研究热点,随着技术的逐渐发展与成熟,知识图谱逐步在语义搜索、智能问答、大数据分析、推荐系统和可解释性人工智能等诸多领域发挥着越来越重要的作用。

本文提出了领域知识图谱构建方法,并将其实际应用于挖掘机故障诊断领域,首次构建了大规模挖掘机故障诊断领域知识图谱,其知识体量和覆盖范围均达到领域之最。

### 2.2 机械故障诊断

故障诊断技术在机械设备领域的应用已有近50年的历史。1967年,美国宇航局首次提出对机械进行故障诊断的概念。在欧洲,英国将其应用于汽车领域,多家机械制造公司联合设计了柴油机故障诊断系统。在亚洲,日本将其应用于民用行业,日本三菱公司独立开发了铁路机械故障诊断系统。20世纪80年代,我国开始对石化开采、钢铁冶炼和电力系统等领域的机械故障诊断进行研究。“中国制造2025”设想被提出后,我国开始由机械制造大国向制造强国方向全面升级。机械故障诊断作为其中重要的部分,已成为一个热点研究领域。

现有的故障诊断方法主要分为3类,基于分析模型的故障诊断方法、基于知识的故障诊断方法和数据驱动的故障诊断方法<sup>[14-15]</sup>。

1971年,Bread<sup>[16]</sup>提出了基于分析模型的故障诊断方法,该方法需要精确的数学分析模型,按特定的数学方法对被测信息进行诊断。该方法适用于构造简单、规模较小的机械设备,但是在实际生产环境中,该方法往往因为难以对机械设备精准建模而具有一定的局限性。Dadon 等<sup>[17]</sup>提出了一种齿轮故障的诊断方法,该方法构建了两个齿轮及中间连接弹簧的分析模型,分析齿面故障对齿轮振动特征信号的影响,模型充分考虑了势能变化、齿轮刚度变化以及故障导致的齿面磨损修正等情况,能够很好地模拟出齿面故障,对齿面的故障诊断和预测具有重要的价值。

基于知识的故障诊断方法不需要对机械设备进行精准建模,在获取诊断对象的基本信息后,综合运用领域经验知识进行推理与分析。但是存储经验知识的专家系统或知识库往往不具备自主学习能力,导致领域知识难以扩展且维护成本较高。Yang 等<sup>[18]</sup>针对故障诊断领域经验知识描述存在多样性和不确定性的问题,提出了一种基于决策和 RS 理论的知识库构建方法,针对数据源进行分析并构建决策表,通过属性约简和值约简来得到故障诊断和诊断规则的最小属性集,并通过案例分析的方式证明了知识构建的有效性。但是这种方案

只适用于与机械数据相关的故障知识,无法对纯文本的故障知识进行建模和抽取。

1994年,Askin等<sup>[19]</sup>提出的数据驱动的故障诊断方法是对当前采样的在线数据和系统存储的历史离线数据进行分析处理。数据驱动的故障诊断方法对数据的要求较高,在工业领域的实际应用中,该方法经常受限于数据的体量和质量而不能达到预期效果。Costamagna等<sup>[20]</sup>对氧化物燃料电池的相关物理量进行了离线分析,构建了支持向量机和随机森林分类器,通过实验验证了支持向量机的故障诊断方法比混合故障诊断系统方法的诊断准确率更高,但是这种方式在多故障混合时的诊断效果并不好。

本文提出了知识驱动的机械设备故障诊断方案,通过构建领域知识图谱,并将其集成进通用的诊断流程中,以应对机械设备的故障诊断问题。有效性验证表明,该方案具有较高的准确性和实用性。

### 3 故障诊断领域的知识图谱构建

工业领域具有细分的业务范围、良好的业务模型和大量的业务数据,但是由于数据结构复杂、质量较差且缺乏有效的数据集成和知识表示方式,数据未能充分发挥其潜在价值,领域知识的积累、扩展和使用也存在瓶颈。

传统远程故障诊断过程依赖于专家经验,要求专家能够熟练掌握挖掘机机理和标准装修手册中的内容,能够根据故障信息进行快速的故障定位,并需要与用户频繁沟通。但是,在这一过程中,一方面要求故障诊断人员具有较多知识储备,才能进行较为准确的故障定位,另一方面故障诊断知识过度依赖于人工,知识积累耗时费力,知识共享和传播受限,无法

进行有效的知识应用和知识积累,新手故障人员仍需依赖翻书查找等原始手段,诊断效率及准确率低。

为了提升故障诊断的效率和准确率,本文提出构建故障诊断领域知识图谱,融合多源异构的故障诊断知识,构建知识有效的组织结构和知识表达方式,为后续故障诊断提供扎实的知识基础。如图1所示,本文提出了包括知识数据源选择、数据预处理、本体构建、知识生成和知识加工5个核心步骤的领域知识图谱构建方法,具体如下。

(1)知识数据源选择:知识数据源选择需要从大量领域数据中划分出具有潜在价值且可以从中抽取领域知识的数据作为知识数据源。

(2)数据预处理:数据预处理是对知识数据源进行数据清洗、模式规范化和数据类型转换等操作,以提高数据的准确性、一致性和完整性,为知识图谱构建提供高质量的数据。

(3)本体构建:知识图谱本体构建需要结合领域专家的专业知识和领域经验,从知识数据源中完成知识抽象和规则定义。

(4)知识生成:知识生成过程主要包括知识抽取和知识存储,该过程从预处理后的知识数据源中完成属性、实体和三元组等知识要素的抽取,在此基础上形成本体化的知识表达。

(5)知识加工:知识加工过程通过知识更新、知识融合、知识推理和知识质量检测等,来消除实体间的歧义,发现实体间潜在的关系,逐步完善领域知识体系。

本文将上述方法应用于挖掘机故障诊断领域,首次构建大规模挖掘机故障诊断领域知识图谱,用于指导实际生产环境中的挖掘机故障诊断任务。下文将详细介绍其构建流程。

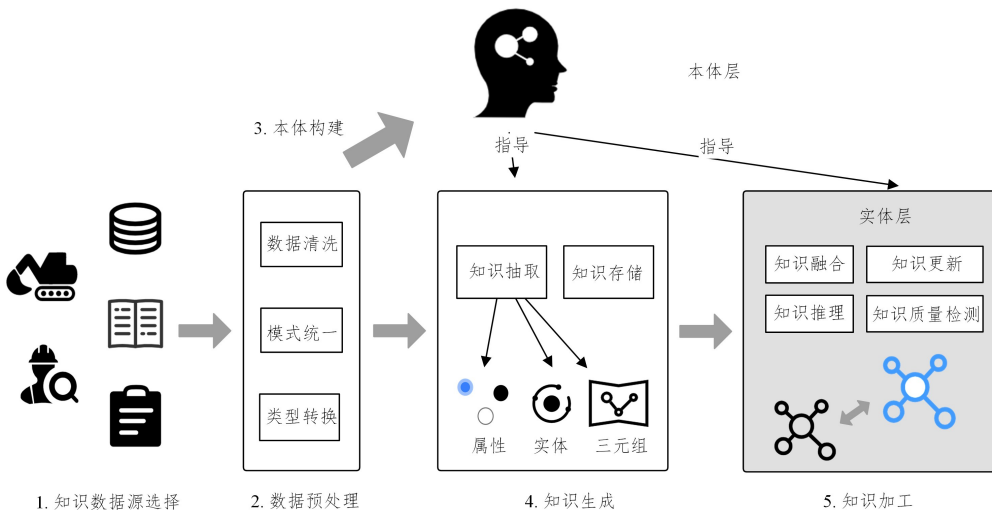


图1 领域知识图谱的构建方法

Fig. 1 Method of constructing domain knowledge graph

#### 3.1 知识数据源选择

根据本文提出的领域知识图谱构建方法,首先需要确定知识数据源。知识数据源主要来自企业发展过程中积累的大量业务数据、领域专家的专业知识和维修人员的维修经验。通过对大量的多源异构业务数据进行综合分析,如图2所示,梳理出5类挖掘机故障诊断领域的核心知识数据源。

挖掘机故障诊断领域的核心知识数据源具体如下。

(1)挖掘机设备信息:设备控制器在挖掘机运行过程中会实时采集设备状态信息,主要包括设备基础信息和故障信息两部分,包括机型、机号、产品结构、零部件、故障代码、故障现象和设备工作状态等。该类数据是具有较高时效性的时序数据,更新较为频繁。

(2)挖掘机故障和维修历史工单:在企业的智能服务数据库中存储了14000余个挖掘机历史工单,工单中详细记录了不同故障挖掘机设备的故障现象、故障类型、故障起因和修复方案等,其中蕴含着大量故障规律、故障诊断方法和故障维修方案知识。在新的故障发生时,可供其参考使用。该类数据是存储在关系型数据库中的结构化数据,数据量级较大且数据表之间的关联关系复杂。

(3)挖掘机故障维修手册:设备制造厂商提供了37本不同型号的挖掘机故障维修手册,其中包含挖掘机的线路图、电路图 and 7800余个分机型的标准故障模式,其中蕴含了大量挖掘机机理、故障诊断和设备维修专业知识。故障维修手册以非结构化的PDF格式存储,由于版本发布时间不同,新旧版本的

故障模式在结构和内容上均存在差异,结构复杂且解析困难。

(4)挖掘机典型故障案例:国内某大型挖掘机维修商提供的950余个挖掘机典型故障案例。故障案例是全国各地代理商根据实际维修经验对挖掘机故障的总结。与故障维修手册类似,每个案例中均包括故障现象、故障起因、故障诊断方法和故障维修方案等信息。该类数据以半结构化的Excel格式存储。由于数据来源不同,该类数据存在模式不统一和质量不一致的问题。

(5)领域专业知识和挖掘机维修经验:部分挖掘机领域专业知识和工作中积累的故障维修经验以非结构化的文本形式和半结构化的Excel形式存储,从该类数据中完成知识归纳和知识挖掘较为困难且准确率和召回率较低。

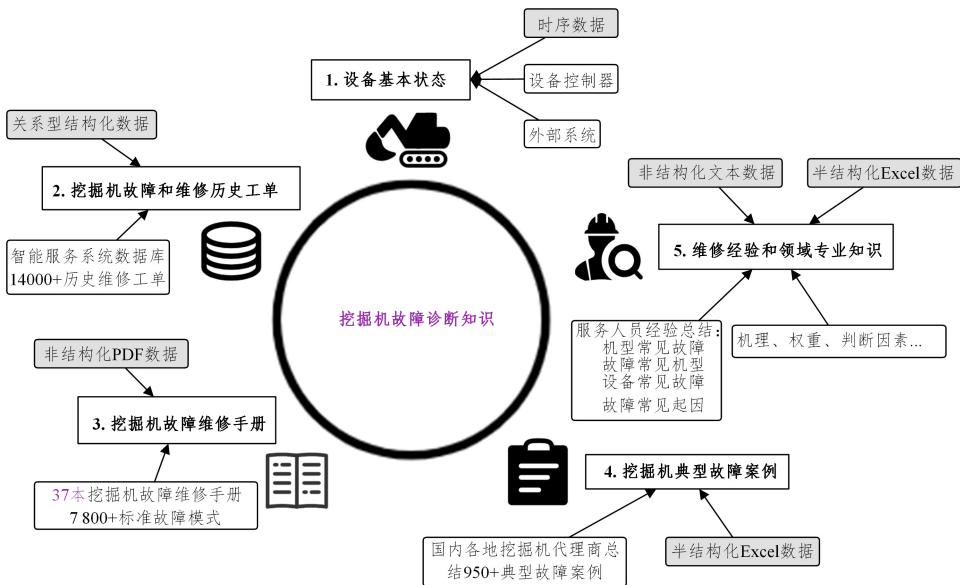


图2 挖掘机故障诊断领域的核心知识数据源

Fig. 2 Core knowledge data in excavator fault diagnosis

### 3.2 数据预处理

由于数据来源多样、业务需求不清晰或数据加工错误等原因,工业数据经常存在数据质量问题。在知识图谱的构建过程中,知识数据源的质量直接影响着知识图谱构建过程的难易程度和所生成知识质量的好坏。如图1所示,本文提出的领域知识图谱构建方法,需要对知识数据源进行数据清洗、模式统一和数据类型转换等数据预处理操作,并辅以人工和自动化协同校对的方式来保证数据质量。

挖掘机领域的5类核心知识数据源均需要进行数据清洗、去除初始数据的重复属性、可忽略字段及数据噪声、完成空值填充和异常值处理等。其中初始的非结构化挖掘机故障维修手册由于初始格式较难解析,需要进行数据类型转换,将其从PDF格式转为半结构化的Excel格式,再与同类型的挖掘机典型故障案例共同进行类别划分,最后基于其所属类别的预定义数据模板进行模式统一。领域专业知识和挖掘机维修经验数据需要进行梳理与筛选,从中挑选具有更高应用价值的数,并将其整理成规范的文本结构用于后续的知识抽取。

以挖掘机故障维修手册的数据预处理为例,原始数据以计算机较难处理的非结构化PDF格式存储,由于故障维修手册

不同版本的发布时间不同,新、旧版本的标准故障模式在结构和内容上均存在差异。该类数据量级较大且结构复杂,针对此问题,本文详细分析了73个机型的不同版本的故障维修手册中的7800余个标准故障模式,采用以下两种方法来协同配合完成该类数据的预处理过程。

(1)针对大部分格式较为统一且规范的故障模式,本文采用数据格式自动转换工具,首先将其从PDF格式转换为Excel格式,再根据其所属的故障代码类、电气系统类、发动机类、液压和机械系统类等8种故障类型预定义统一数据组织模版进行数据清洗、格式修正和模式统一。

(2)针对少部分格式较为复杂且不规范的故障模式,本文设计并开发了数据打标工具,通过点选交互的众包形式来完成特例故障模式的预处理过程,保证数据的准确性、规范性和一致性。由于数据打标工具不是本文的重点,因此不详细介绍该部分内容。

预处理后的知识数据源格式规范统一、数据质量大幅度提高且更容易解析,为后续构建知识图谱打下了良好的数据基础。

### 3.3 本体构建

知识图谱的本体用于在特定领域范围内对真实世界存在

的事物或领域内的术语及概念进行认知建模。知识建模过程是为知识图谱构建本体的过程,由于不同领域的专业知识存在差异,本体的构建过程也不相同。斯坦福大学医学院早在2001年就提出了经典的本体模型构建七步法<sup>[22]</sup>,并将其用于领域知识建模。

本文综合考虑了本体有效性和关系多样性,从经过预处理后的5类挖掘机故障诊断领域核心知识数据源中提炼挖掘机故障机理和故障诊断知识要素,并结合领域专家和维修人员的指导意见,构建挖掘机故障诊断领域知识图谱本体。所构建的本体包括概念类型47类,关系类型51类。其中部分核心本体如图3所示。

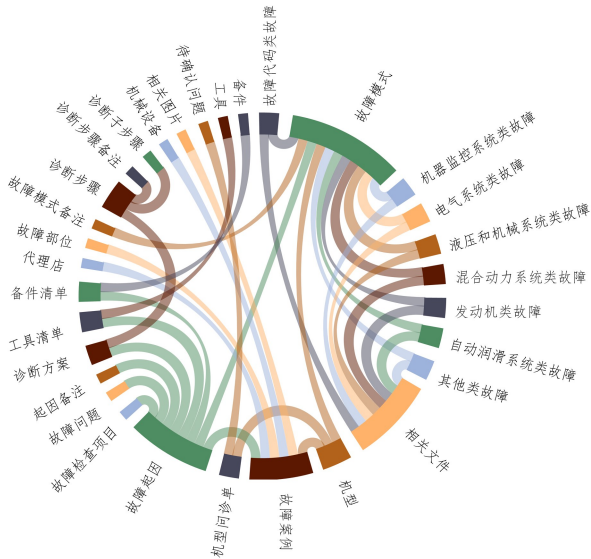


图3 挖掘机故障诊断领域的知识图谱部分核心本体

Fig. 3 Core ontology of excavator fault diagnosis domain knowledge graph

### 3.4 知识生成

挖掘机故障诊断领域的5类核心知识数据源在经过数据预处理后,分别以结构化、半结构化和非结构化形式存储。知识抽取过程是从3类不同来源、不同结构的知识数据源中进行属性、实体和关系抽取,再将所抽取的知识要素组织成本体化的知识结构,构建知识图谱实体层。本文不详细介绍各类数据的具体知识抽取和存储过程,只给出主要类型知识数据源的知识抽取思路以供参考。

(1)结构化的历史工单数据。该类数据存储在关系型数据库中,针对该类数据的知识抽取,本文首先对数据库进行模式探测以获取数据库结构,其中包括数据库中表结构和表之间的关联关系。然后建立用于描述本体与数据之间关系的映射关联模型。由于知识本体和数据库结构的差异性,在实际过程中会出现以下两类知识抽取情况:当实体属性和关联关系在同一张表时,本文基于知识本体和本体与数据之间的映射关系,直接从该表抽取知识要素并将其组织成知识结构;当实体属性和关联关系在不同张表时,本文基于知识本体,根据数据库中表的主外键对应关系,基于数据库结构和映射关联模型构建知识查询路径,从不同表中分别抽取散布在各个数据表中的知识要素

并组织成知识结构。

(2)半结构化的故障维修手册和故障案例数据。在经过数据预处理后,该类数据会根据来源和类型的不同组织成预定义的模式规范。本文针对不同的数据组织模板,采用基于模板的知识抽取方法,首先根据模板中记录的知识要素所处位置从数据中抽取知识要素,再基于本体模型将知识要素组织成既定的本体化知识表示结构。

(3)非结构化的专业知识和维修经验数据。经过筛选和归纳预处理后的专业知识和维修经验数据主要以非结构化的文本形式存储,该类数据中蕴含着故障机理、常见故障、机型特点和故障特征等领域知识,通过自然语言处理领域常用的命名实体识别和关系识别技术对其进行知识抽取并组织成本体化的知识结构。

基于上述知识抽取方案,本文累计从5类核心知识数据源中抽取维修服务类知识260000余个,故障诊断类知识400000余个,初步构建挖掘机故障诊断领域知识图谱实体层。所生成的实体采用NeoTechnology公司提出的Neo4j<sup>[22]</sup>图数据库存储。Neo4j图数据库基于属性图模型进行设计,且提供了完善的数据库操作语言Cypher,支持图遍历、图搜索和图挖掘等算法。

### 3.5 知识加工

初步构建的知识图谱存在属性和逻辑错误、不同来源的相同概念或实体的描述信息冲突以及知识的时效性和完整性问题。因此,如图1所示,要构建高质量的知识图谱需要对所生成的知识进行知识更新、知识融合、知识推理和知识质量检测等知识加工操作。

本文通过本体匹配、实体对齐和知识融合3个知识融合子过程<sup>[23]</sup>来消除实体对象之间的歧义。采用基于约束的知识图谱检测方法,设计36种约束对所构建的知识图谱进行质量检测,历时1年对挖掘机故障诊断领域的知识图谱进行数十轮的迭代更新与测试修复,最终构建的挖掘机故障诊断领域知识图谱的知识体量和知识覆盖范围均达到领域之最。知识分布如表1所列。

表1 挖掘机故障诊断领域知识图谱的知识分布

Table 1 Knowledge distribution in knowledge graph of excavator fault diagnosis

数据来源	知识域名	实体类别	关系类别	实体	关系
故障维修手册/ 故障案例/ 专业知识	故障 诊断	24	27	202058	218408
历史工单	维修 服务	23	24	130494	132951
总计		47	51	332552	351359

## 4 知识驱动的机械设备故障诊断

如图4所示,本文以领域知识图谱为核心,提出了包括故障问诊、故障定位、起因定位和故障维修指导4个步骤的通用机械设备故障诊断流程,并在国内某大型挖掘机维修商实际落地应用。

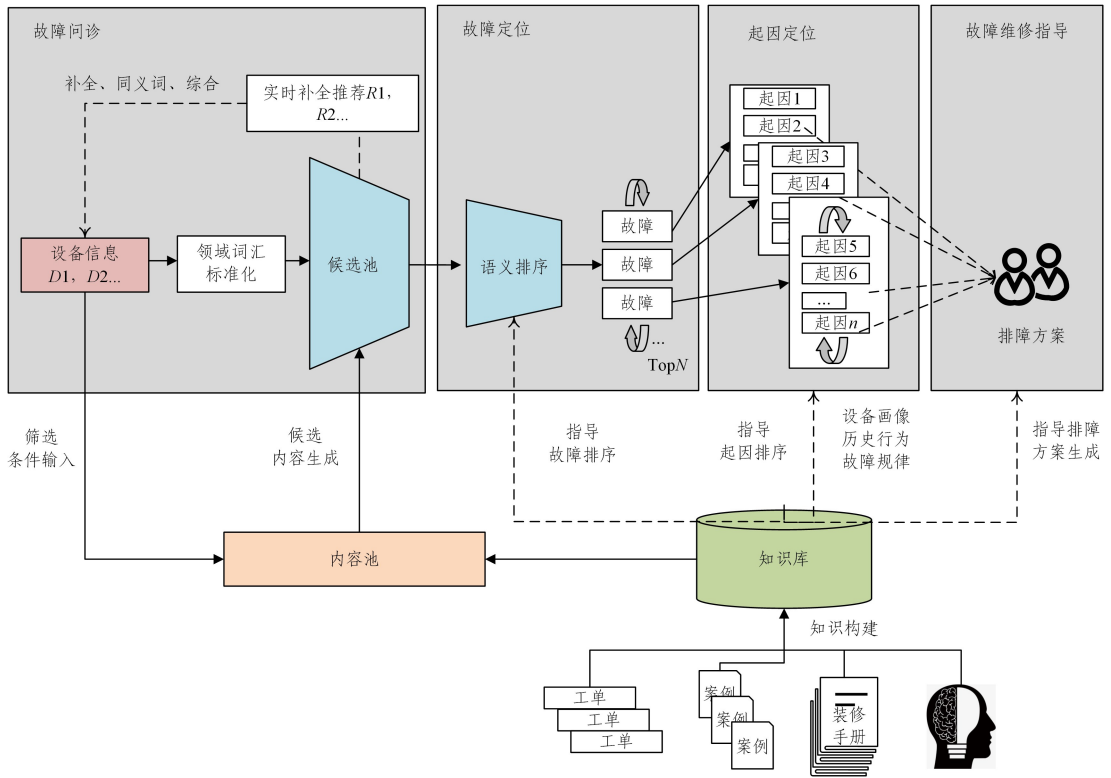


图4 知识驱动的机械设备故障诊断方案

Fig. 4 Knowledge-driven fault diagnosis scheme for mechanical equipment

在机械设备的远程故障诊断中,故障问诊和故障定位是决定故障诊断效果的重要步骤。

故障问诊旨在全面、准确地收集用户机器的故障信息,为后续故障定位提供完备的数据基础。在传统的故障问诊过程中,收集故障现象是问诊的一大痛点。用户在描述故障现象时用词五花八门,包含大量的口语化词汇,这类描述只能依赖于故障诊断专家经验,将口语化的描述和标准故障现象描述对齐,新手故障诊断人员往往会不理解用户的故障现象描述,最终导致无法准确定位故障。因此,本文提出描述标准化和补全推荐的故障问诊方法,能够针对用户有限的故障描述,推荐出知识图谱中相关的标准描述,并能够对用户口语化描述进行标准化处理,在降低对用户的描述要求的同时,充分运用专家经验,提升故障定位过程中故障描述的质量。

故障定位是基于故障问诊获得的信息,定位到知识图谱中的标准故障模式。在实际的应用场景中,用户可以观察到的故障现象可能是多种维度的,如机器本身报出故障代码、可以观测的故障现象、机器控制器的反应、故障的相关信息等,专家必须充分掌握每个故障的相关知识,才能够做到因小见大,做出完善的故障诊断。然而这一过程要求专家具有丰富的经验积累和知识储备,诊断准确性和诊断效率往往依赖专家知识和经验。本文基于构建好的故障诊断知识图谱,提出了基于文本相似性匹配的故障定位算法,针对每个故障进行多维度的相似性打分,从故障现象、故障内容、控制器反应、机器发生的故障等多个维度进行相似性评价,并按照得分对故障进行推荐排序。

#### 4.1 描述标准化和补全推荐的故障问诊

机械设备在发生故障后首先要对其进行故障问诊。故障

问诊主要用于获取设备和故障的基本信息。其中,故障现象描述是反应设备所处状态的核心故障内容。但在实际问诊过程中,由于经验差异性原因,维修人员往往不能对设备实际发生的故障进行准确、全面的描述,这将导致问诊信息描述不准确或关键信息缺失,直接影响最终的诊断效果。为了解决上述问题,本文采用描述标准化和补全推荐的方式进行故障问诊。

描述标准化用于解决故障现象描述不准确的问题。以挖掘机的故障问诊为例,首先基于领域知识图谱中存储的标准故障现象描述,与领域专家协同构建标准/日常故障现象描述同义词表。词表中记录故障现象标准描述词与日常描述词之间的对应关系。在故障问诊过程中,维修人员描述的故障现象会基于标准/日常故障现象描述同义词表进行领域词汇标准化,使故障现象的描述更加准确与规范。

故障现象的补全推荐用于解决对故障不会描述或描述不全面的问题。如图5所示,在故障问诊过程中,实时捕捉维修人员输入的故障现象并对其进行标准化。标准化后的故障现象会影响候选池中待推荐的候选故障现象排序,并实时推荐给维修人员供其输入参考与选择补全。在维修人员重新选择或调整故障现象的描述后,又会生成新一轮的引导与推荐,依次逐步多轮迭代,直至维修人员确定最符合其描述意图的故障现象为止。

描述标准化和补全推荐的故障问诊过程通过引导选择式的故障现象输入替代手动逐字输入,实现了对故障现象的相似查询、引导补全和相关推荐功能。该方法不但可以提高问诊效率,还能使问诊过程中获得的故障现象更加准确与完整,进而提高故障诊断的整体效果。

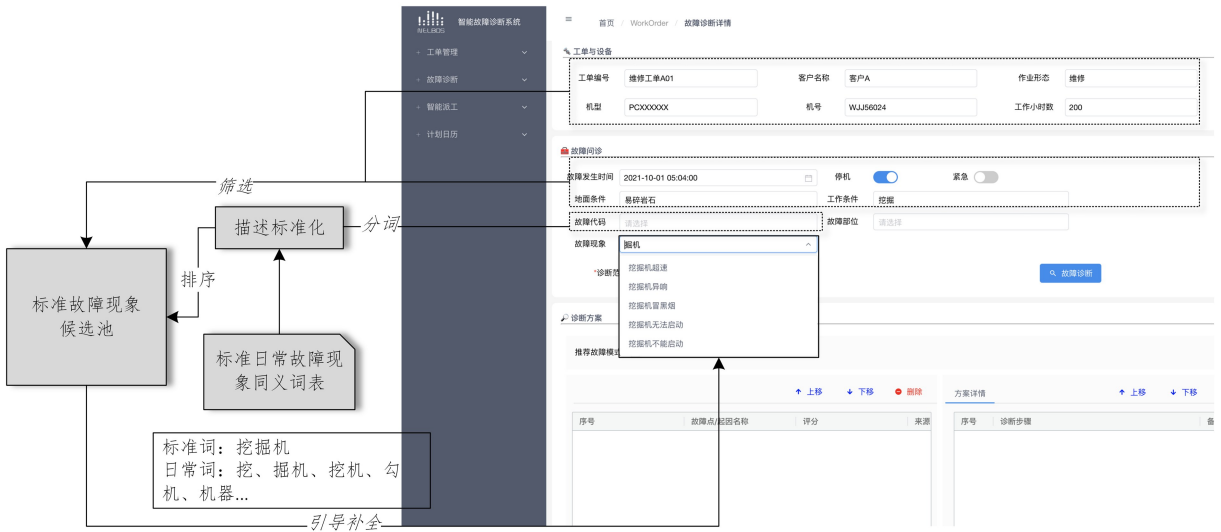


图5 故障现象描述标准化和补全推荐

Fig. 5 Fault description standardization and complementary recommendation

#### 4.2 基于文本相似性排序的故障定位

故障定位是从大量的标准故障中定位设备实际发生故障的过程。该过程是故障诊断的核心,其准确性直接影响着最终的诊断效果。为保证故障定位的准确性,本文提出了基于文本相似性排序的故障定位,通过对问诊信息和标准故障之间进行文本相似性排序,来推荐最符合机械设备实际情况的故障类型以供参考。在实际生产环境中,由于可参考的标准故障数量较多,每次故障定位均对该领域下的全部标准故障进行相似性计算,响应速度无法满足应用需求。为解决上述问题,如图4所示,本文基于问诊过程中获取的设备和故障信息,首先召回部分相关的候选标准故障,再对其进行相似性排序。关于相似性算法的选择,经过相关文献调研后,本文给出如表2所列的4种具有代表性的文本相似性算法供其参考与选择,每种算法均包括特征表示和距离度量两个关键性组件,具体内容如下。

表2 文本相似性算法

Table 2 Text similarity algorithm

序号	文本切分	特征表示	距离度量
算法1	—	—	MED
算法2	分词	Simhash	海明距离
算法3	分词	Word2Vec	余弦距离
算法4	分词	BERT	余弦距离

算法1使用最小编辑距离(MED)计算文本间的相似性。MED是一种经典的距离计算方法,通过计算文本间互相转换所需的插入、删除和替换的最小次数来计算文本间的相似性<sup>[24]</sup>。本文以MED为基线方法,逻辑简单、效率较高且不需要对文本进行分词及特征表示,缺点是该算法只能比较文本间字面的信息,不考虑语义相似性。

算法2使用Simhash<sup>[25]</sup>进行特征表示,并通过计算海明距离来计算文本间的相似性。由于Simhash的局部敏感性,该算法更适用于长文本间的相似性比较,并且使用该算法会出现无关文本对应相同Simhash的特殊情况且Simhash同样不考虑语义信息。

算法3、算法4的整体思路一致,首先基于预训练好的

Word2Vec<sup>[26]</sup>和BERT<sup>[27]</sup>模型对文本分词后的结果进行特征表示,然后通过计算特征间的余弦距离来计算文本间的相似性。算法3、算法4分别使用通过新闻、百度百科和小说数据训练的Word2Vec模型和通过中文维基百科和通用数据训练的BERT模型来进行特征表示。由于词向量在预训练过程中融合了训练语料的上下文语义信息,因此两种算法在计算过程中会考虑语义相似性。

#### 4.3 基于领域知识的起因定位

本文提出了基于领域知识的故障起因定位,通过设备历史行为和领域知识来共同完成故障起因的分析与定位过程。以挖掘机的故障起因定位为例,在挖掘机故障诊断领域知识图谱的构建过程中,实现了5类核心知识数据源的数据集成和知识凝练,其中包括大量标准挖掘机故障类型与故障起因的所属关系建模和知识生成。标准故障类型与故障起因的对应关系一般为1:n,代表通常情况下,引发同类型挖掘机故障的起因有多个。故障起因的定位需要从起因候选列表中筛选并确定挖掘机故障起因。为了更加准确、高效地完成起因定位过程,如图4所示,本文基于故障问诊信息,结合设备画像、历史行为和故障规律等领域知识,综合考虑以下原则对候选故障起因进行可能性评分。

- (1)重复故障起因:重复出现的候选故障起因评分较高。
- (2)高频故障起因:历史上引发故障的高频故障起因评分较高。
- (3)历史故障起因:历史上引发该设备故障的故障起因评分较高。
- (4)典型故障起因:日常积累的典型故障起因评分较高。

经过起因评分后,根据评分的高低对候选故障起因进行重新排序,维修人员基于排序后的故障起因候选列表,依次执行起因诊断步骤,并基于设备实际反馈状态完成最终的起因定位。

#### 4.4 基于知识融合的故障维修指导

在相似的工作环境下,如果同类型机械设备由相同故障起因引发同类型故障,其维修方案可互相参考。本文基于这

一原则设计基于知识融合的故障维修指导过程。如图6所示,挖掘机的故障维修知识主要来源于故障维修手册、故障案例和历史工单中的排障经验、维修步骤和处理心得。

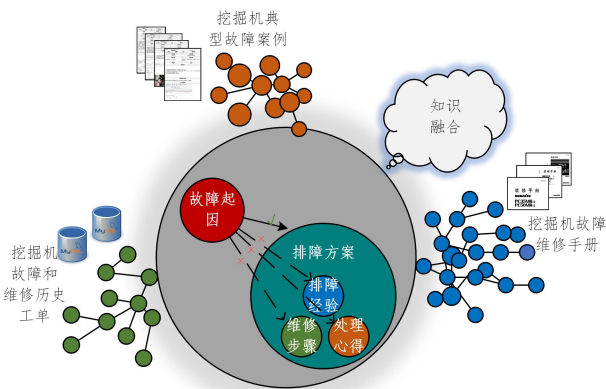


图6 故障维修类知识融合

Fig. 6 Fault maintenance knowledge fusion

在挖掘机故障诊断领域知识图谱的构建过程中,相同故障起因引发的同类型故障在不同知识数据源中会存在不同的维修方法。该过程会综合考虑多源故障维修知识的差异,对相同故障起因的多种维修方法进行知识融合,生成融合后的排障方案以指导维修人员进行故障维修。由于故障维修过程复杂,所提供的排障方案只作为初步的故障维修指导建议,具体的维修过程需要现场维修人员实际调整、确认后选择性地执行。

## 5 实验分析

在故障诊断过程中,诊断的准确性主要取决于3个方面。

(1)问诊信息准确性。在故障问诊阶段,机器的机型、机号、故障代码信息较为规范,获取难度低。但是故障现象的描述参差不齐,同时故障现象是故障诊断中的重要因素之一,故障现象的质量会直接影响故障诊断的质量。因此,本文提出描述标准化和补全推荐的故障问诊,通过主流搜索推荐方法为用户进行补全和推荐,规范用户的故障现象描述。

(2)故障定位准确性。在故障定位阶段主要用机型、故障代码、故障现象描述等信息进行多维度诊断,由于机型、故障代码等信息具有较强的规范性,因此故障定位的准确性主要取决于故障现象,用户的故障现象描述越准确,故障定位就越准确。

(3)诊断方案准确性。诊断方案的准确性实际取决于知识图谱中诊断方案的数据质量。在知识图谱构建的数据源上,本文选取的是企业提供的标准维修手册、专家整理的典型故障案例以及实际处理的历史工单,并在构建过程中进行了数据清洗和数据校验,以保证数据质量及方案的准确性。

本节将介绍基于文本相似性排序的故障定位算法选择,并基于实例和用户研究分析故障定位的有效性和准确性。

### 5.1 故障定位算法的选择

本文提出了基于文本相似性排序的故障定位算法,并给出了如表4所列的4种具有代表性的文本相似性算法以供参考与选择。在实际生产环境中,需要基于领域数据通过算法

选择实验来最终确定采取哪种文本相似性算法进行机械设备的故障定位。

表3 测试数据分布

Table 3 Test data distribution

类别	分组	用例数量	同类占比/%
标准故障描述	故障模式名	12389	32.4
	故障现象	12700	33.2
	故障内容	5819	15.2
	控制器反应	3488	9.1
	机器发生故障	3867	10.1
日常故障描述	历史工单	810	34.7
	故障案例	1522	65.3
数据增广	变长截取	12648	14.7
	关键词提取	37337	43.3
	领域同义词转换	36270	42.0

以挖掘机的故障定位算法选择为例,本文基于该领域中的7800余个标准故障模式、950余个典型故障案例和14000余个历史工单,通过变长截取(CutOut)、关键字提取(KeyWords)和同义词转换(Synonyms)等数据增广技术,模拟维修人员在工业生产环境中面临的差异化故障定位场景,所构建的测试数据集共包含数据126850条。根据数据来源的不同,其主要分为3类:标准故障中相关属性生成的标准故障描述(SFD)、日常故障诊断过程中的口语化故障描述生成的日常描述(DFD)和对前两类数据进行数据增广后生成的数据增广类故障描述(DA),详情如表3所列。

本文分别使用表2中的4种文本相似性算法对DFD和DA类测试数据与标准故障进行相似性排序,并使用推荐系统常用的评价指标MR(Mean Rank)、MMR(Mean Reciprocal Rank)、HR@N(Hit Rate)<sup>[28]</sup>进行评价。

MR为平均排名,该指标越小,说明推荐出的元素在所有内容中排名越靠前,其计算式如下:

$$MR = \frac{1}{|S|} \sum_{i=1}^{|S|} rank_i = \frac{1}{|S|} (rank_1 + \dots + rank_{|S|}) \quad (1)$$

其中,S表示所有推荐的集合,rank<sub>i</sub>为第i个元素在最终推荐顺序中的排名。

MRR为平均倒数排名,该指标越大越好,其计算式如下:

$$MRR = \frac{1}{|S|} \sum_{i=1}^{|S|} \frac{1}{rank_i} = \frac{1}{|S|} \left( \frac{1}{rank_1} + \dots + \frac{1}{rank_{|S|}} \right) \quad (2)$$

HR@N为预测排名小于N的元素个数的比,其计算式如下:

$$HR@N = \frac{1}{|S|} \sum_{i=1}^{|S|} \mathbb{I}(rank_i \leq N) \quad (3)$$

DFD是标准故障现象的高度概括和日常口语化表述。实验结果表明,使用Simhash进行特征表示的算法2对DFD类测试数据的效果最差,其他3种算法的效果相近,MR均小于4且HR@5均在0.9以上。相比之下,使用Word2Vec模型进行特征表示的算法3的效果最好,MR仅为2.480,HR@5达到0.919。DA类测试数据用于评估在问诊数据出现不规范、高度概括或信息缺失的场景下各种文本相似性算法的效果。实验结果表明,算法3在DA测试数据上的表现同样优于其他算法。

4种文本相似性算法在不同测试数据上的实现效果如表4所列。在DA实验数据下,所有算法在CutOut数据集上

都表现得最为出色,原因是这部分实验数据是直接从原始数据中截取获得的,实验数据的相似性和原始数据的相似度最高;所有算法在 Synonyms 上的表现最不佳,因为这部分数据

进行了大量同义词的替换,包含了很多原始数据中没有的内容,因此效果一般;KeyWords 由于是从原始数据中提取出的一些关键词,因此其指标介于上述两种算法之间。

表 4 文本相似性算法的效果对比

Table 4 Comparison of text similarity algorithms

分组	算法	MR	MRR	HR@1	HR@3	HR@5	HR@10	
DFD	MED	2.951	0.845	0.799	0.865	0.906	0.944	
	Simhash	6.655	0.741	0.700	0.739	0.791	0.850	
	Word2Vec	<b>2.480</b>	<b>0.860</b>	<b>0.813</b>	<b>0.878</b>	<b>0.919</b>	<b>0.962</b>	
	Bert	3.346	0.852	0.810	0.883	0.904	0.926	
CutOut	MED	1.700	0.908	0.858	0.958	0.973	0.981	
	Simhash	2.965	0.771	0.675	0.844	0.899	0.946	
	Word2Vec	<b>1.480</b>	<b>0.920</b>	<b>0.874</b>	<b>0.964</b>	<b>0.979</b>	<b>0.989</b>	
	Bert	2.175	0.869	0.805	0.921	0.955	0.974	
DA	KeyWords	MED	8.474	0.503	0.364	0.574	0.670	0.786
		Simhash	6.580	0.586	0.458	0.659	0.745	0.845
		Word2Vec	<b>6.140</b>	<b>0.611</b>	<b>0.484</b>	<b>0.689</b>	<b>0.768</b>	<b>0.862</b>
	Synonyms	Bert	9.890	0.521	0.396	0.584	0.667	0.773
		MED	<b>9.705</b>	0.434	0.295	0.483	0.590	0.739
		Simhash	13.281	0.383	0.251	0.427	0.524	0.667
Synonyms	Word2Vec	10.037	<b>0.486</b>	<b>0.357</b>	<b>0.544</b>	<b>0.629</b>	<b>0.754</b>	
	Bert	10.361	0.453	0.313	0.518	0.615	0.743	

综上所述,使用 Word2Vec 模型进行特征表示的文本相似性算法 3 的综合效果优于其他算法,更适用于挖掘机故障定位场景,具有更高的实用价值,至此完成了挖掘机领域的故障定位算法的选择过程。

5.2 实例分析

知识驱动的机械设备故障诊断方案从故障问诊过程开始,以挖掘机的故障诊断过程为例,如图 7 所示,挖掘机在发生故障后首先要进行故障问诊,以获取设备和故障的基本信息。通过图 7 中的 A 所示的故障问诊过程可知,该报修挖掘机的机型为 PC200-8M0,所属客户张三,机号为 XS202166773,该设备在易碎岩石地面累计执行挖掘任务 200h 后发生故障导致停机,故障现象为发动机不转。图 7 中的 B 为故障定位过程,通过计算故障问诊信息与大量标准故障的相似性后,初步定位该设备可能发生 A(发动机不能启动)、B(发动机动力不足)

或 C(发动机转速波动)3 类标准故障。维修人员通过详细对比分析 3 类待确诊标准故障的详情后,认为标准故障 A 与报修挖掘机当前所处故障状态更相近,因此初步诊断该挖掘机发生了 S-1 类故障。图 7 中的 C 为起因定位过程,通过所构建的领域知识图谱可知,S-1 类故障通常由使用了不合适的燃油、飞轮齿圈损坏、蓄电池故障或老化 3 类原因引起。本文基于设备历史行为和领域知识对 3 类起因进行排序后,由维修人员分别执行其对应的诊断步骤并进行起因定位。在检查到挖掘机发动机的飞轮齿圈时发现飞轮齿圈损坏,因此完成初步确诊并向维修人员推荐对应的排障方案。该报修挖掘机初步诊断结果如下:PC200-8M0 型号的 XS202166773 号挖掘机发动机的飞轮齿圈损坏引发 S-1 类故障,发动机不能正常运转而停机,需要更换发动机飞轮齿圈进行故障恢复。

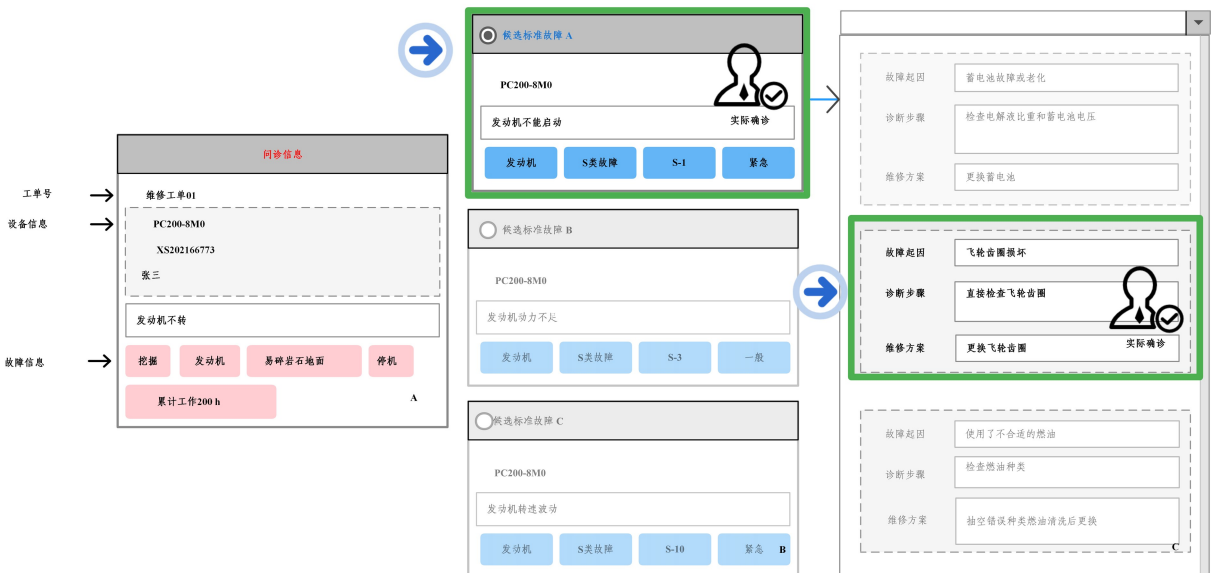


图 7 挖掘机故障诊断实例

Fig. 7 Example of excavator fault diagnosis

### 5.3 用户研究

用户研究的整体目标是调研用户在真实情况下使用本文方案进行故障诊断时的准确度和满意度。为了更好地进行用户研究,我们选择挖掘机故障诊断领域,基于本文提出的知识驱动的机械设备故障诊断方案,设计挖掘机故障诊断系统,并召集4位参与用户实际使用以对以下3个方面进行评估。

(1)采用描述标准化和补全推荐的故障问诊能否提高故障诊断效果?

(2)基于文本相似性排序的故障定位能否有效地完成故障定位任务?

(3)拥有不同程度领域经验的用户使用该方法能否在工业生产环境中独立完成故障诊断任务?

参与用户研究的4位用户均为挖掘机故障诊断领域的相关从业人员,包括2位专家用户和2位普通用户。专家用户在该领域均有5~10年从业经验,普通用户中有1位具有相关工作经历,另1位没有相关工作经历。

整个用户研究过程包括流程介绍、故障诊断、方法评价和评价分析4个环节。

在流程介绍环节,我们向4位参与用户详细介绍该系统的诊断流程、使用方法和评价规则,具体过程与5.2节的实例分析类似。该环节大约持续30min,流程介绍完成后,确保全部参与用户均能够独立使用该系统完成挖掘机故障诊断任务。

在故障诊断环节,首先随机挑选50个挖掘机故障和维修历史工单,参与用户结合工单中的问诊信息使用该系统重新对其进行故障诊断,通过对比使用该系统得出的初步诊断结果和历史工单中记录的最终诊断结果,判断诊断结果的准确性。该过程不限时间并对诊断过程进行详细记录。如图7所示,参与用户需要从多个系统推荐出的候选标准故障中判断最符合报修挖掘机实际情况的故障类型,在执行其对应的起因诊断步骤后,根据设备反馈状态完成故障确诊。

在方法评价环节,本文使用平均交互次数<sup>[29]</sup>和诊断准确率两个指标评价该方案的诊断效果。平均交互次数记录的是用户在一次故障诊断中最终确认故障时点击“故障诊断”按钮的次数,交互次数可以反映故障诊断效率;由于实际故障诊断准确率依赖于故障诊断工单的执行,数据获取困难,因此本文采用用户研究结合专家评审的方式进行准确率判定。在整个过程中,我们鼓励用户根据实际诊断过程中的切身感受对不合理或不满意的方面提出反馈意见,以方便后续对方案进行优化调整。

评价结束后,总结并分析用户诊断过程,结果如表5所列,具体结论如下。

(1)平均确诊交互次数。参与用户使用该系统进行故障诊断的平均确诊交互次数为1.545,这表明具有不同程度领域经验的用户平均与该系统进行了1.545次交互就能够完成挖掘机的初步故障诊断过程。专家用户表示,相比传统的挖掘机故障诊断过程,该方案流程清晰、操作简单且具有较高的实际应用价值。对应缺乏专业知识的普通用户来说,采用传统方式进行故障诊断需要查阅大量的维修手册知识、案例知识等,并且普通用户缺乏经验积累,无法快速确定发生的

故障,本系统综合多源知识,大幅度降低用户机械知识的要求,提升普通用户故障诊断的速度和效率,帮助普通用户在有限的交互次数内,确认发生的故障,简单高效地完成故障诊断。

(2)诊断准确率。诊断准确率是根据测试人员诊断的故障与机器实际发生的故障是否一致计算得到的。专家用户和普通用户使用该系统进行故障确诊的平均诊断准确率为95.3%,普通用户与专家用户的诊断准确率相差4%,在可接受的范围内。这说明具有不同程度领域经验的用户均能够使用该系统独立完成工业生产环境中的故障诊断任务,但最终确诊的效率和准确率会存在差异,这是因为知识图谱中整合了部分专家经验知识,但并不完备,因此专家用户的诊断准确率会高于普通用户。随着知识图谱中专家经验的不断积累,普通用户的诊断准确率会有相应的提升,甚至达到和专家用户持平的效果。专家用户在拥有十几年诊断经验的基础上,能够在故障诊断系统中达到98%的准确率,而普通用户在有很少诊断经验的基础上,也能够达到94%的诊断准确率,这在实际生产环境中具有很高的应用价值,极大降低了对故障诊断人员的知识储备要求,节约了企业培养故障诊断人员的人力、时间资源,减少了实际诊断中出现错误的次数。

表5 参与用户的评价表

Table 5 Evaluation of participating users

用户类型	平均确诊交互次数	诊断准确率/%
专家用户	1.347	98.0
普通用户	1.649	94.0
平均	1.545	95.3

除上分析结果外,用户还针对本文提出的故障诊断方案提出了许多宝贵的反馈建议,如可以采用多轮对话的形式进行故障问诊信息收集,以提高交互友好性和在故障问诊阶段考虑用户的历史行为,从而更好地揣摩用户意图等。这些反馈建议充分反映了用户的实际应用需求和诊断过程中的痛点,对后续方案的改进提供了宝贵的建议。

**结束语** 为了应对机械故障诊断领域中故障机理研究不足和可参考知识匮乏等挑战,本文提出了知识驱动的机械设备故障诊断方案,并将其实际落地应用于挖掘机故障诊断领域。该方案主要包括知识构建和诊断流程两个部分。在知识构建方面,本文提出了领域知识图谱构建方法,并使用该方法构建了大规模挖掘机故障诊断领域知识图谱,其知识体量和知识覆盖范围均达到了领域之最。在诊断流程方面,本文提出了包括故障问诊、故障定位、起因定位和故障维修指导4个步骤的通用机械设备故障诊断流程。有效性实验的结果表明,该方案提高了挖掘机故障诊断领域的知识化程度和智能化水平,并表现出了较高的准确性和实用性。但是由于该方案的诊断效果高度依赖故障问诊信息的准确性,下一步的研究方向为故障问诊环节,以多轮交互的形式对用户意图进行捕捉,并基于领域知识、用户历史行为和设备画像等对故障信息进行引导与推荐,从而进一步提升机械设备故障诊断的整体效果。

### 参考文献

[1] LI Q Y. Engineering machinery (2nd edition) [M]. Changsha:

- Central South University Press, 2012; 12-14.
- [2] LI J G, ZHANG J, GU Y. Mechanical fault diagnosis[M]. Beijing: Chemical Industry Press, 1999; 23-25.
- [3] WANG G B, HE Z J, CHEN X F, et al. Basic research on machinery fault diagnosis: what is the prescription[J]. Journal of Mechanical Engineering, 2013, 49(1): 63-72.
- [4] JI S, PAN S, CAMBRIA E, et al. A survey on knowledge graphs: Representation, acquisition, and applications[J]. IEEE Transactions on Neural Networks and Learning Systems, 2021, 33(2): 494-514.
- [5] WANG H F, QI G L, CHEN H J. Knowledge graph: methods, practices and applications [M]. Beijing: Electronic Industry Press, 2020; 56-67.
- [6] HOGAN A, BLOMQUIST E, COCHEZ M, et al. Knowledge graphs[J]. ACM Computing Surveys (CSUR), 2021, 54(4): 1-37.
- [7] SHADBOLT N, BERNERS-LEE T, HALL W. The semantic web revisited[J]. IEEE Intelligent Systems, 2006, 21(3): 96-101.
- [8] LI C, LI A, WANG Y, et al. A survey on approaches and applications of knowledge representation learning[C]// 2020 IEEE Fifth International Conference on Data Science in Cyberspace (DSC). New York: IEEE Press, 2020; 312-319.
- [9] LENAT D B, PRAKASH M, SHEPHERD M. CYC: Using common sense knowledge to overcome brittleness and knowledge acquisition bottlenecks[J]. AI Magazine, 1985, 6(4): 65-65.
- [10] MILLER G A. WordNet: a lexical database for English [J]. Communications of the ACM, 1995, 38(11): 39-41.
- [11] LIU H, SINGH P. ConceptNet—a practical commonsense reasoning tool-kit[J]. BT Technology Journal, 2004, 22(4): 211-226.
- [12] BOLLACKER K, COOK R, TUFTS P. Freebase: a shared database of structured general human knowledge[C]// Proceedings of the 22nd National Conference on Artificial Intelligence. Menlo Park, CA: AAAI, 2007; 1962-1963.
- [13] VRANDEČIĆ D, KRÖTZSCH M. Wikidata: a free collaborative knowledgebase[J]. Communications of the ACM, 2014, 57(10): 78-85.
- [14] WANG H, DING T Y, BROWN J L, et al. Data Driven Fault Diagnosis and Fault Tolerant Control: Some Advances and Possible New Directions[J]. Acta Automatica Sinica, 2009, 35(6): 739-747.
- [15] GAO Z, CECATI C, DING S X. A survey of fault diagnosis and fault-tolerant techniques — Part I: Fault diagnosis with model-based and signal-based approaches[J]. IEEE Transactions on Industrial Electronics, 2015, 62(6): 3757-3767.
- [16] BEARD R V. Failure accommodation in linear systems through self-reorganization[D]. Cambridge: Massachusetts Institute of Technology, 1971.
- [17] DADON I, KOREN N, KLEIN R, et al. A realistic dynamic model for gear fault diagnosis[J]. Engineering Failure Analysis, 2018, 84: 77-100.
- [18] YANG G, YU S. Construction Research of Diagnostic Knowledge Base Based on Decision Technology and RS Theory[C]// Recent Developments in Intelligent Computing, Communication and Devices (ICCD 2017). Singapore: Springer 2019; 311-323.
- [19] ASKIN R G, DROR M, VAKHARIA A J. Printed circuit board family grouping and component allocation for a multimachine, open-shop assembly cell[J]. Naval Research Logistics (NRL), 1994, 41(5): 587-608.
- [20] COSTAMAGNA P, DE GIORGI A, MOSER G, et al. Data-driven techniques for fault diagnosis in power generation plants based on solid oxide fuel cells[J]. Energy Conversion and Management, 2019, 180: 281-291.
- [21] NOY N F, MCGUINNESS D L. Ontology development 101: A guide to creating your first ontology[J]. Stanford Knowledge Systems Laboratory Technical Report KSL-01-05, 2001(5): 1-33.
- [22] Neo4j. (2021)[EB/OL]. <http://www.neo4j.com>.
- [23] GOU J, JIANG Y, WU Y, et al. A New Knowledge Fusion Method Based on Semantic Rules[C]// 2006 8th International Conference on Signal Processing. New York: IEEE Press, 2006; 3.
- [24] RISTAD E S, YANILOS P N. Learning string-edit distance [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1998, 20(5): 522-532.
- [25] SADOWSKI C, LEVIN G. Simhash: Hash-based similarity detection[R]. Santa Cruz: University of California, 2011.
- [26] MIKOLOV T, CHEN K, CORRADO G, et al. Efficient estimation of word representations in vector space[J]. arXiv: 1301.3781, 2013.
- [27] DEVLIN J, CHANG M W, LEE K, et al. Bert: Pre-training of deep bidirectional transformers for language understanding[J]. arXiv: 1810.04805, 2018.
- [28] BORDES A, USUNIER N, GARCIA-DURAN A, et al. Translating embeddings for modeling multi-relational data[C]// Advances in Neural Information Processing Systems. Cambridge, MA: MIT Press, 2013; 2787-2795.
- [29] CARD S K, MORAN T P, NEWELLA. The keystroke-level model for user performance time with interactive systems[J]. Communications of the ACM, 1980, 23(7): 396-410.



**DONG Jiaxiang**, born in 1994, doctoral student. His main research interests include construction and application of domain knowledge graph, big data analysis.



**ZHANG Li**, born in 1960, Ph.D., professor, Ph.D supervisor. His main research interests include knowledge graph, industrial big data, and real-time computing.