



智能时代

吴军 / 著

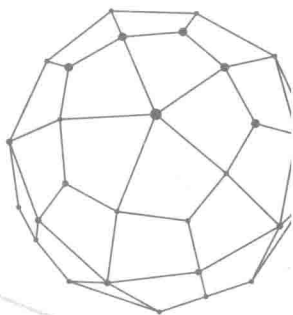
大数据与智能革命重新定义未来



智能时代

大数据与智能革命重新定义未来

吴军 / 著



中信出版集团 · CHINACITICPRESS · 北京

图书在版编目(CIP)数据

智能时代:大数据与智能革命重新定义未来/(美)
吴军著.--北京:中信出版社,2016.8(2016.8重印)
ISBN 978-7-5086-6381-4

I. ①智… II. ①吴… III. ①数据处理-社会影响-研究②人工智能-社会影响-研究 IV. ①TP274 ②TP18

中国版本图书馆CIP数据核字(2016)第144007号

智能时代:大数据与智能革命重新定义未来

著者:[美]吴军

策划推广:中信出版社(China CITIC Press)

出版发行:中信出版集团股份有限公司

(北京市朝阳区惠新东街甲4号富盛大厦2座 邮编 100029)

(CITIC Publishing Group)

承印者:北京通州皇家印刷厂

开本:880mm×1230mm 1/32 印张:12.25 字数:230千字

版次:2016年8月第1版 印次:2016年8月第2次印刷

广告经营许可证:京朝工商广字第8087号

书号:ISBN 978-7-5086-6381-4

定价:68.00元

版权所有·侵权必究

凡购本社图书,如有缺页、倒页、脱页,由销售部门负责退换。

服务热线:400-600-8099

投稿邮箱:author@citicpub.com

谨以此书献给我的家人

推荐序

智能时代，未来已来

最近几年，人类在一些科技前沿领域取得了重大的突破，这些领域包括：人工智能、基因技术、纳米技术等。过去一年，我们看到了许多存在于科幻小说中的内容成为现实：人工智能击败了人类顶尖棋手，自动驾驶汽车技术日趋成熟，生产线上大批量的机器人取代工人……甚至在我们有生之年，也许可以期待看到星际航行技术的成熟。当这些曾经是对人类社会“未来”描述的事情一件件成真，或许我们可以说，已经初露端倪的“智能时代”就是人类想象中“未来”的样子。

《智能时代》这本书展现了吴军博士的真知灼见和前瞻思维，这些都来自于他在大数据和机器智能领域的多年第一线实践经验。全书对大数据与智能革命带来的思维革命、技术上的挑战，以及机器智能如何改变人类社会，都做了全面的讲解。与其他一些写机器智能的书不同，这本书与作者之前的几本书一样，维持了作者对科学生动而易于理解的、有温度感的一贯的表述方式。

试读结束：需要全本请在线购买：www.ertongbook.com

大数据是解决不确定性的良药

“用不确定的眼光看待世界，再用信息来消除这种不确定性”，是大数据解决智能问题的本质。吴军博士在书中提到了世界的不确定性来自两个方面，一是影响世界的变量太多以至于无法用数学模型来描述；二是来自客观世界本身：不确定性是我们所在宇宙的特性。因此，用机械论已经完全无法对未来进行预测。

香农，这位不世出的天才，则通过借用热力学中“熵”的概念，引入“信息熵”，用信息论将世界的不确定性与信息联系在一起。这个建立在不确定性上的理论，正是今天人类研究大数据与机器智能的基石。

解决智能问题，就是将问题转化为消除不确定性的问题，大数据则是解决不确定性问题的良药。可以预见，在这里会诞生无数的机会。

现有产业 + 新技术 = 新产业

吴军博士在书中总结了从第一次工业革命以来历次技术革命中的一个规律，即每一次技术革命都会围绕着一个核心技术展开，第一次工业革命是蒸汽机，第二次工业革命是电，信息革命是计算机和半导体芯片，当下的智能革命则是大数据和机器智能。而在每一次技术革命中，只有率先采用新技术，才能立于不败之地。

在智能革命中，现有产业采用了新技术后，将会全面升级，成为新产业，这将给我们带来无限的机会。

智能革命带来前所未有的不连续性挑战

本书的一个重要观点是：机器智能革命的发生来自大数据量的积累达到质变的奇点。从这个角度来看，机器的学习同人类的学习并没有什么本质的不同。几千年以来，我们人类的知识都建立在归纳法之上，归纳法隐含的假设是“未来将继续和过去一样”，换句话说应该叫连续性假设。但即将到来的这个“智能时代”，可以说人类将遭遇前所未有的“不连续性”。如何在新的时代里生存，跨越底层认知的不连续性，是前进的第一步。

与工业革命相比，人工智能带来的革命程度将更深更广。书中也提到，一些人对变化开始有了一定程度的担心，认为机器智能将在未来危及整个人类的工作机会，大多数人在未来将不再被社会需要。不可避免，每一次大的技术革命都会带来阵痛，但同时诞生的，还有更多新的机会。而要想在智能时代取得胜利，成为“2%的人”，我们需要做的第一步，是打破现有的认知束缚。

如何在智能时代开始跨越思维的不连续性？寻找答案，此书也许是最恰当的一本。

前言

人类的胜利

2016年是机器智能历史上一个具有纪念意义的年份，它是一个时代的结束，也是新时代的开端。这一年距离1956年麦卡锡、明斯基、罗切斯特和香农等人提出人工智能的概念正好过去了60年，按照中国的习惯来说，正好过去了一个甲子。而当年在达特茅斯学院提出这个概念的10位科学家中最后一位科学家明斯基也在这一年的年初离开了人世，这或许标志着人类在机器智能领域第一阶段的努力落下了帷幕。就在明斯基去世后的两个月，Google^①的围棋计算机AlphaGo^②在与世界著名选手李世石的对局中，以4:1取得了压倒性的胜利，成为第一个战胜围棋世界冠军的机器人，它的意义要远远超过1997年IBM^③的深蓝战胜卡斯帕罗夫，因为从难度上讲，围棋比国际象

① 谷歌。

② 阿尔法狗。

③ 国际商业机器公司。

棋要难 6~9 个数量级。这件事不仅是人类在机器智能领域取得的又一个里程碑式的胜利，而且标志着一个新的时代——智能时代的开始。

从计算机发展的角度看，智能机器在所有棋类中战胜人类其实只是一个时间问题，因为机器运算能力的提升是指数级增长的，而人类智力能够做到线性增长就不错了。因此一定存在一个时间点——在所有的棋类比赛中智能机器都会超过人。在 1997 年 IBM 的深蓝战胜卡斯帕罗夫之后，围棋不仅是计算机尚未超越人类的最后一个棋类，而且还蕴含着上千年的东方文化，即棋道。虽然大部分人相信计算机最终可以在围棋上超越人类，但总是觉得那还是几年后的事情。就在 AlphaGo 和李世石比赛之前，李世石本人认为前者的水平和他相差一到两个子，也就是说，即使他让先也能 5:0 获胜。中国围棋界的泰斗聂卫平也认为今天的计算机是不可能战胜人类冠军的。就连曾经在 Google 工作过的 IT^① 行业老兵李开复博士也不相信 AlphaGo 能赢。这并非李开复等人对今天机器智能的发展状况不够了解，而是因为下围棋是一件太难的事情。2015 年年底，AlphaGo 仅仅赢了樊麾二段而已，离九段还差得远呢。但是大家忘记了一件事情，那就是 AlphaGo 水平的提高并不需要人那么长的时间，事实上在 Google 内部，大家在开赛前已经知道 AlphaGo 的水平并不在九段之下。

① 信息技术。

2016年3月9日，AlphaGo和李世石之间的世纪大战开始了。AlphaGo在第一盘出人意料地轻松获胜。当然，大部分人在赞誉AlphaGo水平的同时，依然认为这可能是李世石在试探计算机而已，毕竟那是五盘棋的比赛，用一盘棋试探自己毫不了解的对手未尝不是明智之举。但是当AlphaGo在第二盘获得连胜并且下出了很多人类预想不到的好棋后，对机器智能持怀疑态度的聂卫平等，都对它产生了敬意。在AlphaGo获得第三盘胜利之后，很多超一流的棋手都渴望和它一战，希望以此检验自己的水平，并且能够提高技艺。虽然李世石在第四盘抓住AlphaGo的一个失误打了一个漂亮的翻身仗，但是AlphaGo在最后一盘稳稳地控制着局面，直到胜利。可以讲在那一次人机大战之后，围棋界对机器智能从怀疑变成了顶礼膜拜，大家都意识到，按照AlphaGo在过去几个月里的进步速度，只要Google愿意继续进行科研，很快人类所有的围棋高手都无法和它过招了。

计算机之所以能战胜人类，是因为机器获得智能的方式和人类不同，它不是靠逻辑推理，而是靠大数据和智能算法。在数据方面，Google使用了几十万盘围棋高手之间对弈的数据来训练AlphaGo，这是它获得所谓的“智能”的原因。在计算方面，Google采用了上万台服务器来训练AlphaGo下棋的模型，并且让不同版本的AlphaGo相互对弈了上千万盘，这才保证它能做到“算无遗策”。具体到下棋的策略，AlphaGo里面有两个关键

的技术。第一个关键技术是把棋盘上当前的状态变成一个获胜概率的数学模型，这个模型里面没有任何人工的规则，而是完全靠前面所说的数据训练出来的。第二个关键技术是启发式搜索算法——蒙特卡罗树搜索算法 (Monte Carlo Tree Search)，它能将搜索的空间限制在非常有限的范围内，保证计算机能够快速找到好的下法。虽然AlphaGo的训练使用了上万台服务器，但是它在和李世石对弈时仅仅用了几十台服务器（1 000 多个CPU^①的内核以及 100 多个GPU^②）。相比国际象棋，围棋的搜索空间要大很多倍，AlphaGo的计算能力相比深蓝，其实并没有这么多倍的提高，它靠的是好的搜索算法，能够准确地聚焦搜索空间，因此能够在很短的时间里算出最佳行棋步骤。由此可见，下围棋这个看似智能型的问题，从本质上讲，是一个大数据和算法的问题。

当然，Google开发AlphaGo的最终目的，并非要证明计算机下棋比人强，而是要开发一种机器学习的工具，让计算机能够解决智能型问题。AlphaGo和李世石对弈，实际上是对当今机器智能水平的一个测试。从樊麾到李世石，他们实际上是用自己的专才在帮助Google测试当今机器智能的发展水平。在人机对弈的第四盘李世石反败为胜的过程中，他无意中发现了AlphaGo的一个缺陷。因此，Google的成功里面也有李世石等棋手的功劳。从这个角度来讲，

① 中央处理器。

② 图形处理器。

AlphaGo的胜利标志着人类在机器智能方面达到了一个崭新的水平，因此它是人类的胜利。

AlphaGo无论是在训练模型时，还是在下棋时所采用的算法都是几十年前大家就已经知道的机器学习和博弈树搜索算法，Google所做的工作是让这些算法能够在上万台甚至上百万台服务器上并行运行，这就使得计算机解决智能问题的能力有了本质的提高。这些算法并非专门针对下棋而设计，其中很多已经在其他智能应用的领域（比如语音识别、机器翻译、图像识别和大数据医疗）获得了成功。AlphaGo成功的意义不仅在于它标志着机器智能的水平达到了一个新的台阶，还在于计算机可以解决更多的智能问题。今天，计算机已经开始完成很多过去必须用人的智力才能够完成的任务，比如：医疗诊断，阅读和处理文件，自动回答问题，撰写新闻稿，驾驶汽车，等等。可以讲，AlphaGo的获胜，宣告了机器智能时代的到来。

AlphaGo的获胜让一些不了解机器智能的人开始杞人忧天，担心机器在未来能够控制人类。这种担心是不必要的，因为AlphaGo的灵魂是计算机科学家为它编写的程序。机器不会控制人类，但是制造智能机器的人可以。而科技在人类进步中总是扮演着最活跃最革命的角色，它的发展是无法阻止的，我们能做的就是面对现实，抓住智能革命的机遇，而不是回避它、否定它和阻止它。未来的社会，属于那些具有创意的人，包括计算机科学

家，而不属于掌握某种技能做重复性工作的人。

在AlphaGo取得人机大战胜利之际，我们出版这本书，希望能让大家更多地了解大数据的本质、它的作用、它和机器智能的关系、机器智能的原理和发展历程，以及它们对未来产业和社会的影响。本书一共分为七章，分别介绍了数据的作用，大数据和机器智能，机器智能的原理及其发展历程，大数据思维的核心及其重要性，大数据和机器智能与商业的关系，它们对社会正反两个方面的巨大影响。书中的核心内容来自我在研习社和一些大学商学院讲课的讲义，但是考虑到大家读书和听课毕竟有很大的区别，因此在将讲义改写成书的时候，我在书中增加了大量的案例和历史背景介绍，以方便大家能够系统地了解大数据和机器智能的来龙去脉，以及我们对未来进行分析的依据。

本书的出版，在很大程度上是研习社负责人曾兴晔女士、空无边处出版团队的张娴和郑婷女士，以及中信出版社经管分社的朱虹社长和赵辉编辑等相关人员积极推动的结果。在此我对他们表示衷心的感谢。由于本人水平有限，书中不免有这样或者那样的错误，希望广大读者朋友不吝赐教指正。

吴军

2016年4月25日于硅谷

目录

推荐序 智能时代，未来已来 IX

前言 人类的胜利 XIII

第一章

数据——人类建造文明的基石 001

.....
如果我们把资本和机械动能作为大航海时代以来全球近代化的
推动力的话，那么数据将成为下一次技术革命和社会变革的核
心动力
.....

现象、数据、信息和知识 003

数据的作用：文明的基石 009

相关性：使用数据的钥匙 019

统计学：点石成金的魔棒 025

数学模型：数据驱动方法的基础 030

第二章

大数据和机器智能 037

.....
在有大数据之前，计算机并不擅长于解决需要人类智能的问
题，但是今天这些问题换个思路就可以解决了，其核心就是变
智能问题为数据问题。由此，全世界开始了新一轮技术革
命——智能革命
.....

- 什么是机器智能 040
- 鸟飞派：人工智能 1.0 044
- 另辟蹊径：统计+数据 050
- 数据创造奇迹：量变到质变 057
- 大数据的特征 063
- 变智能问题为数据问题 074

第三章 思维的革命 087

在无法确定因果关系时，数据为我们提供了解决问题的新方法，数据中所包含的信息可以帮助我们消除不确定性，而数据之间的相关性在某种程度上可以取代原来的因果关系，帮助我们得到我们想知道的答案，这便是大数据思维的核心

- 思维方式决定科学成就：从欧几里得、托勒密到牛顿 090
- 工业革命，机械思维的结果 099
- 世界的不确定性 111
- 熵——一种新的世界观 115
- 大数据的本质 124
- 从因果关系到强相关关系 127
- 数据公司 Google 135

第四章

大数据与商业 143

.....
在未来我们可以看到，大数据和机器智能的工具就如同水和电这样的资源，由专门的公司提供给全社会使用
.....

从大数据中找规律 145

巨大的商业利好：相关性、时效性和个性化的重要性 155

大数据商业的共同点——尽在数据流中 164

把控每一个细节 167

重新认识穷举法——完备性带来的结果 176

从历史经验看大数据的作用 186

技术改变商业模式 197

加(+)大数据缔造新产业 202

第五章

大数据和智能革命的技术挑战 215

.....
大数据的数据量大、维度多、数据完备等特点，使得它从收集开始，到存储和处理，再到应用，都与过去的数据库方法有很大的不同。因此，使用好大数据也需要在技术和工程上采用与过去不同的方法
.....

- 技术的拐点 217
- 数据收集：看似简单的难题 230
- 数据存储的压力和数据表示的难题 238
- 并行计算和实时处理：并非增加机器那么简单 243
- 数据挖掘：机器智能的关键 247
- 数据安全的技术 254
- 保护隐私：靠大数据长期挣钱的必要条件 261

第六章

未来智能化产业 273

.....

现有产业+机器智能=新产业，未来的农业、制造业、体育、医疗、律师，甚至编辑记者行业都将迎来崭新形态，新产业将取代旧产业满足人类的个性化需求，大数据将导致我们整个社会的升级和变迁

.....

- 未来的农业 276
- 未来的体育 281
- 未来的制造业 287
- 未来的医疗 292
- 未来的律师业 310
- 未来的记者和编辑 313