

大数据技术在航空发动机中的应用

The Application of Big Data on Aero Engine

■ 王潘 刘魁/中国航发研究院

什么样的数据才是大数据？相较于传统的数据搜集统计，大数据技术究竟有何独特的“魔力”？笔者在解答上述问题的基础上，对大数据技术在航空发动机领域的应用进行了探讨。

大数据是指在一定时间范围内无法用常规软件工具进行捕捉、管理和处理的海量、高增长和多样化的数据集合。大数据技术是一种通过分析和挖掘全量的非抽样数据的辅助决策技术。

工业大数据的产生及特点

工业大数据

工业大数据是指在工业领域中围绕智能制造模式，从研发设计、制造生产、运行维修等整个产品全生命周期产生的各类数据的总称，这些数据与设计、制造、服务等产品研发关键环节深度融合，对提高产品质量、降低研制成本等有重要作用。

大数据技术是在分布式处理技术、存储技术的迅速发展下产生的，传统数据分析基于样本数据，随着Hadoop分布式文件系统（HDFS）和分布式并行计算技术MapReduce的应用，数据存储和分析的成本大幅降低，通过实时处理和分析海量数据提取隐含价值成为可能。大数据分析的技术方法逐渐成熟并应用于工业领域，解决了传统数据挖掘所面临的性能、吞吐量、容量的

问题，并将数据分析由采样数据转变为全量数据，能挖掘出更多的关联关系。

工业大数据的特点

相比传统数据，大数据有4V的特征：一是数量（Volume）大，且每年的增长速度快；二是类别（Variety）多，结构化和非结构化数据种类多；三是数据获取和处理的速度（Velocity）要求高；四是数据价值（Value）密度低，价值密度高低与数据总量成反比，通过数据分析的算法提取数据价值成为关键问题。

工业大数据的发展经历了三个阶段：第一阶段为1990—2000年，以远程监控、数据采集和管理为核心

技术，主要解决的是以产品为核心的状态监控；第二阶段为2000—2010年，以数据分析为核心技术，以及时维修和预测型维护为目标；第三阶段为2010年至今，以大数据分析平台为高级数据分析工具，实现了以应用为主导的服务生态体系。

工业大数据除具有一般大数据的特点外，还具有更强的专业性、关联性、流程性、时序性和解析性等特点，仅仅依靠传统的互联网大数据分析技术是无法满足工业数据的分析需求的。工业大数据与互联网大数据的对比分析如表1所示。

总而言之，互联网大数据的价值发现主要是从数据端驱动，如通

表1 工业大数据与互联网大数据对比分析

	互联网大数据	工业大数据
数据量需求	大量样本数据	尽可能全面使用样本
数据质量要求	较低	较高，需要对数据质量进行预判和修复
对数据属性意义的解读	不考虑属性的意义，分析统计显著性	强调特征之间的物理关联
分析手段	以统计分析为主，通过分析挖掘样本中各个属性之间的相关性进行预测	具有一定逻辑的数据分析手段。强调跨学科的融合，包括数学、机器学习、物理、控制等
分析结果准确性要求	较低	较高

过分析产品属性关联性以形成推荐方案；工业大数据的价值更多的是从功能端驱动，在数据分析时往往需明确待解决的问题。在数据特征的提取方面，工业大数据更注重特征背后的物理意义及特征之间关联性的机理逻辑，而传统的互联网大数据更侧重于挖掘属性之间的关联性。在预测结果方面，工业大数据对预测和分析结果的容错率远低于互联网大数据。

航空发动机大数据

航空发动机数据可谓是典型的工业大数据，数据产生于全生命周期过程（如图1所示），主要包括：设计数据，如研发需求、数据模型、设计方法等；制造数据，如工艺设计、产品实作等；试验数据，如试验设计、测试方法、测试数据等；服务数据，如维修数据、备件数据、故障数据等；资源数据，如情报、材料、成果专利等；环境数据，如气象数据、地理数据、航电数据等；管理数据，如项目管理、业务流程等。

工业大数据的价值产生于发动机的全生命周期（如图2所示），如在研发与设计阶段可实现基于大数



图2 全生命周期的数据价值体现

据的模拟仿真设计，利用虚拟仿真技术，可实现对原有研发设计过程的模拟、分析、评估、验证和优化，减少工程更改量降低成本；在生产过程中实时监控与管理，优化生产流程，现代化生产线安装了数以千计的小型传感器，通过探测生产设备的工作状态（如温度、压力、振动和噪声等）对设备实时监控，同时可将生产制造各环节的数据整合，建立虚拟模型，仿真并优化生产流程；在运行维修阶段，将大量数据用于发动机的预测与健康管理系统（PHM），如故障检测、诊断及剩余寿命预测，同时将健康管理信息转化为备件需求和维修操作建议。

系统（ACMS）在测试阶段记录的数据量相较于之前增长了50倍。空客A350被收集的参数将近60万个，每天可收集到的数据已超1.8TB，2015年采集到的数据达到91TB。在采集的数据中，仅有飞行测试阶段的数据用来存储和分析，在服务阶段只有部分状态监控的数据存储和分析，其他绝大部分参数并未存储和利用。2014年，空客公司成立数据处理与试飞集成中心，与甲骨文公司共同建立了基于Hadoop技术的大数据处理系统及飞行模拟数据分析软件，软件用于收集并分析飞行样机上传感器在试飞过程中产生的数据，实时监控飞行状态，提供优化建议。

波音787飞机仅执行一个航班，所采集的客舱压力、高度、燃油消耗等数据就能达到0.5TB。2015年，波音公司与卡耐基-梅隆大学合作组建了航空数据分析实验室，利用人工智能和大数据来对波音飞机进行全面升级，收集航空飞行中产生的巨量数据，用机器学习的方法来优化飞机的飞行方式，用数据分析的结果来指导未来的设计、制造和运营。例如，基于前一个航段的运行数据，飞机可以自我判断下一个航段中会出现的问题，通知机务进行检查和维修。航空业比其他行业更早进入EB级数据时代，从海量存储中高性能调取，闪电般采集数据、抽取数据、分析数据，并从数据可

大数据在航空领域的应用 飞机制造商大数据技术应用情况

空客公司的新型飞机的状态监控

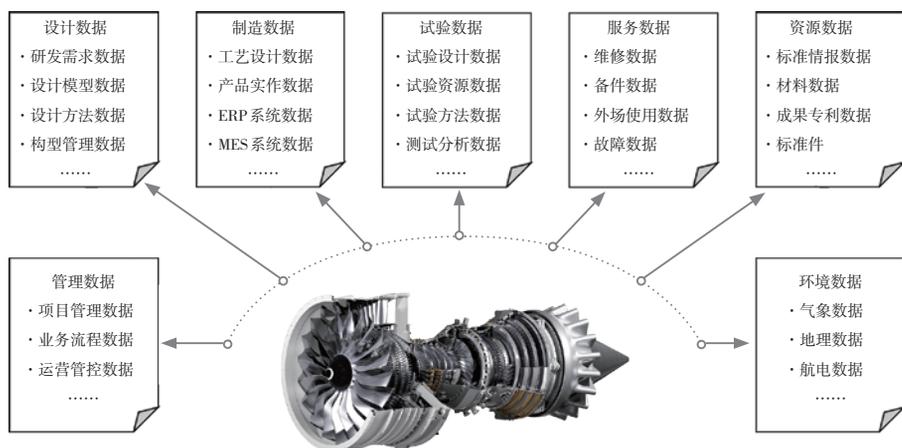


图1 航空发动机全生命周期的数据

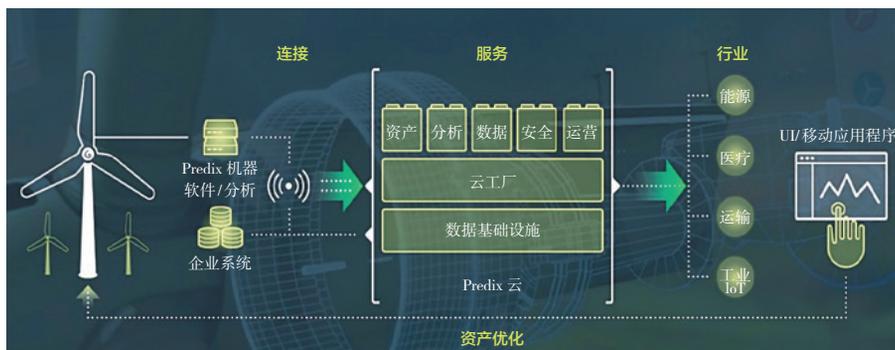


图3 Predix平台架构

可视化展示等几个方面对数据分析提出更高的要求。

发动机公司大数据技术应用情况

通用电气（GE）公司于2013年宣布，推出第一个大数据与分析平台——GE Predix平台（如图3所示），管理由大型工业机器所产生的数据。该平台支持工业互联网并把大数据转化为实时有用的信息，该平台可应用于能源、医疗、运输等领域。对于纷繁复杂的工业设备和工业数据，Predix平台提供了一种相对标准和统一的呈现形式。

2015年年底，GE公司所有发动机的诊断数据从旧系统迁移到Predix平台上，并将其用于发动机状态监控，帮助发动机监控团队捕捉更多的数据，使数据分析变得更精准快捷。GE公司的每台发动机都配有故障诊断系统，每台发动机的参数大概有480个，目前约有35000台发动机将其起飞、巡航、降落等关键飞行阶段的数据包传输至Predix平台。在Predix平台启用之前，GE公司主要依靠工程师组成的“智囊团”对飞行数据进行分析，无法集中不同发动机的复杂数据并快速预测出整个机队的发展趋势，而大数据分析平台擅长对多变量数据进行分析。

例如，在不同的运行环境（如高温、沙漠等）下，Predix平台可为发动机提供调整预警参数的服务，为每台发动机提供异常预警。2015年，Predix平台产生了约35万个警告信息和9000份（CNR客户通知记录单），其中86%的CNR都是准确的。在Predix平台中，尽管CNR的数量在增加，但虚警率却在不断下降，这意味着这一平台通过预测发现的问题数量越来越多，该平台的预测性能越来越好。

同时，GE公司用大数据为中国东方航空公司（东航）的飞机提供远程体检服务，对发动机的关键零部件进行寿命预测。Predix平台利用CFM56的历史运行数据（500多台发动机的维修报告和远程诊断数据）提取发动机与叶片损伤相关联的数百个参数数据。经过筛选，几个关联度最大的参数被最终确定，其中就包括飞机所执飞航线的空气污染程度。GE公司利用这些参数建立了叶片损伤分析预测模型，并对多台东航现役发动机的叶片损伤程度进行了预测，模型预测的准确率达到80%以上。

罗罗公司的发动机被称为“大数据引擎”。罗罗公司较早利用了大数

据技术安排生产和维修，其全球发动机健康检测中心建立了发动机健康管理（EHM），利用机载传感器和实时卫星地面运营中心传送的发动机健康数据，来实时检测上千台迺达发动机的工作状态，协助工程师了解每台发动机的情况，以便合理安排使用和维修时间，同时帮助设计更加高效低耗的发动机。罗罗公司的民用发动机都安装有EHM专用传感器，如迺达发动机上安装了25个永久传感器用于控制和健康管理。发动机运行时，飞机状态监控系统获得EHM数据，形成快照式报告发送给EHM专业分析机构并被自动处理，数据被“趋势化”，以便发现不同飞行架次之间发动机状态的细微变化。基于智能算法将多个传感器和多个飞行架次的数据结合分析，提供敏感的探测能力。与此同时，分析出的趋势化的数据发送到罗罗公司的客户服务网站上，协助供应商查看发动机机队的健康状态。

大数据技术在发动机行业中的应用分析架构

虽然不同阶段的数据价值不同，但大数据应用分析架构相同，主要分为数据层、平台层、应用层。以运行使用阶段的发动机飞行参数数据为例，整体的分析架构如图4所示。

数据层

数据层主要描述的大数据分析平台的数据来源，发动机使用阶段数据包括状态监测数据（性能参数、滑油消耗参数、振动参数、环境状态等）、飞行参数历史数据和使用过程中其他非结构化数据（出厂说明、使用操作、状态记录文档文件等）。

数据层要对所有涉及的待存储的数据，如数据规模、数据增长量、

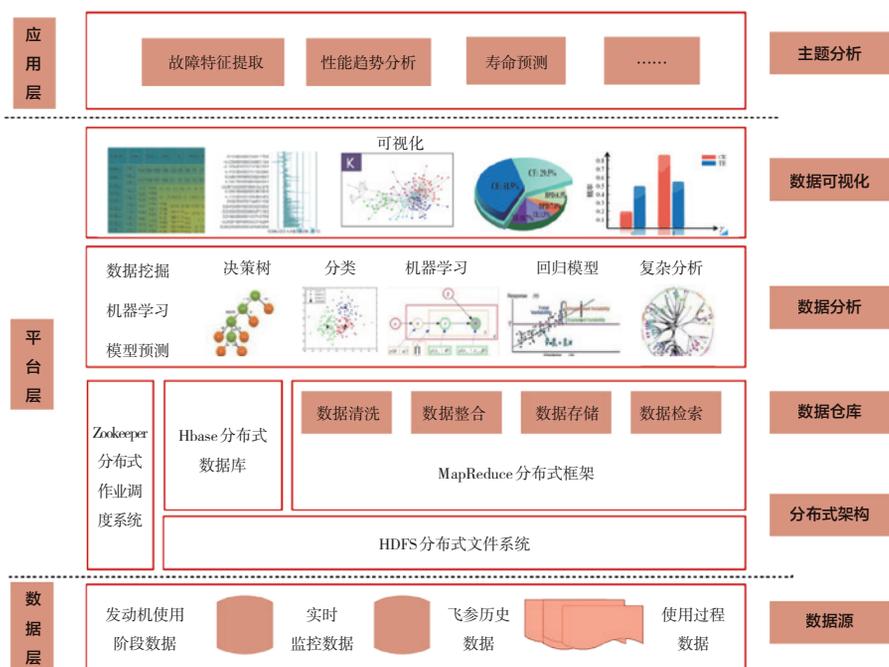


图4 大数据应用分析架构

数据格式、种类、采样周期、噪声水平，以及非结构化数据中关键数据提取的依据等，进行深入分析，为平台层中硬件系统和软件系统的构建提供数据依据。

平台层

平台层主要包括分布式硬件存储架构、数据预处理、数据分析、数据可视化等四个部分。为应对大规模数据的存储、调度和分析性能的需求，大多采用基于Hadoop的分布式硬件存储架构；数据仓库是研究数据源中数据的结构特点，通过飞行参数数据的预处理技术，形成针对主题分析的数据仓库。数据分析部分研究发动机飞行参数的数据分析算法，将数据挖掘、统计分析、机器学习等数据分析的方法集成到平台中。数据可视化技术研究发动机飞行参数数据的可视化技术，如表征相关性关系、聚类关系、统计关系等的可视化技术。

应用层

应用层是针对数据源的基于大数据应用分析平台的主题分析部分，在发动机全生命周期不同阶段的数据源具有不同的分析价值。例如，针对飞行参数数据可对发动机进行基于全量数据的发动机性能趋势分析，突破传统的用单一特征表示的退化特征（如排气温度等），重点分析使用环境（如温度、湿度、循环载荷等）对性能趋势的影响；通过建立多模态的高维数据张量模型，实现发动机状态数据完整表示。

在预测与健康中的应用

预测与健康（PHM）是发动机确保飞行安全、实现视情维修的关键使能技术，而数据挖掘是PHM系统实现发动机健康状态评估、诊断以及剩余寿命预测的核心技术。利用海量数据进行趋势分析是航空发动机预测与健康中的关键问题，

对于把握关键参数变化，并根据设定的报警阈值进行故障早期告警具有重要作用。PHM系统一般采用的是机上数据存储记录、地面保障站进行趋势分析的方案，由于每个飞行架次记录的数据量比较大，且地面保障存储有大量的数据，如何采用高效的数据挖掘算法充分抽取这些数据中短期、中长期、长期趋势是PHM系统的难点问题，如何运用分析的趋势对发动机预测整机寿命，并提出维修意见是研究的重点。

例如，可以选取多份发动机全生命周期内典型飞行阶段（巡航、最大状态等）的测量数据，分析各参数之间的相关性，选取相关性较低的数据，作为分析寿命的关键测量参数。根据损伤传播模型，建立测量参数与健康指标之间的关系，并拟合健康参数方程的未知系数，利用相似性算法，建立全生命周期的发动机健康参数模型，实现寿命预测功能。

结束语

大数据技术解决了传统样本数据的小规模、低吞吐量、低性能的问题，将解决问题的样本扩大到全样数据，并将多源信息快速整合，找出更多的与待解决问题相关联的因素。航空发动机数据具备工业大数据的特点，通过分析大数据技术在发动机领域中的应用架构，并遵循以价值驱动为导向的原则，本文给出了飞行参数数据在发动机寿命预测中的典型应用，希望可以为从业者提供些许借鉴。

航空动力

（王潘，中国航发研究院，工程师，从事航空发动机仿真研究工作。）