

航空发动机故障实体识别方法及应用

张亮¹, 吴闯^{1,2}, 唐希浪¹, 冯少林¹

(1. 空军工程大学装备管理与无人机工程学院, 西安, 710051; 2. 95478 部队, 重庆, 401329)

摘要 故障实体识别是自主获取航空发动机故障知识的基础, 对实现航空发动机故障智能诊断起到至关重要的作用。为准确快速搭建航空发动机大规模故障知识库, 在定义了“单元”“故障状态”“表征信号”“检查方法”和“解决措施”5种航空发动机故障实体类型的基础上, 初步构建了一种以 Bert-BiLSTM-CRF 模型为基础的航空发动机故障实体识别方法。基于某型航空发动机大规模数据集分析抽取了故障实体, 搭建了滑油压力异常故障知识图谱, 验证了该方法识别航空发动机多源异构故障数据的有效性。

关键词 航空发动机; 智能故障诊断; 实体识别; 知识图谱

DOI 10.3969/j.issn.1009-3516.2022.02.001

中图分类号 V37 **文献标志码** A **文章编号** 1009-3516(2022)02-0001-06

A Method of Recognizing Aero-Engine Fault Entity and Its Application

ZHANG Liang¹, WU Chuang^{1,2}, TANG Xilang¹, FENG Shaolin¹

(1. Equipment Management and Unmanned Aerial Vehicle Engineering College, Air Force Engineering University, Xi'an 710051, China; 2. Unit 95478, Chongqing 401329, China)

Abstract Fault entity recognition is a basis of obtaining the knowledge of aero-engine fault autonomously, which plays an important role in realizing the intelligent fault diagnosis of aero-engine. For building up fleetly accurately aero-engine fault knowledge base, on the basis of the five kinds of aero-engine fault entity type defined, i. e. “unit”, “state failure”, “characterization of signals”, “inspection methods” and “solution”, a kind of Bert-BiLSTM-CRF model of aero-engine fault entity recognition method preliminary is constructed. Based on the large-scale data set analysis of an aero-engine, fault entities are extracted and fault knowledge map of abnormal oil pressure is constructed, verifying the effectiveness of the proposed method in identifying heterogeneous fault data of aero-engine.

Key words aero-engine; intelligent fault diagnosis; entity recognition; knowledge map

近年来,随着航空发动机结构和功能复杂程度越来越高,对故障诊断提出了更高要求^[1]。故障发生后,要求一线维修人员根据故障现象或参数变化,及时分析故障原因,定位故障单元,并迅速采取有效的排故措施。如何利用大规模的故障数据信息生成

有价值的排故知识,辅助一线维修人员快速定位故障,能够有效提高航空发动机的战备完好性、任务成功性,降低航空发动机全寿命周期费用^[2]。

在基于知识的故障诊断方法研究领域,Zhou 等利用 Java Expert Shell 系统(JESS)规则引擎对本

收稿日期: 2021-12-01

基金项目: 中国博士后科学基金(2021M693941)

作者简介: 张亮(1982—),男,江西吉安人,副教授,研究方向为装备综合保障。E-mail: zliangmail@sina.com

引用格式: 张亮, 吴闯, 唐希浪, 等. 航空发动机故障实体识别方法及应用[J]. 空军工程大学学报(自然科学版), 2022, 23(2): 1-6. ZHANG Liang, WU Chuang, TANG Xilang, et al. A Method of Recognizing Aero-Engine Fault Entity and Its Application[J]. Journal of Air Force Engineering University (Natural Science Edition), 2022, 23(2): 64-69.

体知识库进行推理,以快速找到风机故障的致因^[3];Zhou 等将故障现象映射到本体中所表示的个体,然后对个体进行本体推理以找出滚齿机的故障原因^[4];Medina-Oliva 等使用本体来检索相同舰船或同类舰船的相似故障诊断案例^[5];Xu 等结合使用本体与规则以推断装载机的故障原因^[6];Su 等提出基于本体对多源、分布式、异构的故障信息进行描述和集成^[7]。这些研究表明,基于知识的方法能够很好地解决系统级别的故障定位问题,应用到航空发动机的故障诊断和监测上具有非常强的潜力,但需要构建完备的故障知识库。以上现有研究成果的故障知识库大多是以人工编辑的方式获取,忽略了自动化获取故障知识的问题,难以完成大规模知识的获取和知识库的构建,制约了该类方法的发展应用。而基于深度学习的知识图谱获取技术^[8-12]被普遍认为是解决上述问题的重要途径。

故障知识图谱的获取主要包含故障实体识别和实体关系提取两个方面^[11]。其中,发动机故障实体识别是构建故障知识图谱、开展发动机智能诊断的基础性工作,是实现航空发动机故障智能维护保障的关键。深度学习可以让计算机自主学习得到相应模式特征,降低对人工特征的依赖程度。基于深度学习的信息抽取方法包括递归神经网络、卷积神经网络模型、双向转换编码(BERT)表示等模型^[13-14]。

综上所述,本文针对航空发动机多源异构的数据特点,提出了一种基于 Bert-BiLSTM-CRF 的航空发动机故障实体识别方法,并在某型航空发动机大规模数据集上进行验证,利用抽取的故障实体构建了滑油压力故障知识图谱。

1 航空发动机故障实体识别任务分析

航空发动机故障实体是指描述各类故障现象、诊断方法、排故过程等故障文本中的关键信息,如故障的表征信号、故障件、故障状态等。实体识别关键任务就是抽取文本中关键信息,且信息抽取质量对未来故障知识库的构建起着至关重要的影响。部队调研发现,航空发动机故障文本广泛蕴含在发动机全寿命周期内积累的各类故障文档数据中,包括故障危害性分析文档、发动机维护保障教材、测试诊断记录、质量控制软件数据、维修保障记录、故障分析与研究报告等,几乎都是以非结构化数据形式存储,行文结构无规则且专业术语较多,见图 1。

图 1 中蓝色表示检查方法,黄色表示表征信号,红色表示故障状态,绿色表示单元。如此定义的实体分类囊括了航空发动机故障文本中大部分有价值

的信息,有利于下一阶段航空发动机故障知识图谱的构建。

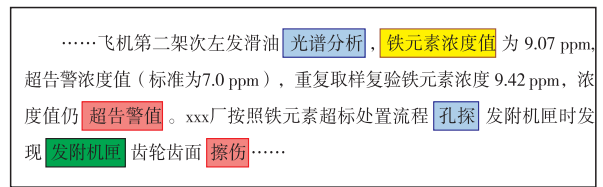


图 1 某型航空发动机故障报告文本片段

针对故障文本的上述特点,本论文计划采用监督下的机器学习。对此,首先必须制定规范的故障实体标注体系,从而保证实体标注较高的一致性,确保实体识别的准确性^[15]。

本论文邀请了数名有丰富航空发动机维修经验的行业专家共同参与标注体系的搭建和研究,其中包括 2 名综合监控室主任与 3 名维修专业人员。经过实践考察和数次技术研讨,并与某型航空发动机的维修实践相结合,将航空发动机各类文本中的故障实体分为 5 类,见表 1。

表 1 某型航空发动机故障关键实体类型

标志符号	实体类别	类别定义	示例
U	单元 (Unit)	发动机各部件、附件等	尾喷口、滑油箱、燃烧室、传感器
S	表征信号 (Signal)	发动机的特征、性能	压力、转速、振动值、温度、黏度
F	故障状态 (Failure)	发动机故障类型	漏油、超温、断裂、起动不成功、变形
C	检查方法 (Check)	采取的检查手段	目视、孔探、油液分析、数据判读
M	解决措施 (Measure)	解决故障的维修措施	清洗、更换、润滑、打磨、串件

2 发动机故障实体识别方法

2.1 发动机故障实体识别方法整体框架

针对航空发动机故障实体识别主要任务,提出了航空发动机故障实体识别整体框架(如图 2)。

首先是故障知识获取模块。发动机的故障知识广泛存在于各家使用单位和维修工厂的维修记录、故障报告和技术使用说明书等故障文本中,未得到有效的利用。为此,在开展故障实体识别研究之前,从某型发动机的 10 余家使用单位和 2 家发动机生产检修的工厂,共收集了相关文本数据 4.6 GB。其次是模型的训练模块。从大量文本中挑选出 600 份故障文本进行人工标注,按照上文定义的 5 种实体类型,分别标注。利用标注的实体对航空发动机故障实体识别模型进行训练,并把训练结果中错误实体进行人工校正后,再次加入标注的实体数据集中,

反复训练模型,直至利用训练模型测试得到的精确率达到 80%以上。最后是模型的应用模块,是从大量故障文本中自动获取故障知识的关键。把模型部署到发动机全寿命数据中心,当维修单位上传新的

故障文本时,能自动提取故障实体,并加入到故障知识库中,形成更加丰富的故障知识,增加故障检测的准确度。

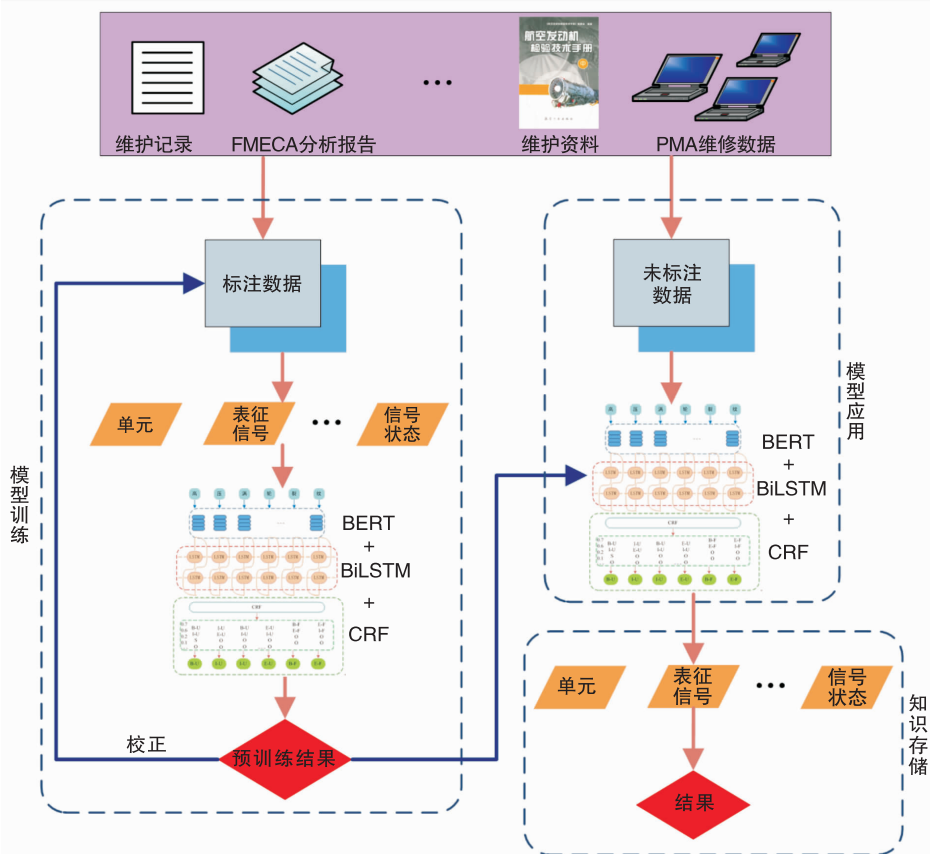


图 2 航空发动机故障实体识别方法架构图

2.2 语料标注

在有监督下的机器学习任务中,必须重视语料标注的一致性和准确性,对模型训练质量起到至关重要的作用。因此制定了完善的语料标注流程,见图 3。首先从 600 份发动机故障文本中任意抽取 180 份,一组标注 120 份发动机故障文本,其中 60 份发动机故障文本一样,用作比较两组语料标注的一致性。当语料标注的一致性检验值达到 0.85 以上时再标注剩余的发动机故障文本^[15];未达标时,则校正 2 组人员的标注规则,重新标注。在语料标注时,遇到与发动机维修无关以及描述错误的内容不进行标注,既能够提高标注的效率,也可以减少冗余数据的产生。

语料标注一致性评价时,将相同数目(60 份)的发动机故障文本分配给 2 组标注人员标注,并对标注结果按下述流程进行评价:将其中组 1 标注的故障实体作为基准参数(A_1),组 2 标注的故障实体结果作为预测参数(A_2)以及 2 组标注一致的数目参数(A_3),计算相应的准确率($P = A_3 / A_2$)、召回率($R = A_3 / A_1$)、F 值($2P \cdot R / (P + R)$),在组 1 和组 2 语料标注的一致性检验结果(F 值)大于 0.85 时,再

进行剩余发动机故障文本的标注。

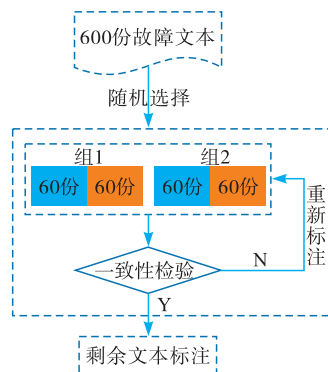


图 3 标注流程

为提高语料标注的效率和准确性,本文语料标注方法采用 BIOES 集^[9],B(Begin)表示实体开始,I(Inside)表示故障内部,O(Other)表示非实体的其他部分,E(End)表示实体结束,S(Single)表示实体只有一个字。对于航空发动机故障本文定义的 4 种实体,标记为“B-(实体类型)”、“I-(实体类型)”、“E-(实体类型)”、“S-(实体类型)”以及无关的其他字符均记为 O。依据航空发动机故障文本的实体标注示例,见图 4。

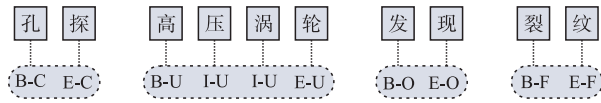


图4 语料标注示例

2.3 发动机故障实体识别模型

为准确识别定义的5类航空发动机故障实体,采用了一种基于BiLSTM-CRF的经典模型,同时通过加入BERT预训练语言模型,将故障文本中的词语转换成词的向量化表示,见图5。

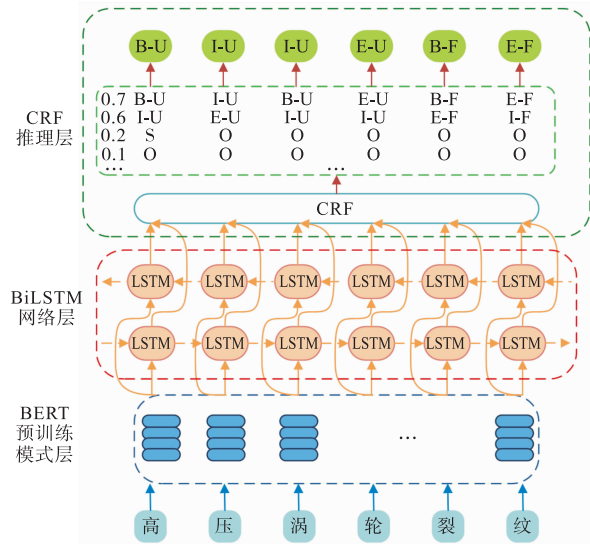


图5 基于BERT-BiLSTM-CRF的故障实体识别模型

模型的第1层是BERT层,它主要负责将原始航空发动机故障文本转换为向量形式,以便后续的双向神经网络进行上下文学习,假设一个句子 W 中含有 n 个字,则该句的向量可表示为 $W = \omega_1, \omega_2, \dots, \omega_n$ 。其中的最主要的内容是得出一个词语在整个句子的相应关系,即用向量的方法表示单个词语在整个语句的关系以及词语和词语之间的相互关系。通过向量形式表达的相互关系在一定程度上,可以表现出这句话中不同词语之间的关联度和重要性^[16]。最终,通过得出的相互关系来改变每个词相应的权重,获得的词向量表征能够更好地结合上下文的语境。利用该方法,得到的词向量既能蕴含其他词与这个词的关系,还能获取一些句子级别的特征,更具有全局性。

模型第2层是双向神经网络层(BiLSTM),用来自动提取句子的特征。能充分利用发动机故障知识文本的上下文依存关系,故障实体识别采用了BiLSTM模型进行建模。BiLSTM以BERT生成的词向量当做输入,通过捕获上下文特征,获取更全面的语义信息。长短期记忆网络^[17](LSTM)是循环神经网络(RNN)的一种变体。和RNN相比较,创造性地引入了记忆单元和门控机制,对上下文历史信息进行有选择性的遗忘、更新和传递。LSTM

单元包含遗忘门、输入门、输出门3种门结构,并通过此种结构来控制LSTM单元的状态。 t 时刻单元状态信息更新可由式(1)、(5)得到。

$$f_t = \sigma(W_f \times [h_{t-1}, x_t] + b_f) \quad (1)$$

$$i_t = \sigma(W_i \times [h_{t-1}, x_t] + b_i) \quad (2)$$

$$o_t = \sigma(W_o \times [h_{t-1}, x_t] + b_o) \quad (3)$$

$$C_t = f_t C_{t-1} + i_t \tanh(W_c \times [h_{t-1}, x_t] + b_c) \quad (4)$$

$$h_t = o_t \tanh(C_t) \quad (5)$$

式中: σ 是sigmoid激励函数; f 分别表示遗忘门; i 表示输入门; o 表示输出门; W 和 b 分别是权重矩阵和偏置项; C 表示单元状态; h 表示隐藏层状态。BiLSTM由一个前向长短期记忆单元与一个后向长短期记忆单元组合,通过把原始有顺序的输入转换成一正一反的2个输入,使得整个模型能同时获得前向和后向的信息,能够更好地获取较长距离的双向语义依赖关系,并解决梯度消失的问题。针对发动机故障实体识别的序列标注任务,拥有更好的优势。

模型的第3层是条件随机场(conditional random field, CRF),是发动机故障实体识别模型的最后一部分,主要用于根据前后文语义之间的依赖关系,约束模型学习到前后文标签之间的内容,在对应输入条件下,得到最大概率输出。CRF是一种条件概率分布模型,可以用 $P(y|x)$ 来表示,为标注的序列赋予对应概率值,有助于选择出最大似然概率的标注序列。其中: x 是输入变量,表示要被标注的观察序列; y 是输出序列,表示与 x 一一对应的标签序列。其核心原理是:

$$P(y|x) \propto \exp\left(\sum_{k=1}^k \omega_k f_k(y, x)\right) \quad (6)$$

式中: f 表示特征函数; ω 表示特征函数对应的权重。

在训练时,利用训练集通过极大似然估计得到条件概率模型 $P(y|x)$ 。在预测时,对于给定的观测序列,利用维特比算法输出可得条件概率 $P(y|x)$ 最大的标签序列 y ,并根据标注序列提取文本中的发动机故障实体。

3 实验运行分析

3.1 评价指标

为了准确验证航空发动机故障实体识别方法的性能优劣,本论文采用准确率(accuracy, P)、召回率(recall, R)以及 F_1 得分(F_1 -score, F_1)来评价模型性能^[15],各评价指标的计算方法如式(7)~(9)。

$$P = \frac{T_p}{T_p + F_p} \times 100\% \quad (7)$$

$$R = \frac{T_p}{T_p + F_N} \times 100\% \quad (8)$$

$$F_1 = \frac{2PR}{P+R} \times 100\% \quad (9)$$

式中: T_p 表示为预测正确的故障实体阳样本数; F_p 表示为预测错误的故障实体阳样本数; F_N 表示为预测错误的故障实体阴样本数。将处理好的数据集导入数据库,针对训练集上的数据进行参数训练,得出最优参数:词向量维度为 300,Drop Out 值为 0.5,一次网络嵌入的数据大小(batch-size)值为 16,代数(epoch)为 50,学习率为 0.001,优化算法使用 Adam。

3.2 实验验证分析

在航空发动机故障关键信息抽取只采用 CRF 模型时,针对不同标注训练样本数进行比较,可以看到当样本数为 50 份时, F_1 值仅为 0.571。因此,在建立提取模型时,通过追加双向神经网络(BiLSTM)、BERT 模型以及样本数量,故障实体识别的 F_1 值提升到了 0.811, F_1 值对比结果见表 2。

表 2 不同模型下故障实体识别 F_1 值对比

样本数	CRF	BiLSTM-CRF	BERT-BiLSTM-CRF
0	0.571	0.583	0.635
100	0.584	0.601	0.702
200	0.591	0.674	0.741
300	0.675	0.721	0.767
400	0.715	0.757	0.804
500	0.753	0.788	0.811

按故障实体类型评价。统计在单元(U)、表征信号(S)、故障状态(F)、检查方法(C)和解决方法(M)这 5 类故障实体识别上的表现,结果见图 6。

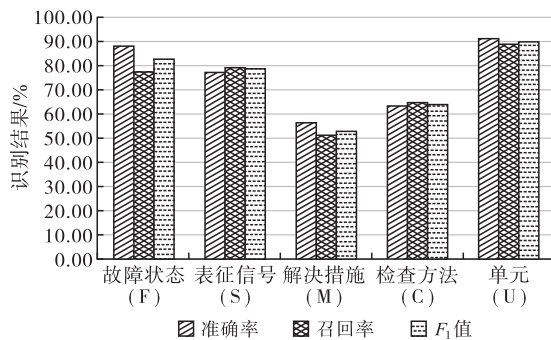


图 6 按类型评估故障实体识别结果

可以看出,单元类、故障状态类和表征信号类故障实体的识别效果较好,能够满足航空发动机故障文本识别任务的准确率要求。但是,检查方法类和解决方法类实体识别的效果仍不够理想,主要原因包括:①没有相关词典对方法概念进行规范;②大量故障文本的表现形式和维修人员群体联系紧密,在不同的维修单位的维修人员描述差异较大;③描述方法的语句过长,故障记录不够精炼。以上几点也是我们下一步要继续提高故障实体识别准确率的主要研究方向。

3.3 实验结果应用

通过上述方法,从收集到的 4.6 GiB 文本数据中准确识别了 15 246 条发动机故障实体。利用抽取到的实体信息,搭建了发动机故障知识图谱,能够将复杂的发动机故障实体关系存储在知识网络中,其中不同颜色的单元代表不同种类的航空发动机故障实体。图 7 为获得的某型发动机滑油压力异常的故障知识图谱。从图中可以得到滑油系统的重要部件单元之间、故障现象和表征信号之间、以及故障现象和故障单元之间密切联系;而且该知识图谱能够根据后续故障数据的不断积累,迭代形成更完备的知识库。

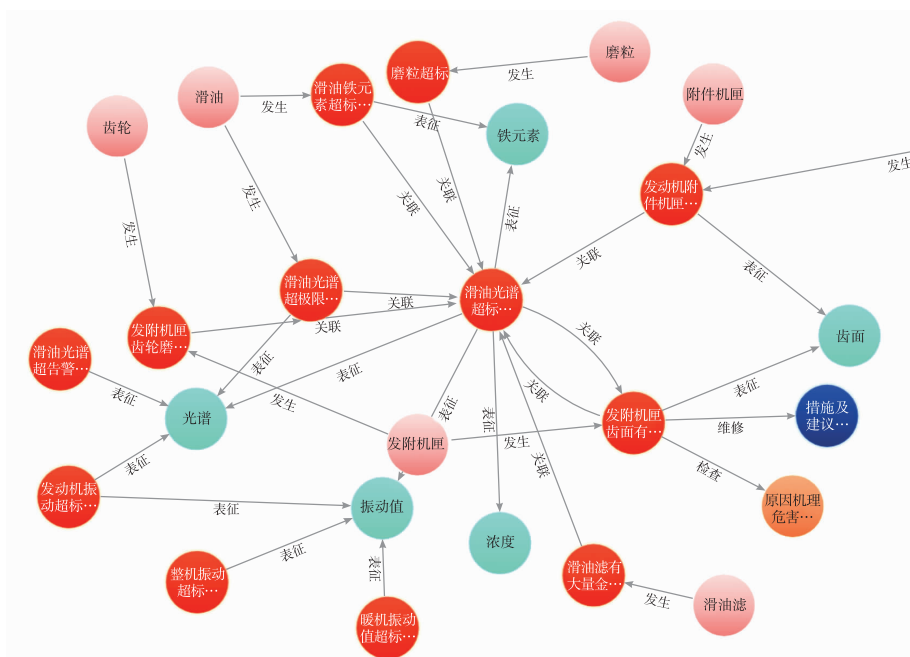


图 7 航空发动机滑油压力异常故障知识图谱

4 结论与展望

为充分利用航空发动机全寿命周期数据,实现基于知识的故障智能诊断,论文以快速准确搭建大规模的发动机故障知识库为目标,初步构建了一种基于BERT-BiLSTM-CRF的航空发动机故障实体识别方法,并通过某型航空发动机故障数据文本进行验证,初步解决了从航空发动机故障文本中自主获取知识并搭建大规模知识库的问题。下一步将在此基础上加强对故障实体识别和故障关系的联合抽取方法研究,并开发航空发动机故障知识图谱和故障智能问答系统等应用产品,助力航空发动机的智能化保障模式发展。

参考文献

- [1] 陈卫,程礼,李全通,等. 航空发动机监控技术[M]. 北京:国防工业出版社,2017:1-10.
- [2] 张凤鸣,惠晓滨. 航空装备故障诊断学[M]. 北京:国防工业出版社,2010:194-203.
- [3] ZHOU A, YU D, ZHANG W. A Research on Intelligent Fault Diagnosis of Wind Turbines Based on Ontology and FMECA[J]. *Advanced Engineering Informatics*, 2015, 29(1):115-125.
- [4] ZHOU Q, YAN P, LIU H, et al. Research on a Configurable Method for Fault Diagnosis Knowledge of Machine Tools and Its Application[J]. *International Journal of Advanced Manufacturing Technology*, 2018, 95(6):1-24.
- [5] MEDINA-OLIVA G, VOISIN A, MONNIN M, et al. Predictive Diagnosis Based on a Fleet-Wide Ontology Approach[J]. *Knowledge-Based Systems*, 2014, 68: 40-57.
- [6] XU F, LIU X, CHEN W, et al. Ontology-Based Method for Fault Diagnosis of Loaders[J]. *Sensors*, 2018, 18(3):729.
- [7] SU Y, LIANG X R, WANG H, et al. A Maintenance and Troubleshooting Method Based on Integrated Information and System Principles[J]. *IEEE Access*, 2019, 7:0513-0524.
- [8] 贾晓霞,谢登峰. 人工智能在航空装备故障诊断中的应用[C]//2019航空装备服务保障与维修技术论坛暨中国航空工业技术装备工程协会年会. 南昌:中国航空工业集团有限公司防务生产与保障部,2019:367-371.
- [9] 吴赛赛,周爱莲,谢能付,等. 基于深度学习的作物病虫害可视化知识图谱构建[J]. *农业工程学报*, 2020, 36(24):177-185.
- [10] 秦健,侯建新,谢怡宁,等. 医疗文本的小样本命名实体识别[J]. *哈尔滨理工大学学报*, 2021, 26(4):94-101.
- [11] 肖仰华. 知识图谱:概念与技术[M]. 北京:电子工业出版社,2019:84-86.
- [12] 马忠贵,倪润宇,余开航. 知识图谱的最新进展、关键技术和挑战[J]. *工程科学学报*, 2020, 42(10):1254-1266.
- [13] 杨飘,董文永. 基于BERT嵌入的中文命名实体识别方法[J]. *计算机工程*, 2020, 46(4):40-45, 52.
- [14] 尹学振. 多神经网络协作的军事领域命名实体识别关键技术研究[D]. 上海:华东师范大学,2020:34-42.
- [15] 许源,葛艳秋,王强,等. 基于CRF与RUTA规则相结合的卒中入院记录医学实体识别及应用[J]. *中山大学学报(医学版)*, 2018, 39(3):455-462.
- [16] 王欢,朱文球,吴岳忠,等. 基于数控机床设备故障领域的命名实体识别[J]. *工程科学学报*, 2020, 42(4):476-482.
- [17] 闵海根,方煜坤,吴霞,等. 基于自动编码器和长短时记忆网络的智能汽车故障诊断方法研究[J]. *四川大学学报(自然科学版)*, 2021, 58(5):053003.

(编辑:姚树峰)