

大数据、大见解 — 数字世界里，知识还是力量吗？

Thomas Gehling 与 Bernhard Wolters, Gen Re, 科隆

“现在，我想要的就是事实”，托马斯·葛雷更在查尔斯·狄更斯的小说《艰难时世》开场白中如是说。如果用“数据”来替换“事实”一词，大数据迷们一定会欣然认同。一些人甚至断言，人们不再需要理论，因为数据模式已经告诉我们一切。这种说法是否缺乏根据？毕竟，数百年来科学家们一直在利用数据进行研究。学术遁辞日益被视为反经验主义，社会科学和自然科学随之应运而生，而经验数据则是其中的关键。

在数据资源极大丰富、垂手可得的世界上，数据被认为可以为人们的行动提供充分的指引，本文探讨在这一背景下，所谓理论已丧失实际用途的说法是否成立。在后续的《保险观察》期刊文章中，我们将讨论大数据对寿险行业可能产生的影响，尤其着眼于人类专家知识在未来的作用。

理论的终结

最后一个宣称重大事物终结的人是美国政治学者弗朗西斯·福山。1989年，他声称历史即将在最后阶段达到顶点，彼时整个世界将被自由民主所主宰。¹ 呜呼，历史根本没有停滞不前，事实已经证明他的错误。

但是数据聚合商（如谷歌）是否就代表着理论的掘墓人？是否到了该问“科学从谷歌那里能学到什么”的时候？² 或者如同福山对上世纪九十年代初期的划时代政治剧变感到震撼一样，大数据迷们是否也会屈服于大数据无可争辩的动力？

大数据将令人们脱离任何理论，通过经验之门即可探知世界，这一观点近年来被反复谈论。³ 事实上，这只是科学与哲学历史上一个由来已久争论的新近表现。著名的理性主义哲学家（如笛卡尔和莱布尼兹）断言，某些知识只能来源于理性与逻辑。相反，培根、洛克和休姆的经验主义学派认为，经验形成知识的终极基础。培根甚至宣称，经验事实本身就足以说明一切。⁴

二十世纪前半叶，知识的基础再次成为一个有争议的话题，令“证伪主义”追随者与“逻辑实证主义”主张者互相对立。这一次是科学知识面临考验，或者更准确地说，是知识观点称得上科学知识所必须遵从的标准。

目录

理论的终结	1
我们是否需要知识？	2
拥有知识的益处	3
结语	3

本刊简介

《保险观察》是Gen Re面向全球寿险及健康险管理人员推出的一份专业季刊，每年出版四期。期刊文章聚焦于精算、核保、理赔、医学以及风险管理方面的问题。重点讨论的产品包括寿险、健康险、失能收入险、长期护理险和重大疾病险。

按照哲学家卡尔·波普尔所言，基于经验观察，科学知识观点（理论）原则上必须是可驳倒的（可证伪的）。⁵ 例如，“明天会下雨”的理论是科学的，因为如果明天不下雨，这个理论是可驳倒的。另一方面，“明天会或者不会下雨”的理论是不科学的，因为没有可想得到的观察论断能够反驳这个理论。

逻辑实证主义者认为，科学知识的本质植根于基础经验观察论断中，归纳法是其标志元素。在这一基础上，只有已经下雨多天，才能说明明天会下雨的论断（理论）是科学的。否则，这个论断只能被视为不科学的臆测。

我们是否需要知识？

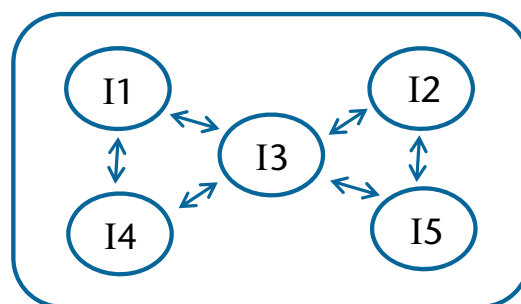
大数据倡导者声称，在数字世界里成功运作不再需要知识。用他们的话来说，“相关性[一系列模式]取代因果关系[阅读：知识]，即使没有不同的模型、统一的理论或者任何机械论的说明，科学也能够进步。”⁶

这一观点反映经验主义学派对于知识的永恒确定性固有的怀疑论。图1描述相互关联的数条信息（I）构成的网络，其中每一条信息都在说明其他信息（蓝色箭头）。这个网络构成知识。如果没有将一条信息与其他信息关联起来的解释和说明，我们所拥有的只是一堆信息碎片，“信息碎片无助于人们理解他们想要处理的现实问题。”⁷

信息可被描述为数据（D）与意义（M）之和，即 $I = D + M$ 。⁸ 大数据倡导者宣称，知识可以被完全免除，因为“意义”可以从这个等式中去掉。按照特定规则（语法）组合且格式良好的数据已经足够（如译文或广告），不再需要有关语言、文化和习俗的任何信息。

要评估知识是否真的不再必要，我们必须思考特定知识与背景知识之间的试探性区别。¹⁰ 特定知识包含可呈现统计相关性的说明，根据我们的背景知识，这些统计相关性看似合理。背景知识除了支持特定知识以外，还能使我们以目标导向型方式收集数据，并筛选有价值的数据模式。¹¹ 此外，背景知识有助于我们利用已经发现以及可能已经说明的相关性。前者使我们的预测更可信，而后者使我们的预测更有用，不仅仅是在商业意义上。

图1 — 以相互关联的信息网络形式呈现的知识⁹



心理历史的例子可以用来说明特定知识的用处。二十世纪前半叶，名为“行为主义”的理论研究法逐渐显现出影响力（文本框1）。

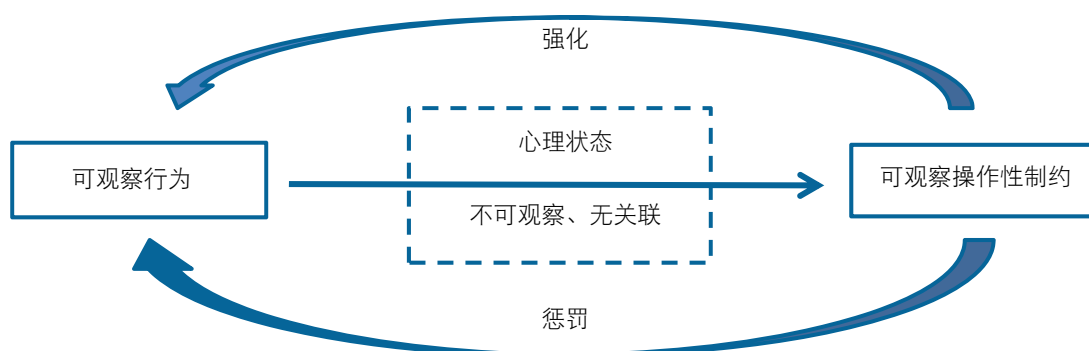
行为主义拒绝说明可观察操作性制约与可观察行为之间的相关性，并预期大数据对于知识的排斥。行为主义的正确性随后受到另一种名为认知心理学的理论方法的挑战。

认知心理学认为，心理状态和头脑运作的方式非常重要。在一项实验中，首先要求参与者们表达对某些话题的真实意见，随后他们需要以小论文的方式反驳这些意见，以此获得不同数额的奖金。¹⁴ 在实验的最后阶段，参与者需要完成问卷，他们被告知问卷为匿名方式，从而可以再次收集他们的真实意见。

文本框1 — 行为主义

佛洛伊德心理动力学方法以（不可观察的）潜意识心理过程为核心，行为主义与此相异，行为主义支持者将其视为“自然科学纯粹客观的一个分支。”¹² 行为主义的理论目标为“对行为的预测与控制。”¹³ 行为科学理论由 B.F.斯金纳充分发展起来，他补充了早期对经典条件反射的关注，即刺激通过操作性条件反射引起反应，即偶然行为受到操作性制约的强化或惩罚，其结果是，这一行为再次发生的可能性增加或减少。按照行为理论，这是学习的本质。斯金纳没有否认与操作性条件反射共存的心理状态，但他认为两者并无关联（见图2）。

图2 – 行为主义



行为主义者认为, 如果受试者反驳其最初意见的小论文可获奖金额越高(强化), 更高比例的受试者在最后问卷中的意见改变就越大, 即他们学到了新的意见。实际上情况恰恰相反: 奖金数额越小, 实验诱导出的意见改变就越大。¹⁵

这一结果被认知心理学准确地预测到, 或者更确切地说, 是认知失调理论(文本框2)。得到更少奖金的学生受其冲突行为的影响更大, 因此, 他们更想要改变自己的意见, 从而使之与先前的行为保持一致。图3显示, 为了说明(可观察)相关性, 理论可能是有意义的。

拥有知识的益处

背景知识的益处下面的例子中显而易见。¹⁶ 在一家美国零售公司Target工作的数据分析员发现, 25种产品的购买模式使他们能够给出怀孕预测分数以及预产期估计。在咨询过神经学和心理学专家对于习惯形成的意见后(尤其是循环提示、常规和奖励), Target了解到, 在重大人生事件发生后(例如怀孕), 人们易于接受改变习惯的提示。

Target推出婴儿相关产品的定制广告, 吸引孕妇和她们的亲属光顾商店。在店里, 店员会说服她们购买其他产品, 试图改变她们的购物习惯。但是当潜在的公共关系危机隐约浮现时, 因为其掌握的文化和心理背景知识, Target能够及时地发现这一危机。

这家公司意识到, 孕妇可能会对怀孕相关产品的定制广告产生警觉。这些女性可能合理地产生疑惑, Target是如何了解到她们的怀孕信息的, 或者她们是否被暗中监视。因此Target开始在这些广告中混入孕妇可能不太会购买的产品, 如割草机或酒杯。

结语

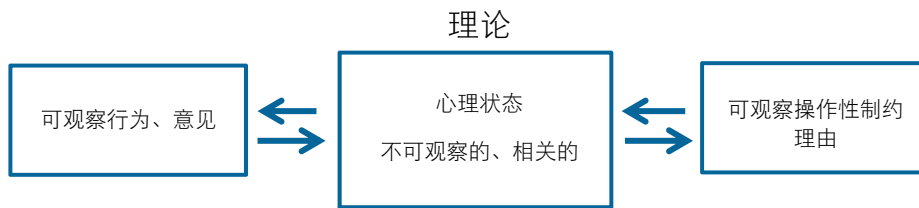
背景知识在大数据主宰的数字世界里仍然具有价值, 看来这一点不容置疑。我们仍然需要提出恰当的问题, 没有背景知识这是不可能做到的。大数据迷们大概也会认同这点。不如说他们的怀疑指向的是以理论形式表现的特定知识, 这些知识说明统计观察, 尤其是相关性。依据他们的观点, 这些知识在过去为了提升我们对于预测的信心是必要的, 当时的预测以有限的观察数量(数据)为基础。理论仅仅是一种辅助手段, 不能得到更高的地位。随着大数据的出现, 采样、归纳推理和假设正在被针对所有人群的实时观察和追踪所替代。¹⁷ 从有限数量的观察到研究针对的目标人群, 理论看似不再是支持这一跳跃的必要元素。在商业应用上, 这似乎是一个有效的观点, 尤其是考虑到几乎为零成本的纠错速度。虽然如此, 仍然有不曾预料到的误区, 正如前述行为心理学实验所展示的。

我们简要地看一下“大数据”一词的可能定义, 这些定义可以反映出我们前面讨论的大多数争议点。其中一个解释是, 大数据指的是数据量大

文本框2 – 认知失调

根据认知失调理论, 两种矛盾意见或者意见与行为之间矛盾引起的心理冲突会诱使受试者要么改变意见, 要么调整行为, 从而恢复心境平和。这一过程也取决于支持新的矛盾意见或行为的理由有多充分(在上述实验中, 理由就是奖金数额)。理由不充分(如奖金较少)意味着更多的心理干扰(认知失调), 因此更有可能改变行为或意见。

图3 – 认知心理学



大, 无法用传统数据库技术管理的数据集。¹⁸ 大数据应用程序无需采样, 只是观察和追踪发生的事件, 就这个意义而言, 数据体量起到额外的作用。¹⁹ 另一种观点认为, 大数据不只关乎计算能力的缺乏, 而且也关乎对可提供附加价值的新模式和相关性的探查。²⁰ 在数据体量与类型之外, 速度是大数据的另一个重要特征, 因为计算通常能够实时完成。²¹

在后续的《保险观察》期刊文章中, 我们将重点关注大数据真正的变革性新特征, 我们还将探讨加速进行的社会变革产生的额外推动力。我们的关注点将转向围绕大数据使用而出现的实际(道德)问题。最后, 我们将回顾大数据已经产生的影响, 并展望大数据将怎样进一步塑造寿险公司应对重大挑战的方式, 即客户需求、变化风险、逆选择和道德风险。

关于作者

Thomas Gehling, 资深精算师, 专门从事德国市场失能险和寿险产品的开发工作。他是德国精算师学会(DAV)会员, 您可以通过以下方式与他联系:
电话: +49 221 9738 746
或电邮: thomas.gehling@genre.com。



Bernhard Wolters 是 Gen Re 寿险及健康险部门副总裁, 多年来一直负责荷兰和斯堪的纳维亚市场业务。您可以通过以下方式与他联系:
电话: +49 221 9738 848
或电邮: wolters@genre.com。

尾注

- 1 Francis Fukuyama, *The End of History?*, The National Interest, 1989, p. 1.
- 2 Chris Anderson, *The End Of Theory: The Data Deluge Makes The Scientific Method Obsolete*, <http://www.wired.com/2008/06/pb-theory/>, accessed on 21 June 2016.
- 3 See for instance Luciano Floridi, *The Fourth Revolution*, 2014, pp. 129-131.
- 4 Bertrand Russell, *History of Western Philosophy*, 2000 (first published in 1946), p. 529. "He hoped that mere orderly arrangement of data would make the right hypothesis obvious, but this is seldom the case."
- 5 Karl Popper, *The Logic of Scientific Discovery*, 2002 (first English edition in 1959; the German original "Die Logik der Forschung" was published in 1935).
- 6 Chris Anderson, *Ibid* at Note 2. The philosopher David Hume already rejected the notion of causation in the 18th century stating that: "Objects have no discoverable connexion together; nor is it from any other principle but custom operating upon the imagination, that we can draw any inference from the appearance of one to the existence of another." David Hume, *A Treatise Of Human Nature* (originally published in 1739), 2003 ed., p. 73.
- 7 Luciano Floridi, *Information: A Very Short Introduction*, 2010, p. 51.
- 8 Luciano Floridi, *Ibid* at note 7, p. 20.
- 9 Source of graphs the authors unless otherwise stated.
- 10 This distinction does not claim any higher dignity. It is simply made for the purpose of our argumentation.
- 11 Luciano Floridi, *Ibid* at note 3, p. 130.
- 12 John B. Watson, *Psychology as the Behaviourist Views it*, *Psychological Review* 1913, <http://psychclassics.yorku.ca/Watson/views.htm>, accessed on 26 June 2016.
- 13 *Ibid*.
- 14 Martin Carrier, Jürgen Mittelstrass, *Mind, Brain, Behaviour. The Mind-Body Problem and the Philosophy of Psychology*, 1991, p. 136.
- 15 Note that the subjects' behaviour (responses to a before and after questionnaire recording their true opinions) was also captured by their essays which, as they were directed, contradicted their original opinions. The responses in all three stages could not be changed.
- 16 Charles Duhigg, *How Companies Learn Your Secrets*, *The New York Times Magazine*, 16 February 2012.
- 17 "Hypothesis" and "theory" are used interchangeably.
- 18 Louis Rossouw, *Big Data – Big Opportunities*, *Gen Re's The Bulletin*, May 2013.
- 19 https://en.wikipedia.org/wiki/Big_data, accessed on 13 July 2016.
- 20 Luciano Floridi, *Ibid* at note 3, p. 16.
- 21 https://en.wikipedia.org/wiki/Big_data, accessed on 13 July 2016.

The difference is...the quality of the promise.

genre.com | genre.com/perspective | Twitter: @Gen_Re

**General Reinsurance AG
Shanghai Branch**
18F China Merchants Tower
161 East Lujiazui Road, Shanghai 200120
Tel. +86 21 6100 6300
Fax +86 21 6100 6366

**General Reinsurance AG
Beijing Representative Office**
Oriental Plaza, East Cahnng An Avenue
Dong cheng District, Beijing 100738
Tel. +86 10 8518 530
Fax +86 10 6517 1267

Editors:
Ulrich Pasdika, ulrich.pasdika@genre.com
Ross Campbell, ross_campbell@genre.com

Photos: © Thinkstock – FactoryTh,
NicoElNino, Vimvertigo

© General Reinsurance AG 2017

This information was compiled by Gen Re and is intended to provide background information to our clients, as well as to our professional staff. The information is time sensitive and may need to be revised and updated periodically. It is not intended to be legal or medical advice. You should consult with your own appropriate professional advisors before relying on it.