

数据 资产论

DATA ASSET THEORY

王汉生◎著

DATA ASSET THEORY

数据资产论

王汉生◎著



中国人民大学出版社

· 北京 ·

图书在版编目 (CIP) 数据

数据资产论/王汉生著. —北京: 中国人民大学出版社, 2019. 6
ISBN 978-7-300-26395-3

I. ①数… II. ①王… III. ①数据处理-信息产业-研究-中国 IV. ①F492

中国版本图书馆 CIP 数据核字 (2018) 第 248527 号

数据资产论

王汉生 著

Shuju Zichan Lun

出版发行	中国人民大学出版社		
社 址	北京中关村大街 31 号	邮政编码	100080
电 话	010-62511242 (总编室)		010-62511770 (质管部)
	010-82501766 (邮购部)		010-62514148 (门市部)
	010-62515195 (发行公司)		010-62515275 (盗版举报)
网 址	http://www.crup.com.cn http://www.ttrnet.com (人大教研网)		
经 销	新华书店		
印 刷	北京联兴盛业印刷股份有限公司		
规 格	148 mm×210 mm 32 开本	版 次	2019 年 6 月第 1 版
印 张	6.75 插页 2	印 次	2019 年 6 月第 1 次印刷
字 数	118 000	定 价	59.00 元

版权所有 侵权必究

印装差错 负责调换

前 言

人们都说，这是一个大数据时代，是数据产生价值的时代，更是数据成为资产的时代。真的是这样吗？

要认真探讨这个问题并不容易。首先，要理解什么是数据。对于数据，可以给出多种不同的定义，各有其道理。但是，哪种定义能够帮助我们更好地理解数据产业？理解它的昨天，今天，还有明天呢？其次，要理解什么是资产。会计学对资产的定义是什么？数据就是资产，这个说法站得住脚吗？根据会计学的定义，资源要成为资产有一个必要条件：带来预期的经济收益。那么什么是经济收益？我认为就是商业价值。由此可见，不是所有的数据都可以称为资产，只有那些能够产生商业价值的的数据才可以称为资产。因此，数据创造价值的过程，就是数据资源资产化的伟大历程！

数据如何产生价值？纯粹而孤立的数据有商业价值吗？没有。数据价值的彰显需要一个必要条件：业务场景。只有合适的的数据配合合适的业务场景，才能产生价值。业务场景可以孤立于数据之外而产生价值，但是数据却无法孤立于业务场景之外产生价值。因此，数据价值之道在于：场景为王，数据次之，算法最后。所以，数据分析的第一步，也是最重要的一步，不是分析数据，不是建立模型，而是分析业务。对业务缺乏充分的理解，不可能让数据产生价值。为此，需要一套规范的方法论，帮助我们从业务角度定义数据分析，从数据分析角度实践业务。这就需要具备一种能力，即把抽象的业务问题转变为数据可分析问题。这是数据分析最困难、最具挑战也是最了不起的地方。为此，你需要重新理解：什么是回归分析。

人们通过各种数据建模的技术手段，对业务的核心指标做出预测，希望可以无限准确。但悲催地发现，无论如何努力都无法做到非常高的预测精度。你可能开始怀疑人生，变得焦虑、苦恼、迷失方向。其实不必慌张，这一切都很正常，这才是数据分析的真实世界。对真实的数据分析而言，预测不准是常态，预测过度准确才是一个让人非常担忧的“变态”。为什么？因为预测过度准确常常是一个假象、一个幻觉。预测过度准确可能是人们犯了愚蠢的错误。所以，任何时候我看到预测过度准确的模型，第一反应是：错了！从未失手！因此，你需要坦然接

受预测不准的常态，这是大千世界无常的一个正常反映。所谓无常就是难以预知的变化，就是不确定性！在实践中，不确定性随处可见。不确定性常常由两部分组成：一是无知；二是无奈。无知可以通过数据分析、模型学习慢慢化解；无奈却让人束手无策，而无奈恰恰决定了预测不准是常态！

既然预测不准是常态，为什么还要做预测？一个预测不准的场景能有什么用处？其实，这才是数据模型最有价值的场景。在这个场景下，由于不确定性的影响，数据模型预测精度可能算不上完美，但是没有数据模型加持的经验方法往往表现更差。因此，数据模型的价值主要体现在预测精度的相对优势上（相对于经验方法），而不是对绝对预测精度的无限追求。事实上，对于一个业务场景，如果数据模型可以预测准确，那么经验方法常常也会表现优异，甚至与数据模型旗鼓相当。此时，数据模型并没有相对优势。深刻理解了预测不准是常态这一道理，你的世界观就变了。虽然仍在努力改进数据，改进模型，提高预测的精度，但已不再排斥不确定性。开始拥抱预测不准这一常态，看待数据商业价值的高度也就变了！

任何事物都有它的两面性。数据带来价值的同时，也带来了伤害：隐私泄露。相信你也深受其害。我家小朋友上了一个英语补习班，其他英语学校的骚扰电话就接踵而来，不胜其烦！因此，全社会对于隐私保护的意识越来越强烈。但为了隐私保

护而保护隐私，就如同医生给病人看病，头痛医头，脚痛医脚，不治根本。隐私保护的**根本是：数据确权与合规。仔细思考，什么是隐私保护？就是保护同隐私相关的数据。为此，你需要得到法律的支持，即我才是这个数据唯一的合法拥有者。因此，需要对数据确立清晰的产权，这就是数据确权。数据确权非常重要，不局限于隐私保护。如果数据的产权都不明晰，它又怎能成为可交易的数据资产呢？但现实中的关键问题是：数据如何确权。数据资源的生产过程很复杂，往往多方参与其中（企业、政府、用户等），而且数量巨大（海量用户）。这一点同实物资产很不一样，需要一些独特的创新，需要法学的智慧！

最后讨论的是，数据资产如何定价？定价的逻辑是什么？同实物资产相比，有什么异同之处？数据资产与实物资产的相同之处在于，资产的价格无法通过一个神奇的数学公式准确计算。因为同样的资产（无论是数据资产，还是实物资产），在不同的场景下价值各不相同，无法给出一个绝对准确的价格。而资产只能遵循最基本的经济学供求关系：高价格提高供给、降低需求，低价格降低供给、提高需求。但是，价格与供需之间到底是什么样的数学关系，无法描述，只能交给市场，通过大量的交易自动碰撞摸索形成。因此，交易才是定价的核心微观基础。要想给数据资产定价，就要促进数据资产交易。那么数据资产应该如何交易？交易的形态应该如何设定？数据资产和

实物资产的不同在于，复制成本几乎为零。因此，如果数据资产模仿实物资产，以金钱为媒介进行交易恐怕欠妥。因为购买方可以零成本立刻复制多份盗版向市场非法售卖，其监管成本极其高昂。由此，我大胆猜测，数据资产交易需要一种独特的交易形态和交易场所。这是本书最后为大家分享的内容。

以上就是本书想要讨论的内容，总结了我对数据产业的思考和瞎想，纯思想性内容，没有任何数学公式。所有的讨论都关注在一个焦点上：践行数据商业价值，也即数据资源资产化。希望探索一条践行数据商业价值的道路，发展一套数据资源资产化的方法。显然，我不能保证书中所有的观点就是正确的，恰恰相反，其中必有荒诞、谬误甚至陷阱。但是，不经历荒诞，不尝试谬误，不蹚平陷阱，哪来的康庄大道？希望你不要袖手旁观，能够为我加油，并提出建设性意见。我想以此书为践行数据商业价值的同行加油喝彩，为我国的数据产业讴歌，为数据资源资产化的伟大历程做忠实记录！

目 录

CONTENTS

第一章	不确定性的无知与无奈	1
	从不确定性开始	2
	什么是不确定性?	4
	不确定性事件广泛存在	5
	不确定性推动科研	10
	不确定性创造商机	13
	产生不确定性的原因	16
第二章	数据的时代特征	22
	数据的朴素定义	22
	数据的时代特征	25
	运营商数据	28

	支付交易数据	31
	手机数据	32
	社交网络数据	36
	位置轨迹数据	39
	浏览日志的数据	41
第三章	朴素的数据价值观	43
	数据分析的核心	43
	价值表现的三个方面	44
	参照系	47
	价值与业务场景	51
	价值与商业模式	55
	可被归因的价值	59
第四章	回归分析的“道”与“术”	64
	统计学不研究统计	64
	回归分析与不确定性	69
	回归分析与决策	74
	回归五式	77
第五章	因变量与业务诉求	87
	确定业务的核心诉求	87
	纵横对比两步法	93
	不存在完美定义的Y	96

	瞄准真实的业务目标	100
	瞄准原因而不是结果	103
第六章	解释性变量与业务洞见	107
	X 就是业务洞见	107
	X 决定竞争优势	112
	运营实践产生 X	117
	制度设计保障 X	120
第七章	预测精度与产品设计	126
	预测不准是常态	126
	预测精度的改进	130
	预测精度的界定	134
	大数定律与产品化	139
第八章	数据资产定价	146
	必要性与挑战	147
	数据确权	153
	交易标的	156
	价值测算	160
第九章	数据资产交易	166
	交易的内容与形式	166
	可替代性与套利交易	171
	一个假想的资产交易所	176

第十章	数据治理与价值创造	184
	数据治理时代	184
	公司治理与数据治理	186
	数据资源资产化	190
	数据确权与合规	193
	价值创造与人才培养	198
	讨论与总结	201

不确定性的无知与无奈

人们都说，这是大数据时代，是数据科学的时代。每个人都在讨论数据，每个人都在研究数据分析。但是，数据分析的目标是什么？其核心内容是什么？不同的学校，不同的学者，不同的专家，可能会有不同的见解。这是一个正常现象。在我看来，在商业环境中，数据分析就一个目的：践行数据商业价值。更具体地说，是在带有不确定性的业务场景下，践行数据商业价值。

在实践中，践行数据商业价值，需要一些必要条件。例如，良好的计算机技能，扎实的经济管理基础，以及丰富完备的分析方法，等等。这些具体的技能与方法，在各个大学的（大）数据科学学院、经济管理学院、计算机学院、数学学院与统计学院等都有标准的课程开设，因此不是本书的讨论重点。本书

要讨论的重点是一套系统地践行数据商业价值的思维方法论。在该方法论的帮助下，你将具备一种驾驭全局的能力。在该能力的帮助下，你掌握的所有相关知识可以融会贯通，进而帮助你洞察业务不确定性，践行数据商业价值，实现数据资源资产化。这是本书的主旨所在。

从不确定性开始

践行数据商业价值应该从哪里开始？应该从认识不确定性开始。

要践行数据商业价值，需要一种独特的能力。该能力的培养，不应该从经济学、统计学、编程等开始，也不应该从领域（业务）知识开始，甚至不应该从分析数据开始，而应该从认识不确定性开始，因为这才是数据商业价值的核心所在。只有建立了对不确定性饱含敬畏的深刻认识，才可能在未来学习的漫长途中不会迷失方向，并获得无限的勇气坚持下去。

不确定性的定义

请注意，不同学者对于不确定性的定义不尽相同，而我的定义可能独树一帜。对此，需要特别声明，我无心追求一个标新立异的定义，这样的做法只能平添烦恼，没有任何意义。但

是，为了清晰地表达我的学术思想，我确实需要一个名词称谓方便讨论。很抱歉，我没有找到更好的称谓，因此就用了这个名词：不确定性（uncertainty）。关于这个名词，在我的理论框架中的确切定义后面会详细讨论。如果该定义跟任何现有领域或者学者关于不确定性的定义大相径庭，请大家包容谅解，因此不确定性非彼不确定性。

不确定性的重要意义

深刻理解不确定性，是践行数据商业价值的必要前提。为什么？因为不确定性对于数据产业实践有着重大意义。俗话说：乱世出英雄。在数据科学的产业实践中，“乱世”就是不确定性，而“英雄”就是数据商业价值。有了不确定性这个“乱世”，才有了成就数据商业价值这个“英雄”的机会。这个观点可能有点异类。因为对于很多从业者而言，极小化，甚至消灭不确定性，是他们努力的方向。而实际上，大千世界，正是因为有了不确定性，才更加精彩；整个数据科学，正是因为有了不确定性，才有了存在的意义。

不确定性产生的原因

不确定性的产生有两大根本原因：一是无知；二是无奈。无知所对应的不确定性，可以通过数据的无限积累、算法的不

停改进、知识的持续增加逐步化解。无奈所对应的不确定性，其根源是人类对稀缺资源的无限博弈，我们对它无可奈何。因此，这种不确定性永远不可能被绝对消除，它将长期广泛地存在，甚至不会衰减。理解这个道理对于数据分析意义重大，从此你将不再对模型预测精度抱有不切实际的幻想。应该认识到数据分析中的可为与不可为：哪些分析结果是可预期的，是值得努力追求的；哪些分析结果是“镜中花，水中月”，是不可预期的、不值得无限投入的。

什么是不确定性？

什么是不确定性？如果一个特定的个体对一个特定的事件无法绝对准确地预测其结果，那么这个事件对于该个体而言是一个不确定性事件（uncertain event）。相反，如果一个特定的个体对一个特定的事件能够做到绝对准确地预测其结果，那么这个事件对于该个体而言是一个确定性事件（certain event）。

根据以上定义，不确定性（或者确定性）事件是有明确指向性的。同一事件，有可能对于个体 A 是不确定的，而对于个体 B 却是确定的。由此可见，以上所定义的不确定性事件，不同于概率论统计学上所讲的随机事件（random event）。随机事件对任何人来说都是不确定的。

打个比方，老王、老李做一个游戏“猜花生米”。老王出一只拳头，让老李猜里面是空的还是抓了一粒花生米。这个事件对老王而言是确定性事件，因为老王自己有没有抓花生米，他心知肚明。这个事件对猜拳的老李来说就是不确定性事件，因为老李无法对老王的猜拳决策做出绝对准确的预测。

再看一个例子：老王掷硬币，正面朝上还是背面朝上？这个事件对老王来说是不确定性事件。但是，假设老李有一个超级厉害的设备，能够在老王出手的一刹那准确捕捉到硬币的角度、速度、角速度、加速度、周围的空气动力环境等诸多重要的物理学指标，那么从理论上讲，老李就可以根据物理学原理，准确预测硬币是正面朝上还是背面朝上。所以，同一事件，对老李来说就成了确定性事件。

不确定性事件广泛存在

在实际生活中，不确定性事件广泛存在。事实上，人生就是由一系列或大或小的不确定性事件构成的。

婚姻选择有着极大的不确定性。一对相爱的男女是否应该接受对方成为自己的终身伴侣，这是一个重大选择。这个选择的后果是什么？是幸福的远航，还是痛苦的开始？这有很大的不确定性。为了极小化这种不确定性，青年男女往往需要一场