

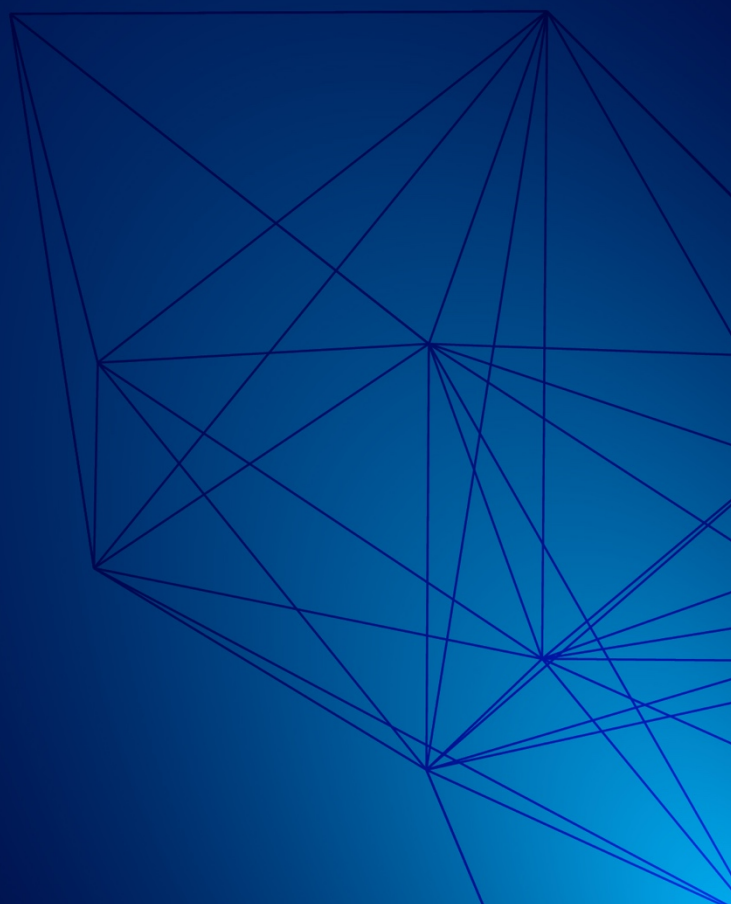


北大汇丰

PHBS FINANCIAL FRONTIER DIALOGUE

金融前沿对话

2018 年第 11 期 总第 11 期



PHBS HFRI
北京大学汇丰金融研究院

主办单位：北京大学汇丰金融研究院

院长：海闻

执行院长：巴曙松

秘书长：本力

编辑：都闻心（执行） 叶静 曹明明 鞠琳琳 方培豪 朱伟豪

北京大学汇丰金融研究院简介

北京大学汇丰金融研究院 (The HSBC Financial Research Institute at Peking University, 缩写 HFRI) 成立于 2008 年 12 月, 研究院接受汇丰银行慈善基金会资助, 致力于加强国内外著名高校、金融研究机构、以及知名金融学者之间的交流, 构建开放的学术、政策交流平台, 旨在提高中国金融理论与政策的研究水平, 促进学术繁荣与发展, 加强与政府决策部门的联系与合作, 为政府决策提供参考意见, 为国际金融体制改革和中国金融业的发展做出贡献。

北京大学汇丰金融研究院院长为北京大学校务委员会副主任、北京大学汇丰商学院院长海闻教授, 执行院长为中国银行业协会首席经济学家、中国宏观经济学会副会长巴曙松教授。

数据背后的人、业务、技术和科学

主持：巴曙松（北京大学汇丰金融研究院执行院长、中国银行业协会首席经济学家、中国宏观经济学会副会长）

嘉宾：胡本立（国际数据管理协会中国区创始人、主席）

数据本身很简单，它是我们每天生活中听到的、感知到的，无处不在的东西，与我们的工作很近，甚至是我们每个人生命过程的一部分。实际上从人一出生，就开始本能的学习对周围环境数据的处理。数据难在它的过程，尤其是人介入的部分，包括它带来的一系列的机会和问题。人和数据这个题目要讨论清楚还有很长的路要走，甚至要等到人工智能讨论清楚以后，但是近年来大数据、机器学习、各种虚拟概念和应用包括区块链、各种币的到来，使反思和理解人与数据成为回避不了的现实问题。把混乱的概念梳理清楚，考虑加入人的因素，虽然不能完全解决问题，但对于认知科学领域里面的人，以及处理模型、数据会有很大帮助。

一、反思数据究竟是什么——一个比什么是“大数据”更根本的问题

图 1 展示了数据与机器之间循环的生命周期图。图中主要的观点是，数据是自然界产生，没有人干预的或者干预很少。另外大量的数

据是人产生的，这当中对数据的理解、解释，通过语言的表达是很复杂的过程。对人的输入来讲，数据是物理信号（是生理刺激），人通过语音、图像、文字、表情输出数据，对他自己或别人，也是信号。图 1 对于大家理解和讨论数据的归属权、隐私等目标和需要都有帮助，但是要把人的整个过程尤其是认知过程和表达过程讲清楚，数据归属权和认知就不是一个简单的技术问题。

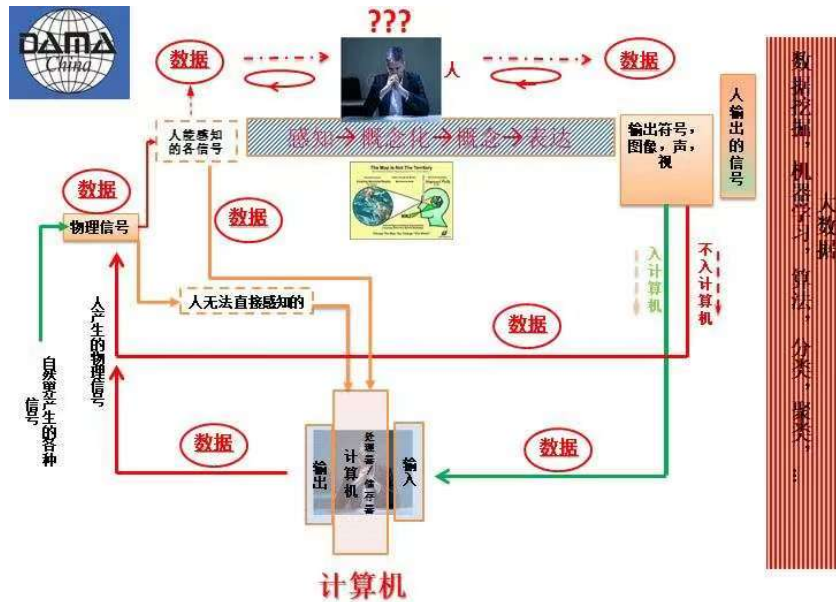


图 1

图 2 描述了对人来讲与数据有关的三个“世界”：实体客观世界，概念世界和概念的表示世界。在哪个“世界”里讨论的问题不应与另一世界的混淆（比喻例外），但它们间（通过人）有复杂的映射和联系。人与数据的互动建立了这三个不同“世界”间的桥梁，包括“虚”与“实”；“主观”与“客观”等重要基本概念。有不少分类把“表示世界”认为是虚拟的，实际上它也是物理的，虚拟的还是在人的脑

中。“表示世界”与“客观世界”的不同是前者是由人产生的，而后者不是。举例来说，人通过学习知道狗的发音、狗的概念，我们通常认为“狗”就是数据，实际上人工智能开始的时候不论中文英文，都不单单是个符号，而是图像化的东西，抽象后成为“狗”的概念。而男孩和女孩、老人脑中狗的概念又和他（她）以前的经历以及不断收集的信息有关，两种概念不完全一致。通过这个例子可以理解为什么我们的沟通这么困难，数据会不一样，都跟过程有关，不可能完全一样。

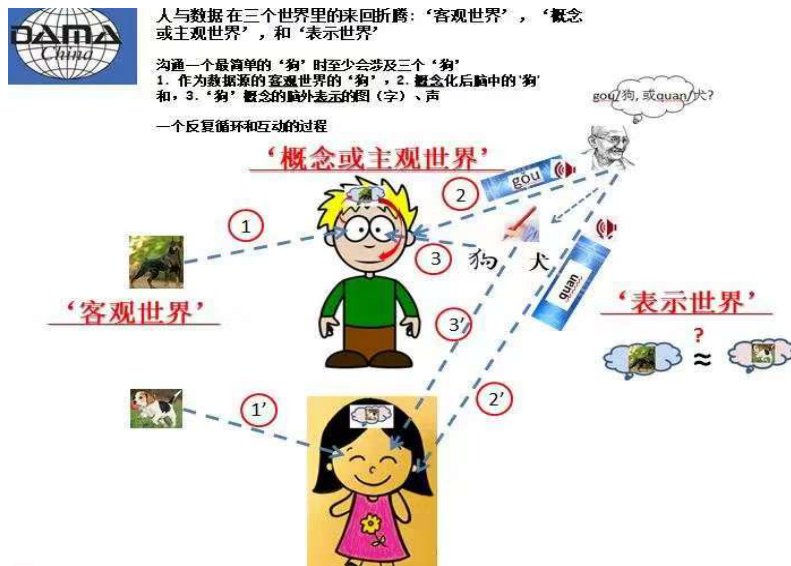


图 2

图 3 描述了对同一对象，不同人会有不同理解和解读（不对称）这一现实。成年以后，通过学习、经历等因素，再来看两个同样世界中的事物，如一堆标签和一篇文章，从物理上来看，这些都是视觉上的信号，看过之后结合自身的知识经验，再回到概念层，两个人的表达可能一样也可能不一致。数据管理中一个很大的问题就是怎么达成

共识，尤其是一个团队要达成共识的话是怎样一个过程，这个过程是效率很低或是很难根本做不成，理解了这个过程，不管是在数据层次还是概念层次，对我们达成共识都会有帮助。



同一客观物理实体，包括文字文章，不同人看后的结果会一样？

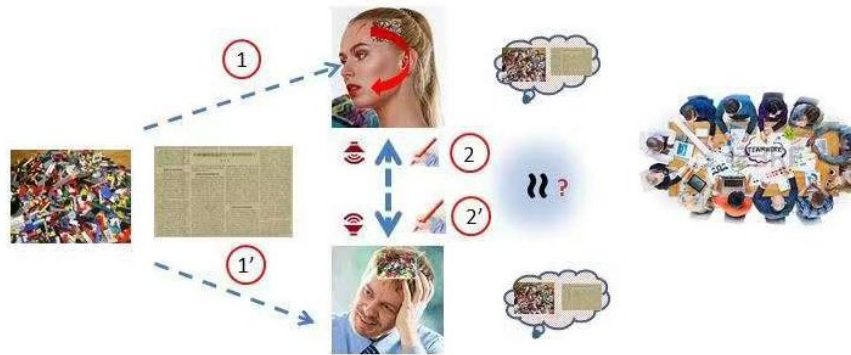


图 3

在传统数据处理的深层次困难如打通数据孤岛、数据共享问题方面，不是一个简单过程。比如数据集（data set）共享，这也是目前大部分共享的层次，但有些时候要比简单共享数据集复杂的多。以杨振宁对电磁学和纤维丛概念术语之间的关联为例，与我们所说的数据共享的距离就很远了。对于一些直接测量的物理数据，人与人之间的沟通相对比较容易，但是要到更抽象的层次，就跟每个人的经验视角有关，这就需要一套新的人跟数据打交道处理的方法论工具，能够帮助大家发现他们的共同之处和分歧。

二、 数据背后的人、业务、技术和科学

大部分人认为数据这个问题是通过技术来解决，更多是 IT 部门的事情，实际上不仅仅是 IT 或技术的问题，还有人和业务的因素在内，这两个不同的专业在关注和处理数据模型的做法，以及模型从公式到数据库的关系怎么衔接融合，里面的缺口怎么填补，我觉得是需要大家更多的参与和讨论的问题。

语言是个很复杂的问题，语言、对话成了数据后对数据管理行业的冲击和影响很大。我们每天都在通过语言沟通、解释，语言变成数据后怎么处理是个比较大的挑战。以风险管理为例，原纽约联储的主席 Jerry Corrigan 曾说过一段话“风险管理，对我来讲归根到底简单的说，就是要把正确的信息在正确的时间提交给正确的人，所以人们能够问正确的问题以作出最好的判断。这是最根本的。”他高度概括了金融风险除了技术以外应该怎么管理，这里好几次都提到了人。一个金融模型再好，如果管理层不能理解模型的背景、算法、统计等，也无法发挥最大的效果。

脑科学和人工智能分别各占一端，传统数据管理在中间的话，那么数据背后到底是什么科学呢。比较起来我觉得最接近数据科学的应该是认知科学或者认知心理学。因为有人的因素在里面，所以我看数据背后的科学在哪个领域，是看这个领域怎样能够把脑科学和人工智能的一些进展联系起来，然后跟传统数据如何结合。

三、 目前全球数据管理面临的几个主要挑战

一是仍较普遍的认为数据问题是个技术的问题，或可被新技术手段或工具来解决。二是行业间数据的割裂，比如传统的，人工智能的，各种币和区块链的，跨领域不同的术语和标准体系等。现在大家都在说数据整合，我觉得要从数据全周期的角度来考虑这些问题。三是数据的共享和交叉过程远比一般理解的要复杂的多，数据的产生者和使用者不一定对称，互联网的产生使信息更加对称解决了部分问题，但还有更深层次的问题没有解决。Feynman 曾说过“the rhythm and pattern of phenomena of nature which is not apparent to the eye, but only to the eye of analysis”，眼睛感官看到的和经过脑处理的数据并不是一回事。四是元数据（描述数据的数据）仍是数据的纠结。用图 1 来解释的话，第三世界里的数据仍是数据，一部分数据是用来描述另一部分数据，这与场景有关，但都是数据。上个月有家媒体分析了特朗普与其他总统风格的区别，以前奥巴马时代发表的话大家听了就会接受，在美国国会和参议院达成共识的过程中，可以有不同的解释或者不同的意见，但都是基于一样的事实（fact）。现在特朗普的智库就提出有 alternative facts，每个人对事实怎么看有不一样的解释。哪些是事实，哪些是数据，哪些是意见（意见本身也是数据）已很难划分，讨论变得困难。这些反映了我们的确进入了另一个数据时代，社会、学术界、业界都需要反思一下数据究竟是什么，为什么这么容易又这么难。

四、 几点看法和建议

一是理解数据是以人为本的过程，人处于中心的地位，数据的问题有一部分也是人产生的，怎么达成共识是一个关键的问题。不论场景是业务还是技术，两种场景如何融合沟通达成共识，还是要回到以人为中心上来。

二是少发明一些不必要的名词。发明了一个新的名词，要解释这个名词是什么，不管是大数据、人工智能，还是其他很多新名词，当发明一个名词，就需要解释这个名词跟原来其他名词的区别、关系，增加了很多信息熵的问题。就像“奥卡姆剃刀”原则，把一些不需要的名词去掉，便于沟通、讨论、深入认识真正需要讨论的。

三是比喻是人的语言和沟通中很重要的一个工具，但不能把比喻当现实。比喻是为了启发语言，便于大家相互之间的理解，但和现实之间是完全不一样的目标对象。在讨论各种币时，往往用物理世界的一些概念套用到概念世界里面去，后果是很严重的，比喻在用过以后就应该收敛回来讨论具体的问题。

四是复杂问题的讨论框架可否借鉴物理学的分析解题方法。现在很多讨论数字货币、电子货币、纸币、物物交换等，我想可以借鉴物理学中的“隔离法”做一些区分。比如货币里面需要密码处理，但是密码学又是另外一个独立的学科，分布式协议是去中心化或半中心化，在技术层次如何处理数据已经有很成熟的讨论，在每个领域互相沟通、介绍，对讨论的效率和质量都会有帮助。

本文根据北京大学汇丰金融研究院执行院长巴曙松教授发起并主持的“全球市场与中国连线”第二百八十四期（2018年12月16日）内容整理而成，特邀嘉宾为国际数据管理协会中国区创始人、主席胡本立先生。

胡本立先生，国际数据管理协会中国区创始人和现任主席。曾任世界银行首席技术官，全面负责世行信息技术架构的设计、开发和运行，支持世行的战略和使命；继后任香港证券及期货事务监察委员会首席信息官、顾问；中国证监会战略及规划委员会委员，期间建议和推动了证监会和国内两个交易所要求国内上市公司信息披露采用的国际标准；中国社保基金理事会高级技术顾问；惠普中国咨询总顾问；曾担任中国和印度二家上市公司的独立董事提供相关战略、业务和技术建议。曾任国际数据管理协会副主席，负责数据标准；近年来主要专注数据，特别是数据在人和计算机之间复杂的交互和认知过程中的地位和作用 - 数据后面的工程和科学。胡先生拥有物理学学士以及计算机科学硕士学位，完成哈佛大学商学院，斯坦福大学商学院和欧洲工商管理学院合办的对企业高级主管关于战略、变革管理、金融和财务的培训。

【免责声明】

“全球市场与中国连线”为中国与全球市场间内部专业高端金融交流平台。本期报告由巴曙松教授和居姍博士共同整理，未经嘉宾本人审阅，文中观点仅代表嘉宾个人观点，不代表任何机构的意见，也不构成投资建议。

本文版权为“全球市场与中国连线”会议秘书处所有，未经事先书面许可，任何机构和个人不得以任何形式翻版、复印、发表或引用本文的任何部分。



PHBS HFRI
北京大学汇丰金融研究院

