

探索走出符合国情的工业大数据自主之路

--工业大数据的范畴、关键问题与实践

文/王建民，清华大学教授、博导、软件学院副院长

2011 年麦肯锡全球研究院大数据报告表明，2009 年美国以装备制造为代表的离散工业领域拥有的数据规模为各领域之首，比美国政府拥有的数据还要多。近年来，随着德国工业 4.0 和美国工业互联网为代表的新工业革命深入发展，以及“中国制造 2025”、“互联网+”行动计划与“促进大数据发展行动纲要”的颁布实施，工业大数据得到了越来越多的关注。这里分享一下我们的思考与实践。

工业大数据三大来源：企业信息系统、装备物联网和企业外部互联网

企业信息系统、装备物联网和企业外部互联网是工业大数据的三大来源。

企业信息系统存储了高价值密度的核心业务数据。上世纪 60 年代以来信息技术加速应用于工业领域，形成了产品生命周期管理（PLM）、企业资源规划（ERP）、供应链管理（SCM）和客户关系管理（CRM）等企业信息系统。这些系统中积累的产品研发数据、生产制造数据、物流供应数据以及客户服务数据，存在于企业或产业链内部，是工业领域传统数据资产。

近年来物联网技术快速发展，装备物联网成为工业大数据新的、增长最快的来源，它实时自动采集了生产设备和交付产品的状态与工况数据。一方面，机床等生产设备物联网数据为智能工厂生产调度、质量控制和绩效管理提供了实时数据基础；另一方面，2012 年美国通用电气公司提出的工业大数据（狭义的），

专指装备使用过程中由传感器采集的大规模时间序列数据，包括装备状态参数、工况负载和作业环境等信息，可以帮助用户提高装备运行效率，拓展制造服务。当前互联网与工业深度融合，企业外部互联网已成为工业大数据不可忽视的来源。本世纪初，日本企业就开始利用互联网数据分析获取用户的产品评价，时至今日，小米手机利用社交媒体数据成功实现产品创新研发。此外，外部互联网还存在着海量的“跨界”数据，比如影响装备作业的气象数据、影响产品市场预测的宏观经济数据、影响企业生产成本的环境法规数据……

工业大数据实施的关键问题：数据质量、多源关联和系统集成

数据质量、多源关联和系统集成是工业大数据实施的关键问题。

拥有大数据不是目的，发掘其价值才是关键。由企业信息化数据、装备物联网数据和外部互联网数据汇聚而成的工业大数据，蕴藏着巨大价值。例如，通过分析用户使用数据改进产品，通过分析现场测量数据提高工件加工水平，通过工况数据进行产品健康管理等。

笔者认为实施工业大数据项目需要关注以下 3 个关键问题。

1.数据质量控制问题

原始数据（生数据）质量决定分析结果的质量。企业信息系统数据质量仍然存在问题，例如 2014 年某大型机车企业 ERP 系统中近 20%物料存在“一物多码”问题。装备物联网数据质量堪忧，某大型制造企业 1 个月的状态工况数据中，无效工况（如盾构机传回了工程车工况）、重名工况（同一状态工况使用不同名字）、时标混乱（当前时间错误或时标对不齐）等数据质量问题约 30%。

2.多源数据关联问题

层次化的物料表(Bill Of Material, BOM)定义了企业信息系统数据的核心语义结构。针对跨生命周期的研制 BOM 和实例 BOM 间结构失配问题,我们提出了中性 BOM 模型,并以此为核心,向前 关联设计制造 BOM,向后关联服务保障 BOM,形成星型结构,极大地降低了数据关联的复杂度。同时,针对装备物联网数据和外部互联网数据,可以根据其绑定的物理对象(零部件或产品)与相应的 BOM 节点相关联。从而以 BOM 为桥梁,关联 3 个不同来源的工业大数据。

3.大数据系统集成问题

工业大数据其来源更加广泛,并且装备物联网数据(半结构化数据)和外部互联网数据(非结构化数据)都要与企业信息系统(结构化数据)进行集成,因此要重构数据支撑平台,甚至替换“旧”系统。

工业大数据实施工程案例：提供故障分析新手段、提升备件需求预测精度

工业大数据分析提升工程装备服务保障水平。这里分享两个我们和国内企业合作实施的工程案例。

案例 1：工业大数据提供故障分析新手段

液压系统是工程机械的关键部件。2013 年我们发现液压系统的油缸密封套腐蚀故障数量异常。于是依据企业信息系统记录的液压系统维修历史数据,通过比对相关状态工况数据(装备物联网数据),搜索推荐与故障车辆关系密切的工况,发现车辆油缸换向频率的波动幅度与这些故障高度相关。

进一步,引入互联网上的行政区划数据和历年工程建设数据(外部互联网数

据)后,发现 2012~2013 年期间这些典型故障均发生在沿海省份,从而推断出盐雾环境是导致密封套腐蚀故障的主要诱因。

案例 2: 工业大数据提升备件需求预测精度

随着工程装备增量市场增长乏力,以维修保障为主的存量市场成为企业盈利新的增长点。我们利用了企业信息系统中的备件销售订单、采购订单和备件库存状态数据,以及工程物联网采集到的工况数据和外部互联网数据(如每个省的 GDP,建筑、交通等规划数据)。

针对 30 个省市区进行了备件需要预测,平均预测精度为 82%,每旬备件需求预测误差在 5 件或真实值的 20%以内。库存水平控制在一个较低的稳定水平,仅为原来库存水平的 48%。同时,因为考虑到了 20 天的配货周期,基于预测的补货策略可以保证现货满足率,消除紧急临时订单。如果按备件库存占有资金 1 亿元计算,可节约库存资金占用 5000 万元。

工业大数据是实现智能制造的基础原料,是提升工业生产力、竞争力、创新力的关键要素。然而必须看到,工业大数据是一个正在发展的学科领域,其内涵外延、模型理论、技术方法及其实施策略等还有待发展与创新。唯有结合中国国情认真实践,才能走出中国工业大数据自主之路,实现制造强国的战略目标。

作者:王建民,清华大学教授、博导、软件学院副院长。国家政府特殊津贴获得者、国家杰出青年科学基金获得者。国家 863 计划先进制造技术领域“面向制造业的核心软件开发”重大项目总体专家组组长(2012 年起),“十二五”国

家 863 计划先进制造领域制造服务技术主题专家组成员，国家卫计委人口健康信息化专家咨询委员会委员（2014 年起），国家“核高基”科技重大专项总体组成员（2008-2013）。2006 年入选国家教育部新世纪优秀人才支持计划，同年获得北京市师德先进个人；1998 年获国家科技进步二等奖（排名 4）；2012 年获得国家教育部科技进步奖一等奖（排名 1）；2013 年获得中国电子学会科技进步一等奖（排名 1），同年入选国家科技部中青年科技领军人才计划；2014 年入选国家“万人计划”，同年获得国家科技进步二等奖（排名 1）。