

doi:10.3969/j.issn.1001-893x.2017.07.020

引用格式: 刘丰恺, 李茜. 航空大数据技术的发展与应用 [J]. 电讯技术, 2017, 57(7): 849-854. [LIU Fengkai, LI Qian. Development and application of big data technology in aviation industry [J]. Telecommunication Engineering, 2017, 57(7): 849-854.]

航空大数据技术的发展与应用^{*}

刘丰恺^{**1}, 李茜²

(1. 中国航空综合技术研究所, 北京 100028; 2. 华北计算技术研究所, 北京 100083)

摘 要: 大数据技术在以互联网、电商、金融为代表的第三产业得到了极大的发展和扩充, 然而在航空工业领域, 大数据技术的价值一直没有得到充分利用与挖掘。事实上, 大数据技术在飞行器设计、系统级故障预测以及航空装备维修与保障领域有重大应用价值, 值得进行深入研究探索。首先介绍了大数据的内涵, 按大数据流转的生命周期分析了相关技术在航空领域的应用, 重点阐述了航空大数据分析技术在系统级故障预测领域的应用, 并结合美国大数据分析技术在航空领域的应用现状, 指出了航空大数据技术的发展趋势。

关键词: 航空工业; 大数据; 系统级故障预测; 应用展望

中图分类号: TP274 **文献标志码:** A **文章编号:** 1001-893X(2017)07-0849-06

Development and Application of Big Data Technology in Aviation Industry

LIU Fengkai¹, LI Qian²

(1. AVIC China Aero-Polytechnology Establishment, Beijing 100028, China;

2. North China Institute of Computing Technology, Beijing 100083, China)

Abstract: Big data technology has developed rapidly in such area as Internet, e-commerce and finance. However, the value of big data technology has not yet been fully realized in aviation industry. As a matter of fact, big data technology possesses great potential in the fields of aircraft design, system-level prognostics and aviation repair and maintenance, which deserves in-depth research and exploration. This paper introduces big data's basic concept and its architecture, followed by analysis of applications of related technology in aviation industry under the context of big data life cycle with emphasis on application of big data technology in prognostics. Finally, it points out the developing trend of aviation big data technology in the future according to the application status of U. S.'s big data technology in aviation field.

Key words: aviation industry; big data; system-level prognostics; application prospect

1 引 言

20 世纪 90 年代中期, SGI 公司首席科学家 John Mashey 在一篇题为“Big Data and the Next Wave of InfraStress”的演讲中提出了“大数据”的概念。所谓的大数据, 指数据量太大, 以至于目前的数据管理工具已不便于管理的数据。通常, 用 4V (Volume, Velocity, Variety, Value) 特性对大数据进行描述^[1]: 大

容量 (Volume) 是指数据体量巨大; 高速率 (Velocity) 从数据产生效率的实时性进行描述; 多形式 (Variety) 从数据的类型进行描述, 包含传统的结构化数据、类似于 XML、JSON 等形式的半结构化和非结构化数据; 高价值 (Value) 指具有高价值的数据。如今, 所谓的大数据已经发展为一个较为宽泛的概念, 是包括数据及其采集、处理、分析、解释等在内的一

* 收稿日期: 2017-04-25; 修回日期: 2017-06-22 Received date: 2017-04-25; Revised date: 2017-06-22

** 通信作者: titikai@163.com Corresponding author: titikai@163.com

系列相关技术、方法、手段的统称。

大数据按其流转的生命周期^[2]可分解为数据源采集、数据存储管理、数据分析和数据展现4个阶段。数据源阶段关心的是数据获取。数据既包含传统企业的结构化数据,也包含互联网用户、传感器等采集的半结构化或者非结构化的数据。数据存储管理解决的是大规模数据的持久存储和管理。首先需要从原始数据中进行数据抽取,以获得数据分析所需的数据信息。对于海量数据而言,需要将数据存入分布式文件系统或者 NoSQL 数据库中。广泛使用的 Hadoop 中,通过 HDFS 分布式文件系统进行数据存储,使用 MapReduce 计算框架处理数据。数据分析是整个大数据处理流程中的核心。在数据分析中,利用分析方法或工具对数据进行检查、变换和建模并从中提取价值,依赖利用数据挖掘、机器学习等复杂技术手段从数据中获取知识。机器学习近年的发展非常迅速,其目的是从数据中自动习得模型,并使用习得的模型对未知数据进行预测。数据展示是大数据处理流程的门户,用于将挖掘出的知识以形象易于理解的形式呈现给用户。其中,可视化作为解释大量数据最有效的手段之一率先被科学与工程计算领域采用。通过对分析结果的可视化用形象的方式向用户展示结果,并且图形化的方式比文字更易理解和接受。

传统意义上的大数据应用在第三产业发展较为成熟,例如互联网、金融、电商和电信服务业,这些行业平日里产生了海量的交易数据和用户行为数据,希望依托大数据技术实行更加精准、高效的商业策略,实现企业价值的提升,因而十分注重相关领域的研发和投入,产生了 Google 公司的谷歌文件系统 (Google File System, GFS)、云计算 (Cloud Computing) 概念、Amazon 公司的个性化推荐策略等一大批世界级的大数据研究成果。第三产业俨然已经成为当下大数据潮流的领军者和倡导者,传统的第一产业,例如船舶、航空以及其他高端制造业似乎正从这股大数据潮流中逐渐被边缘化,并大有被新兴互联网企业蚕食的危险。然而,一直以来为很多人所忽视的是,航空行业本身就是一块孕育大数据的沃土。因为本身的行业特性,航空工业需要大量的标准化数据支撑飞机的设计、维修与安全保障,这些数据因为其专业属性不同而有其严格的规范和格式,由飞机本身安置的计算机和数据记录仪自动记录,误码率较低,承载的信息精确,便于数据清洗和日后的准确分析,这为在航空领域开展大数据应用奠定了很

好的基础。近年来,航空工业巨头波音、空客以及通用电气 (General Electric Company, GE) 均投入巨资开展大数据技术方面的研究,显示出了大数据技术在航空工业领域的良好应用前景。

本文结合航空大数据的特点,从大数据处理的4个基本流程为切入点进行剖析;着重阐述航空大数据分析技术在系统级故障预测领域的应用,并结合美国大数据分析技术在航空领域的应用现状,指出了大数据技术在航空领域应用的发展趋势。

2 航空大数据技术简介

航空大数据技术的总体架构同传统大数据架构类似,需要经过数据源采集、数据存储管理、数据分析和数据展现4个阶段。但是,在每个阶段,航空大数据的相关技术又会同传统大数据技术相区别,体现出鲜明的行业特点。航空大数据技术架构如图1所示。

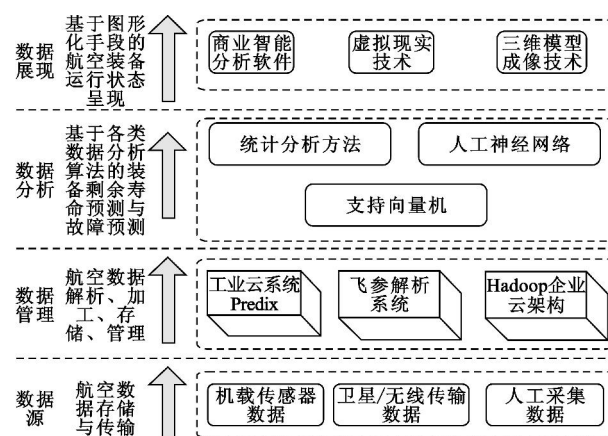


图1 航空大数据技术架构

Fig. 1 Big data architecture for aviation industry

下面结合航空大数据技术架构4个主要组成部分分别进行介绍。

2.1 航空大数据获取技术

航空大数据获取技术是一种利用机载传感器、无线通信设备以及人工手段获取同装备使用状态相关各类数据的综合技术。目前在航空领域,大数据获取技术主要分为3类。

(1) 基于机载传感器的大数据获取技术

主要代表技术又分为民航领域普遍应用的飞参数据获取技术 (Quick Access Recorder, QAR), 航空装备制造厂商如波音、GE、罗罗公司普遍使用的远程诊断数据 (Remote Diagnostics, RD) 采集技术, 以及军用航空装备领域普遍应用的基于飞行数据管理系统 (Flight Data Management System, FDMS) 的数据采

集技术。QAR 飞参数据获取技术通过机载快速存取装置,可同时采集包括飞机飞行姿态、机上设备运行状态、燃油使用量在内的数百种数据。RD 数据采集技术则通过飞机涡轮风扇发动机上的传感器采集发动机在起飞和降落阶段的各种关键参数。例如:GE 的 CFM56 发动机安装有 12 个传感器,每个飞行架次可采集 300~500 条关于发动机温度、压力和振动方面的测量值^[3];基于 FDMS 的数据采集技术则是通过军用飞机内部的飞行数据管理系统收集数据,收集的数据类型涉及飞机的发动机、飞控、大气数据、座舱综合显示和控制、燃油测量、火控、无线电通信、导航和供电等机载系统的各项信息。目前,FDMS 已装备 F-15、F-16、“幻影-2000”等第三代战斗机以及 F-22 第四代战斗机^[4]。

(2) 基于卫星、无线电传输的大数据获取技术

该技术的代表技术为 ACARS 技术。ACARS 全称为 Aircraft Communication Addressing and Reporting System,是一种在航空器和地面站之间通过无线电或卫星传输短消息(报文)的数字数据链系统。通过 ACARS 系统,机组人员和地面可以在飞行过程中实现双向通信,地面人员也可实时了解飞机的飞行状态、气象条件、发动机性能。

(3) 基于便携式辅助维修设备(Portable Maintenance Aid, PMA)以及人工记录的大数据获取技术

基于 PMA 的大数据获取技术允许地勤人员通过便携式计算机现场收集同装备运行状态相关的关键数据,辅助维修决策。基于人工记录的大数据获取技术则多为手工现场记录,记录的数据多为部队用户在实际使用装备过程中得到的数据,这些数据往往通过地勤人员现场记录,汇总后统一录入部队装备信息管理体系。

2.2 航空大数据存储技术

航空大数据存储技术是一种通过利用各种数据库系统对获取的各类航空数据进行解析、加工、存储和管理的综合性技术。该项技术和传统大数据存储技术的区别在于有专门的工业解析协议来解析收集的航空数据。例如:GE 开发的 Predix 工业互联网平台可支持 7 种工业通信协议,并具备相关工业设备运行数据的实时传输与解析能力,其中便内嵌有针对飞机 QAR 参数的数据解析与存储模块。数据存储和管理方面,考虑到在飞机设计、制造、试飞以及运营过程中产生的海量结构化以及非结构化数据,航空制造企业通常采用分布式架构存储和管理航空数据资源。分布式存储架构同时还可满足工业部门对于飞机设计与制造环节数据共享、数据挖掘与分

析,以及可视化展示等特定应用功能扩展的需求,目前在 GE 的 Predix 平台上有专门的航空大数据湖,存储的航空发动机数据以及飞机状态运行数据已经超过 PB 量级且在不断扩充中。GE 同时还应用 SAP、Oracle 等企业资源管理软件,来进一步延伸 Predix 平台对于航空数据规范化管理与使用的功能。此外,航空大数据存储技术对存储装置的可靠性要求较高。例如:QAR 机载飞参记录装置本身就是个小型的数据存储器,能够在飞机飞行过程中以较高精度实时收集和存储各类飞行参数。

2.3 航空大数据分析技术

航空大数据分析技术是利用包括统计分析、支持向量机、人工神经网络在内的多种数据挖掘与分析手段对收集的航空装备状态数据进行分析与挖掘,提炼数据中隐含的反映装备运行状态的规律,从而实现对装备故障的预测,本文稍后将对此部分展开介绍。

2.4 航空大数据可视化技术

航空大数据可视化技术是一种借助图形化手段,清晰有效地传达同航空装备运行状态相关的各项信息的技术,代表技术为虚拟现实技术。例如:在飞行器设计阶段早期,虚拟现实技术利用数字样机将海量数据可视化,以更直观的方式提升设计人员对复杂系统的理解。其他航空大数据可视化技术还包括基于装备损伤数据的三维建模技术。例如:来自加拿大 NGRain 公司开发出的数字化解决工具可允许地勤人员现场实时收集战斗机的损伤数据并进行三维结构建模,极大降低了地勤人员主观误判的可能,有效提高了战斗机维修的效率,该项技术业已应用于美军 F-22 和 F-35 战斗机项目^[5]。

3 航空大数据分析技术在系统级故障预测领域的应用

3.1 飞行器系统级故障的特点

目前,航空大数据分析技术最有前景的应用集中在飞行器系统级故障预测领域。现代飞行器是各种系统的复合体,单独系统的失效往往会牵扯到邻近系统的正常运作,进而导致飞行器的整体失效,因此,对现代飞行器系统级别的故障预测一直是航空业界的研究热点之一。目前,对于飞行器系统级别的故障预测有 3 个棘手的问题:一是个体和个体之间的互相作用(博弈),例如飞行器内部众多的成分和数量及其特征的多样化、系统的层次性、成分间的

交互耦合行为等;二是飞行器自身系统与外部环境之间的相互作用(反馈);三是涌现性。在系统科学中,若干部分按照某种方式整合成为一个系统,就会产生出整体具有而部分或部分总和所没有的东西,系统科学把这种整体才具有、孤立的部分及其总和不具有的特性称为整体涌现性(Whole Emergence)。飞行器自身系统在上述3个要素的共同作用下形成复杂模式(Pattern),对飞行器自身系统行为的预测本质上即为对系统复杂模式的预测。

传统基于故障物理模型或者专家知识库的故障预测方法由于依赖先验知识和系统内部逻辑关系的描述,难以有效处理飞行器复杂系统普遍存在的高维度、非线性以及涌现性问题。而航空装备数据资源的日益丰富以及机器学习方法的再次兴起,则为解决上述问题提供了一条新的思路。基于大数据分析技术的飞行器系统级故障预测,即在系统、子系统、部件、相关使用环境、人为影响因素等方面产生的多维海量数据基础上,运用机器学习以及统计分析的手段对故障样本库进行训练,在历史数据基础上预测系统复杂故障模式演变趋势,从客观数据的角度而不是传统故障机理或者经验知识的角度推测飞行器系统在未来某一时间内发生故障的类型,从而实现飞行器系统级故障的预测。

3.2 美国大数据分析技术在航空故障预测领域的应用现状

美国运用大数据分析技术在航空领域应用的起步较早,理论研究和工程应用结合较为紧密,并注重融合主流的科技成果和理念。早在20世纪70年代,美国国家航空航天局(NASA)便提出了基于传感器数据的航天器综合健康管理概念,并在此概念基础上衍生出了飞机状态监测系统、发动机监测系统以及综合诊断预测系统。20世纪末,随着美军F-35(Joint Strike Fighter, JSF)项目的启动,上述解决方案被正式命名为预测与健康管理系统((Prognostics and Health Management, PHM)系统,实现了基于系统状态数据的故障诊断方法从理论研究向装备型号应用的转变。2000年以后,随着互联网技术和计算科学的飞速发展,西方发达国家开始尝试使用机器学习的手段处理高维海量工业数据,识别复杂装备内部系统的故障发生规律。其中,美国联邦政府科研机构 and 大型航空制造企业下属的研究团队均在不同程度地进行基于大数据和机器学习的装备故障预测研究,目前已知的研究方向包括数据驱动的方法、数据驱动和物理模型混合的方法、基于深度学习的

剩余寿命预测、基于深度学习的故障预测等。下面简要介绍3种不同的航空大数据分析技术并浅析它们各自的应用方向。

3.2.1 NASA 数据驱动的设备剩余寿命预测技术

作为世界顶尖的航空和航天研究机构,NASA一直致力于复杂航空航天装备故障预测领域的研究,并拥有专门的故障预测卓越中心(Prognostics Center of Excellence, PCoE)进行装备故障预测方面的研究,其中一项成果便是由 Kai Gobei 提出的数据驱动的设备剩余寿命(Remaining Useful Life, RUL)预测技术。该项技术强调要在训练数据数量、运算能力、算法鲁棒性、不确定性管理、预测准确度等因素综合考虑与权衡的基础上选取合适的方法进行预测。对应于数据,NASA提出了两种策略:一种是把 n 维特征数据映射为1维健康因子数据,然后采用曲线拟合、外推等方式进行预测;另一种是直接进行 n 维数据的模式匹配,构建特征空间的 n 维数据的退化特征,并给出对象目标的剩余寿命,如图2所示。

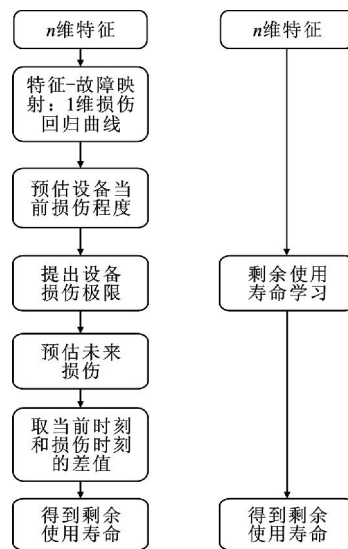


图2 NASA的设备剩余寿命预测策略示意图

Fig. 2 Illustration of NASA's prognostics strategy for equipment's RUL

在对一组太空旋转装置故障前后的时序测试数据研究中,NASA根据装置损伤模型的特性分别提出了基于相关向量机、人工神经网络和高斯回归过程的3种剩余寿命预测方法,其中相关向量机以及人工神经网络方法的结果大部分都位于预测准确度区间内,取得了较好的预测效果^[6]。NASA团队的基于3种剩余寿命预测方法的结果对比如表1所示。其中,负值表明预测的故障发生点晚于实际故障发生点,因此被认为是失败的预测。

表 1 3 种不同方法下设备剩余寿命(RUL) 预测误差的结果对比
Tab.1 Numerical results of three algorithms during RUL prediction
时间单元

剩余寿命	预测误差		
	神经网络	相关向量机	高斯回归过程
1 637	337	207	201
1 137	227	117	17
637	77	17	-83
137	-283	17	-83

3.2.2 UTC 基于深度学习的故障预测技术

美国联合技术公司(UTC)是一家多元化制造企业,同时也是重要的航空装备供应商,旗下的子公司普惠公司是世界领先的飞机发动机设计、制造与服务商。目前,联合技术公司正在进行基于深度学习的故障预测研究。在 2016 年公布的一段视频中,来自联合技术研究中心的工程师通过应用深度信念网络(Deep Belief Network, DBN),在普惠发动机采集的历史传感器数据基础上构建了发动机的运行状态还原模型,并成功实现了对发动机状态参数值的预估。同时,使用深度自编码器(Deep Auto-Encoders)降低了模型重塑的误差,提高了发动机状态预测的效果^[7]。在基于深度学习模型预测的发动机运行状态基础上,机载实时数据分析系统可实时监控飞机发动机的工作状态并预测其变化趋势,并将相关信息传送给飞机任务系统,具体流程如图 4 所示。

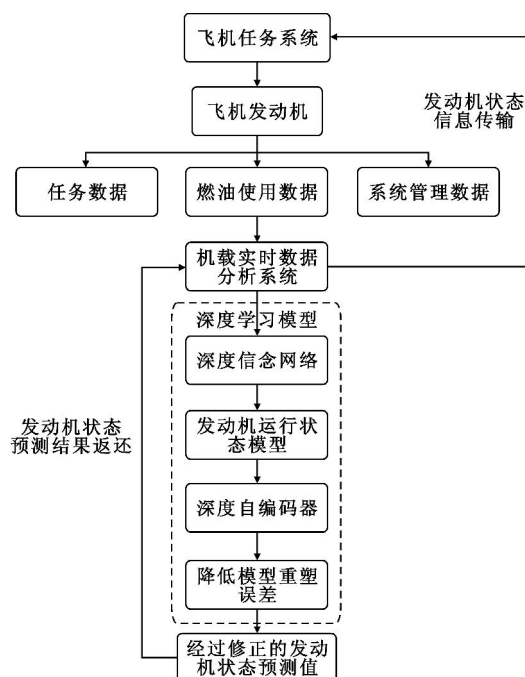


图 4 UTC 基于深度学习的发动机状态预测技术示意图
Fig.4 Deep-learning based approach for engine prognostics

3.2.3 通用电气 GE 的数字双胞胎模型

美国通用电气公司(GE)是世界上最大的提供技术和业务服务的跨国公司,其旗下的 GE 航空集团同时也是世界领先的民用、军用、公务及通用飞机喷气、涡浆和涡油发动机及部件和集成系统的供应商。GE 计划于 2017 年部署数字双胞胎模型(Digital Twin),如图 5 所示。Digital Twin 本质上是一种基于喷气发动机传感数据、数字化发动机物理模型以及虚拟现实技术的综合大数据分析技术,通过将发动机的操作数据添加到对应的数字双胞胎模型中进行虚拟现实操作,数字模型可以给出发动机零部件的更换建议,GE 据此可以得出一个基于环境变化而不是时间变化的发动机维护计划,从而使得维护计划更有针对性,极大地提高了 GE 旗下物理资产的利用价值^[8]。

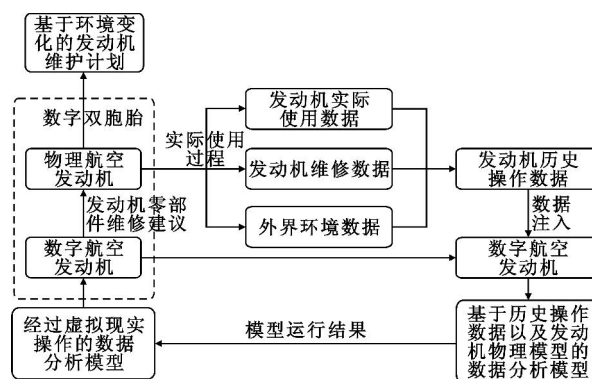


图 5 GE 数字双胞胎技术示意图
Fig.5 Illustration of GE's Digital Twin

4 航空大数据技术展望

根据对国外航空大数据技术发展与应用的总结分析可知,航空大数据技术的发展在未来会呈现以下几个趋势:

一是数据存储量较以往将更加庞大,尤其是新一代民用客机和军用飞机,机体内部的传感器数量多,各系统之间信息交联较以往更加复杂,采集的相关参数种类也更加多样化。航空数据资源的日益丰富,也将进一步扩展航空大数据技术应用的场景与领域。

二是随着分析数据量的增多以及航空装备本身存在的系统耦合性,单纯的统计分析方法将难以准确描述航空装备本身状态的变化,以深度学习为代表的人工智能技术不依赖复杂系统内部逻辑关系的表达,对于航空装备复杂系统级故障的耦合非线性问题有较强的抗干扰能力,在航空大数据分析技术

中将扮演重要角色。

三是随着虚拟现实技术的引入,大数据可视化技术有望有效提升工程人员的感知与分析能力,进而实现更为高效的人机交互,从而推动大数据可视化技术从单纯的数据解释与表达功能向飞行器设计、装备制造与维修过程中的排故等实用功能的转变。

5 结束语

综上,以美国为主的西方发达国家意识到在迅猛发展的信息化社会,数据是重要的战略资源,已率先在航空领域开展基于大数据技术的研究与应用。通过建立规范化收集、利用和整合各方面数据资源的机制,研究深度挖掘海量数据的新手段和方法,构建能够存储和处理海量数据的软硬件能力,西方发达国家在航空大数据技术领域的研究进展较以往达到了一个新的高度。同时,通过集成最新的人工智能技术、虚拟现实技术,美国研究人员极大地拓展了大数据技术在航空工业领域应用的内涵,为处理海量高维航空数据以及复杂系统级故障难以预测的问题探索出了一条新的方向。“十三五”期间,我国将全面启动实施航空发动机和燃气轮机重大专项,同时包括 C919、C929 大飞机在内的一大批航空科研项目也将进入关键节点,亟待大数据技术在上述项目实施过程中的保驾护航。通过以大数据为资源和技术抓手大力提升航空领域的信息化支撑水平,大数据技术将会在航空领域获得更大程度的探索和应用,从而进一步推动大数据+航空的领域化技术研究与发展。

参考文献:

- [1] 马世龙,乌尼日其其格,李小平. 大数据与深度学习综述 [J]. 智能系统学报,2016,11 (6): 728 - 742.
- [2] 张俊林. 大数据日知录 [M]. 北京:电子工业出版社,2014.
- [3] SAFRAN. Diagnostics and management [EB/OL]. [2017 - 04 - 10]. [https://www. safran - aircraft - engines. com/ services/ single - aisle - commercial - jet - engines/ engi- neering - and - consulting - services](https://www.safraan-aircraft-engines.com/services/single-aisle-commercial-jet-engines/engineering-and-consulting-services).
- [4] 张建业,张鹏. 飞行数据的时间序列分析方法及其应用 [M]. 北京:国防工业出版社,2013.
- [5] NGRain. Battle damage assessment and repair for the F - 35 and F - 22 [EB/OL]. [2017 - 04 - 10]. [http://www. ngrain. com/ lockheed - martin](http://www.ngrain.com/lockheed-martin).
- [6] National Aeronautics and Space Administration. Results and Discussions [EB/OL]. [2017 - 04 - 10]. [https:// ti. arc. nasa. gov/ tech/ dash/ pcoe/ data - driven - prog- nostics/ results](https://ti.arc.nasa.gov/tech/dash/pcoe/data-driven-prognostics/results).
- [7] REDDY K K. Applying deep learning for prognostic health monitoring of aerospace and building systems [EB/OL]. (2016 - 11 - 07) [2017 - 04 - 10]. [http://videlectures. net/ kdd2016_ reddy_ deep_ learning](http://videlectures.net/kdd2016_reddy_deep_learning).
- [8] BABCOCK C. GE plans software platform for creating 'Digital Twins' [EB/OL]. (2016 - 07 - 21) [2017 - 04 - 10]. [http://www. eetimes. com/ document. asp? doc_ id = 1330173](http://www.eetimes.com/document.asp?doc_id=1330173).

作者简介:



刘丰恺(1987—),男,北京人,硕士,工程师,主要研究方向为基于大数据的故障预测技术;
Email: titikai@163.com

李茜(1983—),女,山西人,硕士,高级工程师,主要研究方向为大数据技术。