

面向飞机 PHM 的大数据分析与人工智能应用

景 博, 焦晓璇[✉], 黄以锋

(空军工程大学航空工程学院, 西安, 710038)

摘要 针对新一代电子产品密集型的飞机,其系统结构和失效机理复杂、失效模式多样,传统的方法难以对其进行有效健康管理的现状,因此提出了基于海量数据挖掘的飞机 PHM 研究。首先从系统结构、服役环境、数据来源与存储方式等角度分析了新一代飞机对海量数据挖掘的应用需求;其次论述了基于海量数据挖掘的关键技术,包括数据的预处理、集成管理、聚类、分类、关联、预测等;最后提出了一种基于私有云的飞机 PHM 海量数据挖掘平台,详细阐述了该平台的总体框架和软硬件结构,该平台为飞机 PHM 提供了验证平台,对促进飞机 PHM 的集成与工程化实现具有重大的军事应用意义。

关键词 新一代飞机;海量数据;数据挖掘;故障预测与健康管理;私有云

DOI 10.3969/j.issn.1009-3516.2019.01.008

中图分类号 V267;TP277 **文献标志码** A **文章编号** 1009-3516(2019)01-0046-09

Massive Data Mining and Artificial Intelligence Application for Aircraft PHM

JING Bo, JIAO Xiaoxuan[✉], HUANG Yifeng

(Aeronautics Engineering College, Air Force Engineering University, Xi'an 710038, China)

Abstract: Aimed at the problems that the new-generation aircraft is characterized by electronic product-intensive, complex system structure and failure mechanism, multiple failure modes, difficult to carry out effective fault prediction and health management by the traditional methods. A PHM(Prognostic and Health Management)method based on massive data mining of aircraft is presented in this paper. Firstly, the characteristics of the new generation aircraft are analyzed from the aspects of system structure, service environment, and data source and data storage mode. Secondly, the key technologies based on massive data mining are discussed, including data preprocessing, integrated management, diagnosis and prognostic model and so on. Finally, a massive data mining platform of aircraft PHM based on private cloud is proposed. The overall framework, hardware and software structure of the platform are described in detail. The massive data mining platform of avionics system can quickly and effectively carry out data mining, and provide a verification platform for the aircraft PHM, which is an event of great significance in promoting the integration and engineering realization of aircraft PHM.

Key words: new generation aircraft; massive data; data mining; prognostic and health management; private cloud

收稿日期: 2018-08-29

基金项目: 陕西省自然科学基金(2017JQ6034)

作者简介: 景 博(1965—),女,河北邯郸人,教授,博士生导师,主要从事故障预测与健康管理、测试性设计、传感器网络与信息融合研究。

E-mail:jingbo_sensor@163.com

通信作者: 焦晓璇(1990—),男,山西运城人,博士生,主要从事信息融合、故障诊断与预测研究。E-mail:564325155@qq.com

引用格式: 景博,焦晓璇,黄以锋.面向飞机 PHM 的大数据分析与人工智能应用[J].空军工程大学学报(自然科学版),2019,20(1):46-54. JING Bo, JIAO Xiaoxuan, HUANG Yifeng. Massive Data Mining and Artificial Intelligence application for Aircraft PHM[J]. Journal of Air Force Engineering University (Natural Science Edition), 2019, 20(1): 46-54.

随着信息技术的不断发展以及德国“工业 4.0”和“中国制造 2025”战略的进一步深化,未来全空域全天候多元素战争模式对装备高技术超性能的迫切需求,飞机等武器装备的系统结构日趋复杂,功能集成度越来越高,运行环境恶劣,导致飞机失效模式多样,故障传播影响复杂,传统的基于“事后维修”和“计划维修”被动维修保障模式已经难以满足现代化战争对飞机战备完好率的需求。而基于视情维修的预测与健康管(Prognostic and Health Management, PHM)技术是提高飞机安全性、可靠性和经济性的关键支撑技术^[1-2]。

我国在“国家中长期科学和技术发展规划纲要 2006~2020”中将“重大产品和重大设施寿命预测技术”作为前沿技术提出,而且在“十二五”期间,我国已经开始进行飞机 PHM 应用基础研究,虽然已经取得了一定的成果,但是 PHM 的主要研究仍然处于框架研究和理论跟踪的阶段,对 PHM 算法的仿真研究较多,结合实际数据的具体应用较少^[3]。以某型飞机 PHM 为例,当前只能实现机载 BIT 数据集成与集中管理,故障诊断与综合处理以及初步的健康管理能力,距离 PHM 系统的完全成熟化还有很大的差距。

此外,随着信息化和工业化的深度融合发展,为了确保飞机的健康运行,集成了大量的机载传感器的飞机全寿命周期数据规模急剧增加,数据类型繁多,除结构化数据外(状态监测数据),还包括大量的外场维护信息(图片、文本、视频)等半结构化、非结构化的数据。据资料统计,国内某型飞机每小时就能记录超过 20 GB 的数据。海量数据的出现将会引发数据存储、数据集成管理、数据挖掘分析以及可视化等技术的变革,为飞机的维修保障模式改革带来机遇与挑战。目前,飞机外场虽然积累了海量的数据,解决了飞机 PHM 数据源头的问题,但是数据存储分散,各部门之间的信息共享困难,造成了信息孤岛的现象。同时由于缺乏大数据分析方法,数据中隐含的关系和规则难以理解,无法根据现有的状态数据预测未来的发展趋势,不能实现基于状态的主动维修,会造成海量数据的巨大浪费,因此,如何高效地利用这些数据实现飞机智能的故障诊断和预测与健康管,成为新一代飞机 PHM 发展面临的重要问题^[4]。

因此本文针对新一代飞机高综合化、高复杂度和高耦合度导致的传统故障诊断推理策略难以满足现代维修保障需求的问题,在分析飞机系统复杂结构和海量多源异构数据特点的基础上,论述了飞机海量数据挖掘的关键技术,最后设计了基于私有云

的飞机 PHM 海量数据挖掘平台,为飞机 PHM 成熟化和工程化提供了试验验证平台。

1 飞机 PHM 系统海量数据挖掘的必要性

未来战争向着体系化、信息化、网络化方向发展。作战环境日益严酷,飞机在现代化战争中发挥的作用愈发重要,飞机的安全性、可靠性、维修性、经济性等要求进一步提高。除了系统结构复杂外,飞机服役环境的时空多尺度性、全寿命周期数据体量的巨大性以及数据结构的多样性,也给飞机的异常状态检测、故障诊断、剩余寿命预测和健康管带来了巨大的挑战。下面从 3 个方面分析飞机 PHM 系统海量数据挖掘的必要性。

1.1 飞机系统结构的复杂性

随着飞机及机载设备的信息化程度不断提高,系统功能结构耦合越来越高,故障传播影响越来越复杂,发生故障的概率也不断增大。尤其对飞机来说,一旦发生故障,很有可能造成机毁人亡的灾难性事故。目前,通过对飞机外场维护现状的调研可知,外场保障主要采取定期维修和事后维修的保障方式,显然这种被动式的维修保障方式,保障效率低,维修代价高,不能根据系统当前状态,结合历史数据,预测系统未来状态,已经无法满足现代化战争对于机电一体化飞机的发展需要^[5]。而且新一代飞机结构复杂,难以构建其准确的数学或物理解析模型来进行故障诊断、状态估计和寿命预测,因此急需海量数据挖掘技术支撑飞机的视情维修^[6]。

1.2 服役环境的时空多尺度性

飞机及机载设备都属于复杂系统,服役环境极其严酷,通常工作在高低温频繁转换、振动与冲击等恶劣环境中,导致飞机及机载设备容易发生失效^[7]。在气候方面,我国南北地区纬度跨度较大,形成了不同的气候特点,如沿海盐雾地区、高温多雨地区、炎热梅雨地区、温暖潮湿地区、温差大地区、寒冷地区等,不同气候环境下飞机及机载设备的退化失效机理、失效模式会出现较大的差异。除了地域环境复杂外,飞机及机载设备还要承受在飞行过程中的振动、过载、电磁环境等应力因素的影响,而空中出现的故障在地面无法复现^[8]。因此,飞机及机载设备从微观到宏观的退化失效过程和故障诊断、异常监测、寿命预测都必须考虑服役环境的应力因素。

1.3 数据来源和存储方式的多样性

飞机在全寿命周期使用维护过程中产生的所有数据统称为飞机大数据,如图 1 所示,这些数据包括

飞机设计阶段数据、试飞阶段数据和使用维护阶段数据。

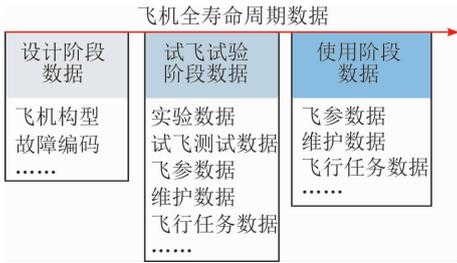


图1 飞机 PHM 大数据

其中,设计阶段数据主要包括飞机构型数据、故障编码数据等。试飞阶段数据主要包括设备在地面试验中采集的数据,在试飞过程中加装测试设备获取的数据,机载飞参记录数据、飞行任务数据等。使用阶段数据主要包括机载飞参记录数据、地面维护保障数据等。

地面维护保障数据是飞机在维护保障过程中生成的数据,如维护保障信息系统中的数据、履历本及维修工艺卡片等信息,这些信息包含了飞机中各个部件的故障数据、故障排除案例、维修记录以及使用环境等信息。

飞参数据主要记录飞机重要系统和部件的工作情况、状态信息等。由于飞机集成了大量的不同类型的传感器,采样周期、采样精度、数据格式、量纲等千差万别,而且没有统一的采样标准和存储规范,形成了海量多源异构的传感器数据。和第三代飞机相比,四代机记录的参数由上百种增加到上千种。随着飞机服役时间的增长,这些数据不断积累,成为飞机大数据的重要组成部分^[9]。

因此飞机的全寿命周期数据除了结构化的数据之外,还包括大量的文本信息、图片、视频等与故障相关的非结构数据,这给数据之间的关联分析带来极大的挑战。此外,由于各个部门之间缺乏统一的数据管理平台,导致大量的数据存储在不同的地方,存在数据孤岛的现象,无法进行数据的关联性分析和进一步利用海量数据实现飞机的故障诊断和故障预测。

2 飞机 PHM 海量数据挖掘的关键技术

新一代飞机结构耦合紧密,失效机理复杂,而且不同服役环境应力对飞机的失效和退化影响也千差万别,传统的基于单一故障假设和单参数的故障诊断和预测方法难以有效、全面地描述系统退化过程和失效模式,急需基于大数据分析和人工智能相结

合的飞机 PHM 系统^[10]。大量机载传感器状态监测数据、总线数据、地面维护数据等为基于大数据分析和人工智能的故障预测与健康管理的提供了数据基础。因此深入挖掘飞机全寿命周期内海量数据的潜在价值,为飞机的故障诊断和寿命预测提供了新的途径,对提高飞机运行可靠性和安全性,降低全寿命周期费用有重要的意义。

飞机全寿命周期数据完全符合大数据的 4V 特性,即数据体量巨大(Volume)、快速的数据流转和动态的数据体系(Velocity)、数据的多样性(Variety)、数据价值密度低(Value)^[11]。但飞机的海量数据与商业大数据不同,商业大数据分析倾向于单纯依赖统计工具挖掘属性之间的相关性,而飞机海量数据分析需要重点关注数据背后的物理意义以及特征之间关联性的机理逻辑。本节针对飞机海量多源异构数据的特点,从海量数据的预处理、集成管理、数据挖掘方法等关键技术展开论述。

2.1 飞机 PHM 海量数据预处理

海量原始数据中存在大量的有噪声、异常、不一致的数据,严重影响了飞机故障诊断和寿命预测的精准度,因此非常有必要进行海量数据的预处理研究。海量数据的预处理主要包括数据清洗、数据集成、数据变换和数据归约,具体内容见图2。海量数据预处理从具体任务、系统边界、参数特性、噪声类型等角度出发,滤除与数据挖掘无关的冗余数据,改善数据质量,提高数据挖掘的效率。

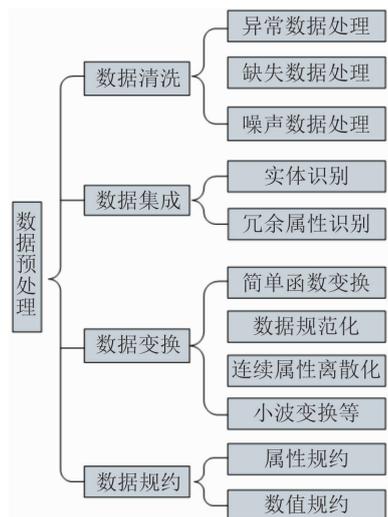


图2 海量数据预处理框架

数据清洗包括异常数据处理、缺失数据处理、噪声数据处理^[12]。针对缺失数据,可以通过删除记录或者插值的方法进行补充。删除记录的方法是通过减少历史数据来获得数据的完备,但这样可能会失去大量隐藏在缺失数据中的信息。常用的数据补充

方法包括基于模型和基于数据驱动的方法。基于模型的数据补充方法需要深入理解系统的工作机理和结构,而飞机系统结构复杂,因此基于模型的数据补充方法并不适用飞机缺失数据处理。常用的基于数据驱动的数据补充方法包括基于时间序列分析^[13]、灰色关联分析、时-空相关性算法等。Wang 等^[14]提出了一种基于聚类和支持向量回归的缺失值处理方法,首先通过聚类将数据进行分类,然后通过支持向量回归估计缺失值,仿真结果证明了该方法的有效性。

在飞机全寿命周期的海量数据中通常包括异常数据点。异常数据的主要来源有:传感器的测量误差、人为操作误差、系统出现故障或其他因素的影响,因此有必要对海量数据中的异常值进行剔除、重构或者挖掘分析。Gu^[15]等人提出一种迭代聚类的方法对异常值进行检测,该方法也适用于高维数据的异常值检测。朱倩雯^[16]等人提出采用 4 位数的异常值检测方法,然后通过多点 3 次样条插值方法重构异常数据。异常数据也可能是由于系统故障导致的结果,因此对具有异常值的数据集进行数据挖掘能够有效检测系统故障。Wei 等人^[17]提出一种自相关核回归的在线异常检测方法。Liu 等人^[18]利用信息论和高斯回归过程对传感器的异常数据检测,通过 NASA 的数据和 2008PHM 挑战数据进行了验证,证明了该方法的有效性。

在数据的采集、传输、存储过程中都可能会因为不同的原因产生噪声数据,常用的噪声数据处理方法有经验模态分解、高斯滤波和回归等方法。严英杰等^[13]人将均值估计的变点思想应用于机械信号的噪声处理当中,该方法简单,处理效果较好。Yadav 等人^[19]针对 ECG 信号,提出一种非局部小波变换的信号去噪方法,利用样本之间的相关性,将样本分成矩阵,然后通过二维离散小波变换系数进行去噪,实验结果表明该方法从定性和定量 2 个方面提高信号去噪的性能。

2.2 飞机 PHM 海量数据集成管理

现有的飞机全寿命周期数据分散,大部分分散存储于计算机、存储设备以及文本资料中,没有统一的数据库管理方案,数据呈现出“信息孤岛”的状态,难以为飞机海量数据的分析和处理提供保证^[20]。分布式文件系统将数据存储于多个物理上分散的存储节点,然后对节点资源进行统一的调度和管理,是海量数据存储的重要技术。目前常用的大数据运行平台包括 Spark、Hadoop 等。Spark 是 UC Berkeley AMP Lab 实验室开发一种开源的基于内存计算的分布式计算系统,能够对大数据进行快速分析处

理^[21]。Hadoop 是 Apache 组织的一个可以在集群上进行海量数据分布式处理的软件框架,具有可靠、高效、可扩展性强的特点,可以通过增加硬件资源的方式实现存储规模的扩展^[22]。Kang^[23]等人构建基于私有云的 Hadoop 平台处理 MapReduce 应用程序,提高了运算效率。Zaharia^[24]利用 Spark 提供的统一建模方法简化了数据挖掘任务。随着网络信息技术的发展,各种云存储也是大数据存储的发展趋势,如阿里云、腾讯云等。

2.3 飞机 PHM 海量数据挖掘算法

飞机 PHM 海量数据挖掘方法包括聚类、分类、关联和预测等方法,分别用来实现相应的功能需求。

2.3.1 聚类方法

聚类是根据最大化类内相似性、最小化类间相似性的原则,将相似的数据、案例分到同一个类别的方法。常用的聚类方法分为:基于划分的聚类方法、基于层次的聚类方法、基于密度的聚类方法、基于网格的聚类方法、基于模型的聚类方法。典型的方法有 K-means, K-中心点, C 均值算法、神经网络等。Tang 等^[25]利用聚类的方法对板级焊点的失效模式进行识别。Yiakopoulos 等^[26]使用基于 K-means 的聚类方法对工业环境中的滚动轴承进行故障检测。

2.3.2 分类方法

分类方法不同于聚类,它建立在已有类标签的数据集上,然后将未知属性的数据映射到给定的类别,它是一种有监督的学习方法,可以用于海量数据的故障诊断。常用的分类方法有:决策树、C4.5、SVM、KNN、ID3 算法等。其他的分类方法还包括基于案例推理、遗传算法、粗糙集、模糊集的分类。Muralidharan 等^[27]将 SVM 分类方法应用于离心泵的故障诊断。Aydin 等^[28]提出一种基于模糊决策树和重建相位空间边界分析的自动化故障诊断方法。Lin 等^[29]利用基于 C4.5 和专家知识数据的方法对卫星网络故障进行故障诊断。

2.3.3 关联方法

数据挖掘的关联方法主要是分析 2 个或多个变量之间存在的但没有直接表现出来的规律性关系,可以用来挖掘飞机系统不同变量之间的内在关系。常用的关联方法有 Apriori 算法、DHP 算法、partition 算法、灰色关联算法等。Huang 等^[30]利用 Apriori 算法分析船舶交通事故的潜在因果关系,能够有效防范事故的方法。

2.3.4 预测方法

预测主要是建立不同变量之间相互依赖的模型,然后对未来的状态进行预测。与分类相似,但分

类预测的结果是类别,不是连续的状态值。当前国内外学者针对预测方法进行了大量的研究^[31],大致可以分为以下3类:①基于失效物理模型的方法;②基于数据驱动的方法;③基于融合的方法。

1) 基于失效物理模型的方法

基于失效物理模型的方法是建立在对产品失效机理、物理模型或数学模型已知的基础上,通过全寿命周期内的载荷和失效机理建立失效模型,从而进行故障诊断与预测^[32]。基于失效物理模型的方法已经在航空结构疲劳寿命预测、电路板级焊点失效等方面广泛应用。典型的失效物理模型方法有:基于 Paris 模型的裂纹扩展建模、基于 Forman 规律的裂纹扩展建模、基于疲劳剥落扩展模型、基于刚度的损伤规律模型等。基于失效物理模型的方法要求的数据量少,结合环境载荷(振动、温度等)和失效模型可以得到精确的预测结果,主要应用于新产品或新系统的研制阶段。但是对于综合模块化的飞机系统,具有较强的结构复杂性和失效随机性,难以建立准确的数学模型或物理模型,大大影响了该方法的使用范围。基于失效物理模型的故障预测方法流程图见图3。

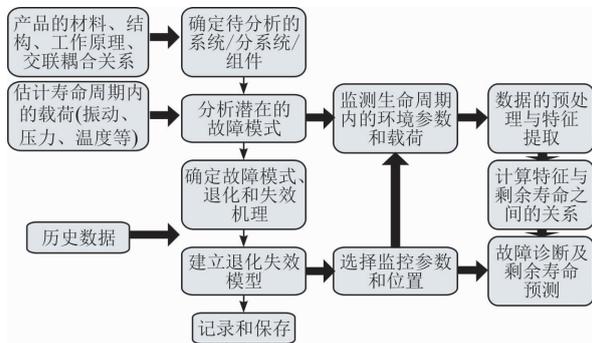


图3 基于失效物理模型的 PHM 方法流程

2) 基于数据驱动的方法

基于数据驱动的方法是利用系统运行过程中监测到的状态数据和全寿命周期的历史数据等数据,采用统计方法、智能机器学习方法或随机过程的方法,从统计和概率的角度出发,对系统的未来健康状态和可靠性进行估计和推断,从而实现对系统的状态评估、故障诊断及寿命预测^[33]。典型的基于数据驱动的方法有:频谱分析、小波分析、隐马尔科夫模型、决策树、贝叶斯网络、神经网络、支持向量机^[34]、扩展卡尔曼滤波、粒子滤波^[35]、ARMA、相关向量机(Relevance Vector Machine, RVM)^[36]等。Tobon 等^[37]提出一种基于高斯隐马尔科夫模型的轴承剩余使用寿命预测方法,该方法分为离线阶段和在线阶段,离线阶段提取传感器数据特征并训练模型参数,然后第2阶段利用模型评估当前轴承的健康状

态。基于数据驱动的 PHM 流程图见图4。

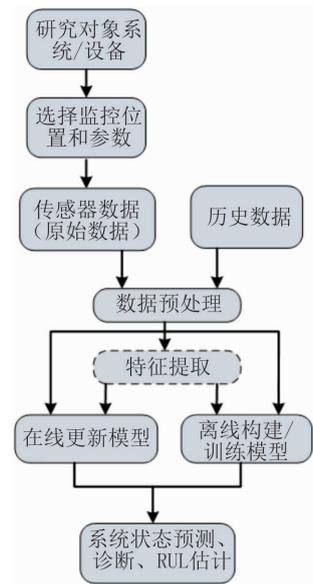


图4 基于数据驱动的 PHM 方法流程

3) 基于融合的方法

由于航空航天领域系统结构的复杂性和系统性,能退化过程的中不确定性、非线性、非稳定性等,导致故障诊断和寿命预测问题复杂。为了克服单一方法的局限,提高故障诊断和寿命预测的准确性,融合型的 PHM 方法成为研究的重点。融合型的 PHM 方法包括基于失效物理模型与基于数据驱动方法的融合和不同数据驱动方法之间的融合。刘月峰等^[38]提出了相关向量机、粒子滤波和自回归模型融合的锂离子电池的寿命预测方法,结果表明,该方法的通用性强,预测精度高,而且能够给出寿命预测的不确定性表达。Liao 等^[39]人提出了一种基于数据驱动和基于模型融合的预测算法框架,结合了大量历史数据和失效模型的优势,结果显示相比于传统的粒子滤波算法,融合型的算法的预测结果准确性更高。Song 等人^[40]提出了一种基于 IND-AR 和 PF 的锂电池剩余寿命预测方法,结合 AR 算法的长期预测能力和 PF 算法的预测结果不确定性,来提高预测算法的准确性。飞机是一个耦合紧密,失效机理复杂的系统,因此基于融合的方法是新一代飞机健康管理的发展重点。

3 基于私有云的飞机 PHM 海量数据挖掘平台

新一代飞机在研制、试验、使用和维护等阶段中积累了海量的数据,为了规范飞机海量数据的存储和管理,进一步充分利用数据中的有用价值,在考虑存储安全性和分析便捷性的基础上,提出了一种基

于私有云的飞机 PHM 海量数据挖掘平台。通过开源 Hadoop 体系的私有云平台技术构建数据存储基础平台,设计一种通用化的数据挖掘算法软件,具有较高的可扩展性和方便的人机交互方式,可以实现对飞机海量数据的相关性分析、故障诊断与寿命预测分析等。

3.1 数据挖掘平台的总体设计

基于私有云的飞机 PHM 海量数据挖掘平台整体框架如图 5 所示。系统由资源层、中间层和应用层 3 部分组成。资源层涵盖了各种来源的数据输入

输出配置、故障案例库、知识库和不同类型的数据处理算法;中间层是根据应用层的需求对资源层进行整合,实现不同的功能,包括数据库的管理、数据挖掘算法的集成等;应用层主要是对海量数据进行数据挖掘,包括相关性分析、异常报警、故障诊断、寿命预测等功能,同时其还能够提供图形化、报表、树状图等结果可视化方式。系统的功能包括:①多源异构数据的标准化管理;②知识库和案例库的构建;③多维数据呈现方式;④相关性分析、异常监测、故障诊断与故障预测。

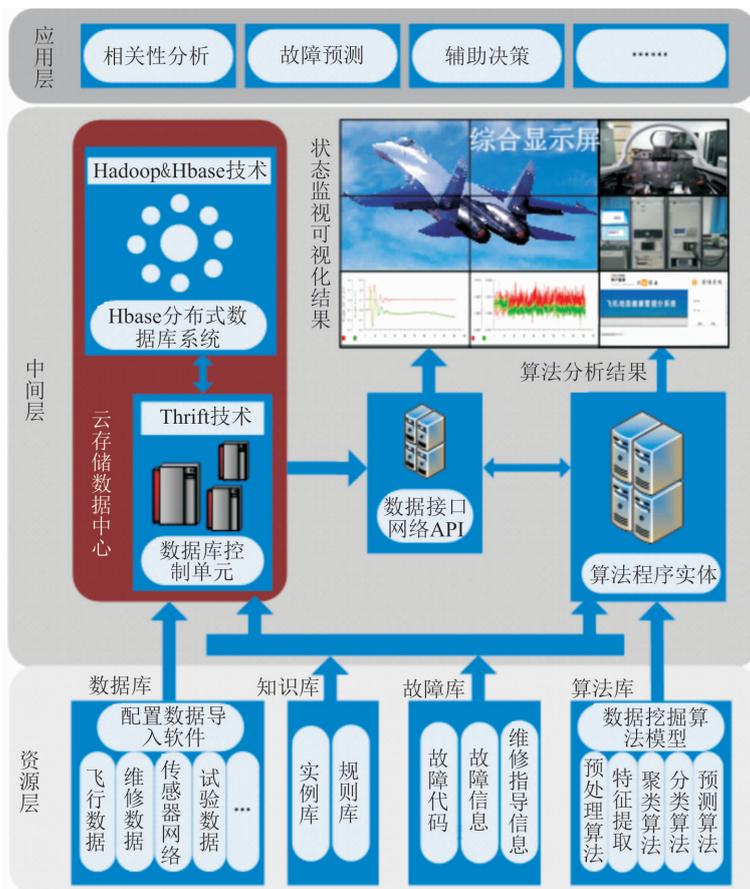


图 5 基于私有云的飞机 PHM 海量数据挖掘平台抽象结构

3.2 数据挖掘平台的具体实现

基于私有云的飞机系统海量数据挖掘平台由云存储中心、数据库控制单元、数据挖掘算法软件和 3×3 的拼接屏等部分组成,具体结构组成如图 6 所示。

云存储中心由 9 个服务器以 Hadoop 的方式构成,能够存储约 200 T 的数据以及高达 1 Gbps 的数据导入导出功能。云存储中心根据飞机系统各类数据的特点进行存储,如:数量规模较小、事务性强、数据更新频繁的数据使用关系数据库进行存储;数据量规模大、事务性弱、数据以增、查为主的数据使用分布式数据库进行存储。

数据库控制单元是整个数据挖掘平台的核心组

件,它可以统一管理底层数据子节点,监测性能、均衡负载、分配数据片段,同时可以提供多种数据检索和提取的接口。数据库控制单元支持对整个试验数据库进行逻辑分区,将整个大的数据库在逻辑意义上分成多个不同类型的子数据库,例如,监测参数数据库,任务库,知识库,故障库,资料库等,这些数据库都是可以由用户自定义生成,数据库中的数据表也可以由用户定义。数据的导入导出接口,支持用户将结构化的状态监测数据,试验数据等以及非结构化的数据存储于 Hadoop 架构的数据节点中。数据在节点中分布式存储,自动备份 2 份,保证了数据存储的安全性。数据接口网络 API 提供了基于 TCP/IP 协议的数据交互接口,数据挖掘算法软件

可以通过该接口实时交互数据。Web端数据检索接口支持用户直接通过Web界面进行数据的检索,避免了Linux和Windows不同操作系统下环境部署等过程。

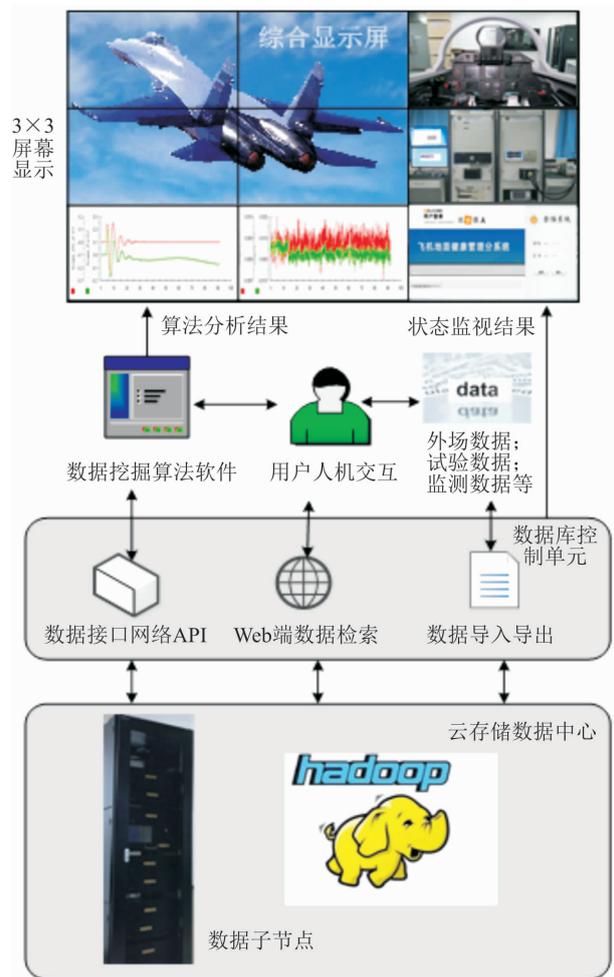


图6 基于私有云的飞机PHM海量数据挖掘平台具体结构

数据挖掘算法软件是一套可编程的图形化算法分析软件,能够对本地数据和数据库中的数据进行挖掘分析。该软件通过算法库接口规范将算法封装成DLL文件,然后在启动的过程中加载相关的DLL文件,目前主要开发了预处理算法、特征提取算法、分类算法、聚类算法、预测算法等,软件结构如图7所示。

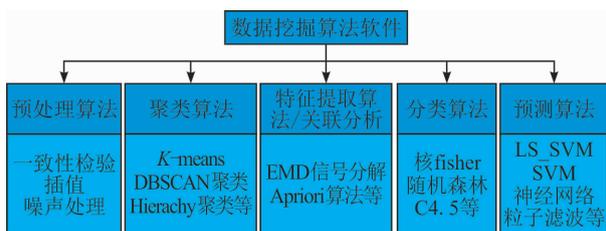


图7 数据挖掘算法软件结构图

该软件以图元化的方式来进行数据挖掘,主要是利用图元来搭建数据挖掘过程,以报表、曲线等方

式显示挖掘结果,提高了针对不同应用的数据挖掘效率。数据挖掘算法软件在启动时首先加载动态算法库,然后加载算法流程图,如果没有算法流程图,则通过图元的方式来进行构建流程图。算法流程图主要由基本图元组成,包括开始图元、结束图元、连接线图元、数据源图元以及一系列的算法图元。算法流程以串行或者并行的方式运行。每个挖掘算法都以开始图元为开始标识,然后选择数据来源图元和挖掘算法图元,最后以结束图元为结束标识。数据挖掘算法软件流程图见图8。

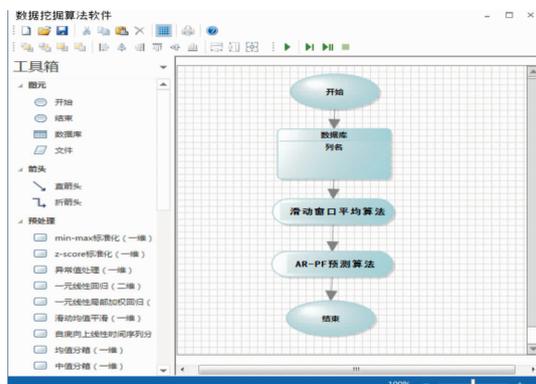


图8 数据挖掘软件流程图

此外,随着数据挖掘软件中算法库的不断丰富,可以使用不同的数据预处理方法、特征提取方法、故障诊断及寿命预测方法对飞机进行数据挖掘。因此该平台能够为飞机的PHM提供数据基础和算法验证平台。

4 结语

本文针对新一代飞机的结构复杂性,服役环境的时空多尺度性、数据来源和存储方式的多样性,以及现有的故障诊断,寿命预测及健康管理模式无法满足现代战争对飞机可靠性、安全性和经济性的需求,论述了基于海量数据挖掘的飞机PHM的关键技术,包括海量数据的预处理技术、集成管理技术、故障诊断及寿命预测技术等,然后提出了一种基于Hadoop私有云架构的飞机PHM海量数据挖掘平台,旨在对飞机全寿命周期的海量数据标准化管理的基础上,利用数据挖掘算法软件方便有效地挖掘隐藏在数据与系统状态、任务之间的关系,在故障或失效发生前及时发出预警,真正实现视情维修。本文所搭建的数据挖掘平台具有广泛的应用价值,可以适用于不同装备系统对象,对于进一步促进PHM的工程化具有重要的意义。但是针对飞机PHM的海量数据挖掘还处于初步阶段,以后还应重点发展方向和研究技术为:

- 1) 数据挖掘过程中的可视化方法;
- 2) 网络环境下的数据挖掘技术;
- 3) 各种非结构化数据的挖掘方法(文本数据、图像视频数据、音频数据等);
- 4) 嵌入式高性能的 PHM 运算平台,全生命周期数据规模的急剧增长,对 PHM 运算平台的计算能力、实时性、准确性和可重构性提出了迫切的需求;
- 5) 基于深度学习的故障预测与健康管理工作。

接下来工作主要是利用现有飞机积累的海量数据,重点研究这些数据之间的关联性,深入挖掘其潜在规律,进一步完善挖掘平台的挖掘算法,提高故障诊断,故障预测的准确性和置信度。

参考文献(References):

- [1] XU J P, XU Lei. Health Management Based on Fusion Prognostics for Avionics System [J]. Journal of Systems Engineering and Electronics, 2011, 22 (3): 428-436.
- [2] SCANFF E, FELDMAN K L, GHELAM S, et al. Life Cycle Cost Impact of Using Prognostic Health Management (PHM) for Helicopter Avionics[J]. Microelectronics Reliability, 2007, 47(12):1857-1864.
- [3] 年夫顺. 关于故障预测与健康管理工作技术的几点认识 [J]. 仪器仪表学报, 2018, 39(8):1-14.
NIAN F S. Viewpoints about the Prognostic and Health Management[J]. Chinese Journal of Scientific Instrument, 2018, 39(8):1-14. (in Chinese)
- [4] YANG H H, HUANG M L, LAI C M, et al. An Approach Combining Data Mining and Control Charts-Based Model for Fault Detection in Wind Turbines[J]. Renewable Energy, 2018, 115: 808-816.
- [5] 彭宇, 刘大同. 数据驱动故障预测和健康管理工作综述 [J]. 仪器仪表学报, 2014, 35(3):481-495.
PENG Y, LIU D T. Data-Driven Prognostics and Health Management: A Review of Recent Advances [J]. Chinese Journal of Scientific Instrument, 2014, 35 (3):481-495. (in Chinese)
- [6] 景博, 黄以锋, 张建业. 航空电子系统故障预测与健康管理工作技术现状与发展[J]. 空军工程大学学报(自然科学版), 2010, 11(6):1-6.
JING B, HUANG Y F, ZHANG J Y. Status and Perspective of Prognostics and Health Management Technology of Avionics System [J]. Journal of Air Force Engineering University (Natural Science Edition), 2010, 11(6):1-6. (in Chinese)
- [7] RAFIEE K, FENG Q, COIT D W. Condition-Based Maintenance for Repairable Deteriorating Systems Subject to a Generalized Mixed Shock Mode [J]. IEEE Transactions on Reliability, 2015, 64(4):1164-1174.
- [8] TANG W, JING B, HUANG Y, et al. Multistate Degradation Model for Prognostics of Solder Joints Under Vibration Conditions [J]. Chinese Journal of Electronics, 2016, 25(4): 779-785.
- [9] 卢海涛, 王自力. 综合航空电子系统故障诊断与健康管理工作技术发展 [J]. 电光与控制, 2015, 22(8):60-65.
LU H T, WANG Z L. Development Review of Integrated Avionics Prognostics and Health Management Technology [J]. Electronics Optics & Control, 2015, 22 (8):60-65. (in Chinese)
- [10] TU J, SUN C, ZHANG X, et al. Maintenance Strategy Decision for Avionics System Based on Cognitive Uncertainty Information Processing [J]. Eksploatacja i Niezawodnosć-Maintenance and Reliability, 2015, 17 (2): 297-305.
- [11] 彭宇, 庞景月, 刘大同, 等. 大数据:内涵、技术体系与展望 [J]. 电子测量与仪器学报, 2015, 29(4): 469-482.
PENG Y, PANG J Y, LIU D T, et al. Big Data: Connotation, Technical Framework and Its Development [J]. Journal of Electronic Measurement and Instrumentation, 2015, 29 (4):469-482. (in Chinese)
- [12] 程学旗, 靳小龙, 王元卓, 等. 大数据系统和分析技术综述 [J]. 软件学报, 2014, 25(9): 1889-1908.
CHENG X Q, JIN X L, WANG Y Z, et al. Survey on Big Data System and Analytic Technology [J]. Journal of Software, 2014, 25(9): 1889-1908. (in Chinese)
- [13] 严英杰, 盛戈皞, 陈玉峰, 等. 基于时间序列分析的输变电设备状态大数据清洗方法 [J]. 电力系统自动化, 2015, 39(7):138-144.
YAN Y J, SHENG G H, CHEN Y F, et al. Cleaning Method for Big Data of Power Transmission and Transformation Equipment State Based on Time Sequence Analysis [J]. Automation of Electric Power System, 2015, 39(7): 138-144. (in Chinese)
- [14] WANG L, FU D, LI Q, et al. Modelling Method with Missing Values Based on Clustering and Support Vector Regression [J]. Journal of Systems Engineering and Electronics, 2010, 21(1): 142-147.
- [15] GU P, LUO X, YANG R L, et al. Outlier Detection Algorithm Based on Iterative Clustering [J]. Journal of Donghua University, 2015, 32(4): 554-558.
- [16] 朱倩雯, 叶林, 赵永宁, 等. 风电场输出功率异常数据识别与重构方法研究 [J]. 电力系统保护与控制, 2015, 43(3):38-45.
ZHU Q W, YE L, ZHAO Y N, et al. Methods for Elimination and Reconstruction of Abnormal Power Data in Wind Farms [J]. Power System Protection and Control, 2015, 43(2): 38-45. (in Chinese)
- [17] WEI M, QIU B, JIANG Y, et al. Multi-Sensor Monitoring Based on-Line Diesel Engine Anomaly Detection with Baseline Deviation [C]//Prognostics and System Health Management Conference. Chengdu: IEEE, 2016: 1-5.
- [18] LIU L, LIU D, ZHANG Y, et al. Effective Sensor

- Selection and Data Anomaly Detection for Condition Monitoring of Aircraft Engines [J]. *Sensors*, 2016, 16: 623.
- [19] YADAV S K, SINHA R, BORA P K. Electrocardiogram Signal Denoising Using Non-Local Wavelet Transform Domain Filtering [J]. *IET Signal Processing*, 2015, 9(1): 88-96.
- [20] 景博, 徐光跃, 黄以锋, 等. 军用飞机 PHM 技术进展分析及问题研究 [J]. *电子测量与仪器学报*, 2017, 31(2): 161-169.
JING B, XU G Y, HUANG Y F, et al. Recent Advances Analysis and New Problems Research on PHM Technology of Military Aircraft [J]. *Journal of Electronic Measurement and Instrument*, 2017, 31(2): 161-169. (in Chinese)
- [21] MENG X, BRADLEY J, YAVUZ B, et al. Mllib: Machine Learning in Apache Spark [J]. *Journal of Machine Learning Research*, 2016, 17(1): 1235-1241.
- [22] SHVACHKO K, KUANG H, RADIA S, et al. The Hadoop Distributed File System [C]// 2010 IEEE 26th Symposium on Mass Storage Systems and Technologies (MSST). Incline Villiage, Nevada: IEEE, 2010: 1-10.
- [23] KANG Y, KANG K W. An Empirical Study of Hadoop Application Running on Private Cloud Environment [J]. *Adv Sci Technol Lett*, 2013, 35: 70-73.
- [24] ZAHARIA M, XIN R S, WENDELL P, et al. Apache Spark: A Unified Engine for Big Data Processing [J]. *Communications of the ACM*, 2016, 59(11): 56-65.
- [25] TANG W, JING B, HUANG Y F, et al. Feature Extraction for Latent Fault Detection and Failure Modes Classification of Board-Level Package under Vibration Loadings [J]. *Science China Technological Sciences*, 2015, 58(11): 1905-1914.
- [26] YIAKOPOULOS C T, GRYLLIAS K C, ANTONIADIS I A. Rolling Element Bearing Fault Detection in Industrial Environments Based on a K-means Clustering Approach [J]. *Expert Systems with Applications*, 2011, 38(3): 2888-2911.
- [27] MURALIDHARAN V, SUGUMARAN V, INDIRA V. Fault Diagnosis of Monoblock Centrifugal Pump Using SVM [J]. *Engineering Science and Technology, an International*, 2014, 17(3): 152-157.
- [28] AYDIN I, KARAKOSE M, AKIN E. An Approach for Automated Fault Diagnosis Based on a Fuzzy Decision Tree and Boundary Analysis of a Reconstructed Phase Space [J]. *ISA Transactions*, 2014, 53(2): 220-229.
- [29] LIN Y, DING S, WANG Y, et al. A Method of Satellite Network Fault Synthetic Diagnosis Based on C4.5 Algorithm and Expert Knowledge Database [C]// 2015 International Conference on Wireless Communications & Signal Processing, Nanjing: IEEE, 2015: 1-5.
- [30] 黄常海, 高德毅, 胡基平, 等. 基于 Apriori 算法的船舶交通事故关联规则分析 [J]. *上海海事大学学报*, 2014, 35(3): 18-22.
HUANG C H, GAO D Y, HU S P, et al. Association Rule Analysis of Vessel Traffic Accidents Based on Apriori Algorithm [J]. *Journal of Shanghai Maritime University*, 2014, 35(3): 18-22. (in Chinese)
- [31] LEE J, WU F, ZHAO W, et al. Prognostics and Health Management Design for Rotary Machinery Systems-Reviews, Methodology and Applications [J]. *Mechanical Systems and Signal Processing*, 2014, 42(1): 314-334.
- [32] ZHU S P, HUANG H Z, PENG W, et al. Probabilistic Physics of Failure-Based Framework for Fatigue Life Prediction of Aircraft Gas Turbine Discs Under Uncertainty [J]. *Reliability Engineering & System Safety*, 2016, 146: 1-12.
- [33] MOSALLAM A, MEDJAHHER K, ZERHOUNI N. Data-Driven Prognostic Method Based on Bayesian Approaches for Direct Remaining Useful Life Prediction [J]. *Journal of Intelligent Manufacturing*, 2014: 1-20.
- [34] LIU R, YANG B, ZHANG X, et al. Time-Frequency Atoms-Driven Support Vector Machine Method for Bearings Incipient Fault Diagnosis [J]. *Mechanical Systems and Signal Processing*, 2016, 75: 345-370.
- [35] YIN S, ZHU X. Intelligent Particle Filter and Its Application to Fault Detection of Nonlinear System [J]. *IEEE Transactions on Industrial Electronics*, 2015, 62(6): 3852-3861.
- [36] WIDODO A, KIM E Y, SON J D, et al. Fault Diagnosis of Low Speed Bearing Based on Relevance Vector Machine and Support Vector Machine [J]. *Expert Systems with Applications*, 2009, 36(3): 7252-7261.
- [37] TOBON-MEJIA D A, MEDJAHHER K, ZERHOUNI N, et al. A Data-Driven Failure Prognostics Method Based on Mixture of Gaussians Hidden Markov Models [J]. *IEEE Transactions on Reliability*, 2012, 61(2): 491-503.
- [38] 刘月峰, 赵光权, 彭喜元. 锂离子电池循环寿命的融合预测方法 [J]. *仪器仪表学报*, 2015, 36(7): 1462-1469.
LIU Y F, ZHAO G Q, PENG X Y. A Fusion Prediction Method of Lithiumion Battery Cycle-Life [J]. *Chinese Journal of Scientific Instrument*, 2015, 36(7): 1462-1469. (in Chinese)
- [39] LIAO L, KÖTTING F. A Hybrid Framework Combining Data-Driven and Model-Based Methods for System Remaining Useful Life Prediction [J]. *Applied Soft Computing*, 2016, 44: 191-199.
- [40] SONG Y, LIU D, YANG C, et al. Data-Driven Hybrid Remaining Useful Life Estimation Approach for Spacecraft Lithiumion Battery [J]. *Microelectronics Reliability*, 2017, 75: 142-153.