

# 基于综合集成方法的设备故障诊断及其应用\*

■ 叶含瑞<sup>1,2,3\*\*</sup> 张玲玲<sup>1,2,3\*\*</sup> 季续国<sup>4</sup>

1. 中国科学院大学经济与管理学院 北京 100190

2. 中国科学院大学数字经济监测预测预警与政策仿真教育部哲学社会科学实验室(培育) 北京 100190

3. 中国科学院大数据挖掘与知识管理重点实验室 北京 100190

4. 中科知程科技有限公司 北京 100081

**摘要:**设备故障的诊断涉及到知识工程、知识管理、数据挖掘、专家系统、可靠性工程等多领域知识,常常需要综合定性和定量方法共同进行诊断。本研究从综合集成法的思路出发,提出了一种定性定量结合的设备故障诊断方法,并以铁路CIR设备为例,综合知识图谱、文本分类、贝叶斯网络等技术,应用于设备故障知识管理、故障定位、故障诊断推理。研究表明,综合集成方法对复杂设备故障知识的管理和诊断实践提供了有效指导,同时能为维修人员和管理人员进行设备健康管理提供决策支持。

**关键词:**综合集成方法 故障诊断 知识图谱

**DOI:**10.11842/chips.20211221001

## 0 引言

对于复杂设备来说,可靠性和操作安全性对于系统的工作性能有直接影响,同时对于工业、国防等领域的关键设备,更是关乎社会财产和人民生命安全,因此对于关键设备的维护一直以来都是各方关注的重点。设备故障诊断是一个传统且较为成熟的话题,近年来随着专家系统、机器学习、深度学习等技术的发展,更是不断取得了新的成果<sup>[1]</sup>。然而过去许多研究大多集中于对结构化数据的建模和探索,对故障模式识别算法的提升和优化,而忽视了在设备故障诊断中占较大比例的文本数据和专家经验知识,同时,针对完整的设备故障诊断方

法论的探讨仍然较少。

智能工业设备的健康运行是一个涉及设备、人、机、料、管理和环境等的系统工程,在发生故障时,常常牵涉到多方面的因素,快速精确的诊断往往高度依赖于专业知识、过往经验以及对多源信息的综合考量。随着传感器、数据库软件和并行计算等技术的发展,通过机器获取、处理和分析诸如振动信号、声发射信号和油粒计数等结构化数据的状态监测和故障诊断技术取得了良好的实践效果<sup>[2]</sup>,但在设备故障诊断的实际操作中,过多地依赖于对该类结构化数据的应用,涉及经验知识等方面,自动化程度低,且多是分阶段采用人工或机器方法

\* 2020年国家自然科学基金面上项目(72071194):基于知识图谱和链路预测的推荐系统及在设备健康管理中的应用研究,负责人:张玲玲。

\*\* 叶含瑞,在读硕士研究生,研究方向:知识管理、数据挖掘;张玲玲(通讯作者),教授,研究方向:知识管理、数据挖掘、创新管理。



的简单组合<sup>[3]</sup>。缺少综合多源信息、多种方法解决故障诊断问题的方法论,由此带来诸如设备健康管理低效、设备知识挖掘和管理水平不足等问题,进而造成大量人财物的浪费,同时不可避免地带来诊断结果可解释性差、可靠性不足等情况。

因此考虑该领域丰富数据中所蕴藏的知识的重要性,以及对设备健康管理、故障诊断的方法论的迫切需要,课题组申请了《基于知识图谱和链路预测的推荐系统及其在设备健康管理中的应用》的国家自然科学基金项目,关注设备健康管理领域知识管理及非结构化数据的挖掘应用,探索指导故障健康管理和故障诊断及应用的方法论,本文是该项目的部分成果,在该项目中起到了基础性的作用,将综合集成方法应用于设备故障诊断,提出了总体研究框架,其中知识图谱、文本分类技术及贝叶斯网络的应用,也为后续的研究提供了重要工具。

本文在探索综合集成方法在设备故障诊断领域的应用时,重点关注基于知识的设备故障诊断方法,该方法将各种人工智能技术应用于工业过程的历史数据和专业知识,这些知识隐式地表示了系统变量之间的依赖性,然后检查所观察到的操作系统行为与知识库之间的一致性,并借助分类器做出故障诊断决策<sup>[4]</sup>,该方法具体又可以分为定性和定量两种类型。定性的故障诊断方法以专家系统<sup>[5]</sup>和定性趋势分析技术(Qualitative Trend Analysis, QTA)<sup>[6]</sup>为代表基于专家系统和定性趋势分析技术的研究是本研究的重要基础,这类相对成熟的定性诊断方法为从故障维修等文本中进行知识抽取及故障诊断框架构建提供了重要思路。而基于知识的定量诊断方法本质上是诊断问题的解决公式化为模式识别问题<sup>[7][8]</sup>,在设备故障诊断中的应用,常常以异常点识别、分类、预测等任务体现。随着神经网络和深度学习技术的应用,基于结构化数据的故障诊断方法取得了先进成果<sup>[9]</sup>,但是基于神经网络的诊断面临最大的问题是缺乏严格的理论支持,黑盒式的模型使得研究者很难了解这些模型如何从监测数据中学习诊断知识,进而也导致诊断结果的解释性差。相对于此而言,统计学习理论(例如支撑向量机<sup>[10][11]</sup>、随机森林<sup>[12][13]</sup>和贝叶斯网络<sup>[14][15][16]</sup>等)具有严格的理论基础,可促进构建具有易于理解的模型参数、特征和诊断结果的诊断模型,计算复杂度相对较低,模型参数的设定可以结合专家知识经验等信息。因此,智能故障诊断中统计学习方法的应用仍然值得研究<sup>[17]</sup>。

相比于以往研究,本文主要存在如下的改进和创

新:首先,将综合集成方法应用于设备故障诊断,构建了一个融合知识图谱、文本挖掘技术、贝叶斯网络等方法的故障诊断框架;其次,结合传统的故障失效分析(Failure Mode and Effects Analysis, FMEA)、故障树分析(Fault Tree Analysis, FTA)方法,构建了故障诊断知识模型,为知识图谱在故障诊断领域的应用提供了参考;最后,以实际的铁路综合无线通信(Cab Integrated Radio Communication, CIR)设备为例,结合相关单位真实的故障相关数据进行了应用探索,取得了较好的实践结果。

## 1 综合集成的故障诊断方法

### 1.1 综合集成方法

20世纪70年代以来,在一些复杂系统中构建处理复杂问题的模型的困难性日益突出,促使人们将解决问题的方式从简单的数学建模转变为对复杂问题的考虑。20世纪80年代,复杂性科学(Complexity Sciences)的诞生,标志着一种理解自然与社会的新理念的形成。然而,“复杂性”所包含的语义同样复杂,复杂性科学又涉及广阔的学科范围,各个领域的学者在研究中有不同的体会,难以对其进行统一明确的定义<sup>[18]</sup>。作为复杂性科学的专门机构——美国圣菲研究所将复杂性科学的对象确定为复杂系统,复杂系统具有非线性性和动态性、非均衡、非周期性和开放性等一系列特征<sup>[19]</sup>。复杂性科学研究的先驱者——霍兰(John Holland)从生物的遗传、变异现象中得到启发,创立了著名的遗传算法,并于1994在其著作《隐秩序》中,用隐喻的方法从适应性(adaptation)视角进行了系统深入的研究和探讨,构建复杂性理论模型,提供了一种分析复杂系统的科学方法。然而要研究复杂系统,隐喻或基于计算机的仿真模拟都显出不足<sup>[20]</sup>,需要综合各种方法的优势,尤其需要关注那些被定量建模忽略的因素,从不同的角度研究关注的问题<sup>[21]</sup>,形成新的研究方法。

在国内,最早明确提出探索和应用复杂性科学的是钱学森先生,通过对系统科学的深入探究,钱学森根据系统的复杂程度和开放性对系统进行了分类,开创性地提出了最复杂的系统,即开放复杂巨系统(Open Complex Giant System, OCGS),开放性是指与外界进行能量、信息或物质交换,该系统内存在大量具有层次结构和复杂相互关系的子系统,并指出社会系统、人类大脑和身体以及地理系统是典型的开发复杂巨系统。在钱学森、于景元、戴汝为、顾基发等国内学者的努力下,系统观念逐渐在社会实践和工程中被认识和应用,其中我

国航天事业的发展就是系统工程技术的典型成功应用案例<sup>[22]</sup>。但传统的还原论方法在处理开放复杂巨系统时适用性较差,尤其是缺乏对子系统间相关作用的考虑,在20世纪80年代末至90年代初,以钱老为代表的中国学者提出将还原论和整体论方法结合起来形成系统论方法,并从社会系统、人体系统、地理系统3个复杂巨系统研究实践中发现,以科学理论、经验知识和专家判断力等定性认识为支撑和基础,结合经验及对系统的实际理解进行建模计算,并反复对比,可以得到我们在现阶段认识客观事物所能达到的最佳结论。经过反复研究讨论,钱老等提炼、概括和抽象出来了综合集成方法(Meta Synthesis Approach)<sup>[23]</sup>。综合集成方法的实质就是把专家体系、数据与信息体系、计算机体系有机结合,构成一个高度智能化的人、机结合系统<sup>[23][24]</sup>。王丹力<sup>[25]</sup>等总结了综合集成法在过去30多年的研究和发展历程,介绍了该方法在应用信息技术、智能技术和社会科学的大量成果。随着思维认知科学、系统科学、信息技术的巨大进步,尤其是大数据、云计算、人工智能的飞速发展,越来越多的复杂巨系统出现了,复杂系统的管理将成为当下的越来越重要的一类新的管理思维、实践与研究范式<sup>[26]</sup>综合集成方法在为解决该类问题提供指导的同时,也与当前的智能科学技术结合日趋紧密,融合发展。

具体而言,当下认识综合集成法需注意如下关键点:(1)把定性研究和定量研究有机结合起来,从多方面的定性认识,上升到定量认识;(2)把科学理论性经验知识结合起来共同解决问题;(3)根据系统思想,结合多学科理论方法来进行综合研究;(4)根据复杂巨系统的层次结构,统一宏观研究和微观研究;(5)借助大型计算机系统的支持,人机结合,人网结合,但以人为综合集成信息、知识和智慧<sup>[23][25][27]</sup>。本文尝试结合复杂系统观念,以综合集成方法为指导,探索出对复杂设备进行智能故障诊断的框架。

## 1.2 设备故障诊断问题分析

### 1.2.1 智能机械设备的特點

随着现代工业和技术的飞速发展,机械设备正朝着自动化、复杂化和体系化的方向发展。一方面,设备本身可能涉及到多个系统,如车、机、工、电、辆等各系统模块,不同设备之间的关联性逐渐提高,同一设备的不同部分也紧密耦合;另一方面,智能机械设备的运行环境中常常有多种因素影响设备正常工作,且这些因素与设备故障之间有着复杂的关联关系,且设备自身的正

常运行,也不可避免伴随着自然折损。通过内外两方面因素分析可知,设备故障的影响和传导机制具有明显的非线性和随机性等复杂特性。

### 2.2.2 设备相关数据特点

智能设备在运行及维护过程中所产生的数据体量庞大且类型复杂,但数据价值需要进一步整理挖掘。数据来源多样具体体现在,设备操作手册和检修指导手册、技术参数、设备运行和维护台账、人工检查记录、访谈研讨记录、故障分析案例等文本类数据,视频监控录像、图像、音频等数据均与设备故障有一定关联。虽然数据规模客观,但也存在结构复杂、专业性强、数据质量参差不齐、可靠性及有效性存疑等问题,在实际应用中,不可完全依赖于可观察到的各类数据,仍然需要领域专家、设备维护人员等协助决策。

### 2.2.3 设备故障诊断问题特点

智能故障诊断是一个以相关数据的采集、清洗、整理、存储为基础,以所提取的专业知识和专家经验为驱动,以维修人员、管理专家和计算机系统为支撑,综合知识管理、信息技术、数据挖掘、可靠性工程等学科知识,涉及知识建模、知识表示、状态监测、诊断推理和辅助决策等多项技术支持的复杂动态过程,可以看作一个系统工程进行分析研究。

从整体来看,故障的诊断是一个跨领域、跨学科、多部门协同的过程,需要系统性的思维来指导实践。首先,需要一线人员做好数据记录、采集,搭建数据管理系统平台;随后,结合实际问题,结合知识管理、故障诊断、可靠性理论等专业知识及经验,设计知识模型整理数据,提取重要知识;最后,在实际故障诊断中,由维护人员结合现场情况,依据数据和过往经验,必要时辅以专家支持。

从人员和技术支持上而言,需要我们将设备运维人员、领域专家与计算机技术结合起来,共同用于故障诊断问题。借助计算机技术来采集、处理相关数据,从中寻找用于故障诊断的经验、模式等等,所提取出来的知识对应于知识管理中的“显性知识”;对于未能在数据中体现或难以通过现有数据所抽取到的经验知识,也即知识管理术语所定义的“隐性知识”,则需要借助专家支持,通过综合评估判断得到可靠的结果。

从数据处理和应用上,需要定性定量方法结合。一方面,设备的故障涉及到多种来源及类型的数据信息,这些数据中所包含的设备故障影响因素也与故障的发生有着复杂的时间空间联系,针对不同类型的数据,需要综合采用定性定量的方法来合理处理;另一方面,对

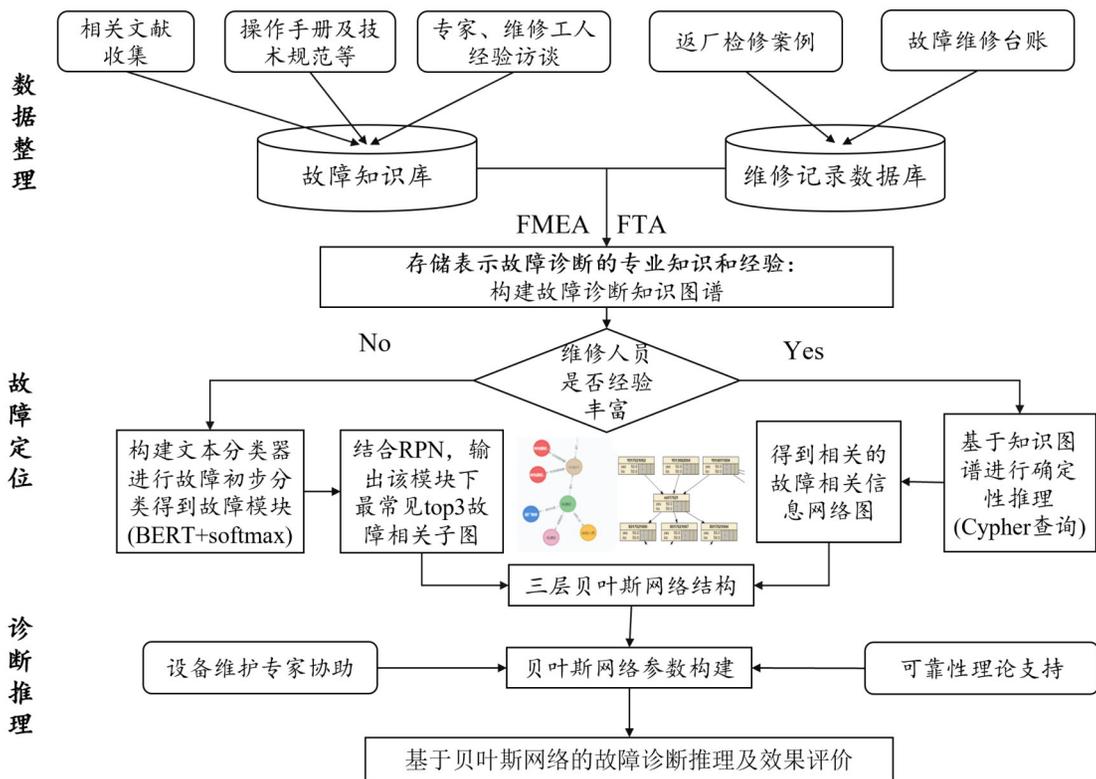


图1 综合集成的故障诊断框架

于最终的设备故障诊断结果,需要在应用各类方法分析的基础上,对感性经验和理性知识进行融合,通过定性的分析认识,逐步上升到定量方法和推理,更好地指导实践。

对于复杂设备的故障诊断,需要综合宏观理论和微观实践多层次分析建模。包括从系统到组件的结构功能梳理,从检测、定位到识别的流程分析,以及在每个环节采取有效的算法和人员支持。

### 1.3 综合集成的故障诊断框架

结合前文对设备、故障、相关数据、诊断问题的分析,本文以综合集成方法为指导方法论,融合知识图谱、文本挖掘技术、贝叶斯网络等多种方法,构建一个定性定量结合的设备故障诊断框架,整体研究架构如图1所示。

文章以铁路CIR设备为例,首先,通过分析设备的组成结构及故障特点,结合现有研究中针对设备故障诊断相关的知识图谱构建过程,进行知识建模,采用自上而下的方法,提出了针对铁路CIR设备的知识图谱构建模型,实现了该领域的知识图谱可视化。充分利用知识图谱复杂关系的表现能力,将设备相关的故障诊断流程、故障原因、故障部件间复杂的关系展现出来。

然后,提出一个针对不同类型维修人员的故障定位方法。通过分析设备故障诊断的台账记录,借助文本挖掘方法,协助经验水平较低的维修人员进行设备故障的初步分类,定位到故障所属的一级模块,并依据过往经验给出可能发生故障的下一级模块排序;对于经验水平较高的维修人员,可直接通过特定故障关键词在知识图谱中查询定位,得到潜在故障的相关信息网络子图。

最后,在已经构建出的知识图谱中,通过故障关键词检索隔离故障,辅以文本挖掘方法的初步定位,可以得到故障、部件原因与现象之间的网络图,为利用贝叶斯网络进行设备故障诊断提供了网络基础,基于此知识子图的网络结构,构建贝叶斯网络的节点和关系结构,并结合可靠性理论、维修数据统计分析、专家访谈等方法,共同确定贝叶斯网络参数。再借助已构建好的贝叶斯网络进行定量推理得到设备故障发生的概率值,协助管理和维修操作人员进行决策。

### 1.4 综合集成的故障诊断特点

(1)结合系统思想解决问题。通过对设备、设备故障、数据及故障诊断问题进行综合分析,明确故障诊断是一个跨学科多领域的研究问题,需要综合知识管理、数据挖掘、可靠性工程、信息技术等多领域知识来共同

解决问题,其中以系统思想为指导,将故障诊断问题看作系统工程来分析处理。

(2)把科学理论和经验知识结合起来。选择人机结合、人机交互、定性定量结合的方式,将专家群体和维修人员、统计数据和信息资料、自动化技术结合起来;借助FMEA和FTA进行故障分析,综合知识图谱、文本分类算法和贝叶斯网络等科学方法,共同梳理、挖掘和应用文本数据中所蕴含的经验知识,进行设备故障诊断。

(3)定性研究和定量研究有机结合起来。方法上借助知识图谱和文本分类算法处理定性的经验知识,利用贝叶斯网络来实现定量的非确定性信息推理。从定性的数据和知识出发,落脚于定量的研究和应用,实现从感性到理性,从定性到定量的转变提升。

(4)分层次体系化地分析研究设备故障诊断问题。从设备本身来看,针对设备功能结构等进行分析,可以得到不同的子系统和模块,更好地理解系统工作原理并做出可靠的故障诊断结果;从问题来看,设备的诊断又分为故障检测、故障定位和故障识别,对每一个环节针对性地进行分析建模,环环相扣,紧密衔接,如对于多源异构的故障相关知识和数据,采用知识图谱进行整理和存储;通过文本分类技术进行故障初步定位,再结合知识图谱确定性推理进一步确定故障相关网络图;根据故障定位所得的网络图构建贝叶斯网络,进行定量的诊断推理,逐步确定故障根本原因。

接下来则是在该综合集成的故障诊断框架指导下进行的具体探索。

## 2 铁路 CIR 设备的故障诊断实践

### 2.1 故障诊断知识图谱构建

在逻辑架构上,我们通常将知识图谱划分为数据层和模式层两个层次,技术架构上,知识图谱的构建主要有自顶向下和自底向上两种方法因为设备故障诊断的数据常常存在非结构化程度较高,数据规模相对较小,异质性较高,记录相对不完整等问题。但其知识内容比较明确,关系比较清晰,而设备故障诊断对于模型准确度要求较高,故而采用了自顶向下的知识图谱构建方法。而知识图谱中所涉及故障诊断的关键实体,即故障原因、故障现象、故障检测方法、故障模块、故障维护方案等,都是故障诊断的重要知识要素,需要在构建图谱前进行有效的定义和梳理。因此,我们结合传统的故障模式与影响分析(FMEA/FMECA)、故障树分析法(FTA)进行故障分析并构建故障知识模型,在此基础上构建用

于设备故障诊断的知识图谱。

#### 2.1.1 故障分析

在为复杂设备设计维护方案并整理经验知识时,首先要准确理解其故障行为,故障模式与影响分析(FMEA)(也称失效分析)和故障树分析(FTA)是目前常用的两种故障分析方法<sup>[28]</sup>。

本文以这两种方法为指导,首先,通过分析设备的功能层次关系,得到如图2所示的铁路CIR设备的功能层次与结构层次划分的对应关系,同时确定了故障模块主要有:主机、A子架、B子架、MMI、数据采集编码器、连接组件、馈线系统等。随后定义故障判据并确定故障模式及最小分析粒度。

根据故障分析,主要确定了故障排查链路、故障原因、故障部件、部件所属模块这几个核心实体,得到故障诊断知识核心如图3所示。

#### 2.1.2 知识图谱构建

基于故障分析所得的故障知识核心,构建了故障诊断知识图谱的模式层,及故障诊断知识模型(图4),其中节点包含了故障判断路径(troubleshooting path)、故障原因(causes)、故障模块(modules by units)、维修方案(solutions)、员工(crew)、部件(units);属性主要是各个节点的名称(name)和编号(id)等;关系包含了原因、故障判断路径的父步骤(fatherOf)、故障部件(troubleUnits)、检修人员(maintainer)等。

在已经搭建好的知识图谱框架下,结合铁路CIR设备操作说明及维护手册、故障分析流程图、厂家故障分析案例、故障维护台账、CIR设备故障知识库、专家及维修人员访谈等数据,提取相应的实体及关系,并先在Excel中以关系表形式整理存储。具体而言,数据层的构建需要经过知识抽取、知识融合、知识加工及知识更新等一系列的基本步骤<sup>[29]</sup>。其中,知识抽取是在模式层知识组织架构的指导下,通过一系列知识抽取方法从非(半)结构化数据中获取实体、实体间关系以及属性等结构化知识;知识融合是对知识抽取所得到的实体进行实体消歧和共指消解处理;知识加工是指对知识进行本体构建、编码和知识连接,构成了在知识图谱中的基本构成元素“节点”与“边”;知识更新则是在知识图谱应用的过程中,对其中知识的质量与时效性进行评估,并结合知识的发展进行更新和修正<sup>[30]</sup>。知识图谱的基本构成元素已经准备完毕,并以图数据库的形式存储。本文采用当前知识图谱构建的主流数据库Neo4j数据库进行知识图谱的构建和可视化的实现,实体关系及属性数量统计见表1。

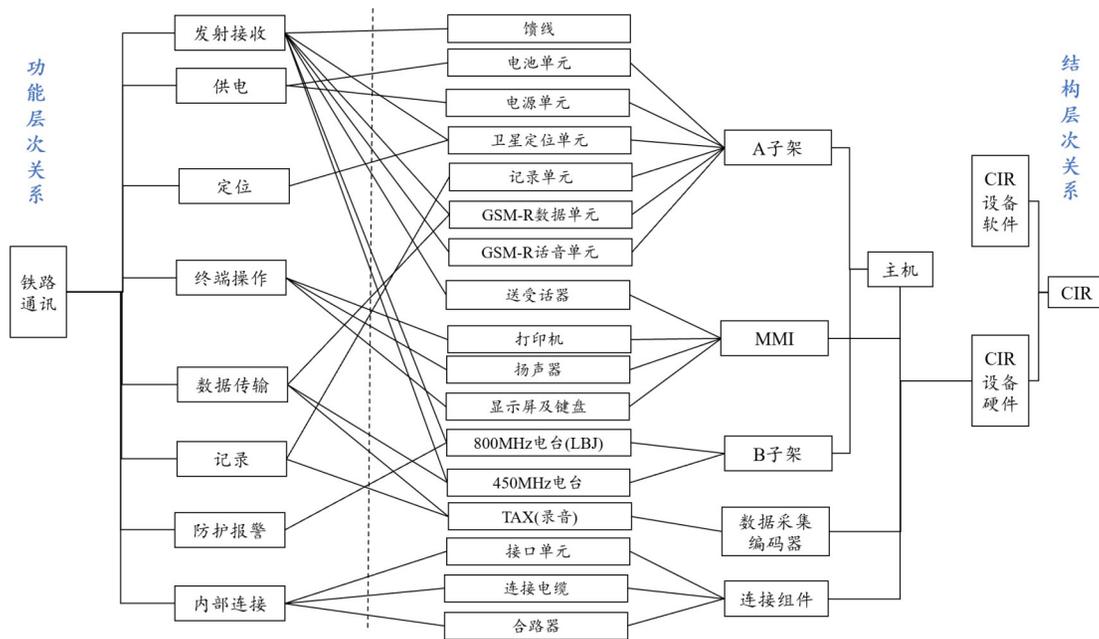


图2 铁路CIR设备的功能层次与结构层次划分的对应关系图

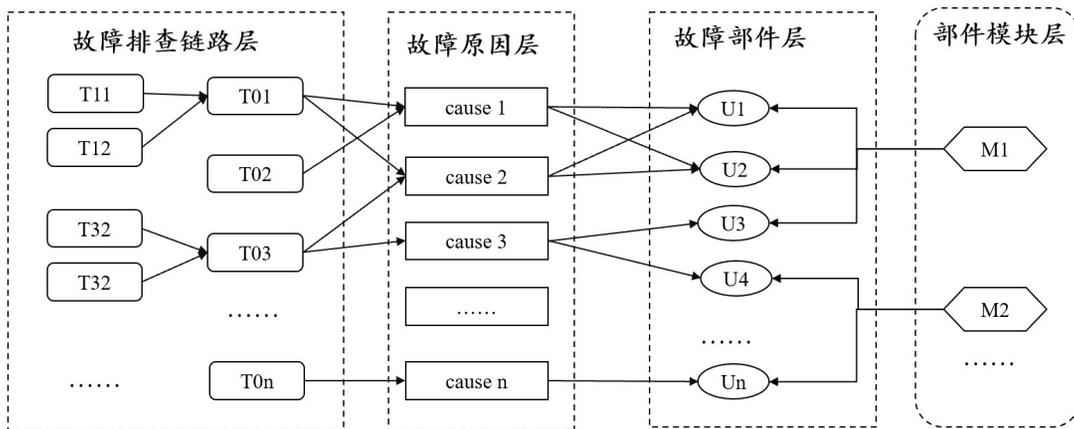


图3 故障诊断知识核心结构图

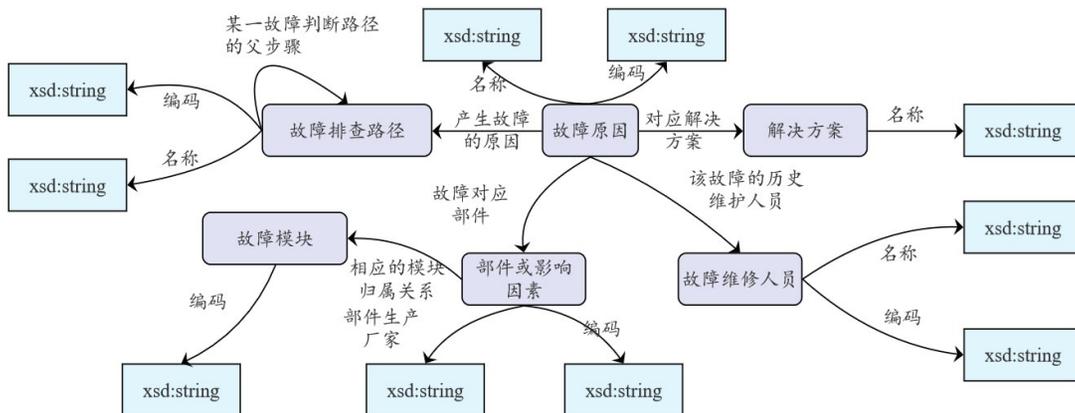


图4 设备故障诊断知识图谱模式层构建示意图

表1 铁路CIR设备故障诊断知识图谱的实体关系及属性数量

类别	实体/关系/属性	图谱中对应名称	数量	合计
实体类	故障原因	causes	284	1593
	故障维修人员	crew	348	
	故障模块	modules by unit	9	
	故障排查路径	troubleshooting path	516	
	解决方案	solutions	236	
	部件或影响因素	units	200	
关系	相应的模块的归属关系	belongTo	200	2077
	产生故障现象的原因	cause	379	
	某一故障判断路径的父步骤	fatherOf	467	
	该故障的历史维护人员	maintainer	362	
	对应解决方案	solution	343	
	故障对应部件	troubleUnits	326	
属性	节点名称	name	1498	3046
	故障原因编码	causeid	284	
	维护人员编码	crewid	348	
	部件生产厂家	factory	200	
	故障判断路径节点编码	symptomid	516	
	部件编码	unitsid	200	

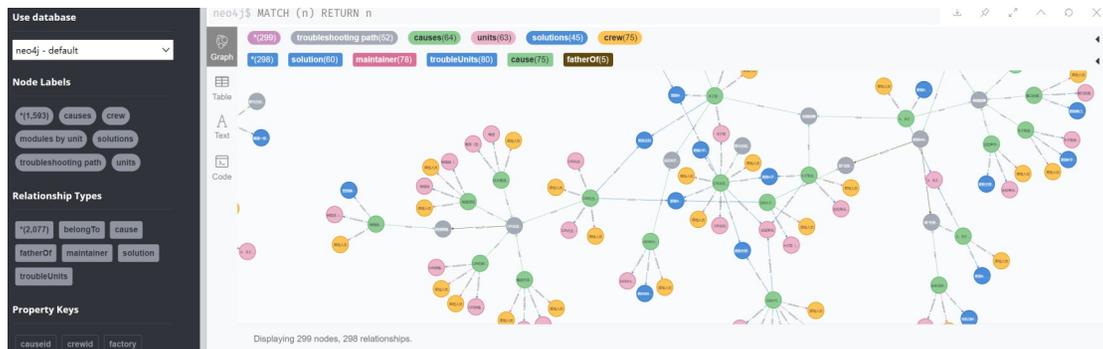


图5 基于neo4j的知识图谱可视化示意(部分)

为了便捷操作,本文使用基于neo4j的python开发包py2neo导入实体关系数据,初步建立CIR维修知识图谱如图5所示。

## 2.2 基于文本分类技术和知识图谱的故障定位

考虑到维修人员的经验知识水平差异,本文通过知识图谱及文本挖掘技术,提出了面向不同类型的故障维修人员的设备故障定位方法。

### 2.2.1 基于文本分类的故障初步定位

在进行设备故障诊断时,经验相对较少的维修人员往往难以直接通过故障现象判断到具体下一步该如何操作,即对应于知识图谱的故障判断路径,对于故障部件及原因判断更为模糊。因此加入一个从故障现象到

故障部件之间的分类器,可起到重要作用。一般而言,对同一数据集的分类标签越多,分类准确率必然越低,为保证较为有效的分类准确率,需选择合适的类别数量。结合前文故障分析所得的故障模块,即对应的故障部件所属的大类,可以构建一个输入为故障现象,输出为故障类别的文本分类器。故障维修台账等文本分类的一般流程如图6所示。

首先进行数据的清洗,包含错别字替换、同义词替换、专业名称统一等问题处理;在分类算法的核心环节,选择Bert模型进行文本表示<sup>[31]</sup>,加入Softmax层来做分类;最后通过召回率、准确率、F1值等指标对算法进行测评。

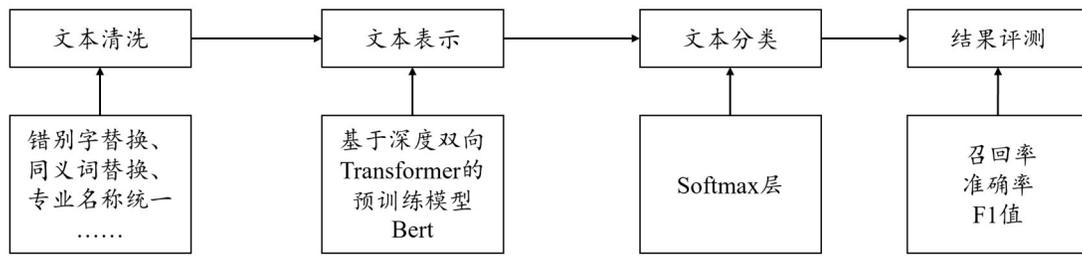


图6 文本分类流程

结合实际中的故障记录情况,在选择类别时,为保证较好的分类效果,常常面临存在同一模块下涉及到的故障原因及部件数量过多问题,在设计相应的设备故障定位系统时,需要考虑用户使用的便捷性及效率问题。因此,我们考虑引入故障模式、影响和危害性分析(Failure Mode, Effects and Criticality Analysis, FMECA)中的风险优先数(Risk Priority Number, RPN),作为故障的相关特性。FMECA是FMEA的拓展,即在FMEA的基础上加入了危害性分析(Criticality Analysis),赋予纯定性的FMEA方法以定量分析的能力。在FMECA中,每种故障模式的严重程度由风险优先数(Risk Priority Number, RPN)量化,该指标将故障严重度、发生概率等级、被检测难度等级均考虑在内,RPN分数越高,表示该项失效模式的风险越大<sup>[32]</sup>。通过为每个故障定义风险优先数,并在通过文本分类算法得到模块时,输出该模块内RPN评分最高的前n个故障,可在保障系统安全性的同时,提高故障定位的效率。

### 2.2.2 基于故障诊断知识图谱的定位

一些较为熟练的维修工人,可以根据所观察到的故障现象判断该采取何种故障排查措施,根据此关键词对应到故障知识图谱中对“故障排查路径”(troubleshooting path),根据如下Cypher查询语句,考虑输出包含“故障原因”(cause)、“故障部件”(troubleUnits)、“故障排查路径”(troubleshooting path)3类关键实体节点的故障网络图,其中“故障排查路径”信息可为维修人员提供有效的故障排查引导,如图7所示。

```
“MATCH (cause0:causes)-[r1:'cause']->(symptom1:
'troubleshooting path')<-[r2*]- (symptom2:'troubleshoot-
ing path')
```

```
WHERE (symptom1.name CONTAINS '电台故障'
OR symptom2.name CONTAINS '电台故障')
```

```
WITH cause0, r1, symptom1, symptom2, r2 MATCH
(unit)<-[r3:'troubleUnits']-(cause0:causes)
```

```
RETURN r1,symptom1,r2,symptom2,cause0,r3,unit”
```

基于文本分类方法的故障初步定位,最终也落脚于故障知识图谱中的3层网络结果,即包含原因、部件、现象信息的网络图,为后续贝叶斯网络的构建提供基础。

### 2.3 基于贝叶斯网络的故障诊断推理

贝叶斯网络可以通过融合多源信息,迭代更新后验概率,从而做出可靠且可解释的判断结果。考虑到贝叶斯网络完善的理论基础,及其在推理机制和处理不确定性信息方面的强大能力,基于前文的故障知识整理所得的知识图谱、故障诊断模型,构建了一个用于设备故障诊断推理的贝叶斯网络。

贝叶斯网络(Bayesian Network, BN)由结构和参数两部分构成<sup>[33]</sup>。图的结构表示变量之间的条件独立性,以及不同层级间的依赖性;节点间的条件概率参数则定量地描述了层级间的依赖关系。

#### 2.3.1 故障诊断贝叶斯网络构建

结合前文的设备故障定位结果,得到故障诊断的知识图谱子图作为贝叶斯网络的结构基础,构建一个包含了故障原因、故障部件、故障判断路径的网络关系图。并以故障原因为核心,由故障部件正向推导出故障原因,故障判断路径在某种意义上相当于故障原因发生后,进行判断时才可以观察到的现象。根据故障原因、故障现象、故障部件之间的关联结构,搭建贝叶斯网络结构。

故障诊断并结合台账记录中所给出的部件维修信息,统计维修频次,同时咨询专家,赋予贝叶斯网络相应的概率参数值:具体包括故障部件发生故障的概率、故障原因出现的先验概率以及3类节点间的条件概率。

(1)针对故障部件发生故障的概率(该节点的先验概率)可考虑借鉴可靠性理论,假设部件的寿命服从指数分布,部件分正常(normal)和失效(fail)两个状态。记 $t_i$ 为部件*i*的使用寿命,假设 $t_i$ 服从指数分布,即

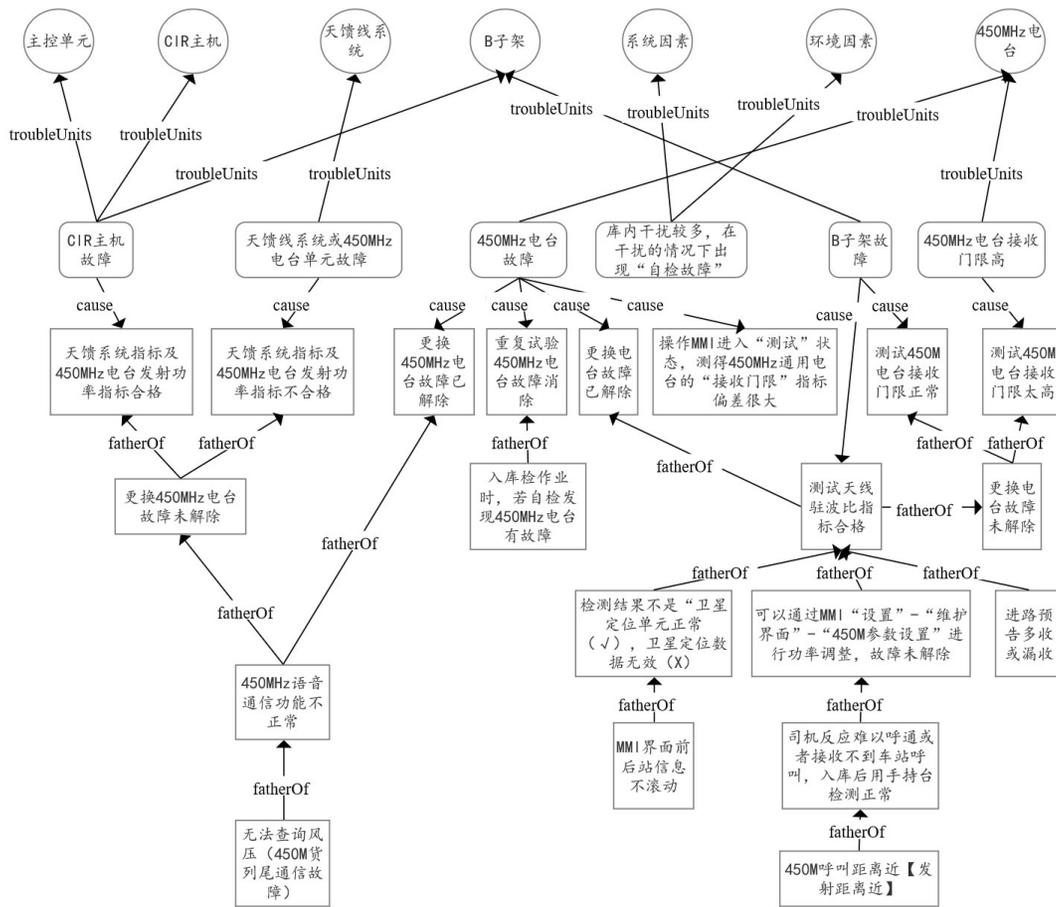


图7 基于neo4j输出的涉及“电源故障”的3层网络关系图及详细信息图示

$t_i \sim E(\lambda_i)$ 。则在  $t$  时刻发生故障的概率  $f(t_i) = \begin{cases} \lambda_i e^{-\lambda_i t}, & t > 0 \\ 0, & t \leq 0 \end{cases}$ 。再根据故障记录台账等数据进行统计

分析,得到相应的参数值。(2)故障原因出现的先验概率,也即需要先确定正常情况下,该故障发生的概率,可通过对过往数据的分析取得,亦可参照Cai等<sup>[34]</sup>的研究中的设置,假定所有故障节点的先验概率都相同,以便通过新的观察结果来强调后验概率,在没有新的故障现象被观察到的时候,故障(原因)不出现的概率为98%,而有2%的可能性会发生。节点间的条件概率,结合过往统计数据及专家访谈给出。(3)节点间的条件概率则常常需要依赖专家协助确定。

### 2.3.2 应用贝叶斯网络进行故障推理

构建完成用于设备故障诊断的贝叶斯网络之后,也可以根据实际可获得数据情况,将操作员经验数据、传感器实时数据、环境数据等分别作为附加信息编码到贝叶斯网络中进行故障诊断。

基于贝叶斯网络的定量分析沿着两个方向进行,即

正向分析(预测)和逆向分析(诊断)。在正向分析中,根据根节点的先验概率和每个节点的条件概率计算任意节点的出现概率。如:已知有某故障发生时,可能会出现故障现象的概率;当已知某部件使用时间时,其对应的故障原因发生的概率。在逆向分析(诊断推理)中,当某子节点的概率发生更新后,可以根据贝叶斯公式更新父节点的后验概率,也即是故障诊断的实际推理过程,具体如图8所示。

通常,故障的先验概率与后验概率之差越大,相应的故障发生的可能性就越高<sup>[34]</sup>。故障诊断方法可以提供发生故障的可能性,但不能得出明确的诊断结果。因此根据工程经验,一般需要设定故障发生概率的阈值来辅助判断,Cai等<sup>[35]</sup>定义了如图8所示的两条判断规则,来输出是“警告”、“故障”还是“正常”。

## 3 结论与展望

本文从综合集成的角度出发,综合了知识管理、机器学习方法、文本挖掘技术、可靠性理论等多个领域的

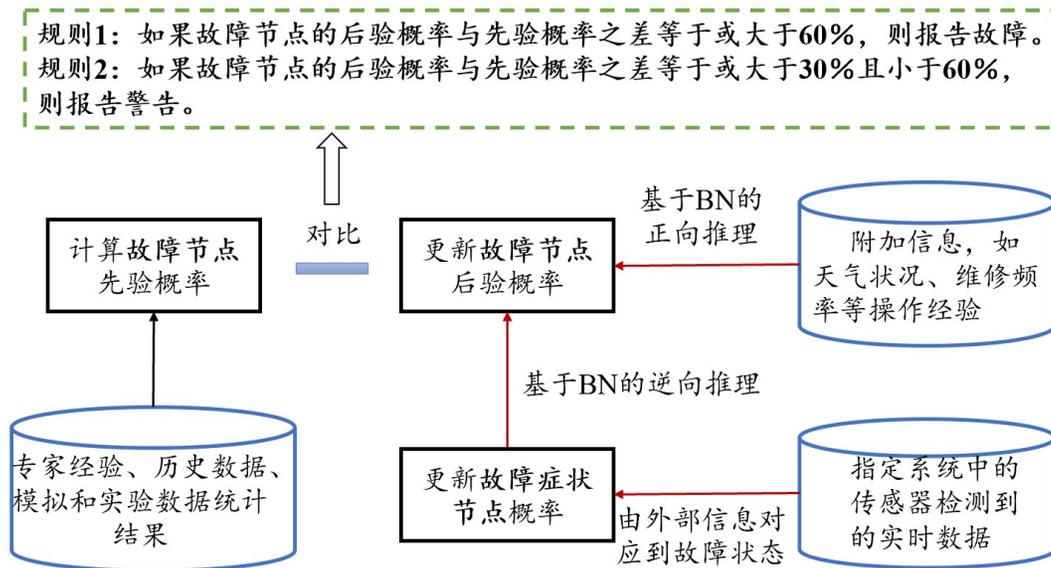


图8 基于贝叶斯网络的设备故障推理

思想及技术方法,着重考虑设备故障诊断中专家经验和维修人员等要素,构建了一个以知识图谱、文本分类技术、贝叶斯网络为主的定性定量结合的设备故障诊断框架,为后续的设备故障诊断方法研究提供了基础,为实际中的设备故障诊断知识的管理和应用提供了参考。研究亦表明,综合集成方法论可以有效指导复杂设备的故障知识管理及诊断实践,进而对于提升设备故障诊断的可解释性和可靠性具有重要基础性作用。综合集成方法对设备故障诊断等提供了较好的方法论指导,然而为进一步保障、落实设备的正常运行,快速准确地诊断设备故障,仍有较多工作需进一步研究探索。

(1)故障诊断知识图谱构建研究。知识图谱对设备故障知识的整理、存储、共享和复用提供了重要支持,然而当前该领域的图谱构建自动化程度低,主要是该领域的数据存在存储分散、样本严重不平衡、记录不完整、专业性且规模一般较小等原因,通用领域的实体识别、关系抽取、知识融合等知识图谱构建技术难以直接发挥作用,需要针对相应问题探索自动化的解决方案。

(2)故障诊断知识的挖掘研究。设备健康管理和故

障诊断领域蕴含了丰富的专业知识和专家经验,对于该类知识的挖掘,不仅仅需要探索用于处理结构化和非结构化数据的算法,更需要研究体系化、结构化的业务人员和专家知识管理方法,为后续的设备健康管理和故障诊断任务提供决策支持。

(3)基于贝叶斯网络的故障诊断研究。贝叶斯网络在设备故障诊断领域已经有了十分成熟的应用,但是未来仍需关注将贝叶斯网络与上下游任务的紧密联系,如故障定位、确定故障诊断结果并推荐维修方案等,推动综合集成方法的实践和应用。

(4)设备故障诊断方法论研究。设备故障诊断作为一个包含了数据采集处理、知识抽取、专家经验应用等方面工作的复杂动态过程,亟需有较好的方法论指导故障诊断的整体工作,本文从综合集成角度出发进行了探索,未来仍需结合工作实践和技术、理论发展,来验证、完善、创新、提升。

(鸣谢:本研究受到“中国科学院大学数字经济监测预测预警与政策仿真教育部哲学社会科学实验室(培育)基金”资助)

## 参考文献:

- [1] 刘鑫.面向故障分析的知识图谱构建技术研究[D].北京邮电大学,2019.
- [2] WANG Z, ZHAO W, DU W, et al. Data-Driven Fault Diagnosis Method Based on the Conversion of Erosion Operation Signals into Images and Convolutional Neural Network[J]. Process Safety and Environmental Protection, 2021, 149(3): 591-601.

- [3] NIU G, LI H. IETM Centered Intelligent Maintenance System Integrating Fuzzy Semantic Inference and Data Fusion[J]. *Microelectronics Reliability*,2017,75(08):197-204.
- [4] 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.
- [5] 李向前. 复杂装备故障预测与健康关键技术研究[D]. 北京理工大学,2014.
- [6] WU H, ZHAO J. Fault Detection and Diagnosis Based on Transfer Learning for Multimode Chemical Processes[J]. *Computers & Chemical Engineering*,2020,135(4):106731.
- [7] GAO Z, CECATI C, DING S X. A Survey of Fault Diagnosis and Fault-Tolerant Techniques Part II: Fault Diagnosis with Knowledge-Based and Hybrid/Active Approaches[J]. *IEEE Transactions on Industrial Electronics*,2015,62(06):3768-3774.
- [8] ROUSSEAU F. Big Data and Data-Driven Intelligent Predictive Algorithms to Support Creativity in Industrial Engineering [J]. *Computers & Industrial Engineering*,2017,112(05):459-465.
- [9] DIEZ-OLIVAN A, PAGAN J A, SANZ R, et al. Data-Driven Prognostics Using a Combination of Constrained K-Means Clustering, Fuzzy Modeling and LOF-Based Score[J]. *Neurocomputing*,2017,241(11):97-107.
- [10] 徐可,陈宗海,张陈斌,等. 基于经验模态分解和支持向量机的滚动轴承故障诊断[J]. *控制理论与应用*,2019,36(6): 915-922.
- [11] 栗蕴琦. 经验小波变换和支持向量机在滚动轴承故障诊断中的应用研究[D]. 西南交通大学,2018.
- [12] 张钰,陈珺,王晓峰,等. 随机森林在滚动轴承故障诊断中的应用[J]. *计算机工程与应用*,2018,54(6):100-104+114.
- [13] 钱力扬. 基于随机森林和XGBoost的大型风力机故障诊断方法研究[D]. 浙江大学,2018.
- [14] 查园园,王亭岭,上官伟. 基于贝叶斯网络的列车车载设备故障诊断[J]. *北京交通大学学报*,2021,45(5):37-45.
- [15] LI T, ZHOU Y, ZHAO Y, et al. A Hierarchical Object-Oriented Bayesian Network-Based Fault Diagnosis Method for Building Energy Systems[J]. *Applied Energy*,2022,306(1):118088.
- [16] PANG T, YU T, SONG B. A Bayesian Network Model for Fault Diagnosis of a Lock Mechanism Based on Degradation Data[J]. *Engineering Failure Analysis*,2021,122(04):105225.
- [17] LEI Y, YANG B, JIANG X, et al. Applications of Machine Learning to Machine Fault Diagnosis: A Review and Roadmap [J]. *Mechanical Systems & Signal Processing*,2020,138(2):106587.
- [18] 黄欣荣. 复杂性科学的方法论研究[D]. 清华大学,2005.
- [19] 张江华,陈中飞,任之光,等. 复杂性科学及其在经济领域中的资助和研究进展[J]. *管理科学学报*,2020,23(11):117-126.
- [20] 武杰,刘煊,孙雅琪. 复杂性科学的主要方法及其基本特征[J]. *系统科学学报*,2016,24(4):28-33.
- [21] GU J, TANG X. Meta-Synthesis Approach to Complex System Modeling[J]. *European Journal of Operational Research*, 2005,166(03):597-614.
- [22] 于景元. 钱学森系统科学和系统工程的成就与贡献——从系统思想到系统实践的创新(下)[J]. *中国航天*,2022(1): 41-45.
- [23] 钱学森,于景元,戴汝为. 一个科学新领域——开放的复杂巨系统及其方法论[J]. *自然杂志*, 1990(1): 3-10+64.
- [24] 于景元,周晓纪. 从定性到定量综合集成方法的实现和应用[J]. *系统工程理论与实践*, 2002(10): 26-32.
- [25] 王丹力,郑楠,刘成林. 综合集成研讨厅体系起源、发展现状与趋势[J]. *自动化学报*,2021,47(8):1822-1839.
- [26] 盛昭瀚,于景元. 复杂系统管理:一个具有中国特色的管理学新领域[J]. *管理世界*,2021,37(6):36-50+2.
- [27] 于景元. 从定性到定量综合集成方法及其应用[J]. *中国软科学*,1993(5):31-35.
- [28] PEETERS J F W, BASTEN R J I, TINGA T. Improving Failure Analysis Efficiency by Combining FTA and FMEA in a Recursive Manner[J]. *Reliability engineering & system safety*,2018,172(4):36-44.
- [29] 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.
- [30] 郭榕,杨群,刘绍翰,等. 电网故障处置知识图谱构建研究与应用[J]. *电网技术*,2021,45(6):2092-2100.
- [31] DEVLIN J, CHANG M W, LEE K, et al. Bert: Pre-Training of Deep Bidirectional Transformers for Language Understanding[J]. *arXiv preprint arXiv:1810.04805*, 2018.
- [32] 宋欣,何乾峰,张鑫鑫,等. 基于FMEA模式护理在严重创伤患者急救中的应用效果[J]. *中国医药导报*,2021,18(27):



170-173.

- [33] 宋宇飞.融合多源不确定性及复杂失效特征的系统可靠性综合评估[D].电子科技大学,2021.
- [34] CAI B, LIU Y, FAN Q, et al. Multi-Source Information Fusion Based Fault Diagnosis of Ground-Source Heat Pump Using Bayesian Network[J]. Applied energy,2014,114(2):1-9.
- [35] CAI B, LIU H, XIE M. A Real-Time Fault Diagnosis Methodology of Complex Systems Using Object-Oriented Bayesian Networks[J]. Mechanical Systems & Signal Processing,2016,80(12):31-44.

## Integrated Equipment Fault Diagnosis Method and Application

YE Hanrui<sup>1,2,3</sup>, ZHANG Lingling<sup>1,2,3</sup>, JI Xuguo<sup>4</sup>

1. School of Economics and Management, University of Chinese Academy of Sciences, Beijing 100190

2. MOE Philosophy and Social Science Laboratory of Digital Economic Monitoring, Forecasting, Early Warning, and Policy Simulation(Cultivation), University of Chinese Academy of Sciences, Beijing 100190

3. Key Laboratory of Big Data Mining and Knowledge Management, Chinese Academy of Sciences, Beijing 100190

4. Visionary Intelligence technology Co. Ltd. , Beijing 100081

**Abstract:** Equipment fault diagnosis is a system engineering problem involving knowledge engineering, knowledge management, data mining, expert system, reliability engineering, and other fields of knowledge. Starting from the idea of the meta-synthesis method, this study put forward a qualitative and quantitative method for equipment fault diagnosis, and took Cab Integrated Radio(CIR) Communication Equipment as an example to apply a combination of the knowledge graph, text classification, Bayesian network, and other technologies to equipment fault knowledge management, fault location, and fault diagnosis. This study shows that the meta-synthesis method provides effective guidance for the knowledge management and fault diagnosis of complex equipment. It also provides decision support for maintenance personnel and management personnel for equipment health management.

**Keywords:** Meta-Synthesis method; fault diagnosis; knowledge graph

(责任编辑:何岸波; 责任译审:毛子英 张述庆)