



CROWDSOURCED HEALTH

How What You Do on the

Internet Will
Improve Medicine

医疗大数据

大数据如何改变医疗

[德] 埃拉德·约姆-托夫
(Elad Yom-Tov) 著
潘苏悦 译

移动互联网、大数据、云计算等多领域技术与医疗领域的跨界融合
提升医院的运营效率、临床决策支持、
医疗质量监管的新思路，辅助科研、个性化治疗的新方式





CROWDSOURCED HEALTH

How What You Do on the

Internet Will
Improve Medicine

医疗大数据

大数据如何改变医疗

埃拉德·约姆-托夫
(Elad Yom-Tov) 著
潘苏悦 译



机械工业出版社
CHINA MACHINE PRESS

我们很多人都在网上查找过有关医疗健康的信息和数据，这些互联网数据可以辅助常规医学研究和诊疗手段，为医学界提供独特的视角与信息，改变医学研究模式。通过本书，作者讲述了针对互联网搜索的研究为科研人员提供的巨大帮助，通过互联网数据，医生可以了解患者在得知悲观的诊断结果后所作的反应，从而更好地诊治患者，并帮助患者更客观地认识自己的病情。故此，本书通过医疗大数据的特征和结构，阐述医疗大数据对未来医疗模式的改变和影响。

Elad Yom-Tov/The MIT Press/1 - 1/by Elad Yom-Tov/978-0262034500

Original English Language Edition published by Massachusetts Institute of Technology/

Copyright © 2016, All Rights Reserved.

This title is published in China by China Machine Press with license from Massachusetts Institute of Technology This edition is authorized for sale in China only, excluding Hong Kong SAR, Macao SAR and Taiwan. Unauthorized export of this edition is a violation of the Copyright Act. Violation of this Law is subject to Civil and Criminal Penalties.

本书由 Massachusetts Institute of Technology 授权机械工业出版社在中国境内（不包括香港、澳门特别行政区以及台湾地区）出版与发行。未经许可之出口，视为违反著作权法，将受法律之制裁。

北京市版权局著作权合同登记 图字：01 - 2016 - 0490 号

图书在版编目 (CIP) 数据

医疗大数据：大数据如何改变医疗 / (德) 约姆 - 托夫 (Yom-Tov, E.) 著；潘苏悦译。
—北京：机械工业出版社，2016.5
书名原文：Crowdsourced Health
ISBN 978 - 7 - 111 - 53663 - 5

I. ①医… II. ①约… ②潘… III. ①互联网络—影响—医疗保健事业—研究
IV. ①R19 - 39

中国版本图书馆 CIP 数据核字 (2016) 第 088429 号

机械工业出版社 (北京市百万庄大街 22 号 邮政编码 100037)

策划编辑：坚喜斌 责任编辑：杨冰 刘林澍

责任校对：赵蕊 版式设计：张文贵

责任印制：李洋

北京圣夫亚美印刷有限公司印刷

2016 年 5 月第 1 版 · 第 1 次印刷

145mm × 210mm · 6.5 印张 · 3 插页 · 92 千字

标准书号：ISBN 978 - 7 - 111 - 53663 - 5

定价：42.00 元

凡购本书，如有缺页、倒页、脱页，由本社发行部调换

电话服务

网络服务

服务咨询热线：(010)88361066 机工官网：www.cmpbook.com

读者购书热线：(010)68326294 机工官博：weibo.com/cmp1952

(010)88379203 教育服务网：www.cmpedu.com

封面无防伪标均为盗版

金书网：www.golden-book.com

谨以此书献给我的先母什洛米斯
(Shlomith) 和父亲约拉姆 (Yoram) , 是
他们教导我探求真知。也以此书献给嘉
莉特 (Galit) 和孩子们。

致 谢

我很荣幸能有机会和众多才华横溢的杰出科研工作者共事，其中一些人的名字将会在本书中出现，非常感谢他们为我答疑解惑。

同时还要感谢我先后任职的几家公司，它们都非常重视长期研究，这对我来说是一大幸事。耶夫根尼·加布里洛维奇 (Evgeniy Gabrilovich)、安德烈·布罗德 (Andrei Broder) 和珍妮弗·蔡斯 (Jennifer Chayes) 是这几家公司的研究负责人，他们给我提出了很多问题和建议，开阔了我的视野，加深了我对问题的理解，没有比他们更优秀的负责人和同事了。

最后还要感谢麻省理工学院出版社的各位编辑和为我斧正书稿的珍妮弗·科林斯 (Jennifer Collins)。

推荐序一

本文系应机械工业出版社之邀请为《Crowdsourced Health: How what you do on the internet will improve medicine》中译本撰写的推荐序。通读了译本全文，虽未能阅读英文原著，但也能感受到该书的新理念、新方法、新观点和新思路，可谓开卷有益。

作者埃拉德·约姆-托夫（Elad Yom-Tov）作为一个非医学专业人士，从个人的特别经历开始，延伸到以民众的健康与疾病诊断、治疗需求为导向，探索研究如何将互联网查询、日志等海量数据应用于改善医疗，如：网民搜索药物的药效、不良反应，偏头痛的症状，肥胖的标准，厌食症、癌症、性病患者需要的资源信息及其产生的数据，都能够提供独有的互联网数据信息，辅助传统常规医学研究、疾病诊疗，甚至可能用于流感、

“非典”（SARS）等传染病的预警和监测，为公共卫生、医学研究提供独特的视角、信息和研究模式，同时提出了医疗互联网面临的挑战（隐私保护等）。

移动互联网、大数据、云计算、认知计算、人工智能等技术在医学领域的应用一定会促进对疾病的发生、发展和转化规律的认识，推进医学模式的创新及医疗生态的重构。该书的理念、思维和观点将为应用互联网技术改善医学提供助力。

北京大学人民医院原院长、外科教授、
医学大数据研究中心主任
中国医师协会副会长、中国医院协会副会长
王 杉

推荐序二

我们正朝着精准医疗的时代大步迈进。

斯坦福大学的 Snyder 教授以自己作为实验对象，收集和分析了自己的基因组、转录组、表观组等组学数据，医学影像数据、可穿戴设备数据和生活环境数据，甚至社交数据，希望给自己建立一个全时全景的健康模型；美国、欧盟、中国等都宣布要投入巨额的研究经费来实现精准医疗；美国国家癌症研究所（NCI）代理所长 Douglas Lowy 和国立卫生研究院院长 Francis Collins 在新英格兰医学杂志（NEJM）上借美国总统奥巴马的“癌症登月计划”，宣告我们正处在“走向彻底治愈癌症之路”；美国、日本等国家将大量的科研经费投入到脑科学领域的研究，试图破解人类的智力极限。

希望在不久的将来，“AlphaGo”不仅会越来越多，也越来越强，超验骇客所描述的“硅元智能”很快就会

到来。所有的这些都将辅助我们正在进行的医疗大数据“拼图游戏”，最终将利用物理、化学、计算机、生物和医学等学科的知识和技术，解决人类的健康问题。未来，人类健康活到 120 岁将不是梦想！

埃拉德·约姆-托夫先生具有深厚的搜索引擎技术研究背景，他的遭遇和我认识的或听说过的一些朋友一样：计算机背景，家人或自身遇到严重的健康问题，在目前的医疗条件下感到无助和恐惧，都想依靠自身的技术背景，借助互联网大数据解决问题。

该书围绕“与现实生活紧密相关的互联网数据可以改变医学研究模式”展开讨论。无疑，互联网正以前所未有的速度改变我们的生活。书中包含大量的实例，如互联网大数据对新药物及其副作用的研究；互联网作为工具用来研究病患家庭的心理状态；互联网对人们的日常社交生活和自身隐私数据的保护；互联网对人类的罕见病研究的作用（PatientsLike Me 只是众多医疗健康社交网站中的一个）；互联网在流行病学的研究方面起着重要的作用；Google、Yahoo、百度等企业在这方面做了大量工作；以及互联网在健康数据的标准制定以及数据

众筹上非常出色的表现 [GA4GH (Global Alliance for Genomics and Health) 就是一个很好的例子]。

正如埃拉德先生所引用的狄更斯的名言，医疗大数据“既体现了这个时代的愚蠢，也体现了这个时代的智慧”。就像核能的开发利用一样，一切事物总有其两面性。合理利用医疗大数据，充分保护个人隐私，确保医疗大数据的标准和质量，对促进医疗大数据更为广泛的使用至关重要。同时，政府对数据的有效监管，避免因互联网医疗大数据的利用而产生的医疗事故和纠纷，也是医疗大数据应用的重要的环节。

一些题外话：每个人的医疗健康大数据都是个人的宝贵资产。不可能在没有人奉献个人医疗健康大数据的情况下，就指望有医疗健康大数据，或指望医疗健康大数据对整个医疗健康体系的促进。我们可以考虑利用“区块链”的概念，像比特币一样运营个人健康医疗大数据，这将是一个有趣的、值得探讨的问题。

深圳华大基因研究院副院长、信息中心主任

方 林

推荐序三

毫无疑问，这是一个数据的时代，一个半天产生的数据量就相当于人类社会自诞生到公元 2003 年所产生的数据量总和的时代！数据渗透到当今每一个行业和业务领域，成为重要的生产要素，改变着产业结构，甚至重新塑造着行业的边界，一个“大数据”的时代或已来临。

全球著名咨询公司麦肯锡称：“人们对于海量数据的挖掘和应用，预示着新一波生产率增长和消费者盈余浪潮的到来。”有经济学家甚至预测，统治经济增长理论长达六十年之久的索洛增长模型（Solow Growth Model）会被改写——除了资本、劳动力和全要素生产率之外，推动经济增长的要素还应该包括“大数据”。

作为中国医疗健康领域的探索者和实践者，泓华医

疗正在探索“大数据”在“以患者为中心”的医疗健康领域应用的广阔前景，并形成了自己的一些理解，应机械工业出版社邀请，在此做一个抛砖引玉的分享，希望能对各界同仁有所启示。泓华医疗以“让健康更简单”作为办医的核心理念和追求目标，其内涵就是希望在全国连锁医院、诊所和家庭护理服务的基础上，通过包括移动医疗、大数据、人工智能在内的新技术的应用构建线上线下医疗服务无缝对接平台，让医疗健康服务需求者获取医疗服务时更简单，让医疗服务提供者提供医疗服务时更便捷，并让利益相关各方能获得实实在在的良好场景体验，从而做真正专业、靠谱的医疗服务。

医疗健康服务的本质在于“通过使人类更健康来创造价值”，必须充分尊重患者、医生、护士、社会的利益和价值立场才能实现。这是办医者应该坚守的价值理念，也是办医者在生死面前应该秉持的敬畏之心。医疗无小事、人命关天，是医学人文精神的真实体现。遵循这样的理念，收集、整理、挖掘和利用患者在诊疗过程中产生的数据并以患者更健康、更方便等权益为出发点，采

集和分析包括患者基本数据、电子病历、诊疗数据、医学影像数据、检验数据、临床管理数据、医疗设备及仪器监测数据、诊疗保健过程的行为数据等信息，为人们更好地理解疾病产生的机理、寻找更好的治疗路径，并最终使人类更加健康的前景提供了广泛的想象空间。

同时，正如本书作者一再强调的，“大数据”的来源远远不局限于医院和诊所。过去，与健康相关的数据信息大部分来自医疗机构，只能反映人们健康状况的局部与片段，大部分仅为疾病症状描绘以及对相关因素的浅层次分析。移动互联时代，通过对人们行为模式的观察与数据挖掘，借助智能手机、智能手环、其他可穿戴设备等，我们能搜集人们的日常生活数据，研究视角从而从医学转向健康。这些健康数据反映了人们日常起居、生活习惯、生理变化、潜在病灶、病情发展、用药频次、康复状况等多个方面的状况，同样具有医学上的临床价值。《医疗大数据：大数据如何改变医疗》一书向读者展示了怎样将这些数据与常规的医学研究方法结合，更好地理解疾病的成因，寻找更为有效的治疗方案，为人

类健康服务。对于医疗大数据这样一个新兴的交叉学科，作者更是大声呼吁，强调跨界融合的开放性思维的重要性——只有医学专业研究者和计算机专业研究者联手合作，才能从互联网数据中挖出宝藏，从而为医学发展提供独特的创新型解决方案。这些论述、分析和具体案例的讨论会给医护人员、医疗机构、管理部门、公众健康服务部门乃至大健康领域的创业者们带来深刻的启示。

然而，技术可以习得，如何坚守医学人文精神却是一个现实而紧迫的难题。写下这篇推荐序之时，正是媒体、公众、医疗从业者和监管部门热烈争论、深刻反思“谁该对魏则西之死负责”之际。得知患者罹患“滑膜肉瘤”这种不治之症后，魏家人辗转不同医疗机构，最终找到通过竞价排名靠前的医疗机构，花费巨资接受境外已经被质疑的所谓“生物免疫疗法”。这种号称效率很高的治疗方法，不但没能挽救魏则西的生命，还让他在生命的最后时刻痛感某些医疗机构的欺骗和虚伪。同为“大数据”的应用，在这个过程中我们看到的却是办医者理念的扭曲、贪婪和自私：企业的社会责任被抛弃、

监管的混乱和医学人文精神的丧失。痛定思痛，追责当事责任人固然重要也很简单，但我们都应该从这次事件中吸取教训。大数据的应用同样是一把双刃剑，我们要从体制机制上寻找根本原因，通过这次事件引发的全民大讨论，从而为行业未来发展和医疗大数据的应用构建制度基础和理性的认识。

《医疗大数据：大数据如何改变医疗》一书，向我们展示了那些充满悲天悯人情怀的医学和计算机科学研究者如何用他们的智慧为人类健康服务，也引发了我们对新技术对我们医疗保健切身利益的影响的多角度思考。我热情推荐这本书，因为它同时也在提醒我们，在这个阶段，除了技术之外，我们更需要医学人文精神的回归！

泓华医疗集团创始人、董事长兼 CEO

曾人雄

推荐序四

互联网+时代，海量数据的挖掘应用成为重要命题。该书作者是多年从事搜索引擎相关工作的互联网人士，结合自己的家人对疾病信息的需求，探讨了在互联网上如何获取有用的医疗信息，从而帮助患者获取诊疗数据，帮助医学专家开展疾病研究。互联网医疗的应用是一个全新的模式，尚处于探索研究阶段，同时涉及医学、信息技术、工程技术等专业交叉学科，其研究并不是一蹴而就的，在应用上也有着方法、理论、技术、安全隐私甚至伦理问题。国内对“互联网大数据”概念的热炒和严格意义上的学科研究、尚缺乏更多专业人士和体系支撑的现状形成了鲜明的对比。当然我们也欣喜地看到越来越多的医学专家和互联网行业相关人士加入这一新兴产业，推动中国全面健康促进的工作中来。

从信息技术（IT）时代到数据技术时代（DT）的大趋势已成为科技界的共识。如何开展大数据研究工作？哪些工具和材料能够让我们挖掘出数据背后的价值，哪些医疗专业知识适合通过互联网来实现变革？该书不是一本纯技术或者纯理论创新类的书籍，而是为了吸引医学和计算机科学领域、医疗服务领域的跨界者认识到基于互联网研究常规医学难以触及的一种新方法、新思维，并通过自身工作经历，结合案例娓娓道来，可谓是互联网医疗行业盛宴的一道“前菜”。

现代医学从传统医学到社会医学模式的转变，使得人们对医学的研究关注点向病前预防和病后康复转移。越来越多的人通过互联网了解健康知识、与自身疾病相关的专业知识，或求医问药。互联网运营商和服务公司掌握了大量数据，这些数据如果能安全有效地应用，可以对公众健康状况产生积极影响。比如，对药物反应特别是副作用的长期追踪调查，就是常规医药研究不可企及的。很多药物在上市前小样本测试时没有问题，但很难保证上市后对大规模人群不存在严重副作用。基于