

干货清华陆薇：释放工业大数据价值

7月14、15日，首届大数据应用大会在成都召开。大数据文摘作为战略合作媒体，今日带来干货内容--清华大学数据科学研究院·工业大数据研究中心的常务副主任陆薇题为《释放工业大数据价值》的演讲。



在首届中国大数据应用大会上，清华大学数据科学研究院·工业大数据研究中心的常务副主任陆薇作了题为《释放工业大数据价值》的演讲。（大数据文摘记者魏子敏 摄）

◆ ◆ ◆

演讲原文如下：



很高兴有这样一个机会和大家分享我们中心对工业大数据的理解和一些工作内容。



首先，简单介绍一下清华大学数据科学研究院工业大数据研究中心。清华大学数据科学研究院成立于 2014 年春，结合了清华在信息技术，特别是分布式系统和数据管理分析技术方面的优势，以及工科、经济、人文、健康等大数据应用领域的深厚积淀，致力于发挥学科交叉的协同作用，推进大数据系统研究与应用实践，培养人才，同时也参与贡献大数据相关国家战略。



众所周知，第四次工业革命以美国的工业互联网、德国的工业 4.0 为代表，根据各国制造业不同的发展优势而各具特点。美国制造业大量外包生产环节，比如波音公司的飞机部件是在全球多个国家制造，但是美国制造企业牢牢占据高知识产权和附加值的产品设计和服务环节，同时把控整个生态链的上下游为其服务。因此，美国提出的工业互联网的智能核心在云不在端，关注利用互联网技术的资源优化配置作用提升对整个产业生态链的效率和价值。德国的情况则正相反，德国制造业优势在于实体制造，有精良的生产设备和工艺手段。因此，德国提出的工业 4.0 重点在于智能工厂本身的建设，并以制造为核心，带动上下游业务发展。

我国工业具有自己的特点。一方面，我国是世界工厂，实体制造比重大，同时低水平、劳动密集、高资源消耗制造的比重也大，产业升级迫在眉睫。另一方面，我国互联网产业发展具有领先优势。过去十多年消费互联网的高速发展使互联网技术得到长足发展，互联网思维深入人心。我们需要充分发挥这一优势，并将之和制造业紧密结合，促进制造业升级和生产性服务业的发展。

因此，我国制订的“中国制造 2025”战略从我国实际国情和发展需求出发，兼顾智能制造和制造服务，用互联网技术驱动制造全生命周期从设计、制造到交付、流通、服务、回收各个环节的数字化、互联化、智能化升级，推动制造全产业链智能协同，优化生产要素配置和资源利用，消弭低效中间环节，从而整体提升中国制造业发展水平和世界竞争力。

The image is a composite of three parts. At the top is a banner for the 'BDA6 中国大数据应用大会' (China Big Data Application Congress 2016) held in Chengdu from July 14-15, 2016. The banner title is '第四次工业革命以数字化、互联化、智能化为特征，大数据是关键支撑手段' (The 4th Industrial Revolution is characterized by digitalization, interconnection, and intelligence, with big data as a key supporting means). Below the banner on the left is a circular infographic titled '德国工业企业调查：工业4.0的机遇与挑战' (German Industrial Enterprise Survey: Opportunities and Challenges of Industry 4.0). It lists ten points: 1. Industrial internet transforms the entire company and must be part of the CEO agenda; 2. By 2020, European industrial companies will invest 4.5 billion annually in industrial internet applications; 3. In five years, more than 80% of companies will have digitized their value chain; 4. The industrial internet opens the way for new, often disruptive digital business models; 5. Digital products and services generate approximately 11 billion of additional revenue per year for the European industry; 6. The industrial internet creates better production and resource efficiency - an 18% increase in efficiency after four years; 7. The industrial internet enables the entire value chain to be digitized; 8. The industrial internet opens the way for new, often disruptive digital business models; 9. Horizontal co-operation allows for improved satisfaction of customer requirements; 10. The industrial internet transforms the entire company and must be part of the CEO agenda. A callout box states '数据的整合分析和使用是关键能力' (Data integration and use is a key capability). On the right is a diagram titled 'GE工业互联网白皮书' (GE Industrial Internet White Paper) showing three pillars: 1. Intelligent Machines (connecting world's machines, factories, fleets, and networks with advanced sensing, controls, and software applications); 2. Advanced Analytics (combining power of physics-based analytics, predictive algorithms, automation and deep domain expertise); 3. People at (connecting people to work or on the move, any time to support more intelligent design, operations, maintenance and higher service quality and safety). A callout box states '先进的分析能力是关键要素' (Advanced analysis capability is a key element). Below the GE diagram is a quote: '“互联网+”行动计划将重点促进以云计算、物联网、大数据为代表的新一代信息技术与现代制造业、生产性服务业等的融合创新，发展壮大新兴业态，打造新的产业增长点，为大众创业、万众创新提供环境，为产业智能化提供支撑，增强新的经济发展动力，促进国民经济提质增效升级' ——摘自《政府工作报告中相关词语的注释》' (The 'Internet+' action plan will focus on promoting the integration and innovation of new-generation information technologies such as cloud computing, IoT, and big data with modern manufacturing and producer services, etc., to develop and expand emerging business forms, create new industry growth points, provide an environment for mass entrepreneurship and innovation, provide support for industrial intelligence, enhance the new economic development momentum, and promote the upgrading and improvement of the national economy's quality and efficiency). The quote is attributed to '摘自《政府工作报告中相关词语的注释》' (Quoted from the explanation of related terms in the Government Work Report).

虽然发展重点各有侧重，但大数据被公认为是不可或缺的关键技术。

德国的《工业 4.0 十大挑战与机遇》报告指出，数据的整合分析与使用是实现工业 4.0 的关键能力。工业 4.0 有两大关键支撑技术。一个是 CPS，这是智能工厂比较偏硬的部分，讲究用更加智能的基础设施来降低车间复杂度和提高灵活性。另一个是数字化企业平台，跨生产“shop floor”和经营“top floor”、贯穿 CAD/PLM/MES/ERP 等生产、经营信息系统的信息集成与数据融合贯通，建立伴随产品制造过程的完整信息流，做到全数字化、数据全采集、数据通路全打通、数据流动过程不落地。同时基于对这些数据的分析，使企业可以全面深入把握和优化提升产品质量、生产效率、资源利用率。

美国通用电气公司的《工业互联网白皮书》中指出工业互联网实现的三大要素是智能联网的机器、人与机器协同工作及先进的数据分析能力。工业互联网的核心是通过智能联网的机器感知机器本身状况、周边环境以及用户操作行为，并通过这些数据的深入分析来提供诸如资产性能优化等制造服务。

我国在今年推出“中国制造 2025”“互联网+制造”战略的时候，也特别强调利用大数据为产业智能化提供支撑。

有一个对工业 1.0 到工业 4.0 变迁历史的总结非常精辟，借用一下：工业 1.0 是“工厂+机械”，利用机械替代人力，解放了生产力；工业 2.0 是“工厂+电”，通过电气化进一步提升生产效率；工业 3.0 是“工厂+电脑”，通过信息系统替代人工管理生产经营过程。那工业 4.0 是工厂+什么呢？顺着这个思路，3.0 完成信息化之后，工厂经营生产的方方面面都已经有数据积累，再进一步的提升自然是基于这些数据进一步分析优化生产经营和探索数据驱动的新业务模式，因此，工业 4.0 就是“工厂+大数据”。



- 清华大学数据科学研究院工业大数据研究中心简介
- 工业大数据是新工业革命的核心技术要素
- 典型工业大数据应用需求及核心技术
- 工业大数据应用案例

工业大数据应用前景广泛， 对国计民生具有重大价值

制造



- 智能工厂
- 智能联网产品运维服务
- 绿色制造

电力



- 广域电网测量WAMS
- 能源装备的远程运维
- 新能源分析与优化
- 停电预测和过载预警

航空



- 飞行安全监控
- 航空器维护
- 航路管理，能耗优化

石油



- 物探大数据管理和共享
- 油气生产物联网
- 管道完整性管理、能耗优化

轨道交通

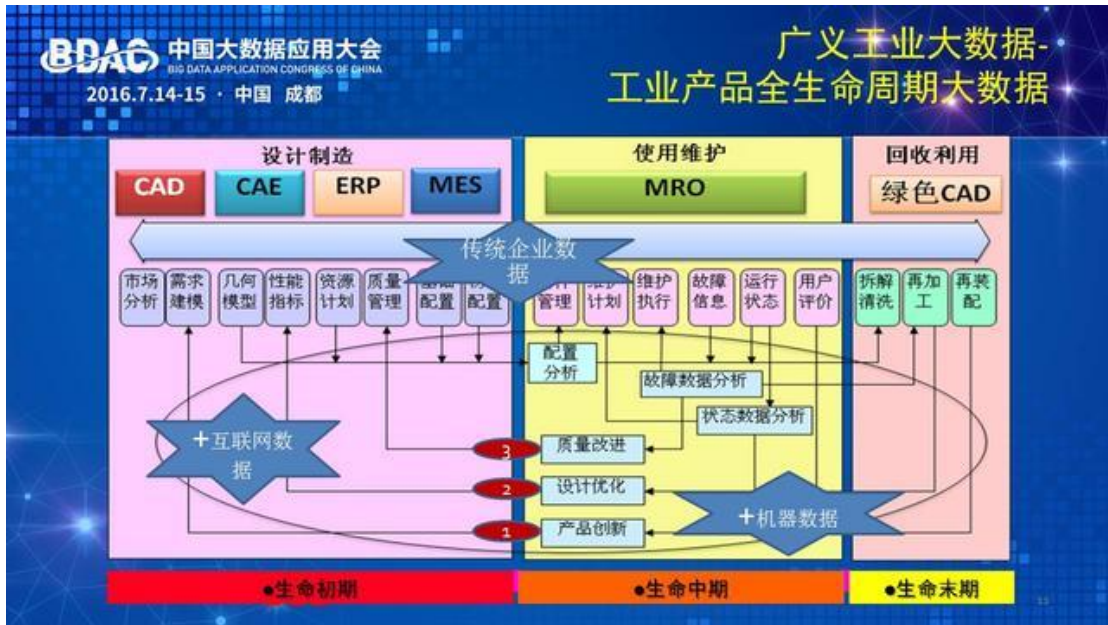


- 列车运维
- 行车安全
- 环境安全

船舶



- 设备综合保障
- 降低船舶能耗
- 航海安全数据服务



工业大数据从哪里来？来源于产品生命周期的各个环节，包括市场、设计、制造、服务、再利用各个环节，每个环节都会有大数据。“全”生命周期汇合起来的数据更大。当然，企业外、产业链外的“跨界”数据也是工业大数据“不可忽视”的重要来源。



机器数据（物联网数据）

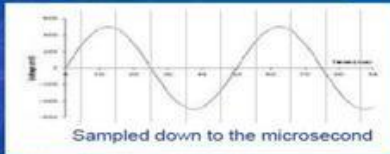
1 定义

时间序列
位置轨迹
实时安全

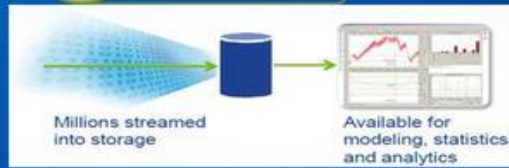
2 来源多样



3 密集采样



4 存储和分析



工业数据分析特点

	工业大数据	商业大数据
研究对象	以物理实体与环境为中心	以互联网支撑的交互
已有的基础	中/微观机理模型与定量领域知识, 在当前基础上前进很困难	宏观理念与定性认识, 存在广阔的提升空间
新的驱动力	新的感知技术、产品的服务化转型	新的交互渠道 (如社交媒体)
对分析的期望	因果关系才有用 模型的高可靠性 (很难接受概率性的预测)	相关性关系就非常有帮助 大数原则



◆ ◆ ◆

要利用好机器大数据，需要突破几类核心技术

首先是数据怎么管理的问题。海量的机器型数据如时间序列、时空数据等高速采集完成后，需要把它存下来，这涉及到数据有效打包、压缩、放置的问题。数据存下来是为了被利用，这需要支持快速定位查询到应用需求的数据，而这又是一个如何建立高效的时空数据索引的问题。

数据存好管好了，下一个问题就是如何支持各种分析。做过实际分析的人都知道，分析绝不仅仅是开发一堆算法的问题。算法只是一小部分工作，大部分的工作是根据对业务问题的理解选取需要的数据，理解数据的特征，然后根据特征设计一个合适的模型和算法。这中间数据特征的理解对机器大数据来说是很难的。因为机器数据不能为人所直观理解，需要交互特征工程。此外，从模型和算法的层面，机器数据往往是对一个物理世界系统的感知结果，而物理世界有许多机理性的原理存在，比如机械领域涉及力学原理，冶金领域涉及化学原理，因此机器大数据的分析需要有机结合机理模型和数据统计模型。还有

一个常常被忽略的问题是数据质量的问题——如何把握数据质量，如何修正数据质量。

再谈应用的角度，如何更简单地访问数据和使用分析，特别是对领域专家。在多源异质数据，屏蔽数据集成关联的问题，使得领域专家不需要懂复杂的大数据技术和编程。



大纲

- 清华大学数据科学研究院工业大数据研究中心简介
- 工业大数据是新工业革命的核心技术要素
- 典型工业大数据应用需求及核心技术
- 工业大数据应用案例



应用案例：工程机械在线运维系统

- 基于主机工况的在线健康监控，故障诊断与预测；了解用户使用习惯，改善用户体验；提升企业经营效率
- 几万台工程机械，秒级数据采集，峰值每秒20万条数据



全国主机在线情况

图例：
■ 主机
■ 主机
■ 主机
■ 主机
■ 主机
■ 主机
■ 主机
■ 主机
■ 主机
■ 主机



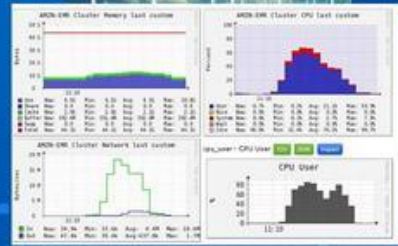
支持研发

支持质保

支持服务

应用案例：风力发电机能效提升

- 定期调整风机偏航角以优化风机发电功率
- 每14秒回传一次的风速、风向、功率、桨叶角等近200个指标，上万台风机
- 基于全量历史数据的分析优化



Thank you!