

被算法

From Facebook and Google to Fake News and Filter-bubbles
The Algorithms That Control Our Lives

深入谷歌、脸书、推特、亚马逊的算法黑箱
揭示科技巨头塑造的智能时代隐形秩序

操控的

重新定义精准广告、大数据和AI

[瑞典] 大卫·萨普特 著

David Sumpter

译

生活

湖南科学技术出版社

OUTNUMBERED

被算法

From Facebook and Google to Fake News and Filter-bubbles
The Algorithms That Control Our Lives

操控的

重新定义精准广告、大数据和AI

生活

[瑞典] 大卫·萨普特 著
David Sumpter
易文波 译

CSK 湖南科学技术出版社

图书在版编目 (CIP) 数据

被算法操控的生活：重新定义精准广告、大数据和 AI / (瑞) 大卫·萨普特

(David Sumpter) 著；易文波译. — 长沙：湖南科学技术出版社，2020.1

ISBN 978-7-5710-0301-2

I. ①被… II. ①大… ②易… III. ①数据处理②人工

智能 IV. ① TP274 ② TP18

中国版本图书馆 CIP 数据核字 (2019) 第 195197 号

Outnumbered: From Facebook and Google to Fake News and Filter bubbles – The Algorithms That Control Our Lives

Copyright © 2018 by David Sumpter

Simplified Chinese edition copyright © 2020 by **Grand China Publishing House**

Published by arranged with Bloomsbury Publishing PLC through Big Apple Agency, Inc Lauban, Malaysia.

本书中文简体字版通过 **Grand China Publishing House (中资出版社)** 授权湖南科学技术出版社在中国大陆地区出版并独家发行。未经出版者书面许可，本书的任何部分不得以任何方式抄袭、节录或翻印。

BEI SUANFA CAOKONG DE SHENGHUO CHONGXIN DINGYI JINGZHUN GUANGGAO DASHUJU HE AI

被算法操控的生活——重新定义精准广告、大数据和 AI

著 者：[瑞典] 大卫·萨普特 (David Sumpter)

译 者：易文波

策 划：中资海派

执行策划：黄 河 桂 林

责任编辑：汤伟武

特约编辑：林 晖 陈 彬

出版发行：湖南科学技术出版社

社 址：长沙市湘雅路 276 号

<http://hnkjcs.tmall.com>

湖南科学技术出版社天猫旗舰店网址：

<http://hnkjcs.tmall.com>

印 刷：深圳市雅佳图印刷有限公司

(印装质量问题请直接与本厂联系)

厂 址：深圳市龙岗区布吉坂田大发埔村大发路 29 号 C 栋一楼

邮 编：518116

版 次：2020 年 1 月第 1 版

印 次：2020 年 1 月第 1 次印刷

开 本：710mm×1000mm 1/16

印 张：16

字 数：212 千字

书 号：ISBN 978-7-5710-0301-2

定 价：59.80 元

(版权所有·翻印必究)



GRAND CHINA

人类思想内容提供商

唯一类似人类的智能形式 只会是人类自己

OUTNUMBERED

打破 AI 崇拜的诚意之作 颠覆大众对人工智能的流行认知

这是一个“算法世界”：建立在数据之上的算法指导社会的运行、决定我们能在网上看到什么；它更是自动驾驶、智能管家、未来医疗以至智慧城市的基石。如果我们不了解算法如何使用数据，就无法知道它和依赖它运行的人工智能将如何改变我们的生活。

在本书中，大卫·萨普特将引领读者展开一段有关算法的趣味旅程，了解数学阴暗的一面。通过采访那些在算法研究领域工作的科学家，以及动手进行有趣的数学实验，他介绍了那些能够分析我们、影响我们，甚至会变得更像我们的算法。你将由此明白谷歌搜索是否存在种族歧视和性别歧视，为什么选举预测有时会与真实的选举结果间存在巨大的偏差，使用算法来找出罪犯是不是一个严重的错误，以及我们把决策权交给机器时，未来又会发生什么。

萨普特还将深入浅出地告诉我们，到底是什么样的算法在支撑着最能聊天的AI——水谷、围棋中的无冕之王阿尔法狗——零和传说中的写作机器人。他所道出的真相将刷新我们的观念：直到今天，AI 的智能都仅与大肠杆菌相当，和某些科技巨头的激进预测相反，取代人类的超级 AI 其实离我们非常遥远。

计算机语言和开发平台日新月异，但万变不离其宗的是那些算法和理论……如果你把计算机的发展放到应用和数据飞速增长的环境下，你一定会发现：算法的重要性不是在日益减少，而是在日益加强。

——李开复

OUTNUMBERED

权威推荐

《科克斯书评》(*Kirkus Reviews*)

《被算法操控的生活》独具特色……对互联网数据操控的分析颇具洞察力和趣味性，对现在的流行观点抱有怀疑态度。

《出版商周刊》(*Publishers Weekly*)

一本给人启迪的书。

《金融时报》(*Financial Times*)

萨普特的专业分析和质疑观点因本书中的趣味案例而不再枯燥乏味，这些案例包括寻找街头艺术家班克西，以及训练神经网络玩《太空入侵者》游戏。

《科技新时代》(*Popular Science*)

《被算法操控的生活》传递了非常重要的讯息。

《卫报》(*The Guardian*)

萨普特对操纵我们数字生活的庞大系统进行了深入的思考，抽丝剥茧地解开了这些系统背后的数学秘密，清晰地向我们展示

了这些看起来十分高级的运算，这些运算控制着系统输出的结果，同时也制约着这些系统本身。

约翰·默多克 (John Burn-Murdoch)

英国《金融时报》数据新闻记者

当数以百万计的人慢慢意识到自己落入了陷阱，正在将数字生活的信息交给科技巨头时，萨普特通过采访算法研究领域的内部人士，结合引人入胜的亲自演示，向读者揭示了数据“炼金术士”们试图说服和预测我们的方法，并为我们验证算法是否真的像他们吹嘘的那样无所不能。

蒂曼德拉·哈克尼斯 (Timandra Harkness)

《大数据》(*Big Data*) 作者

你一定听说过有些算法能够操控你的生活，所以你很可能想要知道两件事：算法是如何操控你的生活的？对于这种操控，我们到底应不应该担心？萨普特以坦诚的态度和大多数人未曾接触的深度知识回答了这两个问题。

凯特·耶茨 (Kit Yates)

英国巴斯大学数学科学系高级讲师

《被算法操控的生活》是一本关于数学在现实世界中如何应用的一流作品，写得十分引人入胜。每一章都讲述了一个回味无穷的故事，而萨普特热情和诙谐的风格也表明，数学家不仅仅是一台把咖啡变成定理的机器。

OUTNUMBERED

推荐序 I

窥视我们隐私的互联网， 背后有何真相？

《科克斯书评》

关于数据如何入侵我们的生活这一话题，本书中的深入研究让人由衷信服。

专家们总是不厌其烦地警告我们：如今的数字技术能够窥探到我们内心最深处的秘密。本书的内容也与数字技术相关，而且写得极为精彩，但与其他同类书籍相比略显不同的是，作者萨普特（瑞典乌普萨拉大学应用数学教授，著有《足球数学》）认为上述看法虽然不无道理，但也存在着严重的缺陷。本书并非一场唇枪舌剑的论战，它更多的是在进行调查性研究，通过事实推导出结论，对人工智能中常用的数学方法和工具进行深入浅出地讲解。它向我们展示了社交媒体、搜索引擎和商业公司如何收集我们在网上留下的痕迹（这些痕迹往往能透露出我们的品味、个性和政治立场），并通过一系列的规则（也就是算法）对其加以利用。

通过为自己量身定制的广告完成购买行为，或许让一些读者感到舒适便利，但在阅读本书之后，他们可能就会对这类广告退

避三舍了。因为他们将意识到，自己每一次点击鼠标，包括上传一张照片、转发一则笑话或在短片下面点一个“赞”，都是在为建立与自己相关的庞大数据库添砖加瓦，而其揭示出的人格特质真实得让人恐惧。“我们的朋友往往会忘记这些细节，并且不太在意他们从我们身上得出的结论，”作者在书中如是说，“但脸书（Facebook）却不同，它正在系统地收集、处理和分析我们的情绪状态。它在数百个维度上分析我们的人格，因此它能找到最冷静、最理性的角度来观察我们。”

通过操纵数据来诱导我们购物似乎无关紧要，不会带来太多的负面影响，但是互联网中同时还充斥着经过精心炮制的、试图影响我们投票的假新闻。正因为如此，萨普特才再三提及特朗普出人意料的胜选和英国脱欧公投的震惊世人的结果。作者运用他的数学工具，向我们一步步揭示了目前算法对大数据的利用情况，并且告诉我们，算法并非如我们眼见的那么神奇。民意调查仅仅能够计算单次事件的概率，它们根本无法“预测”任何事。真正相信假新闻的人对其来者不拒、照单全收，但它们对广大读者思想的影响微乎其微。因此，在算法和人工智能面前，我们还不必谈虎色变。

美国读者们不应该过多地纠结本书语言中的英式风格，只有这样，才能更好地欣赏本书对互联网数据操纵所做的颇有见地又略带批判性的分析。

OUTNUMBERED

推荐序 II

科学技术的危机

《卫报》

一位数学家认为，算法无法操控我们，因为我们远比算法聪明。但数学没有道破我们所面临的危机。

偏见、排名和假新闻

道格拉斯·亚当斯（Douglas Adams）在《银河系漫游指南》一书中写道：“宇宙浩瀚无垠，广阔得让你难以置信、叹为观止。你可能会觉得去药店的路远得走不到头，但我告诉你，这段距离在宇宙中微不足道。”在我读萨普特的这本书时，亚当斯的话不断在我脑海中回响。

本书旨在对管理着我们数字生活的系统的全貌一探究竟。仅仅是几组简单的数字就足以让你晕头转向：脸书拥有 20 亿用户，他们每小时会写下上千万篇帖子，还有数以百万计的照片、点赞和互相关注，脸书利用这些数据在数百个的维度上对我们进行建模。在这种模型面前，我们人类的大脑显得微不足道，因为我们

的大脑最多只能具象化四个维度。

同样地，谷歌翻译（Google Translate）的系统能将几百种语言分解成多维词义矩阵，这些矩阵会创造出我们无法理解的、只属于它们自己的形而上语言，并将各种语言的内隐偏见包含其中。

这些系统的设计初衷是理解词语和概念之间的相互关系，然而我们在其中输入英国最常用的婴儿名字时，将会得到这样的答案：“奥利弗之于聪明正如奥莉薇娅之于轻浮。”作者由此表达了他的担忧：“我们后代的性别角色已经被算法提前设定好了。”

作为一位应用数学教授，萨普特对此类问题的本能反应是在数学模型中对它们进行重建，然后消除导致这些问题的因素。在这个过程中，我们对算法难以言喻的恐惧或许能够得到缓解。他使用了大量论文中的原始数据，抽丝剥茧地解开了这些系统背后的数学秘密，清晰地向我们展示了这些看起来十分高级的运算，这些运算控制着系统输出的结果，同时也制约着这些系统本身。

采用复杂的回归模型拟合脸书的数十亿个“赞”，这种做法本身无可指摘，但其结果却并不那么值得称道：没错，“民主党人士比共和党人士更有可能喜欢哈利·波特（Harry Potter）”，但“这并不一定意味着其他哈利·波特粉丝都是民主党人士”。

作者采用同样严肃的方法颠覆了那些我们懒于认真思考的常见论断：我们都在被假新闻愚弄、被困在不断强化偏见的过滤气泡中难以脱身等。我们显然还没有愚蠢到看不出问题的地步。

并非只有所谓的“社交媒体傻瓜”在承受这种压力。萨普特不断以其他系统作为例证，粉碎了当前社交媒体将悲哀都归咎于脸书等网络巨头的谬论。作为学术论文搜索引擎的谷歌学术（Google Scholar）也会根据每篇论文的被引用次数来对其学术影响力进行排名，致使人们对热门的研究领域趋之若鹜，科研经费也一哄而上投入其中，而那些真正可能改变世

界的研究却可能因为没有足够的资金支持而止步不前，甚至胎死腹中。在线预测市场中的那些所谓的专家也只不过是滥竽充数，其预测水平和普通人并无差别，得出的投票预测数据和胡猜没什么区别。

萨普特举了一个又一个例子，旨在告诉我们，不管是在精英群体中还是在更易接近但自我意识稍弱的普通民众中，过滤气泡、肤浅的排序机制和算法推荐的毒害都有迹可循。然而，它们的危害程度并不像媒体向我们灌输的那么严重：“只要有时间，我们科学家仍然可以把研究做好。我遇到的大多数科学家都是出于对真理的永恒追求和对正确答案的真切渴望而献身科学的。”

复杂系统自有它的局限性

本书的面世适逢其时，书中的调查研究让作者早在公众对剑桥分析公司（Cambridge Analytica）的愤怒彻底爆发之前，就开始质疑这家企业的角色，并且想尽办法采访到了该公司的首席研究员亚历山大·科岗（Alexander Kogan）。作者最终认为剑桥分析公司在美国大选中所起的作用微不足道。他写道，我们之所以相信剑桥分析公司的方法行之有效，更多的是因为我们相信了该公司管理层的夸张宣传，而并不关心其背后的数学支持是否牢固可靠。

但是仅仅关注数学本身会使我们片面地看待这些新兴的复杂系统在日常生活中所起的作用。作者深度探究了美国和法国选举期间的假新闻背后的数学逻辑，以及机器人在英国脱欧公投中扮演的角色，却忽略了更值得我们担忧的问题，比如肯尼亚、斯里兰卡或缅甸等国家的网络仇恨言论所带来的影响。后者不是误用数学的结果，而是这一趋势的后果：由于技术决定论造成的文化盲区，人们只关注主流社会和发达国家，却对第三世界的情况视而不见。

脸书和谷歌的研发人员拥有美好的初衷，他们相信在一种情境下能有效运行的算法到了另外一种情境下同样会有效运行，但他们未能预见这些平台会被别有用心者的政治煽动者利用，或者被用来监控整个国家、审查每个网民。凡此种种，都让人不得不怀疑：也许我们遭遇的算法危机根本就不只是一个数学问题。

当然数学家们也不难意识到这一点，而且萨普特还煞费苦心地强调了自己作为数学家的优势和局限性。他的上一本书《足球数学》(*Soccermaths*)带领我们在美妙的足球运动所蕴藏的数学逻辑中进行了一次引人入胜的有趣探索。但我们都知道，足球毕竟只是一项体育运动，尽管充满了喧嚣与愤怒、机会与混乱，但它终究还是由两个半球缝合而成的，它的目标只有一个，那就是进球。量化球场上发生的事情虽然有些难度，但很可行。可球场外所发生的事件无法被量化，而相信可以量化的人也遭遇了数不胜数的意外后果。

尽管萨普特一直坚持认为，在当今的算法系统中，人类的监管必不可少，但他的分析似乎无法解释捣蛋鬼的行为，以及那些无视所有证据，继续生活在阴谋论、偏见和权力欲中的人的行为。正如一位研究人员指出的那样，对于语言翻译系统中的偏见，“如果不首先解决现实社会中的种族主义和性别歧视，就没有办法真正解决由无监督学习造成的问题”。

在本书的开头，有一段简短的文字提到了谷歌开发的一款能在 YouTube 上生成定向广告的软件。这款软件可以让品牌根据用户的个人资料在视频中插入文字和图片，将其广告生成为无数种不同版本，进而增强品牌的影响力。这款软件的名字叫沃冈 (Vogon)——毫无疑问这是道格拉斯·亚当斯创造的，他用这个名字来称呼那些要毁灭地球的像鼻涕虫一样的外星人：“沃冈人是银河系中最不令人愉快的种族之一，但事实上他们并非邪恶的种族，而是脾气暴躁、官僚作风严重、爱管闲事并且冷酷无情的种族。”

理解这些复杂系统背后的数学原理会让你受益良多，而作者真诚的叙述同样揭示了它们的局限性。但是，再多的数学知识也无法帮助我们重新获得日渐失去的生活主动权，也无法击败那些故意利用数学的复杂性来对付我们的人。

被算法

操控的

生活

OUTNUMBERED

目录

第一部分

算法在监视我们

互联网对个人数据做了什么？

算法不停地旋转和降维你的数据集，直到它能读懂你、透视你。它使脸书能用点“赞”情况预测你的性格，从表情、照片甚至你与屏幕的互动来评价你的精神状态。算法的高维理解完胜你对自己的了解，但它们并不具备完美的预测能力和公平公正的态度。

- 3 第 1 章 | 数学不仅大煞风景，还会败坏道德？
 - 3 谷歌人：统计数据越多越好
 - 7 “伯明翰 ISIS”：数学杀伤性武器引发的血案
- 11 第 2 章 | 算法无处不在，我们却对它一无所知
 - 11 黑箱操控——互联网巨头不想说的事
 - 13 给重定向广告加点“噪声”
 - 15 算法也会歧视人
- 18 第 3 章 | 你以为自己了解朋友？脸书比你更懂他们
 - 19 社交网络的黑箱世界
 - 22 主成分分析法：15 个帖子就能让他人读懂我们的生活

27 第4章 | 脸书比你更了解你的
人格和行为

- 28 你点的“赞”在暴露你的智商
- 30 你的情绪已被脸书编号

35 第5章 | 剑桥分析公司
如何玩转总统选举

- 36 人格决定选票
- 39 为政治操控每一种情感
- 42 算法告诉共和党，该去星巴克拉选票
- 46 脸书很了解你，但它未必能预测你的行为

51 第6章 | 要算法毫无偏见？
不可能！

- 52 定义偏见是道数学难题
- 56 相同的误报率和相同的漏报率不可兼得
- 60 只有悖论，没有公平

63 第7章 | 数据炼金术士

- 64 算法不知道的事情
- 68 精准广告：互联网公司滥用了私人信息吗？
- 70 人肉预测与模型预测，谁更胜一筹？