

数据驱动的飞行器智能故障诊断系统研究

李智¹, 姜悦², 蔡斐华², 褚厚斌², 张丽晔²

(1. 北京东方仿真软件技术有限公司, 北京 100029; 2. 中国运载火箭技术研究院, 北京 100076)

摘要: 飞行器装备及试验技术的快速发展带来试验数据的指数级增长, 传统的试验数据管理与分析手段已经无法满足装备研制与优化设计的需求; 提出一种数据驱动的飞行器智能故障诊断技术, 通过将人工智能技术与故障诊断技术相融合, 能够从飞行器历史测控试验数据库中, 自动获取参数的阈值、数据链特征与关联关系, 形成准确的飞行器参数知识模型库; 通过构建数据驱动的飞行器智能故障诊断系统, 将知识模型库中的知识与实时试验数据进行比对分析, 实现实时数据的自动化、智能化判读, 进而发现航天飞行器可能存在的潜在故障, 提前预测, 将风险降低到最小。

关键词: 数据挖掘; 故障诊断; 飞行器

Study on a Data-driven Intelligent Fault Diagnosis System of Spacecraft

Li Zhi¹, Jiang Yue², Cai Feihua², Chu Houbin², Zhang Liye²

(1. East Simulation Software Technology, Beijing 100029, China;

2. China Academy of Launch Vehicle Technology, Beijing 100076, China)

Abstract: The rapid development of spacecraft test technology brings exponential growth of test data. The traditional data management and analysis methods cannot meet the requirements of equipment development and optimization design. A data driven intelligent fault diagnosis technology of spacecraft is proposed, which can automatically obtain the threshold value, data link characteristics and correlations of parameters from the historical aircraft test data by integrating AI technology, and form an accurate spacecraft parameter knowledge model base. By constructing a data driven intelligent fault diagnosis technology of spacecraft, real-time data can be interpreted automatically and intelligently.

Keywords: data mining; fault diagnosis; spacecraft

0 引言

近年来, 数据挖掘技术迅猛发展, 能够借助算法从大量数据中快速获取隐藏信息, 在金融、医疗、零售、制造等行业都得到了广泛应用。在飞行器地面试验阶段, 测控通信速率最大包络数据可达到 100 Mbps, 一次总检查所获得的原始数据可达数十 GB, 按照解析后数据量增至 50 倍的数据容量统计, 每次总检查数据达到上千 GB, 在综合试验、匹配试验、热平衡、热真空、总装测试过程中累计需要经历上百次总检查, 产生的总数据量级为上百 TB; 在飞行器在轨运行阶段, 测控通信速率最大包络数据可达到 2 Mbps, 每天传输 24 h, 每天产生总数据量约为 1.15 TB。按照飞行器试验周期一年(按照 365 天)计算, 产生的总数据为 420 TB。可以看出, 地面试验和在轨试验产生的总数据已经达到 PB 级。由于测试数据信息量大、参数数量多、种类杂, 依靠人工进行数据判读及复查, 不仅工作量大、效率低, 而且容易造成人为判读遗漏和偏差。利用数据挖掘技术能够从大量无序数据中自动抽取其隐藏特征及关联关系, 及时发现飞行器故障征兆并采取有效措施, 从而避免重大故障发生。

目前, 国外航空航天领域已经率先开展数据挖掘技术

在故障预测与诊断方面的研究, 并取得了一定的成果。如美国空军研究实验室开发的 EHM 样机系统、CFM 公司的 CASSIOPEE、英国航空公司的 ARIADNE 等, 它们将数据挖掘技术与故障诊断技术相融合, 建立发动机维修决策支持专家系统。数据挖掘在我国航天领域的应用研究起步较晚, 尚处于理论方法探索阶段, 成功的应用案例较少。

1 飞行器故障诊断研究现状

1.1 测控试验数据的特点

飞行器在进入靶场发射前需要经过长时间的功能及性能试验, 通过对试验过程产生的遥测数据进行判读和诊断, 可以评估各系统相关设备的电气功能和性能指标, 验证飞行器上软、硬件设备的可靠性。由于飞行器体系庞大、结构复杂、分系统众多, 每次试验过程都会产生大量数据。这些数据具有以下特点。

1) 多源性: 飞行器测试或在轨运行中各系统产生的数据相互交叉, 多通道、多波段测控数据同时下传, 导致数据具有多源性。

2) 多样性: 测控数据包括不同的数据类型, 如源码、解析后物理量等结构化数据, xml 等非结构化数据, 图片、图像、文本等非结构化数据, 即使对于结构化数据, 也包

收稿日期: 2020-07-23; 修回日期: 2020-09-02。

作者简介: 李智(1988-), 男, 吉林梅河口人, 硕士研究生, 主要从事虚拟仿真软件开发以及仿真软件平台方向的研究。

引用格式: 李智, 姜悦, 蔡斐华, 等. 数据驱动的飞行器智能故障诊断系统研究[J]. 计算机测量与控制, 2021, 29(1): 50-53.

含整数型、浮点型等多种数据类型, 导致飞行器测控数据具有多样性。

3) 复杂性: 飞行器测试阶段要开展静力试验、热平衡、热真空、振动、电磁兼容等多项大型地面试验。不同试验状态下飞行器采用不同的工作模式, 导致产生的数据具有较高的复杂性。

当前, 我国飞行器发射频率越来越高、周期越来越短、地面测试发射控制技术向着具备快速响应发射的能力发展, 在这样的新要求下, 研究快速、先进的数据处理及故障诊断方法是十分必要的。

1.2 传统诊断方法存在的问题

传统的飞行器故障诊断方法包括门限判断法、专家系统方法等。目前国内应用较为广泛的自动判读系统利用现有计算、网络和人工智能技术, 实时分析或事后分析的遥测数据, 依据已录入计算机的知识和规则自动进行数据的判读分析, 在火箭测试中发挥了很大的辅助决策作用。但是, 判读规则以及故障模式仍然需要工程师人工梳理完成, 梳理过程工作量大、也很难全部覆盖火箭所有工作状态。

目前, 试验数据管理分散, 试验数据异构性强, 缺少规范统一的数据存储格式, 数据共享困难。由于缺乏规范的试验流程与统一的数据存储格式, 使得试验数据比分析的过程中人为干涉的步骤和因素较多, 不能及时、准确、规范的反应测试环境、各类参数及测试结果, 或记录的数据无法被相关人员正确理解, 很大程度上影响了测试结果的置信度与重用性, 使得跨平台数据共享困难。同时目前测试数据的挖掘与利用水平偏低。现有的试验数据管理更多地表现为数据堆积、数据仓库的概念, 数据之间未能有效地关联, 数据资源的利用率非常低下, 数据的价值, 特别是海量数据的整体价值没有得到体现。试验完成后, 数据经常仅进行基本的分析, 通常是“一次测试, 再不使用”, 未进行数据挖掘和更深层次的数据分析, 不能持续提供实时、深入的数据服务, 难以通过试验数据量化地全面验证产品各属性的内在关系。

随着航天飞行器装备及试验技术的发展, 面对过去常年累积的、目前不断产生的、未来持续呈指数形式增长的海量试验数据, 传统的试验数据管理与分析手段已经无法满足装备研制与优化设计的需求。为充分利用试验数据价值, 迫切需要开展试验数据管理与挖掘技术研究。

2 飞行器智能故障诊断方法

测控试验数据量大、参数多, 人工实时判读效率低下, 开展基于数据挖掘的飞行器故障诊断技术研究对解决构建参数知识库、实时异常监测、预测问题具有决定性作用。其核心思想是研究利用数据挖掘、人工智能算法, 对历史试验数据进行深度学习, 通过分析参数数据在每次试验中的数据链条特性, 以及与其它参数之间的关联性概率, 参数的数据趋势特征, 形成飞行器参数知识模型库, 与人工梳理的故障模型库结合, 采用混合模式对飞行器当前实验数据进行故障诊断和预测。

基于数据挖掘的飞行器故障诊断技术首先对参数进行预处理分析, 然后利用聚类分类算法对参数之间的关联关系及影响概率进行分类, 之后对利用神经网络、贝叶斯等机器学习算法对参数值及参数间变化关系进行学习, 经过对历史数据的反复学习后, 确定参数的阈值及数据链特征, 再对当前试验数据进行对比分析, 给出参数故障诊断的结果。具体分为离线训练学习建模阶段和在线实时诊断阶段。

2.1 离线训练学习建模阶段

离线训练学习建模阶段主要采用基于数据挖掘的算法建立飞行器参数知识模型库, 学习过程如图 1 所示。

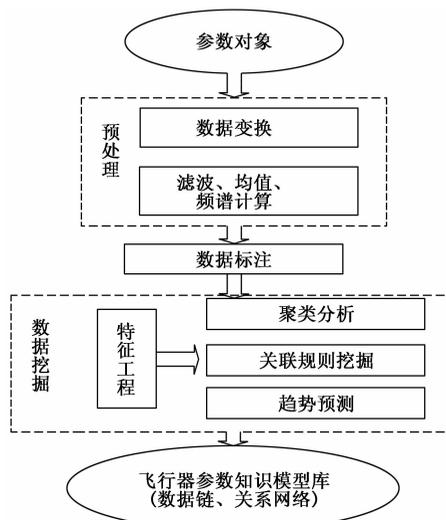


图 1 离线学习阶段框架

2.1.1 数据预处理

数据预处理主要完成试验数据的处理, 将参数进行变换, 形成可直接用于挖掘和分析的参数, 数据转换主要包括求均值、方差、斜率、频谱等, 实现数据的初步加工。

2.1.2 数据标注

离线训练数据需要进行标注。需要注意的是, 不是所有的历史数据都需要标注, 因为这样会带来极大的工作量并且对机器学习模型效果的提升有限, 适当选取有代表性的数据标注即可。最新的技术也支持通过少量标注来自动标注历史数据, 甚至无需标注就可以进行机器学习。将标注好的历史数据分为训练数据和测试数据, 常用的比例是 9:1 或者 4:1, 也可以根据项目的实际情况进行调整。

2.1.3 数据挖掘

挑选数据的特征是数据挖掘中的重要步骤, 由于特征过多可能会带来过拟合的问题, 所以需要特征工程技术对特征进行选取, 合并、裁剪一些特征。根据不同的需求可以采用不同算法, 如图 2 所示。

1) 利用基于 Kmeans、Clusteam、决策树等聚类算法, 对历史数据学习, 获取单个参数值的变化范围, 进而获取参数判据及变化数据链特征知识;

2) 基于贝叶斯、支持向量机、Apriori 等关联规则模型学习算法获取参数与参数之间的状态关联概率特征, 获取

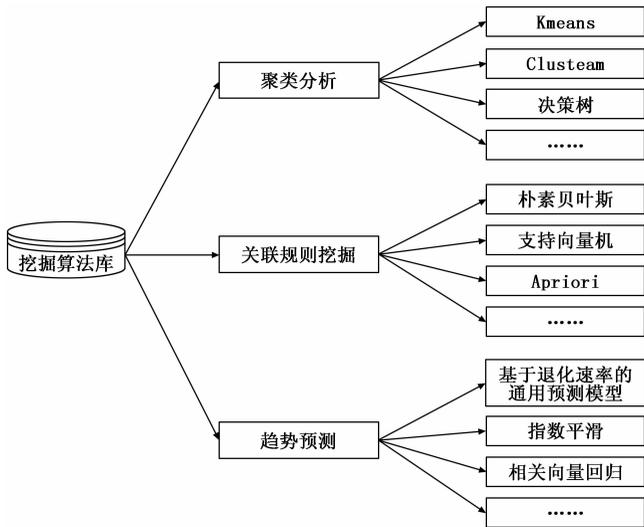


图 2 挖掘算法库

参数状态强关联规则知识；

3) 基于退化速率的通用预测模型、指数平滑、相关向量回归算法，对历史数据进行学习，获取参数的趋势特征模型，建立参数的趋势特征知识库。

结合 Spark、MapReduce 等分布式计算的框架，可以有效地处理海量数据，通过训练和不断地参数调优得到效果最好的分类模型。

2.2 在线实时诊断阶段

在线实时诊断阶段结合故障模型库进行综合决策诊断。通过将诊断知识库中的知识与实时实验数据进行比对分析，分析当前试验数据的是否存在异常或趋势性变化，同时运用数据知识与故障模型知识进行决策诊断，从而发现航天飞行器可能存在的潜在故障，提前预测，将风险降低到最小。

飞行器实时故障诊断业务场景如图 3 所示，实时接收火箭前端测试设备发送的数据流，对数据流进行预处理，根据聚类关系对数据流进行分类，根据参数数据链特征对参数值结果进行分析判定；根据关联规则模型知识，对比参数的状态与其它参数状态关联规则是否匹配，进行分析判定；根据参数趋势特征知识，对参数趋势变化进行判定。在诊断处理异常参数后，与故障模型库的参数特征进行综合诊断，实现对故障的定位和预测。

3 实例

飞行器故障诊断问题可以看作一个二分类问题，使用

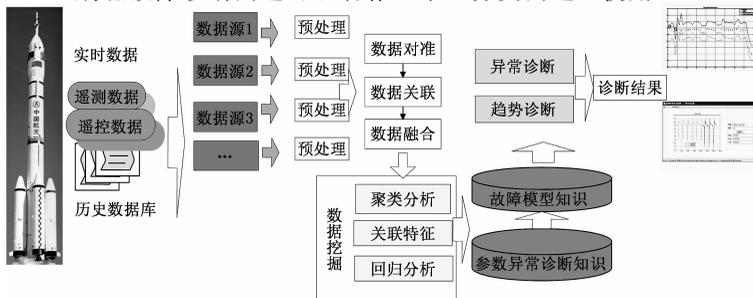


图 3 基于数据挖掘的飞行器故障诊断业务执行图

常用的 SVM、逻辑回归等可以达到较好的分类效果，但是由于这些算法不能捕捉到数据中的非线性特征，所以存在一定的瓶颈。树模型可以较好地捕捉到非线性特征，被广泛应用到现代的机器学习系统中。常用的树模型包括随机森林和梯度提升决策树（gradient boosting decision tree，简称 GBDT）。

这里我们借助 XGBoost（一个优化的分布式梯度增强库）作为梯度提升决策树的实现。我们以 5 次飞行器总检查试验中的配电器数据作为数据源，共计 117987 帧数据，分别被标注为正常数据和异常数据，每帧数据包括当前指令 ID 以及 30 个配电器参数值。随机将数据集拆分为训练集（64%）、验证集（16%）和测试集（20%）。在特征工程阶段，我们将当前指令 ID 做了 one-hot 编码（单热编码），生成了 22 个特征，加上原来的 30 个连续值特征，共 52 个特征。

在这里，我们对逻辑回归和树模型两种算法。我们借助 XGBoost（一个优化的分布式梯度增强库）作为梯度提升决策树的实现。在树模型中，我们设置 XGBoost 中树的最大深度为 6，“objective”参数设置为“binary: logistic”，即作为二分类问题，训练轮数为 50（即生成 50 棵决策树）。以 AUC（area under ROC curve）作为评估指标，得到结果如表 1 所示。

表 1 算法指标评估结果

	逻辑回归	树模型
训练集 AUC	0.99520	0.99999
验证集 AUC	0.99626	0.99994
测试集 AUC	0.99161	0.99997

在测试集上，我们选用不同的阈值进行故障判别，不同算法在不同阈值情况下的精确率、召回率和 F1 值如表 2 所示。

表 2 不同阈值下两种算法对比结果

指标类型	逻辑回归	树模型		
		0.5	0.1	
阈值	0.01-0.5	0.5	0.1	0.05
混淆矩阵	[[22823 0] [13 762]]	[[22822 1] [11 764]]	[[22822 1] [8 767]]	[[22813 10] [2 773]]
精确率	1.0	0.9986928	0.9986979	0.9872286
召回率	0.9832258	0.9858065	0.9896774	0.9974194
F1 值	0.9915420	0.9922078	0.9941672	0.9922978

可以看到，树模型的 AUC 指标表现更优，召回率更高，对故障更敏感，模型的整体表现更好。

4 基于数据挖掘的飞行器故障诊断系统探究

4.1 飞行器故障诊断系统框架设计

飞行器故障诊断系统分为数据存储层、计算层、服务层和应用层，如图 4 所示。

数据存储层：负责存储系统需要的各类数据，涵盖结构化数据、半结构化数据及非结构化数据。本系统中，采用关系型数据库存储结构化数据，采用文件的形式存放半结构化数据和非结构化数据。

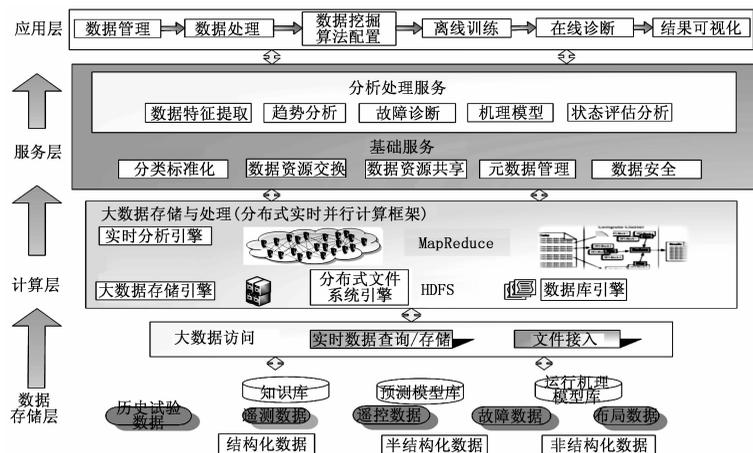


图 4 基于大数据的多源异构海量试验数据融合、存储与管理图

计算层: 通过大数据存储与处理框架满足大数据的处理能力, 并实现与现有数据库的集成, 实现数据抽取再存储, 实现多源异构数据集成。基于大数据技术的数据管理, 主要有大数据采集、大数据存储、大数据处理三大部分内容。所有这些功能都围绕大数据存储的分布式文件系统展开。本系统拟采用 hadoop 分布式文件系统 (HDFS: hadoop distributed file system) 构建 HDFS 集群, 并在 hadoop 接口基础上构建整个文件管理系统。随着 SQL-on-Hadoop 数据库的出现, 表明 hadoop 与关系型数据库的有机整合已进入技术成熟期, Hadoop+RDBMS 的混合通用型数据库是数据挖掘领域在数据存储和服务方面的发展趋势。

服务层: 包括基础服务和分析处理服务两部分。基础服务包括数据分类标准化、数据资源交换、数据资源共享、元数据管理、数据安全等为数据的处理提供服务; 分析处理服务提供数据分析服务, 包括特征提取、趋势分析、故障诊断、机理模型、状态评估分析等, 为装备的故障诊断与评估提供必要的服务算法。每个核心模块需提供相应的 API 接口, 包含数据查询接口、数据上传、下载接口、集成第三方软件接口, 这些接口主要提供给试验分析控件、可视化控件等外部系统调用。

应用层: 应用层为系统提供与用户交互的功能, 以及基于安全机制的上线运行。应用层主要实现产品数据管理、数据处理、数据挖掘算法配置、离线训练、在线诊断、结果可视化等功能。

4.2 基于数据挖掘的故障诊断服务流程

基于数据挖掘的故障诊断主要是调用离线训练形成的飞行器参数知识模型库, 对实时存入的新试验数据进行故障诊断与趋势分析, 分析故障原因, 记录故障处理过程。

1) 数据加载: 将挖掘所要使用到的数据从 HDFS 文件系统和 Hbase 中加载至实时缓存区, 供上层数据快速调用。

2) 数据处理: 从 HDFS 文件中, 获取输入的数据。并将处理结果也保存在 HDFS 文件系统中, 为数据挖掘发展所经过的程序进行数据加载、提取、清理、转换, 对数据进行融合处理, 形成规范化数据格式。

3) 数据挖掘算法配置: 在线完成对各类常用数据挖掘算法参数配置修改, 与数据库中数据接口关联匹配, 支持算法包括 PLS、PCA 等多元统计学习方法、KMean、Clusteam、决策树等聚类算法, 朴素贝叶斯、支持向量机、Apriori 等关联规则算法, 最大特征相似集和粒计算的诊断决策技术和广义高斯核函数决策类算法, 退化速率的通用预测模、相关向量回归预测类算法等。

4) 离线训练: 根据选择的数据算法、参数配置值以及关联的实验数据集, 在线运行算法模型进行学习训练, 利用 MapReduce 并行计算框架, 可以大大缩短训练时间, 在获得参数故障诊断的知识后, 存入数据库中。

5) 基于数据挖掘的在线诊断: 在飞行器测试之前, 根据训练结果, 选用表现较好的模型知识库, 与训练所得的知识进行比对决策, 对实时存入的实验数据进行在线故障诊断。

5 结束语

本文提出了利用数据挖掘技术、机器学习算法开展飞行器故障诊断的方法, 采用对历史试验数据进行机器学习, 获取形成飞行器参数知识模型库, 分析当前试验数据的异常和趋势, 可以实现对飞行器开展快速、有效的故障诊断, 对飞行器健康状态进行评估具有重大意义。通过探讨基于数据挖掘的飞行器故障诊断系统框架构成, 可以实现数据驱动的飞行器智能故障诊断技术的系统化、自动化、成熟化, 为下一步的工作提供指导。

参考文献:

- [1] 李志芳, 冯秀芳, 赵红旗. 基于测控数据挖掘的飞行器故障诊断技术研究 [J]. 计算机测量与控制, 2011, 19 (3): 500-502.
- [2] 周 枫. 基于大数据的飞行器性能退化分析方法 [D]. 南京: 南京航空航天大学, 2015.
- [3] 王良栋. 数据挖掘云服务系统支撑框架关键技术研究 [D]. 南京: 南京大学, 2013.
- [4] Arita M M, Syrmos V L, Vian J L, et al. Multi Source Data Integration for Aircraft Health Management [R]. IEEE Aerospace Conference, 2008.
- [5] 邵新杰, 曹立军, 田 广, 等. 复杂装备故障预测与健康管理工作 [M]. 北京: 国防工业出版社, 2013.
- [6] 王 亮, 吕卫民, 滕克难, 等. 基于数据驱动的装备故障预测技术研究 [J]. 计算机测量与控制, 2013, 21 (8): 2087-2105.
- [7] 吴明强, 房红征, 伊大伟. 复杂系统故障预测方法与应用技术研究 [J]. 计算机测量与控制, 2010, 18 (1): 70-73.
- [8] 徐子伟, 张陈斌, 陈宗海. 基于数据驱动的故障检测和诊断方法概述 [J]. 系统仿真技术及其应用, 2015 (16): 452-456.
- [9] 何一帆. 基于多源异构数据融合的深度神经网络故障诊断 [D]. 郑州: 河南大学, 2019.
- [10] 于博文. 数据挖掘技术在故障诊断中的应用 [D]. 大庆: 东北石油大学, 2015.