

工业大数据： 挖掘“不可见世界” 中的价值

李杰 (Jay Lee)、刘宗长 | 文

“有之以利、无之以用”是出自老子《道德经》中的一句话，其中的智慧放在当今工业大数据的价值模式中依然十分适用。





Order Pick Time

00:05:15

Order Pick Time

002:09:27

随着新工业革命时代的序幕徐徐拉开，物联网、工业互联网、智能 ICT 技术、人工智能等技术成为舞台上最受瞩目的新星。在这些新兴技术的推动下，工业领域中的大数据环境正在逐渐形成，数据从制造过程中的副产品转变成为备受企业关注的战略资源。如果大数据是工业价值转型的核心驱动力，那么我们该怎么去定义和使用大数据？关于这个问题，我们在《工业大数据》这本书中曾表达过一个观点：大数据并不是目的，而是分析问题的一种视角和解决问题的一种手段。通过对数据的洞察，可以预测需求、预测制造、挖掘不可见世界的价值、解决和避免不可见问题的风险，以及利用数据去整合产业链和价值链，这些才是大数据的核心价值和目的。

什么是工业大数据？

一提到大数据，人们首先会想到在互联网和商业等环境中，利用大量的行为数据来分析用户行为和预测市场趋势等应用。但是对工业大数据的定义和应用却很难直观地理解和想象。现在对大数据最为流行的定义，来自于维克托·迈尔-舍恩伯格（Viktor Mayer-Schönberger）和肯尼斯·库克耶（Kenneth Cukier）编写的《大数据时代》中提出的 4V 特性，即 Volume（数据量大）、Velocity（流动速度快）、Veracity（准确性）和 Variety（来源多样）。这个定义是针对互联网和社会环境中的大数据，从数据工程的技术挑战方面所提出的。而工业大数据的挑战和目的则要通过“3B”和“3C”来理解。

工业大数据应用的“3B”挑战：

Bad Quality——在工业大数据中，数据质量问题一直是许多企业所面临的挑战。这主要受制于工业环境中数据获取手段的限制，包括传感器、数采硬件模块、通信协议和组态软件等多个技术限制。对数据质量的管理技术是一个企业必须要下的硬功夫。

Broken——工业对于数据的要求并不仅在于量的大小，更在于数据的全面性。在利用数据建模的手段解决某一个问题时，需要获取与被分析对象相关的全面参数，而一些关键参数的确会使分析过程碎片化。举例而言，当分析航空发动机性能时需要温度、空气密度、进出口压力、功率等多个参数，而当其中任意一个参数缺失时都无法建立完整的性能评估和预测模型。因此对于企业来说，在进行数据收集前要对分析对象和目的有清楚的规划，这样才能够确保所获取数据的全面性，以免斥巨资积累了大量数据后发现并不能解决所关心的问题。

Below the Surface——除了对数据所反映出来的表面统计特征进行分析以外，还应该关注数据中所隐藏的相关性。对这些隐藏在表面以下的相关性进行分析和挖掘时，需要一些具有参考性的数据进行对照，也就是数据科学中所称的“贴标签”过程。这一类数据包括工况设定、维护记录、任务信息等，虽然数据量不大，但在数据分析中却起到至关重要的作用。

工业大数据分析的“3C”目的：

Comparison（比较性）——从比较过程中获取洞察，既包括比较相似性，也包括比较差异性。比较的维度既可以

是在时间维度上与自身状态的比较，也可以是在集群维度上与其他个体的比较。这种比较分析能够帮助我们将庞大的个体信息进行分类，为接下来寻找相似中的普适性规律和差异中的因果关系奠定基础。

Correlation (相关性)——如果说物联网是可见世界的连接，那么所连接对象之间的相关性就是不可见世界的连接。对相关性的挖掘是形成记忆和知识的基础，简单地将信息存储下来并不能称之为记忆，通过信息之间的关联性对信息进行管理和启发式的联想才是记忆的本质。相关性同时也促进了人脑在管理和调用信息时的效率，我们在回想起一个画面或是情节的时候，往往并不是去回忆每一个细节，而是有一个如线头一样的线索，你去牵它一下就能够引出整个场景。

这样的类似记忆式的信息管理方式运用在工业智能中，就是一种更加灵活高效的数据管理方式。

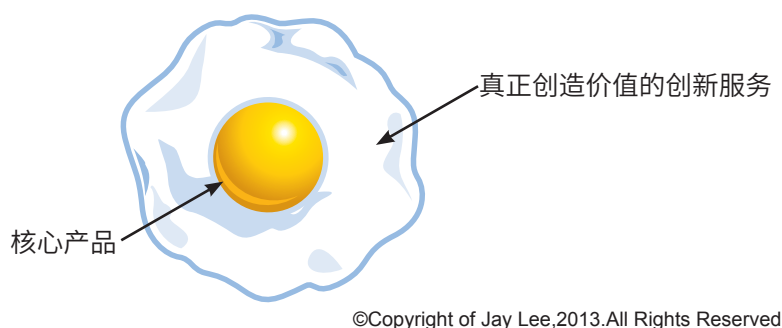
Consequence (因果性)——数据分析的重要目的是进行决策支持，在制定一个特定的决策时，其所带来的结果和影响应该被同等地分析和预测。这是以往的控制系统所不具备的特性，也是智能化的本质。工业系统中的大部分活动都具有很强的目的性，就是把目标精度最大化，把破坏度最小化的“结果管理”。结果管理的基础是预测，例如制造系统中，如果我们可以预测到设备的衰退对质量的影响，以及对下一个工序质量的影响，就可以在制造过程中对质量风险进行补偿和管理，制造系统的弹性和坚韧性就会增加。

总结而言，互联网和商业大数据与

表 互联网 / 商业大数据与工业大数据之间的区别

	 互联网/商业大数据	 工业大数据
数据需求	庞大的样本数据量	数据样本的变量全面性
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，注重统计分析的显著性，意义的参考性多为机会导向	强调特征之间的物理关联和因果性关系，意义的参考性多为精确的效益导向
分析目的	以发散式的分析目的为主，通过挖掘数据中各个属性之间相关性的统计意义进行决策支持	目的更具收敛性，强调利用数据对具备一定理论和经验基础的现象及规律进行精确的定性和定量分析与关系挖掘
分析结果准确性要求	较低	较高

图一 工业价值的煎蛋模型新思维



工业大数据在技术挑战、数据属性和分析目的等方面有很多区别，这也决定了两者技术手段的不同。

工业大数据的价值及意义

在工业和商业从分离到融汇的过程中，我们观察到，有四个值得关注的重要转变，客户需求的转变，生产系统的转变，商业模式的转变，工业系统运转模式（决策模式）的转变。客户需求的转变，是从无到有、从有到精、再从精到个性化。生产系统的转变，是从简单到复杂、从复杂到规模化、从规模化到精益化、从精益化到柔性化、从柔性化到智能化，最终目的是实现无忧的生产环境。商业模式的转变，是从卖产品到卖能力、从卖能力到卖服务、从卖服务到卖价值，目的是去挖掘用户不可见的价值。工业系统运转模式的转变，从应激式的解决问题，到基于经验的预防问题，现在正利用数据向基于线索和事实来避免问题的模式转变，最终目的是实现知识的获取和传承。这一系列转变背后的核心是工业价值的转型，大数据在这个转型的过程中将起到举足轻重的作用。简单

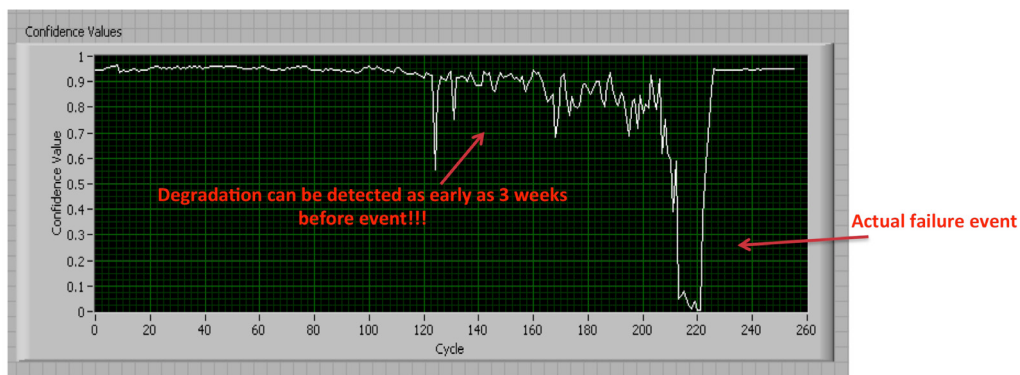
来说工业大数据的目的是实现从自造到制造，从制造到智造，从智造到传承。

利用大数据挖掘“不可见世界”中的客户价值

“有之以为利、无之以为用”是出自老子《道德经》中的一句话，其中的智慧放在当今工业的价值模式中依然十分适用。这句话可以理解为：一切事物的实体为我们提供可以凭借的可见的基础条件，而其中所隐藏的空间和可变化的无限可能才是被我们真正使用并创造价值的所在。

我们在《工业大数据》一书中，曾用煎蛋模型来阐述产品与服务价值之间的关系（见图一）：蛋黄代表的是产品自身，其差异性和客制化程度并不明显，例如一台电视机在挡住了 Logo 之后就很难被区分出来是哪家公司生产的。而蛋白所代表的增值服务却是差异化和客制化的重要体现，也是企业的品牌和可持续性价值的所在。这些价值存在于用户的使用场景、隐形因素的相关性和产品被制造和使用的全生命周期这些“不可见世界”中。数据将成为挖掘这些价值的重要手段，主要体现在：利用大数据挖掘在使用中获得新的知

图二 机器人健康衰退预测



识和技术对现有产品进行改进；利用数据去发现和定义用户未知的需求；以数据为媒介向用户提供增值服务。

以风力发电为例，风机本身的差异化并不明显，用户的定制化需求也并不强烈，但是风机在运行过程中的发电能力、运行稳定性和运维成本等却是用户价值的核心。利用风机的运行大数据可以对风机进行健康管理、对潜在的运行风险进行预测和对风场的运维进行优化，从而提升风机的可用率、改善发电效率和降低运维成本。风机的制造厂商也可以不再仅仅通过卖出装备获得一次性的盈利，还可以通过向用户提供使用过程中的增值服务实现持续性的盈利。

以数据分析为驱动的智能运维服务也成为许多企业进行价值转型的重要方向。

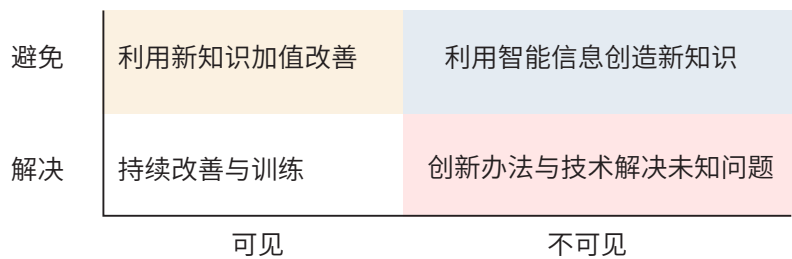
IMS（智能维护系统）中心与日产公司（NISSAN）合作，在工业机器人健康管理方面引进了工业大数据预测分析模型。由于工业机器人被大量使用，且生产环境十分复杂，因此不适合安装外部传感器，而是使用控制器内的监控参数对其健康进行分析。日产的工业机器人中有相当一部分是六轴机械臂，任何一个轴发生故

障都会造成机械臂的停机。在使用伺服轴的转速信号对机械臂的工况进行区分后，再对每一个工况内的状态参数（如扭矩和温度等）建立健康评估模型，从其分析结果中可以发现在故障发生前的三个星期前就可以预测到早期故障特征（见图二）。

随后日产开始在六轴伺服机械臂上推广预测分析模型，在分析和采集大量机械臂数据后，对不同种类和运行工况的机械臂进行了聚类分析，形成了一个机械臂的“虚拟社区”，通过集群建模方法对社区机械臂进行数据分析，通过比较每一个机械臂与集群的差异性来判断其异常程度，并对集群内所有机械臂的健康状态进行排序。

在对机械手臂的健康状态进行量化分析之后，日产对分析结果进行了网络化的内容管理，建立了“虚拟工厂”的在线监控系统。在“虚拟工厂中”，管理者可以从生产系统级、产线级、工站级、单机级和关机部件级对设备状态进行垂直立体化的管理，根据设备的实时状态进行维护计划和生产计划的调度。该系统还能够每天生成一份健康报告，对产线上所有设备的健康状态进行排序和统计分析，向设备

图三 可见与不可见的转型思维



管理人员提供每一台设备的健康风险状态和主要风险部位，这样在日常的点检中就可以做到详略得当，既不放过任何一个风险点，也尽可能避免了不必要的检查和维护工作，实现了从预防式维护到预测式维护的转变。

人类社会在经历了 200 多年的科技革命后，已经积累了巨大的工业产品存量，工业的基础设施和大量基本生产要素，如机床、电力设施、动力设施、制造装备、交通装备等需求都已逐渐趋于饱和。以德国为例，其工业出口产值从 2006 年开始已经连续 6 年没有增长，根本原因就在于发展中国家已逐渐完成工业化升级，对工业装备的需求已经基本饱和。因此德国的“工业 4.0”战略中将面向制造系统的集成和软件服务作为重点，具体表现在“纵向集成”、横向集成”和“端到端集成”。同样发现这个问题的还有美国 GE 公司，他们意识到装备销售过程中的获利远远不及在产品使用过程中的价值服务，客户需要的价值也远不止对产品状态的保持，更在于如何去使用这些能力来实现更高效的价值再创造。

以数据为核心使产品发挥最大的能力，归根结底是利用数据建模实现对状态、环境和任务的精确评估，对管理和控制活

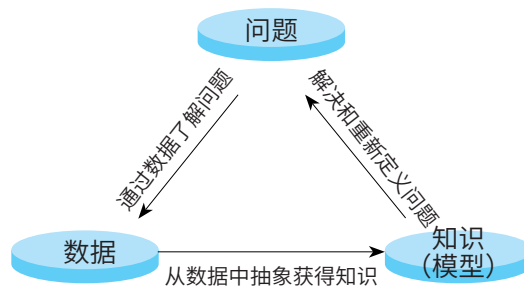
动进行实时的决策优化，并协同和调度相关产品高效率运行的过程。

利用大数据实现无忧虑的制造环境

制造系统中的问题同样也有“可见”和“不可见”之分，我们对待这些问题的方式既可以在问题发生后去解决，也可以在问题发生前去避免。生产系统中存在的“不可见”问题包括设备性能的衰退、精度的缺失、易耗件的磨损和资源的浪费等，可见的问题往往是这些不可见因素积累到一定程度所引起的，比如设备的衰退最终导致停机、精度的缺失最终导致质量偏差等。就如同冰山一样，可见的问题仅仅是冰山一角，而隐性的问题则是隐藏在冰山下面的恶魔。通过大数据对“不可见”问题获得深刻的洞察，是实现无忧虑制造环境的基础，也是智能制造的本质。

制造改进与转型的机会空间可以被分为四个部分（见图三），第一个部分是去满足用户可见的需求和解决可见的问题，这个空间内依然有中国制造需要补的课，比如质量、污染和浪费等问题，需要的是持续的改善与不断完善的标准化。第二个空间在于避免可见的问题，需要从使用数据中挖掘新的知识对原有生产系统和产品做加值改善。第三个空间在于利用创新的

图四 问题、数据、与知识的关系及管理方式



方法与技术去解决未知的问题与创造新的竞争力，例如具有自省能力的设备，以及利用传感器与大数据使不可见的问题透明化，进而去管理和解决不可见的问题。第四个象限是寻找和满足不可见的价值缺口，避免不可见因素的影响，这部分需要利用大数据分析产生的智能信息去创造新的知识和价值与传承力，这也是工业 4.0 的最终目标。

通过分析数据，预测需求、预测制造、利用数据去整合产业链和价值链，这就是工业大数据的思维。工业 4.0 是一场在不可见世界中的战争，而工业数据分析的竞争力则是连接可见与不可见世界的桥梁。

利用大数据实现无忧虑的制造环境有三个方向，数据在每一个阶段中扮演的作用也并不相同。第一个方向是在解决可见问题的过程中积累经验和知识，从而去避免这些问题。在这个过程中，数据可以作为经验和知识的载体。第二个方向是依靠数据去分析问题产生的隐性线索（evidence）、关联性和根原因等，进而利用预测分析将不可见问题显性化，从而实现解决不可见问题的目的。现在的制造系统正在经历从第一个阶段到第二个阶段的转变过程，在完成这个过程后，制造系统将不再有“surprise”，使得所有隐性

问题在变成显性问题和影响之前都可以被提前解决。第三个方向是通过对数据的深度挖掘，建立知识和问题之间的相关性，从数据中启发出新的知识，并能够利用知识对制造系统进行精确的建模，产生能够指导制造系统活动的镜像模型，从系统的设计端避免可见及不可见问题的发生。

这三个方向对企业都非常具有借鉴意义，但是需要对不同的情况也适用于不同的方向。总地概括，这三个方向分别适用于以下几类情况中问题的解决：

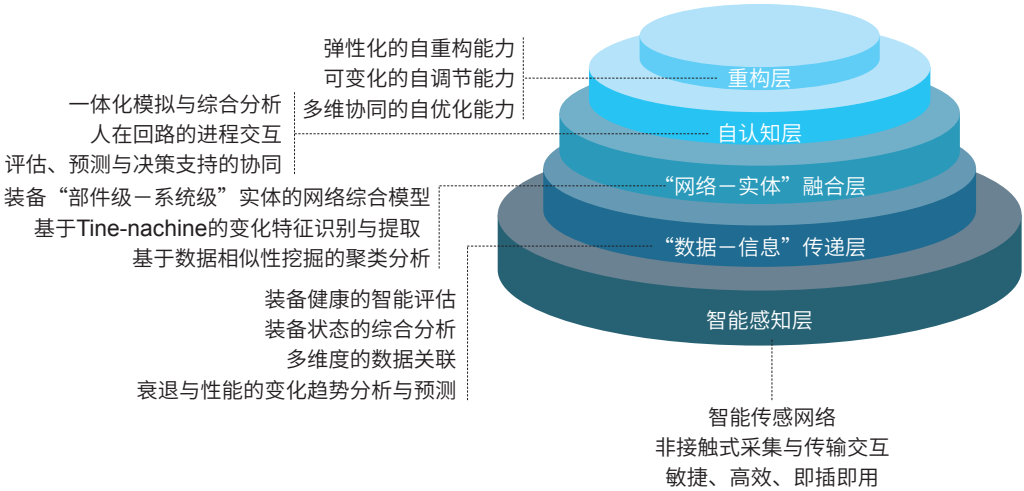
第一个方向：适合在某一个领域已经经营了很久，有了一定的经验积累，但是却很难总结出为什么做得好或是不好。

第二个方向：在解决了可见的问题之后，仍然存在一些不可见问题对制造系统造成的影响，希望能够了解不可见因素的变化过程和相互的关联性，积累更加深入的制造知识。

第三个方向：在制造基础还较为薄弱的领域，并没有形成太多有效的数据，但是拥有非常丰富的使用数据和经验，则可以借助使用过程中积累的知识对制造系统提出设计的要求。

图四简明地阐述了大数据与智能制造之间的关系：制造系统中“可见”和“不可见”问题的发生及解决的过程中会产生

图五 CPS 的 5C 构架



大量的数据，通过对这些数据的分析可以了解问题产生的过程、造成的影响和解决的方式。通过对数据的深度挖掘和抽象化建模后可以形成知识，进而利用知识去认识、避免和重新定义问题。数据在其中起到的作用，是使这个过程从以往依靠人的经验（Experience based）转向依靠挖掘数据中隐性的线索（Evidence based），使得制造知识能够被更加高效和自发地产生、利用和传承。因此，问题和知识是目的，而数据则是一种手段。今天我们来谈利用大数据实现智能制造，是因为大数据正在逐渐成为易得的资源，而在制造系统和商业环境变得日益复杂的今天，利用大数据去解决问题和积累知识将是更加高效和便捷的方式。

以 CPS 为构架的工业大数据分析技术

Cyber-Physical System (CPS)，译为“信息 - 物理系统”，是美国科学基金会（NSF）在 2006 年提出的新技术

概念，并将此项技术体系作为新一代技术革命的突破点。同时，德国的“工业 4.0”战略也将 Cyber-Physical Production System (译为信息 - 物理生产系统) 作为核心技术，实质是 CPS 在生产系统中的应用体系。

CPS 是一个具有清晰架构和使用流程的技术体系，针对工业大数据的特点和分析要求，能够实现对数据进行收集、汇总、解析、排序、分析、预测、决策、分发的整个处理流程，实现对工业数据进行流水线式的实时分析能力，并在分析过程中充分考虑机理逻辑、流程关系、活动目标、商业活动等特征和要求。因此可以作为工业大数据分析中的智能化体系的核心。

为应对工业大数据的分析要求和技术挑战，我们提出了工业大数据环境下的 CPS 技术体系架构（见图五），包括了 5 个层次的构建模式：智能感知层（Connection）、信息挖掘层（Conversion）、网络层（Cyber）、认知层（Cognition）和配置执行层（Configuration）。

这个 5C 的分析构架设计的目的是为了满足不同工业大数据环境中的分析和决策要求，其特征主要体现在以下几个方面：

①智能感知：从信息来源、采集方式、和管理方式上保证了数据的质量和全面性，建立支持 CPS 上层建筑的数据环境基础。

②信息挖掘层：从低价值密度的数据到高价值密度信息的转换过程，可以对数据进行特征提取、筛选、分类和优先级排列，保证了数据的可解读性。

③网络层：重点在于网络环境中信息的融合，将机理、环境与群体有机结合，构建能够指导实体空间的网络环境，包括精确同步、关联建模、变化记录、分析预测等。

④认知层：将机理模型和数据驱动模型相结合，保证数据的解读符合客观的物理规律，并从机理上反映对象的状态变化。同时结合数据可视化工具和决策优化算法工具为用户提供面向其活动目标的决策支持。

⑤配置执行层：根据活动目标和认知层中分析结果的参考，对运行决策进行优化，并将优化结果同步到系统的执行机构，以保障信息利用的时效性和系统运行的协同性。

结语

工业大数据的目的并不是追求数据量

的庞大，而是通过系统式的数据收集和分析手段，实现价值的最大化。所以推动工业价值转型和智能制造的并不是大数据本身，而是大数据分析技术所带来的洞察，行动的准确性与速度。在新制造革命的转型中，能否更加有效地积累和利用数据资源与知识的传承，决定了企业能否在新竞争环境中脱颖而出。

工业大数据定义了制造价值的新主张，这个价值的应用既可以外向，也可以是内向。内向是利用大数据去解决和避免制造系统中的“不可见”问题，实现无忧无虑的制造环境。外向是利用大数据在产品的使用过程向用户提供智能增值服务，实现制造价值的延续。这两者对于中国制造而言，一方面是解决制造“大而不强”的挑战，另一方面是改善制造附加值较低的瓶颈。中国应该利用好使用数据的资源，不断提升企业对制造的理解和知识积累速度，才能弥补中国在装备制造和核心零部件等方面的弱势，逐步弥补这些弱势领域造成的短板，让世界看到中国在工业大数据中创知和创值的成功经验。

数据本身不会说话，也并不会直接创造价值，真正为企业带来价值的是数据分析和挖掘之后产生的洞察和行动的价值，是数据经过实时分析后及时地流向决策链的各个环节，是让数据成为面向客户创值服务的媒介和依据。■

作者简介

李杰

美国辛辛那提大学特聘讲座教授，美国国家自然科学基金会 (NSF) 智能维护系统产学研合作中心 (IMS) 创立主任，专注于全球工业大数据分析与智能维护系统技术研发

刘宗长

北京天泽智云科技有限公司技术研发副总裁，专注于工业大数据分析和“信息物理系统” (CPS) 构架及核心技术研究