



算法之美

指导工作与生活的算法
Algorithms to Live By

[美] 布莱恩·克里斯汀 (Brian Christian) 著
汤姆·格里菲思 (Tom Griffiths)

万 慧 胡小锐 译

图书在版编目 (CIP) 数据

算法之美 / (美) 布莱恩·克里斯汀, (美) 汤姆·格里菲思著; 万慧, 胡小锐译. -- 北京: 中信出版社, 2018.5

书名原文: Algorithms to Live By

ISBN 978-7-5086-8688-2

I. ①算… II. ①布… ②汤… ③万… ④胡… III. ①计算机算法-应用-经济学-通俗读物 IV. ①F0-39

中国版本图书馆 CIP 数据核字 (2018) 第 039429 号

Algorithms to Live By

Copyright © 2016 by Brian Christian and Tom Griffiths

Simplified Chinese translation copyright © 2018 by CITIC Press Corporation

ALL RIGHTS RESERVED

本书仅限中国大陆地区发行销售

算法之美

著者: [美] 布莱恩·克里斯汀 汤姆·格里菲思

译者: 万慧 胡小锐

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)

承印者: 北京诚信伟业印刷有限公司

开本: 880mm × 1230mm 1/32

印张: 11.75 字数: 240 千字

版次: 2018 年 5 月第 1 版

印次: 2018 年 5 月第 1 次印刷

京权图字: 01-2017-0143

广告经营许可证: 京朝工商广字第 8087 号

书号: ISBN 978-7-5086-8688-2

定价: 59.00 元

版权所有·侵权必究

如有印刷、装订问题, 本公司负责调换。

服务热线: 400-600-8099

投稿邮箱: author@citicpub.com

假设你想租房子，正在旧金山四处寻找房源。旧金山可能是整个美国最难找房子的城市了。由于技术产业的蓬勃发展，再加上城市区划法律严格限制建造新住房，旧金山的房租已经与纽约不相上下，甚至比纽约还高。房源清单列出来几分钟，房子就会被人们一抢而空。通常情况下，只有第一个把定金支票塞到房东手里的人，才能拿到房子的钥匙。

理论上讲，认真调查、仔细斟酌是理性消费者的一大特征，但是旧金山的残酷市场并没有为他们留有权衡考虑的机会。在购物中心或者网上购物时，人们可以反复权衡再做出决定，但是将要入住旧金山的租客没有这个特权，他们必须迅速做出决定：要么舍弃其他所有可能的选择，就选定当前正在看的这套房子，要么掉头就走，再也不要回头。

简单起见，我们姑且假设，你唯一关心的就是尽最大可能增加挑中最理想公寓的机会。你的目标是把“看过的好房子被人挑走”与“还有好房子没来得及看”这两种遗憾的发生概率降至最低。于是，你立刻发现自己陷入了两难境地：如果没有衡量的标准，如何判断一

套公寓是否是最合适的呢？如果你不先看一些公寓（这些公寓将被你放弃），又如何确定衡量标准？你收集的信息越多，越能在最合适的机会出现时准确地认出它，但是你已经与最合适的机会失之交臂的可能性也越高。

那么，到底该怎么办？如果收集信息的行为会危及结果，那么怎样才能掌握足够多信息的基础上做出明智决定呢？这个令人极其为难的情境近乎于一个悖论。

在被问及此类问题时，大多数人凭直觉给出的回答可能大致如此：这需要在继续挑选与立刻下手之间达成某种平衡。也就是说，你必须先看足够多的房子，确定一个标准，然后接受符合这个标准的房子。事实上，平衡概念正是解决这类问题的关键。但是，大多数人根本无法确定这个平衡点在哪里。好消息是，这个平衡点已经被找出来了。

答案就是 37%。

如果你希望选中最合适公寓的可能性达到最大，那么在看前 37% 的房子时不要做出任何决定（如果你准备花一个月的时间挑选房子，那么在前 11 天不要做出决定）。这段时间你是在为制定标准做准备，因此看房子时把银行卡放在家里吧。但是，过了这个时间点之后，你就要做好随时签约的准备（包括准备好定金等），一旦你对某套房子的满意程度超过之前看过的所有房子，就立刻下手。在继续挑选与立刻下手之间做出的这种妥协，并不仅仅是一种直觉，而是已经得到证明的最优解。

我们知道这个答案，是因为找房子问题属于数学上被称作“最优

停止”（optimal stopping）的一类问题。37%法则明确了解决这些问题的一系列简单步骤（计算机科学称之为“算法”）。事实证明，找房子仅仅是最优停止问题在日常生活中的表现形式之一。在面临一连串选择时如何做出决定的难题，经常会改头换面，以不同的形式出现在我们的生活当中。在驶入停车位之前，需要绕整个停车场多少圈？在商业风险中何时套现脱身？在买房子或者停车时，何时是结束观望、做出决定的最佳时机？

在约会这个更加令人头疼的问题上，人们也经常要面对这样的难题。最优停止理论是一夫一妻婚姻制度催生的科学。

每天，人们都要面临最优停止问题的困扰（当然，诗人更愿意追逐的话题肯定是求婚带来的烦恼，而不是停车时的两难境地），有时甚至会因此而痛苦不堪。不过，我们大可不必如此，因为这类问题至少可以通过数学方法来解决。借助并不繁复的算法，我们不仅可以解决找房子的问题，生活中遭遇的所有最优停止问题都可以被妥善处理。

从本质上讲，我们身边经常出现因为租房子、停车、求婚而感到苦恼的人，这些人其实就是在自寻烦恼。他们需要的不是治疗师，而是一种算法。治疗师告诉他们要在冲动与多虑之间找到一个正确的、舒服的平衡点。

算法告诉他们这个平衡点就是 37%。

※—※—※

由于我们生活在有限的时间和空间之中，因此所有人都会面临一

系列特定的问题，诸如在一天或者十年里，有些事必须做，有些事应当放手？如何在尝试新的体验与从事自己喜爱的活动中取得平衡，才能生活得惬意自在、心满意足？

这些问题看起来似乎都是人类特有的，其实不然。半个多世纪以来，计算机科学家苦苦思考的问题（有很多已经得到妥善解决）与这些日常难题在本质上并无区别。例如，处理器在执行用户请求时应该如何分配自己的“注意力”，才能降低费用、节省时间？在什么情况下应该在不同任务之间来回切换？刚开始应该接受多少任务量？如何利用有限的存储资源取得最佳效果？应该收集更多数据，还是根据已收集的数据采取行动？对人类而言，如何把握今天可能不是一件易事，但是我们身边的计算机可以轻轻松松地把握每一毫秒。显然，计算机有很多值得我们借鉴的地方。

将算法与人类生活相提并论似乎是一件很奇怪的事。在很多人看来，“算法”这个词意味着神秘莫测的谋划与操作，与大数据、大政府、大企业有密切的联系，正在逐渐变成现代社会基础架构中一个越来越重要的部分。其实，算法指的就是解决问题的一系列步骤，其含义远不限于计算机，存在的历史也远远长于计算机。在计算机开始使用算法之前，人类早就将算法应用到生活当中了。

“算法”（algorithm）一词得名于波斯数学家花拉子密。公元9世纪，这位数学家写过一本书，讨论用纸笔解决数学问题的技巧。[书名为“al-Jabr wa'l-Muqabala”，其中的“al-jabr”就是后来“algebra”（代数）这个词的前身。]不过，最早的数学算法早于花拉子密。在巴格达附近出土的4000年前的苏美尔人泥板文献上，就刻有一幅长除

法示意图。

但是，算法不仅限于数学。在按照食谱介绍烤面包时，食谱上的所有步骤就是一个算法。按照图样编织毛衣时，这份图样就是一个算法。使用鹿角的末端连续精确地敲打，使石器形成锋利的刃的过程（这是制作精密石器的一个关键步骤），也遵循着一个算法。从石器时代开始，算法就已经是人类生活的一部分了。

※—※—※

本书将探讨人类事务算法设计这个概念，以帮助人们更好地处理日常生活中遇到的难题。将计算机科学的研究方法应用于日常生活，可以在多个层面上产生深远的影响。首先，它可以提供切实有效的建议，帮助我们解决具体问题。例如，最优停止理论可以告诉我们何时应该小心观察，何时应该果断行动；探索-利用平衡理论教会我们如何在尝试新事物与因循守旧之间找到平衡点；排序理论可以帮助我们判断出是否需要以及如何整理办公室；缓存理论可以帮助我们合理地填充橱柜；日程安排理论则可以提供合理安排时间的高招。

其次，计算机科学还为我们理解这些领域的深层次运行规则提供了一套语汇。卡尔·萨根指出：“与其说科学是大量知识的汇总，不如说它是一种思考方式。”即使生活中的某些情况非常复杂，我们无法进行严格的数值分析，找不到任何现成的答案，我们也可以考虑这些问题的简单化表现形式，从而得出某些直觉和概念，帮助我们理解其中的关键环节并取得进展。

从更广泛的意义上看，借助计算机科学，我们可以了解人类思想

的本质和理性的意义，学会回答如何度过一生这个最古老的问题。把认知视为一种解决周围环境所造成的问题（从本质上看，都是一些计算问题）的手段，并认真地加以研究，就有可能彻底改变我们对人类理性的理解。

认为研究计算机内部运行机制能够帮助我们学会思考与决策、判断某个事物是否可信、选择行为方式的观点，在很多人看来，不仅把问题过于简单化了，而且具有误导性。即使计算机科学告诉我们应该如何思考、应该采取哪些行动，我们愿意接受吗？读一读讲人工智能和机器人的科幻小说就会发现，那样的生活似乎都不是我们所向往的。

之所以如此，部分原因是我们把计算机看成了机械呆板的确定性系统——这些机器借助严谨的演绎逻辑，通过穷举所有可选方案，无论花费多少时间、问题难度如何，它都可以给出完全正确的答案。事实上，在阿兰·图灵当时的想象中，计算机就应该是这样。这位第一个设想出计算机的人通过类比的方式给出了计算的定义，而类比的原型就是认真钻研的人类数学家——他们通过长长的计算步骤，最终得出绝对正确的答案。

因此，当人们发现现代计算机处理难题的方式与他们对计算机的认识并不一致的时候，他们也许会大吃一惊。当然，简单的算术对现代计算机而言没有任何难度。目前，计算机科学面临的最难解决的问题其实是人机对话、修复破损文件、下围棋取胜，这些问题都具有规则不明确、所需信息不全，或者需要考虑无数种可能性才可以找出正确答案的特点。研究人员已经开发出各种算法，使计算机在解决难度

极大的问题时不需要完全依赖穷举计算。要解决这些来自现实世界的任务，就必须正确处理好可能性问题，利用粗略估算，在时间与精确度之间做出某种妥协。

随着计算机处理现实任务的能力不断增强，计算机算法不仅对于人类自己的生活具有借鉴意义，同时还为人们理解人类认知提供了一个更好的比较标准。在过去的一二十年里，行为经济学对人类进行了非常具体的研究，结果发现，人类是不理性的，很容易犯错误，而问题的源头在很大程度上就是大脑这个古怪而独特的硬件。这种自我贬低的认识越来越普遍，却无法解释某些令人困惑的问题。例如，在完成包括想象、语言、因果推理在内的大量认知任务时，4岁儿童的能力仍然超过成本高昂的超级计算机，这到底是什么原因？

从计算机科学为日常问题提供的解决方案可以看出，人类思维具有另外一种特点——人生充满了难以解决的问题。人经常犯错误，虽然这可以说明人类大脑容易出错，但是也表明这些问题具有难以解决的本质特点。通过算法来思考我们周围的世界，了解我们所面临问题的基本结构以及计算机给出的解决方案的特性，可以帮助我们真实地了解我们自己，更好地理解我们所犯的那些错误。

事实上，人类需要不断面对计算机科学所研究的一些高难度问题，在不确定性及时间有限、信息不全、情况瞬息万变等不利因素的干扰下做出决定。针对一些问题，即使最前沿的计算机科学也没能开发出永远不会犯错误的有效算法，有的情形似乎是任何算法都无法解决的。

不过，尽管有的现实问题异常复杂，人们还没有开发出完善的算

法，但是一代代计算机科学家一直在与这些难题斗争，并且在这个过程中得出了深刻而独到的见解。这些来之不易的真知灼见与我们对理性的直觉认识并不一致，与数学家对周围世界的精确描述也迥然不同——数学家一心想要把这个世界变成整齐划一的线条。计算机科学告诉我们：不要总是考虑所有的可选方案；不必每次都追求最佳结果；偶尔犯点儿错误；放下包袱，轻装前进；有的事情可以暂时放一放；相信自己的本能，不要过多思考；放松自己；采用抛硬币的方式；要体谅，但是不能忘记；忠于自我。

用计算机科学的智慧指引自己的人生之路，这似乎是一条不错的建议。毕竟，与大多数建议不同的是，这条建议有据可依。

※—※—※

当初，算法设计在各学科的夹缝中找到了立足之地，它是数学与工程技术糅合而成的怪异混合体。现在，为人类设计算法的工作也面临相同的境遇——找不到一个现成的归属学科。今天的算法设计不仅需要借助计算机科学、数学和工程技术，还需要得到统计学、运筹学等相关领域的帮助。此外，我们不仅需要考虑计算机算法设计与人类思维活动之间的关系，还需要认真研究认知学、心理学、经济学等学科。

本书作者都有跨学科工作与研究的经历。布莱恩学习的是计算机科学和哲学，研究生阶段学习的是英语，毕业之后从事的是与这三个学科都相关的工作。汤姆学的是心理学和统计学，在加州大学伯克利分校从教期间，他主要研究人类认知与计算之间的关系。但是，人类

算法设计涉及多个领域，任何人都不可能是所有领域的专家。因此，在探索研究方便人类生活的算法时，我们还与过去 50 年最著名的算法专家进行了交流，询问这些全世界最聪明的人，他们的研究对他们自己的生活（包括寻觅配偶、收拾衣帽鞋袜）到底产生了什么样的影响？

接下来，我们就将开始引领大家游览这个神秘的领域。首先，我们将讨论计算机与人类大脑都需要面对的巨大挑战：如何应对有限空间、有限时间、有限注意力、未知的未知事物、不完整的信息与不可预见的未来给我们造成的麻烦，如何镇定自若、充满自信地面对这些麻烦，如何与其他人一起，共同面对这些麻烦，我们将讨论这些难题的基本数学结构，了解计算机解决大多数难题的设计原理（有时，这些设计甚至与我们的想象背道而驰）。此外，我们还将了解人脑的工作原理，了解人脑在解决相同类型问题、应对相同限制条件时有哪些独特且密切相关的处理方式。最终，我们不仅将得到有助于解决身边问题的一系列具体建议，学会在面临最复杂人类困境时有助于我们看清其脉络结构的新方法，还可以清醒地认识到人与计算机深度融合过程中的痛苦与艰辛。此外，我们还会有一些意义更加深刻的收获：一套描述周围世界的全新语汇以及一个从全新角度了解自己的机会。

序 言 VII

01 /001

最优停止理论

如何选择停止观望的时机?

秘书问题 004

37% 从何而来? 007

情场上的出手时机 012

掌握候选对象的完整信息 015

卖房子的时机 019

最优停车位置 023

见好就收的时机 027

随时准备停止 031

02 / 035

探索与利用

要最新的还是要最好的？

- 什么是探索与利用 038
- 如何利用剩余时间？ 041
- 赢留输变 044
- 基廷斯指数 045
- 遗憾与乐观 052
- 网上“土匪” 056
- 试验中的临床试验 059
- 不安分的世界 065
- 探索——孜孜不倦 069
- 走出探索和利用的两难困境 071

03 / 075

排序

建立秩序

- 排序狂潮 078
- 排序带来的苦恼 081
- 大 O 符号：衡量最坏情况的标准 082
- 平方时间：冒泡排序与插入排序 085
- 打破平方时间的魔咒：分治算法 087
- 超越比较法：比对数更好的算法 090
- 排序是搜索的准备工作 094
- 排序与体育 097
- 发牢骚的权利：噪声与健壮性 102
- 杀戮排序：啄食顺序与优势等级 105
- 以竞争取代争斗 108

04 / 113

缓存 忘了它吧

- 分级存储器体系 117
- 缓存清理与未卜先知 120
- 重整图书馆藏书 124
- 本地需求 126
- 家庭生活中的“高速缓存” 130
- 归档与堆存 132
- 遗忘曲线 136
- 经验暴政 139

05 / 143

时间调度理论 要事先行

- 安排时间是一门科学 146
- 处理时限 148
- 把事情做好 151
- 找出问题所在 154
- 优先级反转和优先约束 156
- 减速带 159
- 放弃所有：抢占和不确定性 162
- 抢占并不是随意的：关联转换 164
- 颠簸状态 166
- 中断合并 170

06 / 175

贝叶斯法则

预测未来

贝叶斯牧师的倒推理 178

拉普拉斯定理 181

贝叶斯法则与先验信念 183

哥白尼原则 185

贝叶斯与哥白尼 187

真实世界的先验…… 189

他们的预测规则 191

小数据与思维 197

我们的预测体现出我们自己 199

机械复制时代的先验 201

反对复杂性案例 207

数据崇拜 211

过度拟合无处不在 212

检测过度拟合：交叉验证 216

如何应对过度拟合：惩罚复杂性 218

启发法 220

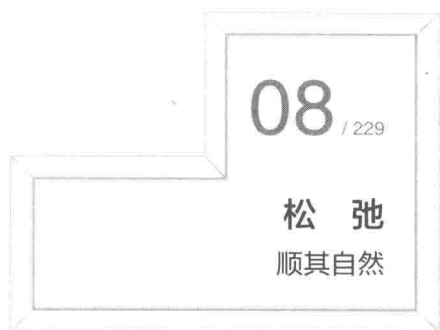
人类进化中的过度拟合 222

何时应该想得更少？ 226

07 / 203

过度拟合

不要想太多



最优化的难度 233

定义的难度 235

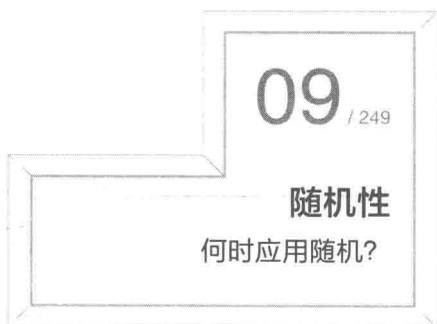
放松吧 236

无数灰色地带：持续的松弛 239

只是一张超速罚单：拉格朗日松

弛算法 242

学会松弛 245



抽样 252

随机算法 255

抽样的优势 260

三部分的权衡 264

山、谷和陷阱 266

局部最大值之外 269

模拟退火算法 270

随机性、进化和创造力 273