

网络技术——工业互联网专题

工业大数据技术

王建民 清华大学软件学院副院长,教授,博士生导师

摘要:工业大数据是指工业领域相关的海量数据,包括信息化数据、物联网数据以及跨界数据,已成为新工业革命的核心动力。本文分析了工业大数据的主体来源,阐述了工业大数据“多模态、强关联、高通量”的数据特点和“物理信息、产业链、跨界”多层次融合的应用特点,归纳了工业大数据管理与分析技术的关键问题,给出了我国工业大数据典型应用案例,最后对自主发展的工业大数据提出了建议。

关键词:大数据;工业大数据;数据特点;应用特点;关键技术;应用案例

1 工业大数据是新工业革命的核心动力

金融危机以来,为了寻找经济增长的新出路,特别是在云计算、大数据和物联网等新兴技术的推动下,全球掀起了以制造业转型升级为首要任务的新一轮工业变革,其中以德国工业4.0和美国工业互联网为典型代表。方兴未艾的新工业革命以数字化为基础、网络化为核心、智能化为目标,将新一代信息技术与现代制造业、生产性服务业等深度融合,以推动产业转型升级。根据麦肯锡全球研究院发布的研究报告表明,美国制造领域拥有的数据规模为各领域之首,大数据作为工业从自动化到智能化跨越发展的核心动力,其重要性不言而喻。

随着大数据技术与开源社区的不断成熟,大数据软件生态系统繁荣发展,商业配套工具日益丰富。IBM、EMC、华为等企业都在开源大数据软件之上封装了各自的商业版数据管理产品。这些技术的成熟为数据技术向工业界渗透提供了必要的条件,同时也为高端制造企业提供了巨大的市场机会。近年来,国际知名工业企业、软件公司和科研机构纷纷研发面向制造业转型升级的大数据产品和系统。

美国通用电气(GE)公司联合Pivotal向全球开放工业互联网云平台Predix,将各种工业资产设备接入云端提供资产性能管理(APM)和运营优化服务;美国PTC公司收购了物联网云平台公司Axeda,打造智能互

联产品ThingWorx;丹麦维斯塔斯(Vestas)公司联合IBM基于BigInsights大数据平台分析气象、传感器、卫星、地图数据支持风场选址、运行评估等工作;德国西门子公司面向工业大数据应用,整合远程维护、数据分析及网络安全等一系列现有技术和新技术,推出Sinalytics数字化服务平台,作为其实现工业4.0的重要抓手;德国SAP公司开发了面向物联网应用和实时数据处理的HANA大数据平台,并利用其在传统企业信息化ERP系统上的优势,推动HANA与信息化系统的集成;美国航空航天局(NASA)对外开放自身数据,帮助进行火星生命探测和天文观测等。此外,硅谷新兴创业公司也在积极投入工业数据的技术和产品研发,典型代表有Uptake Tech公司,为建筑、航空、采矿行业提供分析与预测软件服务。

国内工业大数据平台建设方面也有一定进展,主要依托国内互联网应用的基础,面向轻资产设备数据接入搭建通用平台,例如中国移动物联网开放平台、腾讯QQ物联平台等;在高端装备方面仍然是以龙头企业自建方式为主,例如陕鼓动力的鼓风机远程监测平台、三一集团工程机械物联网平台、远景格林威治风电云平台、红领制衣板型数据平台、南方航空公司航空大数据平台等。这些工业领域的大数据平台存在技术架构差异大、建设水平参差不齐、应用效果不明显等瓶颈问题。

未来,随着我国人口红利逐步消失,环境压力日益

加大,工业数据作为战略核心资产将成为我国制造业转型过程中实现价值留存和新价值创造的关键要素。在此背景下,国家相继出台《国务院关于加强发展生产性服务业促进产业结构调整升级的指导意见》等指导性文件,制定《“互联网+”行动计划》,颁布《中国制造2025》战略规划,特别是国务院《促进大数据发展行动纲要》与《中国制造2025重点领域技术路线图》都将工业大数据作为重点发展方向。

2 工业大数据的主体来源

工业大数据是指工业领域所产生的海量数据,包括信息化数据、物联网数据以及跨界数据。

首先,企业信息系统存储了高价值密度的核心业务数据。20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、供应链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。

以某装备企业2013年PLM系统数据为例,其管理零部件总数达165.5637万个,每周增长量为8182个;管理的Word文档数量共162.7929万个,每周增长量为9149个;管理的图纸共235.9777万张,每周增长量1.1486万张;数据总量为15.8T,日增长约11GB。

其次,随着物联网技术的快速发展,装备物联网成为工业大数据新的、增长最快的来源,它实时自动采集了车间内生产设备和交付给用户的产品状态与工况数据。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,拓展制造维修服务(Maintenance Repair and Overhaul, MRO)。

以风机装备为例,IEC61400-25标准由IEC TC88技术委员会起草制定,是IEC 61850标准在风力发电领域内的延伸,专门面向风电厂的监控系统通信,旨在实现风电厂中不同供应商设备之间的自由通信,通过对风电厂信息进行抽象化、模型化、标准化,实现各设备

之间的相互通信,使各设备之间具有互联性、互操作性和可扩展性。根据该标准,单台风机每秒产生225k字节传感器数据,每台风机按每年7000运行小时计算,单台风机每年生成6TB传感器数据,金风科技目前拥有2万台风机,6.0TB/台/年 \times 2万台=120PB/年。

最后,今天的互联网与工业深度融合,企业外部跨界数据已成为工业大数据不可忽视的来源。21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,今天小米手机利用社交媒体数据成功实现产品创新研发;此外,外部企业还存在着海量的“跨界”数据,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等。

以风电场运营业务为例,来自气象部门的跨界数据、模式数据与卫星数据,其每天处理数据增量为200TB,年70PB;来自环保部门的跨界数据、实时监控数据、环评数据、审批业务数据年增量为1PB。

3 工业大数据的主要特点

工业大数据本身不仅具有广义大数据的3V或4V特点,还呈现出“多模态”、“强关联”和“高通量”3个特点。

(1)多模态

所谓多模态,是指非结构化类型工程数据,包括设计制造阶段的概念设计、详细设计、制造工艺、包装运输等15大类业务数据,以及服务保障阶段的运行状态、维修计划、服务评价等14大类数据。例如,在运载火箭研制阶段,将涉及气动力数据、气动力热数据、载荷与力学环境数据、弹道数据、控制数据、结构数据、总体试验数据等。

(2)强关联

所谓强关联,一方面是指产品生命周期的设计、制造、服务等不同环节的数据之间需要进行关联,即把设计制造阶段的业务数据正向传递到服务保障阶段,同时将服务保障阶段的数据反馈到设计制造阶段;另一方面,在产品生命周期的统一阶段会涉及到不同学科、不同专业的数据。例如,民用飞机预研过程中会涉及总体设计方案数据,总体需求数据,气动设计及气动力学分析数据,声学模型数据及声学分析数据,飞机结构设计数据,零部件及组装体强度分析数据,多电系统模

型数据,多电系统设计仿真数据,各个航电系统模型仿真数据,导航系统模型仿真数据,系统及零部件健康模型数据,系统及零部件可靠性分析数据等,这些数据需要进行关联。

(3)高通量

所谓高通量,即工业传感器要求瞬时写入超大规模数据。嵌入传感器的智能互联产品已成为工业互联网时代的重要标志,是未来工业发展的方向,机器数据已成为工业大数据的主体。以风机装备为例,风机故障状态其数据采样频率为50Hz,每台平均125个测点,金风科技公司拥有2万台风机,其最高瞬时数据写入量超过1亿数据点/s。

工业大数据应用特点集中体现在物理信息、产业链以及跨界3个层次的融合,这与其他领域大数据应用具有明显差异,因此需要从数据模型、语义、查询操作3个层面对工业大数据进行一体化管理。

●“物理信息融合”表现在设计开发阶段主要管理数字产品,而在制造服务阶段主要管理物理产品,跨生命周期管理需要融合数字产品和物理产品,从而构建工业信息物理融合系统(Cyber-physical System, CPS)。

●“产业链融合”表现在互联网大数据环境下,以资源整合优化为目标的云制造模式得以迅速发展,智能产业链需要突破传统企业边界,实现数据驱动的业务过程集成。

●“跨界融合”是指在“互联网+”环境下,企业需要将外部跨界数据源进行集成,如美国某农机公司将天气数据、灌溉数据、种子数据以及农机数据进行综合利用,为农场提供粮食增产服务。

4 工业大数据的关键技术问题

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,如波音公司通过对737型号运行数据分析创新翼尖小翼;通过分析现场测量数据提高工件加工水平,如高端机床设备通过实时测量数据动态优化进刀参数;工况数据进行产品健康管理,如罗罗公司通过对发动机工况数据支持售卖飞行小时新型商业模式等。

工业大数据技术可以分为两个层次,即大数据管

理技术和大数据分析技术。

(1)工业大数据管理技术的关键问题

●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,某大型制造企业一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(如当前时间为1999年,或时标对不齐)等数据质量问题约30%。

●一体化管理问题

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。具体实现机制可以分为3个层面:逻辑层负责统一数据建模,定义数字与物理对象模型,完成底层数据模型到对象模型映射;概念层实现数据语义层面的融合,通过语义提取与语义关联,形成RDF形态的知识图谱,提供基于SPARQL的查询接口;操作执行层负责异构数据管理引擎的查询协同优化,对外提供SQL以及RESTAPI形式的统一查询接口。

●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,因此要重构数据管理平台,甚至替换“旧”系统。

●工业大数据安全问题

工业大数据及其系统是工业互联网的知识与大脑。由于暴露在无所不在、无所不控、虚实结合、多域融合的工业信息物理融合系统中,工业大数据系统安全威胁和风险控制技术,是国家工业主权与产业安全的重要保障。

(2)工业大数据分析技术的关键问题

●多时间尺度数据综合分析

与传统商务智能(Business Intelligence, BI)应用场景不同,工业大数据包括秒/小时尺度的机器工况数据、天/周尺度的车间调度数据和月/年尺度的管理决策

数据,在工业大数据应用中不同时间尺度的工业数据需要综合使用才能驱动智慧企业。例如,智慧矿山企业需要根据跨界数据预测矿石品质需求,依据品质需求生成装备作业计划,根据装备作业计划实时控制装备操作。

●专业分析算法与构件

针对时间序列、时空等弱结构化数据,应提供丰富的特征模板库,方便对典型物理事件(如风速平稳时段、发电机转速快速下降、环境温度逐渐上升等)进行描述;另外,还应提供丰富的时间序列、时空模式、序列模式的深度挖掘算法库,提升工业数据分析的建模效率。

●实时交互分析

工业大数据分析必须能够满足大规模、分散控制和交互迭代等需求。在实时处理上,传统的商业数据分析系统不能有效支持面向大规模数据状态下的低等待时间复杂事件检测。在离线分析上,查询检索与分析建模应紧密协同,同时前台探索展现与工业大数据平台要无缝整合,支持大数据集上的交互挖掘。

●遗留算法并行化

针对重要的应用需求,工业企业通常具有一定的分析工具和科学计算软件积累,然而这些工具通常没有考虑大数据架构。如何有效重用这些分析算法工具,是工业大数据分析应用不能回避的技术问题。

●领域知识自动化

工业大数据中存在着大量的非结构化数据,将当前深度学习、自然语言处理等人工能力融入到工业大数据分析中来,可以有效辅助把专家知识进行有效的沉淀、萃取和自动化。

5 工业大数据的典型工程案例

金风科技股份有限公司通过将风机大数据技术作为抓手,实现企业的工业化和信息化的深度融合,通过每台风机上安装的百种传感器,金风科技每天收集的数据超过TB的量级。通过这些数据技术,可以在实现全量风机的在线预警以及定期体检上,寻找更迅速、更有效的解决方案,提升设备的可靠性,而风场运营商通过这些数据,可以做到更精准的预测、检测和数据采集以节约人力和时间成本,实现风场效益最大化。通过机器大数据技术对风机数据进行管理,为金风创造了

可观的经济效益,典型案例如下:

●快速响应客户需求

在美国德州 Rattlesnake 风场建设项目投标中,金风公司基于大数据平台成功以3小时47分的速度完成招标方提出近250万个文件的处理要求,加工出风速、温度、发电机转速、功率、桨距角、风机状态等结果,使回应速度由100多天缩短到2天。

●风速仪对风优化

在风场对风机的变桨数据进行调优。基于平台在5小时31分完成基于全量历史数据(包括1.7万台风机、300多亿条历史记录)的对风偏差估计和优化,使得原本需要400多天才能完成的运维调优任务得以周期性实施。据测算,每年由此调优获得额外发电量可达2亿元人民币以上。

●齿形带断裂预警

齿形带是变桨系统的一个重要部件。齿形带断裂会造成非计划性停机及叶片失控等次生灾害。因为没有直接的传感器监测,目前检测手段根据断裂后引起的振动、桨距角异常等次生故障判断,通常有10余秒的滞后。大数据分析基于大量风场的历史故障信息,对SCADA(Supervisory Control and Data Acquisition,监控和数据采集系统)瞬时数据的时间序列模式提取,挖掘桨距角一致性、变桨过程曲线模态、振动模式、变桨电机温差、ng5充电电流差异等断裂征兆模式,通过多模型融合和深度学习,提前90h进行断裂预警,通过预防性维修消除重大故障隐患。大数据还对20ms数据进行分析,分析断裂后发电机转速、风速、功率、电磁扭矩等多指标时间序列模式变化,将当前PLC超短时报警提前1s,降低故障可能的破坏风险。

●机组健康管理

建立并完善系统安全体系;建立整机、大部件健康评估模型,实现预警功能;建立风电场级健康安全管理及优化体系;建立解决方案中心,实现模型引擎的远端运行,达到既能做出风力发电机组健康综合评估诊断也能做出解决方案的“医院”级功能,成为健康评估模型中探索模型的“试验平台”,可实现各类模型的快速迭代优化。

6 结束语

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。中国不仅是制造大国更是使用大国,在装备应用过程中产生的海量机器数据正是工业大数据的主体,为我国智能制造创新发展提供了独特而丰富的战略资源;同时,对于国计民生息息相关的装备制造业而言,工业大数据更是事关国家安全与主权。工业大数据是一个正在发展的

学科领域,在内涵外延、模型理论、技术方法及其实施策略等方面,我国和国际巨头处在同一起跑线上。必须紧密结合中国国情认真实践,加大投入,快速研发,走出中国工业大数据自主发展之路,实现国产化工业大数据软件技术的弯道超车与应用的自主可控,支撑制造强国战略目标落地实现。

Industrial big data technologies

WANG Jianmin

Abstract: Industrial big data consist of enterprise data, IOT data and external data, whose volume is very huge and which is a primary driven-force of the new-round industry revolution. Where do the industrial big data come from is explored first. Next, the data characteristics, “heterogeneous, interrelated and high-throughput”, and the application characteristics, “physical-cyber, inter-enterprises and cross-domain” data integration, are described. The technical challenges on data management and analysis are presents thereafter. An industrial big data use case on wind turbine farm are given. Finally we suggest we should build an industrial big data software platform by ourselves.

Key words: big data, industrial big data; data characteristics; application characteristics; key technologies; use case

(收稿日期:2016-07-21)

爱立信携手 SmarTone 在香港推出 Wi-Fi 通话服务

近日,爱立信携手香港运营商 SmarTone 部署了全球首个多终端 Wi-Fi 通话服务,可在多个智能终端包括平板电脑、个人电脑、智能手表上,使用同一个手机号码。

SmarTone 的多终端 Wi-Fi 通话服务可以让用户同时配对 5 个没有 SIM 卡的智能终端,支持通过 Wi-Fi 进行语音通话。用户无需在终端上额外安装 APP。这些智能终端可以通过不同的 Wi-Fi 热点接入网络,比如用户将手机留在家中,可以用笔记本电脑接听电话。即使在通话过程中也可以从一台终端转换到手机上继续通话。此外,用户还可以使用同一个电话号码,同时使用手机和其它终端进行通话。

爱立信携手沃达丰展示新的 5G 概念验证

爱立信与沃达丰在今年巴展上宣布联合开展 5G 创新后,双方日前携手展示了新的 5G 概念验证,共同打造了准 5G 核心网原型,以及 5G 智能网络边缘原型,并使用机器视觉应用展示了网络切片和分布式云技术的优势。

在爱立信亚琛研发中心举行的创新日活动展示的现网演示中,爱立信与沃达丰合作展示了 5G 智能网络边缘如何提升效率。例如,网络延迟降低,云技术人脸识别应用的识别率大幅提高。需要通过广域网传输的视频流量大幅减少,并且敏感数据保存在本地,从而能够更好地防止非法访问。

沃达丰德国公司沃达丰创新园主管 Sonja Graf 表示:“只用了短短 3 个月,我们就通过双方的实验室‘对接’,打造出 5G 智能网络边缘原型。人脸识别用例作为一个展示 5G 技术如何满足各行各业各种需求的示例。”