



# 陈阳 杨熙坤

## 《算法帝国》提要

北京市陈经纶中学 陈阳 杨熙坤

### 目录

导读 .....	2
关键词 .....	2
前言 .....	3
第 1 章 华尔街，第一张多米诺骨牌 .....	5
第 2 章 人与算法的简史 .....	13
第 3 章 机器评出的前四十榜单 .....	19
第 4 章 计算机的秘密高速公路 .....	24
第 5 章 系统里的博弈论 .....	26

第 6 章 呼叫机器人医生.....	28
第 7 章 人的分类.....	30
第 8 章 华尔街与硅谷的较量.....	31
第 9 章 华尔街的损失让我们大家获益.....	33
第 10 章 未来属于算法和它们的创造者.....	35
参考文献.....	37

## 导读

已经公布的《普通高中信息技术课程标准》（2017 年版）指出：“课程内容紧扣数据、算法、信息系统和信息社会等学科大概念”，“学习算法，可以从系统的角度描述和解决问题，有助于学生未来专业的发展”，“学生应该理解利用算法进行问题求解的基本思想、方法和过程，掌握算法设计的一般方法；能描述算法，分析算法的有效性和效率，利用程序设计语言编写程序实现算法；在解决问题的过程中能自觉运用常见的几种算法”，“掌握贪心、分治、动态规划、回溯等常见算法”，“体验算法效率的差别”，“掌握算法分析的一般方法和过程”，“能有意识地把算法及算法思想迁移应用于实际生活和学习中”，.....，等等。

认真阅读《算法帝国》一书，无论是对于提高教师的施教能力还是对于提高学生的学科素养来说，都是至关重要的。

今天，算法涉足的领域已经远远超出了其创造者的预期。特别是进入信息时代以后，算法的应用涵盖金融、医疗、法律、体育、娱乐、外交、文化、国家安全等诸多方面，显现出源于人类而又超乎人类的强大威力。《算法帝国》是《纽约时报》畅销书作者斯坦纳的又一力作，通过一个又一个引人入胜的故事，向读者介绍了算法掌控世界的真实情况，揭示了“机器人革命”是如何悄然在我们身边发生的。斯坦纳常在各种会议就科技、算法和大数据等主题发表演说，阐释算法与大数据如何塑造、变革和掌控人类生活。

这篇《〈算法帝国〉提要》是两位作者在认真研读《算法帝国》后合作写成的。文章浓缩了原著的精华，对其中的案例、观点、叙述、解释和分析进行了提炼。

## 关键词

算法 计算机科学 编程 华尔街交易 数学 数量分析 物理学 工程学 机器人 博弈论

## 前言

作者举了《高得离谱的书价》的案例。2011年4月初，加利福尼亚大学伯克利分校的进化生物学家迈克尔·艾森登录到亚马逊网站，想要为自己的实验室购买一本课外书。他要找的书是彼得·劳伦斯所著的《The Making of a Fly》，讲的是苍蝇怎样由一个单细胞卵长成为一只嗡嗡响的飞虫。这本1992年出版的书虽已绝版，但还是很受专业学者和研究生的喜爱。艾森习惯花35~40美元买旧书来读。但4月8日这天，亚马逊网站上有两个知名商家提供这本书的新书，要价却远远高于他愿意出的钱，一家卖1730045美元，另一家卖2198177美元。

艾森猜想这个天价肯定是误报或开玩笑。他认为没有人（包括作者本人）会对这书估价如此之高。第二天他再次登录亚马逊查看，却发现这本书非但没有回归到正常价格，反而分别涨到2194443美元和2788233美元。第三天，这本书又涨到2783493美元和3536675美元。此后连续两周，书价一直在涨，4月18日到达最高价23698655.93美元，买家还得支付3.99美元的运输费！4月19日，价格下降到106美元。

一本研究苍蝇遗传学、晦涩难懂的书，亚马逊为什么要将它的价格设定为近2400万美元呢？是它突然受到了亿万富翁级收藏家的追捧吗？是书里有找到宝藏的线索吗？还是它成了图书界“1869年的拉菲”？事情真相是这样的：两家书商都在亚马逊经营上万种书籍，并且都用算法给图书定价，然而不受监管的算法打了一场价格战。第一家书商利用算法将自己的价格设定为另一家的0.9983倍，而第二家的算法则始终让自己的价格保持为第一家的1.270589倍。就这样的相互攀比、反复交手，最终让这本书的价格赶上了曼哈顿顶层豪华公寓的价格。若非人为干预、覆盖系统设置，价格战还会继续升级。如果说这是一个意外事件，那倒也是一个关于算法失控但却没有造成什么损害的奇事。但差不多在一年以前，还发生过一件比它奇怪得多的事。

作者又举了《剧烈动荡的美国股市》的案例。2010年5月初，由于政府出台财政紧缩计划，希腊爆发全国性大罢工。2010年5月6日早上，全球股市受希腊动荡局势的拖累而走低。很多人担心希腊债务违约在所难免，并且会引发全球经济萧条。纽约时间上午10点左右，美国股市大跌2.5%，损失惨重。

但事情很快由糟糕透顶发展成令人费解。

美国东部时间下午2点42分，股市剧烈波动后进入自由落体状态。2点47分，仅仅300秒之后，道琼斯指数下跌达998.5点，为其有史以来最大单日跌幅。全球最受重视的股票指数道琼斯工业平均指数的走势看起来就像是被人恶搞了一样。近1万亿美元的财富就这样蒸发了。

道琼斯指数在下跌了998.5点触底之后开始迅速反弹，反弹过程就跟下跌过程一样快。实际上，反弹还要快一些。一分钟之内，日波幅难超300点的道琼斯指数，暴涨了300点。

某些股票的每股购入价狂跌至1美分，让好多资产数十亿美元的公司一名不文，但又在数秒之内反弹回30美元或40美元的股价。还有些股票猛然飙升。某一刻苹果的每股成交价高达10万美元（从约

250 美元一股开始上涨)。美股陷入剧烈动荡,无人知晓原因。不管问题出在哪里,总之不能全部归结到某个有严重错误的交易单或是某个流氓交易员。这场动荡来得如此迅疾,有些交易员或市场观察者或许就在上个厕所或喝杯咖啡的时间里完全错过了千点大跌又急速反弹的一幕。

关于闪电崩盘的确切成因众说纷纭。有人把矛头指向堪萨斯城的一位财富投资经理人,他的算法过快出售掉价值 40 亿美元的股指期货,导致其他算法跟风。有人指责一伙不明交易商合谋共同利用算法打压股价。有人认为这不过是老式的恐慌造成的,跟 1929 年的大崩盘不无相似之处。然而可以肯定的是,如果不是因为不受人类管束的算法占据了市场,不到一秒内独立下单并完成交易,股价波动幅度不可能如此之大,波动速度也不可能如此之快。但这样的算法确实统治了市场。

作者说到,闪电崩盘只是预示了一个更大的趋势:算法掌控一切。

关于算法最经典的定义就是:通过所掌握的信息使算法使用者得到某一针对性问题的答案,或掌握输出信息的一系列指令的策略。

在本书中,作者通常把旨在执行一个任务的多元联系算法称为计算机程序。这些计算机程序包含成千上万个输入、因数及函数。其中最复杂的程序就像是你大脑的神经元网络:根据需要上下传递处理信息,它们是动态的,而且能够自我改善。

随着算法及其背后的数学成为华尔街的标准,其他受算法影响较小的领域也吸引了数学家、工程师和物理学家的注意,华尔街把这一类人称为宽客,这个名字来源于数量分析专家 (quantitative analyst)。这些宽客和计算机程序员正在寻找存在弱点的新产业,他们可以利用算法推翻该产业的旧秩序,从中谋取巨额利润。

在线音乐服务软件 Pandora 用算法来识别用户的音乐品味。

如果使用了某些先进的计算机科学技术,比如机器学习和神经网络,算法甚至能够基于观察结果创造出新的改良的算法。

如今,最有社会地位上升机会的就是能够利用代码构建算法的人。

作者谈到,在现代词汇中,黑客一词有两种定义。在某些人看来,黑客行为本就属于犯罪——程序员自由出入本属禁区的电子财产领地。20 世纪 80 年代描绘的黑客形象就是飞快敲打着键盘,侵入政府绝密信息库、银行及其他各种不属于他们领地的不法之徒。

然而在现代科技世界,这个词却表现出来其好的一面。这也是作者在本书中使用的定义。用硅谷的话来说,黑客就是能熟练编写计算机代码并把白板上的理论转化成算法的人。他们构造的算法能做出重大决策、交易股票、开车、为大学申请者分类,还能跟世界上最优秀的人类玩家打扑克。

为需要创造力的领域编程,构思并写出解决人类难题的优美算法需要特殊的才能。

作者在 Y Combinator 遇到的很多同道中人在 14 岁时就学会了编程,到他们进大学时(通常是精英大学)就已经能够写出上千行代码,编写性能稳定的应用程序,并且几个小时就能设计出一个原创网站。2011 年夏天在 Y Combinator 的三个月让作者了解到马克·扎克伯格这类人是如何产生的:他们并

非天赋异禀，而是长年累月在电脑屏幕前练就了对代码的敏锐感知力。正如经验丰富的文字编辑对习语、标点及字体的用法烂熟于心，他们对代码也是再熟悉不过了。

这些代码的设计师和算法的撰写人是这一代最杰出的企业家。新王朝的缔造者不再来自商学院，而是来自工程学和计算机科学研究室，他们在这里夜以继日地研究编码作业，终于练就了一身编程本领，可以构建出有革新意义的算法，并创办由算法驱动的公司。有时候，这些算法构造者早早离开大学，甚至根本不去上大学。他们早在 15、16 和 17 岁的时候就已经知道怎样把自己想象的东西构建成型。

构造创新的算法并不是一件容易的事，但其难度却与日俱减。所以有能力开发计算机程序的人——最聪明的数学、科学以及计算机人才前途无量。

## 第 1 章 华尔街，第一张多米诺骨牌

作者举了《世界上第一个全自动算法交易系统》的案例。1987 年年初的一天，纳斯达克交易所的工作人员琼斯出现在世贸中心的电梯间。他来例行拜访一位快速增长的客户。

公司前台接待了琼斯，然后走进房间去请主人。她带来一位矮个子、满头银发、衣着整洁的男人。他叫托马斯·彼得菲，有着一双蓝色的眼睛，说话带着口音。

彼得菲带琼斯走进交易厅，琼斯迷惑不已。他预想的是骚动的人群、吵闹的电话声、打印机的声音以及交易员向纳斯达克交易终端输入交易指令时此起彼伏的叫买叫卖声，但他只看见一台纳斯达克交易终端。

“其他交易场所在哪儿呢？”琼斯问，“你的交易员呢？”

“那就是，全都在这儿呢。”彼得菲指向房间里唯一的纳斯达克交易终端旁边的 IBM 电脑说，

“我们的交易都是用它完成的。”交易终端和 IBM 电脑之间有一团连接线，电脑里装有指示交易品种、交易时间和交易数量的代码。琼斯没有想到，他刚刚见到的是世界上第一台全自动算法交易系统。

琼斯惊得目瞪口呆。彼得菲把这一切看做是创新的交易方式，琼斯却认为他用临时配置的终端作弊，违反了交易规则。

“你不能这么做。”琼斯说。

纳斯达克没有交易场所，所有交易都是通过电话或电脑完成的。电脑网络接收从独立的纳斯达克交易终端的键盘上传来的交易指令。彼得菲整合了本应该连接到交易终端的数据线，将它连接到他的程序员和物理学家团队自制的嵌入 IBM 个人电脑主板的一块电路板上。IBM 电脑运行的是彼得菲自己编写的软件程序。电脑通过纳斯达克的数据线获取信息，利用算法分析市场，迅速做出交易决定，然后将交易单通过一团连接线再传回纳斯达克终端的内部。

“交易终端和 IBM 电脑的连接必须切断，你得像其他交易者那样，通过键盘逐条地输入交易指令。” 琼斯说。

纳斯达克给了彼得菲一周的改正时间，其交易必须符合监督员的要求。

彼得菲询问了自己的工程师，是否能够造出某种设备像照相机那样从屏幕直接读取信息，然后将它们转化成电子字符，传送到等候指令的 IBM 电脑里。答案是：能。

在接下来疯狂的一周里，彼得菲和他最好的工程师忙着焊接金属、编写代码、焊接数据线。他们在纳斯达克终端屏幕前安装了一个大型菲涅耳透镜，放大屏幕字体。又在离透镜一英尺远的地方安装了一部相机。从照相机牵出一根数据线连接到旁边的电脑上。彼得菲和他的程序员仅用了几天时间就编写出了一套软件，可以解码照相机传来的可视化数据。那些数据从特定程序软件流到彼得菲已经做好的算法里。

现在 IBM 电脑有了一根新的连接线，它不是接入到纳斯达克终端机箱，而是连接到了悬在终端键盘上空的那一堆密密麻麻的金属棒、金属塞和手柄上。这个装置是一个从零组装的自动化打字机。手柄断断续续敲打着键盘，执行从电脑传来的交易指令，不到 30 秒就有几十个交易单输入终端。

纳斯达克说交易单必须得有输入终端，可并没有规定谁来完成输入。彼得菲的团队花了六天时间创造了一种输入交易和指令的半机器人。

琼斯一周后如约而至。彼得菲带琼斯穿过大门，骄傲地指向自己的创作。

“这是什么？”琼斯问。

彼得菲解释他的交易机器如纳斯达克要求的那样运作——键盘输入，一次一单。就在这时，交易市场活跃起来了，机器也跟着忙碌起来。彼得菲的程序交易速度如此之快，输入设备就像一把全自动机关枪不停开火。交易单不断涌入，手柄噼里啪啦打在键盘上，噪声之大甚至淹没了谈话声。机器每次停下来，仿佛要安静一会儿，谁知转瞬之间再次启动，更加气势汹汹地弹出比上次还要多的单子。

琼斯摇摇头，彼得菲扮了个鬼脸。他造出了世界上运行速度最快的交易机器，他也预料到这个交易机器可能面临被拆卸的命运。琼斯沉思了几分钟，然后一言不发地走出了彼得菲的办公室。彼得菲做了最坏的打算，那就是纳斯达克禁止他的发明参与交易。但是琼斯没有再回来，而且彼得菲担心的那个电话也没有打来。他的交易所完好无损。彼得菲几年前不到 10 万美元起家，在 1987 年赚到了 5000 万美元。

构建算法模仿、超越并最终取代人类，是 21 世纪最重要的能力。

作者提到，第一次从事金融业工作时，彼得菲构建了一种方便投资者和交易者一次性同时比较证券不同特性和价值的算法。

作者写到，1969 年，受雇于杰里克的彼得菲引进了华尔街的第一台黑盒。黑盒用于接收市场数据，进行精密的运算处理后，指示客户买卖操作。任何交易决策的算法大部分由模型、函数、决策树构建并模仿人的决策。

杰里克让彼得菲创建一种可以准确预测期权价格的算法。

期权指一种能在未来有效期限内以特定价格买入或卖出一定数量的某种特定商品的权利。

20 世纪 70 年代初期，期权交易很频繁。杰里克和彼得菲总结出期权定价的三大要素：执行价格、合约到期时间及股市或金属商品市场的价格波动量。对于价格波幅很大的股票，期权定价也应该更高，因为很有可能会在合约期内达到极高的执行价格。其他影响因素包括提振看涨期权价格、打压看跌期权价格的现行无风险汇率等。彼得菲需要提出一种合理考量评估每一个因素的简单算法。

苦心钻研一年后，彼得菲构建出一种巧妙地考虑了所有因素的微分方程式算法。算法统治金融界的第一阶段是：将人编写好的算法装入电脑，然后分析输入数据，给人发出交易指令。

彼得菲在工作中运用算法大概一年之后，发生了一件震惊华尔街的大事。1973 年，同为芝加哥大学教授的费用·布莱克和迈伦·斯科尔斯发表了包含布莱克-斯科尔斯模型的论文，阐明怎样用该模式计算期权的确切价值。基于布莱克-斯科尔斯模型的算法会在数十年后彻底改变华尔街，并将一群志趣相投的人——数学家和工程师聚集到金融界的前沿。布莱克-斯科尔斯公式与彼得菲的算法相似，并为迈伦·斯科尔斯赢得 1997 年的诺贝尔奖，而布莱克于 1995 年去世。

改变并非在一夜之间发生。布莱克-斯科尔斯公式是一种奇妙无比的偏微分方程，但大多数交易者不会仔细研究学术期刊。即便是他们读了这篇文章，运用这个公式也并非易事，需要有很深厚的数学功底方能驾驭它。那时候几乎没多少人明白布莱克-斯科尔斯定价模型，更没有人知道一个匈牙利人已经编写出了相似的算法，几乎单单盈利。彼得菲和杰里克在这件事上守口如瓶，毫不张扬。懂得布莱克-斯科尔斯公式的交易者可以利用它推算期权交易的确切价格。这个公式为交易者提供了在金融市场作弊的小抄。任何人只要能准确推算出布莱克-斯科尔斯公式的每个因素值并将它运用于实时期权交易，就能够赚钱。交易者会卖出实际价格高于公式推导价格的期权，买进实际价格低于公式推导价格的期权。

作者表示，20 世纪 70 年代末期，算法开始进入人们的工作，这一趋势席卷了世界各地的金融市场，标志着华尔街黑客时代已然来临。华尔街逐渐吸引了美国越来越多杰出的数学家和科学家投身于编写交易算法的工作。

金融业不仅成功地吸引到研究学府和科技公司前途无量的年轻人才，也使许多功成名就的资深工程师和科学家放弃学界显赫又轻松的职位来到华尔街。

能够率先抓住交易机会的就是赢家，电脑运行的算法总是能抢得先机，击败寻找相同交易的人类。20 世纪 70 年代末期，用代码构建算法的能力还很罕见。

在精明的杰里克和精通算法构造的彼得菲的带领下，莫卡塔赚取了数百万美元，成为世界上最有影响力的商品交易商之一。到 1975 年，他们雇用了 50 名程序员，成为华尔街编程宝库的中流砥柱之一。

算法可以看做是由一个又一个的二元决策组成的巨大的决策树。二叉决策树可以为复杂的对象生出数百万甚至数十亿的节点。算法的决策树接收输入，然后将输入值带入方程和公式中运行，再将得到的答案作为新的输入，层层迭代演绎，创建出细节复杂的一串串长字符串。德国数学家戈特弗里德·莱布尼茨 300 年前就为这一精密学科建立了理论：生命可以分解为一长串连续的二元决策。

彼得菲的算法成为数百个输入、变量以及依赖性微分方程和积分的复杂矩阵，就好像一个密密麻麻的蜘蛛网。

1976 年，彼得菲的黑客军团已经发展到了 80 人，是当时世界上最大的金融交易编程团队。

影响股价的很多因素不可量化，比如商业信誉、发展前景、诉讼隐患和行业竞争。因此不存在股票定价的神奇公式。然而期权价格却可以单纯用概率与数据反映。只有极少一部分包括彼得菲在内的精英人士知道这一点。对懂数学的人来说，机会就如滔滔江河般奔腾不息。

1977 年，彼得菲花两千美元购置了他的第一台 Olivetti 牌家用电脑。他白天在莫卡塔工作，晚上用这台意大利电脑编程，夜以继日地构建可以攻下股票期权市场的算法。

作者指出，1977 年，彼得菲带着 20 万美元存款辞掉了安稳的工作，在期货交易刚刚起步的美国证券交易所买了一个交易席位。购买交易席位花掉了他 3.6 万美元，剩下的 16.4 万美元用来进行交易。

为了能在美国证券交易所的第一个交易日有一个好的开始，彼得菲整天面对着电脑苦干 18 个小时完善算法，绘制出指导他在交易厅买卖期货的表格。他苦心钻研那些他认为期货定价与实际价值相差较大、定价可能存在问题的公司。对每一个公司，他都要做出一系列关于该公司不同股价所对应的期货价值表。在一天中股价波动的时候，彼得菲迅速查询他做的那些表格，推算出股票期权的合理价格，根据这个价格进行买卖操作。这些表格就是他的纸质版算法。

彼得菲严格遵循表格上的价位，如果期权价位不符合他保守的赢利方针，他就不会买入。

他一直坚持算法系统交易，很少有大损失。虽然布莱克-斯科尔斯公式早在七年以前就发表了，但它对市场的撼动力还不够，不足以影响到彼得菲和其他交易者依靠他们自己的聪明才智赚钱的做法。

尽管彼得菲的算法和表格很有成效，但他始终都是单枪匹马一个人。他需要更多人和他一起在交易厅奋斗。他逐渐聘请了一些交易员。为了避免损失并且能够掌控交易员的操作，彼得菲训练他们根据表格开出买卖报价，他每晚都要根据算法更新表格上的数字。由于彼得菲在美国证券交易所的交易量越来越大，他把自己的交易机构重新命名为原木山(Timber Hill)，这个名字来源于纽约某乡村的一个路名，是他曾经度假的一个地方。他买下了更多的交易席位，开始尝试诸如配对交易和套利的其他交易策略，而这两种策略他在莫卡塔制定交易战术时就已经很熟悉了。

彼得菲提出了一个想法，交易者可以通过连接到中央计算机的手持设备输入买单卖单，电脑再根据价格-时间优先的原则自动公平地匹配报价单。但这个想法很快就被控制这一行的人——专家经纪人（也称庄家）扼杀了。



既然不能用技术改变整个市场的运行方式，彼得菲决定加入一个更有权势的群体，成为美国证券交易所的期权做市商。他想只要严格地根据算法在合理的价格波动区间进行交易，扩大交易量就会降低风险、提高利润。但他首先得想法子让控制交易的专家经纪人能一致、迅速地确认他的买单卖单。他不像其他的交易者会饶舌闲聊、互换消息，他更像是一个局外人。专家经纪人不喜欢和彼得菲做交易，所以他们并不经常接受他的买卖申报。彼得菲决定雇一些专家经纪人会喜欢的人。

彼得菲招来的女人不懂交易，更不要说算法了。但那时候彼得菲的交易员并没有依靠自己做交易，也没有用表格做指导。因为彼得菲已经设计出了让每个人都可以漂亮完成交易的新系统。

彼得菲尤其感兴趣的、也是他编写算法所寻找的是称为 **delta** (对冲比率) 中性的交易。在这样的交易里，彼得菲卖出权利金过高的看涨期权，同时买入权利金过低的看跌期权，这样不管行情大涨还是大跌，他都处于一个安全位置。

电脑一旦整合了 **delta** (对冲比率) 中性交易，就立刻打印出来。彼得菲叫来他的场内员工将打印出来的交易单交给他的美女军团。4 个月后，彼得菲赚的钱比以往任何时候都多，而秘诀就是在市场对手毫不知情的情况下从科特龙获取的数据流。当今世界的很多热门公司已经证明了数据是一个公司在某一行业成败的关键所在。彼得菲开创了对大型数据库进行自动化编译及使用的先河，自此华尔街便开始了对数据库的挖掘。

作者说到，彼得菲告诉朋友们，“任何人跟我做交易都能赚钱。”

为了证明自己的话，他请来一位最不可能出现在交易厅的人——马文·范·皮布尔斯，彼得菲旧东家杰里克的朋友。

尽管已经非常有名而且工作也相当繁忙，这个剧作家还是抵挡不住华尔街的诱惑，于 1982 年 11 月来到彼得菲的交易机构原木山，和其他女交易员一起为彼得菲工作。

范·皮布尔斯也赚到了一大笔钱。

作者说到，彼得菲被专家经纪人指派为做市商，这样他就有机会率先得到交易厅新开的单。从技术角度来说，不管市场走势如何，做市商必须总是同时保持买卖单。彼得菲像大多数做市商那样违反了规定，只是按照算法的指示挑选交易来做。他没有一刻连续稳定地报出买卖价。

专家经纪人厌恶了每次交易都输给彼得菲的感觉，他们告诉他至少得对最少数量额度的期权保持公开稳定的买卖报价，否则就要取消他的做市商资格。彼得菲不失谨慎地对此进行了争辩，同时思索着既可以保持利润不受这些人的破坏，又能平息他们怒气的方法。

因为彼得菲被要求在某些股票品种上长时间保持买卖单，所以他不能让交易员迅速到电话处听候每个指示。时刻保持公开报价意味着交易员不得不待在交易厅，时刻警惕地盯着市场走势。

解决方案就藏在他过去对于交易的一项提议中：手持电脑。彼得菲曾经提出在交易所配置某种设备以摆脱专家经纪人组织交易的过时交易方式，但当时他并不知道这种特定的设备是什么（这种设备还

不存在），也没有认真考虑过交易所独断专行的处事方式以及谁握有生杀大权，而这些人正是他想要用某种设备来取代的人。

彼得菲的新计划不是要让整个交易厅都配备手持电脑，只是给他的交易员配备。

交易所监管人员对彼得菲的这一招并不感兴趣。大多数反对者声称机器会挡住不停走动的交易者的路，而且很有可能在市场走势对彼得菲不利的时候莽撞下单导致他没有能力平仓。

美国证券交易所不顾这些反对，同意彼得菲的交易员把小型平板电脑带到交易厅来。

早上，市场交易铃声响起之前，彼得菲会打开设备，从里面拉出一簇连接线，接上从科特龙专线收集数据的个人电脑。手持设备从电脑上下载最新市价和数据，指示只能待在交易场内的交易员如何提出股票和期权报价。交易员输入现行市场价格，平板电脑通过指示灯告诉他们这笔交易是否值得做。所以现在专家经纪人让彼得菲的女交易员提出买卖报价时，她们能迅速开出符合彼得菲的算法的价格。

彼得菲把目光移到了交易量最大的芝加哥期权交易所，但监管人员反对他的提议：交易所的成员绝对不能容忍在交易时使用任何电脑，很明显是害怕这种设备会给用户带来交易上的优势。

彼得菲随即转战纽约证券交易所，该交易所在证券交易领域有统治地位，同时也想发展期权业务。

平常发号施令的交易所为了扩大交易量，急切地满足了彼得菲的要求。为了更好地协助他的交易员们，彼得菲在他们所站的位置上空安装了一组由不同颜色的灯泡组成的灯条，然后他又编写了一个可以通过这些彩灯传递交易指令的系统。对于已经配备了场内电脑的的交易员来说，这真是如虎添翼。

彼得菲用 C 语言重新编写了程序。一切如常，电脑不停地处理数据，寻找好交易。当它找到了合适的交易机会，就立即向纽约证券交易所天花板上的彩灯传送一系列电脉冲，原木山的交易员从开市到闭市一直盯着那组彩灯看。不久之后，彼得菲成了交易所最大的做市商之一。

彼得菲于 1977 年花了 36000 美元在美国证券交易所买下了原始交易席位，所以他的办公总部就一直安在美国证券交易所。如果想要扩大交易领域和交易量，美国证券交易所能够提供的空间已经不够了。所以，他于 1986 年将总部搬迁到世贸中心。在那里，他有足够的空间指挥分散在各地交易所的交易员。

世贸中心离交易厅有好几个街区那么远，平板电脑很难更新数据，因为提供数据的个人电脑也被搬到了新的总部。彼得菲的解决方案也是华尔街过去流行的一种做法：送信人在新动向传开之前跑步到市郊，告知市郊的人。彼得菲雇了好几个跑步快的员工，除了在办公室打打杂，主要任务就是携带数据更新好的手持设备从世贸中心全速奔跑到交易所。

跑步送情报的人最终还是被传输数据的电话专线取代了。20 世纪 80 年代中期，一些精通技术的交易商开始使用电话专线。彼得菲也是引领这一潮流的人。他租赁了电话线在世贸中心办公室和他的交易厅战场之间传输数据，安置在交易厅的电脑处理总部传来的数据，然后将数据一股脑儿导入手持设备。但即使有了这一步升级，交易员一天也要在场内和电脑间跑几次才能让手持设备接收到数据。为

了免去这样的麻烦，彼得菲和他的工程队在手持设备和原木山的电脑上安装了小型无线电发射机。完成了这一步，数据便可以轻轻松松传到交易厅的手持设备上了。

作者说到，1987年，诸如标准普尔 500 指数之类的指数基金不仅在普通民众之间广为流行，专业交易员也对其青睐有加。

彼得菲推测，如果指数构成大同小异，那么它们的波幅及其期权期货走势也应该相同。而事实上，这些金融衍生工具在不同的交易所价格迥异。

为了充分利用这些把握十足的交易机会，彼得菲需要在旧金山的交易所、芝加哥的两个交易所还有纽约的两个交易所安置人手。他和他的工程队制造了新的手持设备，他们雇用的一波波交易新兵将会利用这些设备为原木山占领这个国家的其他交易所。彼得菲购入了几十台计算机，还租赁了横越全国的电话线以保证数据传输畅通无阻，使得计算机网络可以保持实时通信，实现全国范围内交易价格的更新。新的交易所也安装了无线电发射机，这让彼得菲在芝加哥、旧金山和纽约的交易员可以看到相同的信息。现在在纽约下的卖单可以立即用芝加哥的买单实现对冲。

当金融衍生品的价值回归到期望区间，彼得菲的电脑会指示交易员平掉仓位、锁定利润。指数在所有市场的交易量都很大，每天的套利机会没有几千次也有几百次。有些交易商在纽约和芝加哥之间安装有电话线路，方便员工互通价格，利用大的价格差获利。但彼得菲的自动化交易系统大小通吃，不仅能够收获大价差盈利，也能抓住小价差获利机会，而且总是比别人抢先一步。彼得菲创建了第一个统辖东海岸和西海岸的算法交易机构。

手持设备完成的所有交易活动通过无线电传输到每一个交易所等候交易数据的终端，然后电脑将数据打包整理通过租赁电话线直接送达世贸中心原木山的总部。数据到达总部后，由一个简称为相关器的大型主算法进行处理。相关器运行大批密集代码方阵分析市场、定位风险，然后发出指令，调度分散在各个市场的原木山交易员通过相关操作抵消风险。相关器分析十几个证券及金融衍生品市场的实时价格，发出一个个轻而易举稳赚的交易指令。已经完成的交易通过数据线传入，相关器接收了原木山交易员所有已建头寸信息，发出对已建头寸进行风险对冲的交易指令。相关器的某些交易是 100% 全自动的——交易指令直接发送到纳斯达克终端，这些纳斯达克的终端会自动键入交易指令。

相关器触发交易指令，纳斯达克终端自动执行交易，这是算法统治的第二阶段。这一阶段，算法处理数据、分析市场，发出不再是由人而是由一台机器执行的交易指令。彼得菲从根本上创建了一个在机器之间而非在人与人之间传送重要信息的华尔街。第二阶段的算法交易尚需十年以上才能完全统治美国交易市场，但是彼得菲和纳斯达克开启了第二阶段算法交易的先河。在彼得菲将他的科技窍门发展到这一步后，只剩下第三阶段——算法摆脱人的控制实现自动调整，并在某些情况下实现算法自编——来彻底实现计算机程序自动化模式。

彼得菲经常坐在原木山位于世贸中心的总部办公室里看着交易员下的单不断地流入相关器。屏幕上列出了已经完成的交易及相应的对冲风险的交易，总部电脑会将交易指令传达给相关市场的原木山交

易员。对冲风险的交易一旦完成，也会在相关器的屏幕上显示出来。彼得菲从来不是一个赌徒，他的游戏策略是零风险稳赚，不管盈利多小。当大笔交易从相关器的屏幕弹出，彼得菲会目不转睛地盯着屏幕，看他的交易员对冲掉风险。

作者谈到，1986年到1987年两年间，彼得菲赚了5000万美元，这一令人震惊的数字标志了他的巨大成功。然而在某些时刻，被科技超越的华尔街玩家又赶了上来——那是曼哈顿下城掀起一股宽客热的时候，科学家和工程师开始被吸引到金融界。

彼得菲的交易机构在将总部搬到了康涅狄格州的格林威治后更名为盈透证券集团，在20世纪90年代持续领跑华尔街，至今依然如此。

彼得菲的交易机构能够领跑华尔街的原因在于：无论以前还是现在，它都是工程师和程序员的交易机构。在盈透证券集团，工程师就是整个公司，这让彼得菲的机构类似于华尔街版的谷歌——一个工程师生产产品并作出重大决策的地方。盈透证券的员工75%以上都是程序员和工程师。

彼得菲指出盈透证券集团没有一个人有商学院学位，这让它在华尔街显得那样格格不入。彼得菲坚称他永远不会雇用MBA。

从20世纪90年代一直到21世纪，盈透证券的业务不断扩张，在美国获取越来越多的市场份额后，彼得菲把算法用到了欧洲市场。2007年5月4日星期五，盈透证券首次公开发行股票。

他的公司市场定价为120亿美元。

彼得菲保留了盈透证券85%的控股权，只公开出售10%的股份。即便这样，这次IPO也让彼得菲将11.8亿美元的美元尽收囊中。盈透证券的上市亦成为当年美国年度第二大规模IPO。

作者提到，当今证券世界，人们已经成为饶有兴致地观察机器交易的观察者。当今的市场已是算法的天下。道琼斯通讯社和彭博社设有专供交易机器阅读的新闻专栏。

当机器成为主导力量，我们永远也不能确定市场的走势。

在某些交易日美国所有交易所日交易量的40%都是由两家中西部公司完成的，很多人甚至连金融界的人都没听说过这两家公司。一家叫Getco，位于芝加哥；另一家叫Tradebot，位于堪萨斯城。两家公司都雇用了世界顶级的电脑黑客和工程师。他们专注于每一个获利机会，其中很多交易每股净收益甚至不到1美分。这样的公司大小还有成千上万个，它们唯一的工作就是用自动运行的计算机程序在市场上赚钱。总的来说，算法就是交易市场，美国超过60%的交易都是由算法执行的；欧洲市场和亚洲市场也快赶上这个潮流了。曾经由咬牙切齿、成群结队、高声叫卖的交易大军主宰的市场现在已由互相竞争的算法所控制。

我们的证券市场已经成了专业化的战场，以至于算法可以潜伏几个月，一旦让它看穿对手的设计缺陷、过时代码或是识别出对手的交易模式，就会即刻攻击。很多从事证券市场交易的算法被构造出来模拟随机性。但我们不能在随机事物上下注赌博、制订计划或是榨取利益。

彼得菲认为算法交易玩过了头，他已经收敛了公司的做市行为。

彼得菲认为在光速交易的时代，买卖报价都应至少稳定维持一段时间，即使少于一秒的时间也足以消除市场上无处不在的导致股价时而跌落悬崖时而冲上云霄的弄虚作假和阴谋诡计。

他最担心的还是流氓算法引发的一系列大亏损交易会导致它的用户无法清偿债务。因为某些高频交易算法使用杠杆交易，也就不难想象瞬间完成的一连串糟糕的交易会导致清偿危机，使交易者的经纪商及其代理的客户破产。

华尔街和我们的钱已经成了算法交战的虚拟战场。算法的脚步不会仅止于此。华尔街上演的机器传奇让我们得以一窥未来世界的发展蓝图。

## 第2章 人与算法的简史

作者写到，精妙绝伦的算法背后的高等数学正在复兴。从来没有任何一个时代有像现在这么多的人懂它，也从来没有任何一个时代有像现在这么多的人通过讨论与研究传播自己对于它的理解。你只需要登入硅谷创业公司孵化器 YCombinator 的黑客新闻留言板，这个已经发展成为世界上最有影响力的网站，就能了解一切。

首页上总有一、两个帖子在讨论高斯函数、布尔逻辑或其他数学分支，正是这些数学理论成就了今天的算法世界。

无论是会计室还是客户服务台，算法都无处不在。早在人类知道数学会让这一切成为可能的时候，算法统治世界的趋势就悄然逼近了。

作者表示，千百年来，人类一直在设计、修改并分享着算法，这一活动早在算法这个词出现之前就开始了。算法无需用到高等数学，甚至可以和数学领域无关。巴比伦人处理法律事务时会用到算法，古时候拉丁语老师检查语法时会用到算法，医生靠算法来预测病情，无数遍布全球的普通人曾试图用算法预测未来。

算法的核心就是按照设定程序运行以期获得理想结果的一套指令。

算法一词来源于公元 9 世纪的波斯数学家阿布·阿卜杜拉·穆罕默德·伊本·穆萨·花刺子密，他写了第一本关于代数的系统著作《还原与对消的科学》。代数的名字就直接来自于书名中的“还原”。中古世纪的学者们用拉丁语传播花刺子密的学说时，他名字的拉丁语音译“算法”便成了任何程序化运算或自动运算方法的统称。

人们在靠近现代巴格达的舒鲁帕克地区发现了算法最早的文字记录。统治了幼发拉底河流域 1500 年的苏美尔人把公元前 2500 年的泥板文书留给了世界，文书阐释了在人数变动的一群人中均分谷物的可重复使用的方法。该方法使用的是小型测量工具，这在当时很有用，因为粮商没有一次性称量几千磅食物的巨型天平。那块用符号描述测量算法的泥板现在保存在伊斯坦布尔博物馆。

几千年前构建的某些算法在现今电子化的世界依然起着非常大的作用，而这些作用是它们的创造者无法想象的。很多网站、无线路由器和任何需加密用户名和密码的地方，都使用了一个名叫欧几里得的希腊数学家在 2000 多年前构思出来的算法。

欧几里得算法在现代社会各行各业的应用程序编制中无处不在，它还可以被用来提取大部分现代音乐的节奏样板。约公元前 300 年，欧几里得写了《几何原本》，在其后的 2300 多年间，此书一直是几何教科书的典范。书中包含了寻找不同数值的最大公约数的算法。

作者指出，12 世纪的欧洲刚刚引进阿拉伯数字和十进制，研究生物学、植物学、天文学甚至是农学的人对随之发展起来的一个概念应该都不陌生。尤其是数学家开始玩味一个神奇的比率 1.618——也称黄金分割率——大自然中无处不在的一个数字。蕨类植物的叶片形状、DNA 的分子结构以及银河系行星的运行轨道无不包含着黄金分割比。看起来“正确”的伟大建筑总是在空间与图案的设计上体现了 1.618 这个比率。20 世纪最著名的建筑大师勒·柯布西耶在自己的很多建筑作品中体现了这一比率，比如位于巴黎附近嘎尔什的始建于 1929 年的加歇别墅。平面设计师发现很多苹果的产品也应用了这一比率，包括它的商标。

列昂纳多·斐波那契发明了黄金分割数列，并将阿拉伯数字系统引入欧洲。他被许多历史学家公认为中世纪最重要的数学家。类似近代银行的机构在意大利中部崛起以前，比萨的列昂纳多广泛游历了地中海沿岸诸国，甚至到了东方。斐波那契是他为大众熟知的名字，他发现阿拉伯数字比笨拙的罗马数字好用得多，尤其是涉及数学运算的时候。

1202 年，斐波那契发表《珠算原理》。这位年轻的数学奇才在书中解释了小数怎样转化为分数，怎样运用小数、分数简化记账法以及解决实际问题。斐波那契解决了当时常见的难题，他举例说明了如何分发胡椒、动物毛皮和乳酪之类的物资。斐波那契引用花刺子密的观点，发明了一系列西方文明几百年来频繁使用的算法，用于未来现金流流量现值计算和类似现代抵押贷款的利息结算。

斐波那契在销售量将近 1 亿的小说《达芬奇密码》中备受瞩目，作者丹·布朗将《珠算原理》提到的神奇的斐波那契数列作为一大重要元素来挖掘。斐波那契数列从第三项开始，每一项都等于前两项之和：1、1、2、3、5、8、13、21、34.....数列后一项与前一项的比值无穷趋近黄金分割 1.618。

最近 20 年来，市面上有超过 100 本书鼓吹在商品、股票和外汇等各类交易市场中无处不在的黄金分割以及它在交易中所起的神奇作用。

作者认为，戈特弗里德·莱布尼茨和同时代的艾萨克·牛顿一样，也是一位博学的通才。他涉猎的领域遍及欧洲大陆绝大部分有趣的学科。莱布尼茨曾说过，在哲学上只有两条绝对真理：神和虚无。万物皆由此二者而生。

莱布尼茨发明了一种用 0 和 1 两个数字来表示数和加减乘除运算的数制。他在 1703 年发表的论文《关于只用两个记号 0 和 1 的二进制算术的阐释》中给出了二进制语言的定义。

莱布尼茨 1646 年生于莱比锡城，从小就是优等生，15 岁上大学，20 岁完成博士学位课程。他坚信二进制的应用范围会非常广，他不满足将其仅局限于博大精深的数学定理。

莱布尼茨说一切物理变化皆有因。在某种程度上，人甚至也有外力作用于自身形成的既定轨道。博弈论，这个晚于莱布尼茨时代很久以后发展起来的学科就基于这一事实。莱布尼茨相信通过分析因果关系，绝大多数事物的未来都可被预测。

莱布尼茨是历史上第一个接近人工智能的人。这位数学家明确规定了认知思维和逻辑可以被简化为二进制表达式。越复杂的思想越需要简单朴素的概念来描述。复杂的算法同样也是由一系列简单算法组成的。莱布尼茨说，逻辑总是能被无情地简化到其骨干，就像纷繁复杂的国家铁路网追根溯源是由一系列简单的铁道双向岔口组成的。如果逻辑能够被分解为一连串的二元判定，即便这样的二元判定树无限延伸，那么不用人也可以执行。莱布尼茨梦想着将所有逻辑思维简化为机械运算，他开始在自己设计的一台机器上实践着这个梦想。

听闻布莱士·帕斯卡造了一台加法器后，莱布尼茨下决心超过他，开始着手制造自己的机器。他的机器可以更加流畅地进行加减运算，并能解决乘除运算，这是帕斯卡的机器做不了的事情。莱布尼茨画了设计图，聘请到一位巴黎的钟表匠制造机器，终于在 1674 年造出一台更完善的机械计算器。

他的设计在之后的 200 年间发挥了重大作用，一代一代的计算器在此基础上相继问世，他的设计图也被收进文献记录。

莱布尼茨是机器计算的先驱，他相信通过将人类逻辑推理分解为数学运算，能够找到一种演算推论器，即一种可以解决争议的算法。

至于微积分，是他而不是牛顿创造了所有学微积分的学生至今都在使用的积分与导数的优美符号。

微积分实现了对变化的精准研究与建模，莱布尼茨对微积分符号系统和理论的发展完善为数学家提供了构建强大算法的武器，数学家们利用微积分构建的算法制造半导体、实现无线电通信并以激光的精准度将人造卫星发射进轨道。微积分与算法的历史、价值和力量交织在一起。数学家戴维·伯林斯基在他的著作《Infinite Ascent》中很好地总结了这一关系：“人类思想史上爆发了回声阵阵的超音爆！在微积分被发现以前，数学是一门趣味盎然的学科；而在微积分被发现以后，数学变成了一门力量无穷的学科。只有 20 世纪出现的数学理念（计算机）算法的影响力可与之媲美。微积分和算法是西方科学界的两大主导思想。”

莱布尼茨用三种方式推进了算法科学的发展，他是微积分的发明者之一；同样重要的是，他引进了数学建模算法，可以用一系列简单的二元区表示复杂难题。

他对算法的第三个贡献在于，他发现了存在于最简单的语言片段及其表达的人类情感之间的联系。莱布尼茨认为，语言及人类如何使用语言应该用一种严谨、科学的方式来研究。这位哲学数学家猜测人的措辞是依靠他们的情感和认知来选择的。

华尔街或其他地方上演的算法推动的革命，核心在于一个永恒的目标：预测。说得更准确些，就是预测其他人会做什么，这就是赚钱的方法。莱布尼茨的预感——人类个体的活动可按照被预测的方式进行——比任何人知道的都准确，这已经成了驱动华尔街算法发展的事实。

今天，所有的计算机语言都是因为二进制才得以发展，而计算机语言也仅仅是方便编写算法的工具而已。同样得益于二进制的还有运行算法的芯片和电路。

作者提出，只有在待考量和操作的数据确定可知的情况下算法才会起作用。如果不是确知股票期权的价格完全取决于其历史波动率、利率及期权执行价，我们就无法构建出可以推导其真实价值的算法。如果彼得菲或是布莱克和斯科尔斯不知道这些是影响期权价格的关键因素，也就很难创造出一种能够有效交易并击败人类的算法。

有时，发现算法对象关系中真正重要的变量，就跟构建一种自动运用这些发现的算法同样困难。预测性的算法要正常运行，需要得到跳过噪声数据、瞄准价值因素的指示。提取关键因素是通过分析海量数据来实现的。筛选数据，让趣味盎然、无法凭直觉获取的关系重见天日，最常用的方法就是回归分析。回归分析法就是利用历史数据推测现在。物理学家、统计学家或工程师都可以基于历史数据对现在进行准确预测。在发展构建符合历史数据的预测模型的方法上，英国皇室做出了一定贡献。

乔治四世 1817 年委派卡尔·弗里德里希·高斯负责汉诺威公国的测量工作。高斯发现当时的测量工具测出的结果不准确，于是发明了日光反射仪，它可以利用镜子将光束反射至远距离。高斯知道，即使有了这种新工具，测量那么大面积的土地也难免不准。而 20 年前，18 岁的高斯就发现了处理类似问题的一种方法——最小平方方法。

最小平方方法为基于已观察到的结果构建预测模型提供了方法。高斯建立了求实际值和预测模型的计算值的偏差的最小平方和的方程式。最小平方方法是现代统计学和算法建模的基础。比如要构建一个预测未来股价的算法，测试算法准确度的方法就是用确定的历史日期与价格运行算法。人们基于历史数据提炼出一个函数或一条曲线，希望借此预测未来数据。数据分析专家、工程师或是物理学家讨论数据拟合或是回归分析的时候，他们使用的方法很可能就是高斯的发明。

高斯研究了测量误差分布，总结出背离（误差）符合正态（也称高斯）分布。高斯和同时代的数学家观察到绝大多数可被测量的事物符合正态分布。这样的分布用曲线表示出来很像一口钟。钟形曲线边缘被称为尾部。高斯分布的尾部逐渐趋近于零。现实中的情境往往受人的不理智行为所影响，所以曲线的尾部也就不那么纤细了。一旦涉及人的心理和行为，异常事件发生的可能性就会变大，便产生了所谓的胖尾分布，那也就不是高斯分布了。人们凭着高斯分布在华尔街赚了多少钱，就因信奉高斯分布却忽略胖尾分布赔了多少钱。

编写符合正态分布规律的算法更为简单。虽然历史不断证明人的行为从来不能以常态来判断，一些黑客还是选择只考虑正态分布。1987 年的黑色星期一、1998 年的俄罗斯债务违约以及 2010 年的股市闪电崩盘，对于指望用高斯分布设计的算法赚钱的人，这些事件能让他们倾家荡产。



正态分布的应用改变了人类，开辟了现代统计学领域。

高斯认为欧拉和牛顿是最伟大的数学家，但大部分专家认为高斯本人也在其中。很多学者说阿基米德、牛顿和高斯是数学界不朽的三巨头。

高斯 21 岁的时候获得博士学位，他的博士论文首次为代数基本定理提供了完整的论证。高斯在数学领域的很多方面都有重要发现，包括数论、二元方程式以及质数分布定理。

作者说到，现代金融学的很多方面，从养老金到保险再到算法交易，都是基于概率论发展起来的，博彩、建筑和飞机制造行业也是如此。现代医学最有潜力的发展动向便是基于概率论建立其检测和诊断的方法论。政治党派提名候选人、建立其政治理念导向的依据并非实用主义或者事实，而是概率论。足球教练查阅小记录本，根据比赛剩余时间场内发生其他事件的概率做出决定。

所有概率论在现代的应用都可以追溯到 1654 年两个法国人布莱士·帕斯卡和皮埃尔·费马之间的一封信。

这两位著名的数学家横跨法国就概率论这个问题互通书信长达五年之久。帕斯卡写给费马的信使得农民的粮食可以免于遭受市场风险。也正是由于这封信，100 年后欧洲财政部门才能发布准确的死亡率表。而伦敦之所以跃升为世界上最重要的城市并至少持续了一个世纪之久，主要原因之一就在于这个航海国家善用保险，在鼓励商人开辟新贸易市场的同时免于遭受投机活动可能带来的巨大风险。

在帕斯卡研究的基础上，雅各布·伯努利以及他之前的克里斯蒂安·惠更斯将严密的概率论应用到纸牌和骰子等概率游戏中。通过对概率游戏的研究，伯努利提出了大数定理。

高频交易者会保持很大的交易量，目的是使赢利概率趋近一个平均值水平。概率论也是现代棒球队总经理在挑选种子选手时会参照的一个理论。而在医疗诊断领域，出现了依靠基于概率的算法去诊断病情而忽略任何先人之见的一类医生。

对股票交易者来说，一旦市场行情有波动，以买卖点差为盈利目标的算法可能就行不通了，交易者和他的算法可能陷入错误方向的行情。如果算法在 50 美元的价位买入了微软，却没能在股市下跌之前在 50.02 美元的价位卖出，而在新的卖出执行价 49.99 美元卖出，就会遭受损失。

交易算法是用来预测市场走势的。成功的算法只需有 51% 的正确率就够了。但如果高频交易者一天只交易 10 次，51% 的概率也很可能让他陷入亏损。可能有好些日子他 10 次交易里有 7 次都输了。但是高频交易者之所以被称为“高频”是有原因的。他们一天做一万多次交易，几百万股的交易量。

正如伯努利指出的，随着交易次数增加，有盈利的交易会趋近于 51% 的概率。

除了对华尔街的贡献，喜爱数学的人们应该感谢伯努利定义了计算复利的公式。伯努利研究计算连续复利的算法时还发现了数学常数  $e$ 。在与莱布尼茨的通信中，他把这个数字称作  $b$ ，或许取自他自己的姓。后来这个数字以莱昂哈德·欧拉的命名  $e$  著称于世。

作者谈到，促进统计学发展的思想巨人之一、法国数学家皮埃尔-西蒙·拉普拉斯称赞莱昂哈德·欧拉“是我们所有人的大师”。

欧拉毕业于瑞士最古老的巴塞尔大学，他是那个时代最出众的数学家。

欧拉是俄罗斯圣彼得堡科学院的教授，他着手解决困扰当地人的著名的哥尼斯堡七桥问题。哥尼斯堡被普雷格尔河分为几块陆地，包括两座岛，有七座桥将陆地与岛连接起来。当地散步的人希望找到可以穿过这个城市的一条路：每座桥都能走到且只走一次。

为了说明这个问题，欧拉在纸上画了一些点（也称叉点），用线（也称连线）连接起来，分别代表陆地和桥。他注意到线的长度和形状可以任意改变，叉点也可以移动，前提是所有的线（桥）都是完好无缺的。欧拉当时所画的内容在现代数学中被称为图形，这个图形最终证实了不存在一次走遍七座桥且每座桥只通过一次的走法。欧拉解决这个问题的同时创造了图论。

欧拉创造的图论和我们经常看到的股票走势图或销售报告图不是一个概念。欧拉的图是树形图，可以用来表示自然界的网络、微芯片的电路或同城不同人之间的关系。专门为图论编写的算法为现代计算机科学的发展翻开了令人振奋的崭新一页：生物学家可以通过它建立 DNA 链和生理特征之间的联系，教授们可以解码披头士的音乐，中央情报局可以掌控恐怖分子网络，华尔街观察员可以找出看似不相关的事物之间的关系。图论尤其对分析 Facebook 的人脉关系网很有用，通过分析图形中哪些叉点（人）的连线（关系）最多，以及哪些连线在吸引眼球和评论方面最活跃，就可以知道哪些人最有影响力了。工程师要应聘 Facebook 的工作，经常会被考到图论的题。

欧拉的专著、论文、教科书和技术手册多达 886 种，产出量是他有生之年欧洲所有数学出版物的 1/3。为了避免太多发现用同一个人的名字命名，很多数学定理和公式都是以在欧拉之后第一个发现或使用它的人的名字命名的。

很多人可能还记得上学的时候学过的著名欧拉公式： $V - E + F = 2$ 。数不胜数的算法基于这个公式编写。

这个公式并不仅仅局限于如立方体、角锥体、圆锥体和球体之类的基本刚性图形，它在分析碳分子的几何图形、看似随机的天气系统、光学、磁学和流体力学时都被用到。通过公式背后的理论，欧拉开始思考非刚性图形，也就是现在大家熟知的拓扑图形。拓扑学属于混沌理论的一个方面。在过去的 20 年里，很多数学家在华尔街用混沌理论算法赚了不少钱。

作者提到，正是由于布尔发明的计算系统和创新的代数形式，网页才得以工作，我们才能上传照片到 Facebook，在博客里码字。

17 岁的布尔在草坪上散步时突然灵光一现，想到某些代数符号可以用来定义一门逻辑语言，一门思维语言，即解构人类理智内在工作机制的一种方法。

布尔 20 多岁时开始攻克高等数学难题，并从 1841 年起在期刊杂志上发表文章。1854 年，他发表了《An Investigation of the Laws of Thought, on Which Are Founded the Mathematical Theories of Logic and Probabilities》一书。

布尔代数赋予了计算机电路内涵与思想。

阿达·洛芙莱斯是一位数学学者，还被认为是世界上第一位计算机迷。1842 年，她在为查尔斯·巴贝奇的分析机写说明时，设计了几组不同的输入。理论上讲，这些输入会启动机器执行特定的计算和任务。这样洛芙莱斯就成了首次为机器编写算法（程序）的人。

20 世纪 30 年代后期，麻省理工学院研究生克劳德·香农把莱布尼茨的二进制计算数制与包括 **and**（与）、**or**（或）、**not**（非）、**not-or**（或非）和 **if**（如果）在内的布尔运算符结合起来。香农发现所有的表达式都可以写进电子电路，而且能够解决几乎所有的数学难题、存储数据、编辑包括图像和文字在内的任何信息。香农被誉为信息论之父。

### 第 3 章 机器评出的前四十榜单

作者写到，艺术创作、发明创新、著书立说、为公司出谋划策、生产改变世界的产品的工作，向来被认为是算法无法企及的。这些高薪而又自由的职位被人们羡慕。有人称从事这些职业的人为创意阶层，也有人叫他们“在生活中学习的人”或聪明人。

这些聪明人认为骇人的机器革命不会影响到他们，因为他们的想法是算法所不能创新、机器所不能创造的。现在我们意识到这些都是危险的臆断。

算法可以学会评估人的作品的质量和独创性，它们也能自己创作。

现在确实存在能够进行音乐创作的算法，其中一个名字叫 **Annie**。它们创作的作品能如大师勃拉姆斯、巴赫、莫扎特的作品那样豪迈激昂、富有创造力，也能像超市放的音乐那样朗朗上口、耳熟能详。

作者举了《时来运转的本·诺瓦克》的例子。本·诺瓦克是搞音乐创作的，2004 年他还只开得起 1993 年产的尼桑蓝鸟。诺瓦克最不喜欢的地方就是它的收音机，奥克兰那么多的广播电台，它却只能搜到两个。但他又没钱买新车，唯一的选择就只能听听 **BBC** 了。

一天，**BBC** 的一篇报道谈到西班牙发展起来的一项技术，被采访的人声称他能够预测哪些歌会成为热门流行歌曲。

诺瓦克开车回家，坐到了电脑前。他搜索到当时叫 **Polyphonic HMI** 公司的网站。只要花 50 美元，就可以用这家公司开发的算法分析他上传的任何音乐文件。有潜力成为热门金曲的歌会得高分，不吸引人的歌就得低分。

诺瓦克几年前写了一首歌叫《Turn Your Car Around》，他觉得这首歌很有潜力，便上传到网站。

不久，网站跳出了一个分析结果给诺瓦克。它使用算法打出分数，高于 6.5 分的歌有相当大的潜力成为热门，高于 7 分的歌有机会进入流行歌曲排行榜。诺瓦克的歌得了 7.57 分——这个得分和有史以来很多最热门的摇滚金曲得分一样高。

Polyphonic HMI 公司总部在西班牙，维护算法的电脑工程师会把高分歌曲从服务器上下载下来，拿到办公室去放。

“这首歌明显有它的可取之处，我们的人一遍又一遍地放着。”经营着这家公司的麦克·麦卡克雷迪自己就是个音乐人，他给欧洲唱片公司的联系人打了电话。

诺瓦克把他的歌上传到麦卡克雷迪的网站两个星期以后，他的电话响了。打给他的是英国音乐制作人阿什·霍维斯的代理人。霍维斯手下的年轻流行歌手黎·莱恩的专辑正缺曲目，霍维斯觉得诺瓦克的歌正好可以放到这张专辑里。

诺瓦克很快同意了这个利润丰厚的交易：这首歌在广播、电视或广告中插播时他可以得到 50% 的版权费。他的歌不仅进了黎·莱恩的专辑，还被指定为首支单曲，并在首次播出时排行英国流行歌曲榜第 12 位，连续两个月成为在英国播放次数最多的单曲。它在整个欧洲大陆也一样流行。

这首歌的成功改变了诺瓦克的生活，但他一直没有换掉那辆旧的尼桑蓝鸟。“它还过得去，而且我现在再也不会关掉 BBC 了。”他说。

在音乐的世界里，一丁点儿运气就可能成就一个艺术家，可要是没了那点儿运气，艺术家们也就此被埋没了。慧眼识珠的算法能够改变这一切。

作者又举了《一部作品如何斩获八项格莱美大奖》的例子。身为美国人的麦卡克雷迪选择了一条与众不同的道路，成为改变未来音乐界的技术权威。

麦卡克雷迪还是学生的时候在西班牙加泰罗尼亚语区待过。说加泰罗尼亚语的人表达时间的方式别具一格。

麦卡克雷迪和朋友设计了一种表盘，表达时间的方式和加泰罗尼亚语一模一样。他找到法国一家钟表厂来制作这种表。不久，麦卡克雷迪经营的表成了轰动西班牙的时尚。

麦卡克雷迪爱好音乐，他先后有两首歌位列加泰罗尼亚语流行曲排行榜榜首。麦卡克雷迪的成功使他跻身正牌音乐创作人之列，而他卖表的故事亦成了西班牙人津津乐道的营销神话。1992 年巴塞罗那举行了第二十五届夏季奥运会，承包奥运场馆修建及使用的公司聘请了大名鼎鼎的麦卡克雷迪。他们想聘请一位营销总监用新的方式来让场馆赢利。

正当麦卡克雷迪把场馆生意做得风生水起的时候，硅谷的科技热潮传到了西班牙。

麦卡克雷迪遇到了巴塞罗那的一家小型技术公司，这家公司开发算法来分析流行音乐的基本结构、模式和风格。麦卡克雷迪和这家公司的工程师见面详谈后，坚信搞技术是可以有一番作为的。他提议组建一个针对音乐人和唱片公司开发技术的新公司，他们把它叫做 Polyphonic。

Polyphonic 公司开发的算法对输入其中的音乐有着令人称奇的解析。算法基于新兴光谱反卷积技术，用不同序列的傅里叶变换和数学函数解构歌曲，分辨出歌曲的调子、节拍、节奏、音高、和弦进程、声音饱满度、声音质量和抑扬顿挫。软件将算法得到的数据构建成三维模型，算法通过观察而非聆听歌曲的三维结构的方式，绝对客观地将其与过去的热门歌曲进行对比。把刚分析过的歌曲与过去的冠军单曲在屏幕上用点表示出来，形成云结构。基本结构相似的热门歌曲汇聚成带。靠近热门歌曲带中心的歌即便不一定能红，也一定是首不错的歌。

麦卡克雷迪一边完善算法，一边用他的机器分析了尽量多的即将发行的专辑，以此测试算法的能力。据算法分析，有一张专辑的 14 首歌里有 9 首都可能成为热门歌曲，这是披头士乐队的水平。专辑的艺人没什么名气，但这张专辑《Come Away with Me》问世后，销量高达两千万张，并为歌手诺拉·琼斯斩获八项格莱美大奖。

作者表示，算法可以引进新面孔，但它判断一首歌是基于过去的流行歌曲，因此我们得到的结果很可能与这些过时的流行歌曲类型一样。这么多年来，这样的歌曲都会成为算法分析判断的对象，很明显这是该技术的瑕疵。

麦卡克雷迪希望自己的算法会在未来几十年里成为音乐界的焦点。他认为算法带给我们的流行歌曲榜类型将会更多样化，而不是越来越单一。有些艺人虽然声音很另类，但因为所唱歌曲内在的节拍、旋律和节奏靠近热门歌曲带，所以这样的艺人会被挖掘出来而不是被埋没。麦卡克雷迪还可以帮助音乐监制和艺人通过微调歌曲的节奏和节拍使其更靠近热门歌曲带，但这样做的结果是：算法会带来一个被迫同化的音乐世界，而不是引领多样化的音乐爆炸性地发展。

这已经变成了现实。前四十流行榜单的大部分歌都是由同一队人马操刀的。瑞典歌曲作家马丁·桑德伯格，业界称他马克思·马丁，从 20 世纪 90 年代开始为邦乔维、后街男孩和布兰妮写过一系列排行榜冠军单曲。2008 年以来，他写了 10 多首冠军单曲、20 多首排行前十的单曲。某些人对写出吸引大众的歌很拿手，麦卡克雷迪的算法又熟悉流行歌曲的一般性特征，因此不难推断流行音乐的世界很快会被机器统治。不管当下大众喜好什么，唱片公司必定会迎合大众的口味。但要想持续推出热门歌曲，算法再适合不过了。

作者指出，热门歌曲带不仅适用于流行音乐，同样适用于昔日的音乐大师贝多芬、莫扎特、巴赫和亨德尔的作品。麦卡克雷迪的软件显示，某些举世闻名的交响曲自成一派。莫扎特和贝多芬虽然对麦卡克雷迪的高科技玩意儿或者说算法一无所知，但也肯定懂得某些特定的音乐形态和旋律对听众更有吸引力。这就是为什么有些古典名曲听起来如此相似。成功的音乐家会引来其他音乐家的竞相模仿，他们都在追寻同一类声音，以求获得赞誉。

你不会把挖洞看成是铁铲的功劳，你会归功于挖掘者。

作者再举了《披头士（甲壳虫）乐队的秘密》的例子。算法也会告诉我们过去听的是什么样的音乐。一位加拿大教授开发的一套算法引起了世界上最多的死忠粉丝群——披头士狂的躁动。

杰森·布朗不是一般的披头士粉丝。他能将音乐的方方面面与背后的数学原理联系起来，他的专业论述在全世界范围内被广泛引用。

布朗凭借着自己的数学和编程背景，开始思考几十年来围绕披头士的文化遗产所产生的一系列著名疑问，这些疑问以一种独特的声音开始：《A Hard Day's Night》的开场和弦。

《A Hard Day's Night》的起始和弦或许是摇滚史上录制的认知度最高的和弦弹奏。虽然它仅持续了一秒多，但即便是漫不经心的广播听众也知道接下来会发生什么。它是披头士创作的最难以磨灭的声音之一。长期担任披头士音乐制作人的乔治·马丁承认了此和弦的效果和重要性。

但是任何出色的吉他手都会告诉你，不可能还原开场和弦。不同的乐曲书为这首歌的开场一枪列举了不同的和弦和音符，但它们全都错了。

有不少理论把它归功于哈里森在 1964 年披头士首次在美国巡演时得到的一把 12 弦吉他 Rickenbacker，但这把吉他并不能说明问题的全部。很多摇滚迷说他们听出和弦声里有钢琴声，还有人说听到了约翰的 6 弦吉他声，甚至还有保罗的贝斯低音。

哈里森 2001 年去世前接受雅虎聊天室的访谈时提供了一半的答案。

a\_t\_m98：哈里森先生……请问在《A Hard Day's Night》中，你用的是什么开场和弦呢？

乔治·哈里森：那是 F 调还有最上面的 G 调（12 弦吉他上的）。

乔治·哈里森：但是你得问保罗他弹奏的贝斯音符才能全盘通晓。大多数歌迷都没法接触到保罗·麦卡特尼，所以这个谜团仍然未解。

多年以后，身为达尔豪斯大学数学教授的布朗利用傅里叶变换构建的算法解决了披头士和弦之谜。

1807 年，约瑟夫·傅里叶发表了他的热传导方程。周期函数的傅里叶变换是基于欧拉和丹尼尔·伯努利的一些研究成果发展而来的一系列三角级数，现称为傅里叶级数。傅里叶方法被证实也可用于解码音乐。20 世纪 90 年代中期，布朗无意间在一本数学学刊上看到这一实例，他对此非常感兴趣，下意识地记住了这个小点子。

与其几个小时抱着吉他沉浸在还原和弦声的幻想中，布朗想，为什么不用科学解决这个难题呢？要理解布朗试图做的事，就得初步掌握音乐和声音背后的科学原理。所有声音都是以音调为构成要素。音调都有一个频率和振幅。频率与音高有关，振幅与音量有关。音调组合起来就构成了和弦。要想得到一个和弦的数学函数，只需对其内在音调的函数求和。音调的频率与振幅可用正余弦函数实现数学建模，这正是傅里叶变换所使用的方法。

录制乐器的演奏时，录进的音不仅包括乐器的纯音调，还有所有乐器都会在发声后发出的小音量余音以及当时反射到录音筒的噪声。布朗利用傅里叶变换的算法分析和弦。

布朗首先把他能找到的杂音最少的歌曲样本传入电脑，然后从大概维持了 1 秒的开场和弦中间剪切了一段声音。这短暂急促的谐音包含了他所要知道的所有信息。他将这段谐音输入专门为解码音乐构建的傅里叶变换算法中运行，和弦被分解为 29375 个不同的音频。

布朗从中筛选出音量大于.02 的音频，这样就把音频数减少到 48 个。布朗又构建了第二个算法，把音频转化为披头士乐队弹奏的音符。

布朗开始分析手上的 48 个音符。

有一个音符 D3 弹出的声音几乎比其他任何音符都高出一倍。布朗曾一度认为哈里森用到他的 12 弦吉他，因为在这首歌后面确切无误可以听到 12 弦吉他声。但是吉他琴弦大小一致，发出的音符音量也大致相同，也就是说不可能弹出超音量的 D3 音符。同样的原因，这音符也不会是列侬的吉他弹出的，只能是麦卡特尼的迷你贝斯发出的声音。

布朗得到的数据里一共有四个 D3 音符，所以还有三个音量稍小一点的没找到发声源。即便有一个来自哈里森的 12 弦吉他，另外一个来自列侬的 6 弦吉他，也仍然还有一个多余的 D3 音符。恼人的是 D3 音符上还飘荡着三个 F3 音符。如果哈里森用 12 弦吉他弹出了其中一个，那他的吉他也会产生 F4 音符——但是根本没有任何 F4 音符的痕迹。

布朗突然想到这 3 组音符都是钢琴而非吉他发出的。任何看了大钢琴内部构造的人都知道，钢琴靠的是琴键拉动弦槌敲击琴弦来发声。实际上，3 根琴弦发出的是高音符（2 根发出中音符，1 根发出低音符）。这是一个高音符，所以钢琴琴弦可以解释为什么同一音符会以三种音量一组的方式出现，以及为什么没有 F4 音符。布朗现在知道了弹奏开场和弦的人还有乔治·马丁，乐队使用主乐器的同时马丁用钢琴伴奏。

布朗的算法分析出哈里森用 Rickenbacker 弹出来的真正音符:A2、A3、D3、D4、G3、G4、C4、C4。列侬在他的 6 弦吉他的上弹奏了高音量的 C5 音符，马丁则在钢琴上弹奏了 D3、F3、D5、G5、E6。

有了这一背景知识，哈里森在 2001 年网上聊天时的评论也就更容易理解了：“那是 F 调还有最上面的 G 调（12 弦吉他上的）.....但是你得问保罗他弹奏的贝斯音符才能全盘通晓。”

那的确是 F 调音符上面还有 G 调音符。正如哈里森在屏幕上打出的答案，12 弦吉他弹奏了 G 调音符。正如哈里森希望的，大多数人想当然地以为吉他也弹奏了 F 调音符。但是 F 调音符实际上是马丁的斯坦威钢琴弹出的，而保罗弹的则是另一个全然不同的音符。布朗说，马丁在混音时把钢琴声的音量调到和吉他声的音量一致，这样就把钢琴声隐藏在更尖锐的吉他声里了。

对布朗来说，解析《A Hard Day's Night》的开场和弦还不算完。乔治·哈里森在这首歌里用吉他弹出了堪称完美绝妙的跳音独奏，完美得有点不可思议。布朗一直觉得吉他独奏有可能是以半速录制的，然后在最终混音时加快了速度。

众所周知，那段吉他独奏实际上是哈里森用 12 弦吉他和马丁用钢琴模仿哈里森同时弹奏的。如果原本真是以半速录制的，那么他们弹奏的时候应该要降一个八度，因为把磁带音速提高一倍会使音乐升高一个八度。

因为有钢琴的参与，布朗有个方法可以一试。他取样一小段独奏，输入算法运行。他选取的这一小段独奏包含一个哈里森的吉他和马丁的钢琴都弹奏的 G3 音符。布朗检查录音里表现出来的频率时，发现 G3 音符的确有三个音频。其中一个是哈里森的 12 弦吉他弹奏出来的，另外两个来自钢琴。这就有了意外收获。如果这段独奏是以全速弹奏的，那么钢琴应该弹出来三种音频的 G3 音符。

前面提到过，钢琴的高音符有三根琴弦，弹中音符时弦槌只需击两根琴弦。如果这段曲子是以合适的八度全速弹奏的，那么马丁应该击打三根弦弹奏 G 调才对。然而如果在钢琴上降一个八度，则只会产生两根弦发声的 G 音符。这就是答案——布朗无可辩驳地证明了录音带的速度的确提高了一倍。

后来布朗回放了英国广播公司 1964 年播出的一档叫做 Top Gear 的节目，那是披头士《A Hard Day's Night》在录音室录制完成三个月后的现场表演。布朗很早以前就注意到节目里有一处不太一致的地方：就在乔治独奏之前，视频在明显的胶接处有短暂的停顿。布朗说，现在看来好理解了，原因就是哈里森那个时候还不能以常速表演独奏，所以这档节目在那一部分用了配音。

布朗迅速地指出，处理过的作品丝毫没有贬低哈里森的能力。实际上哈里森在 BBC 这档节目播出后几个星期就已经掌握了全速独奏。布朗在一篇数学期刊上写到：“哈里森知道很快他将登台演出，在全世界的注视下全速表演吉他独奏，他还是有胆量以半速录制如此高难度快节奏的独奏，这是他的荣耀（也令披头士迷很高兴）。”

布朗现在着手解决另外一个披头士之谜，这涉及比较敏感的领域。很多狂热的披头士迷都知道麦卡特尼和列侬都曾经宣称《In My Life》大部分是自己写的。

为了查明谁更有可能创作了这首歌曲，布朗用算法工具分析了麦卡特尼和列侬单独创作的歌曲。基于输入的歌曲，算法分别构建了两位作曲家风格特征的基本结构。

布朗做这项工作时用到了图论。布朗输入指令到算法，把重复副歌定义为点，之间的主歌定义为线。

作者写作本书时，布朗还没有把《In My Life》输入算法。从这个项目耗费的时间来判断，也许布朗只是不愿意侵入两位披头士成员的争议领域，况且其中一位尚还健在。

## 第 4 章 计算机的秘密高速公路

作者认为，算法的价值全部体现在它的速度上。

能够预测你可能喜欢看什么电影的算法之所以目的明确，是因为它正确、快速地衡量了成千上万个输入因素并反馈给你分析结果。

在算法化的计算机世界里，用最少的时间完成最多的任务称为消除延迟。

20 世纪 90 年代，华尔街到处都是黑客人才，越来越多的机构运用了相似的策略和算法。既然在战术和代码上不能做到与众不同，交易机构就在硬件上精益求精。



作者举了《Spread Networks 的秘密高速公路》的例子。2008 年，美国一家对冲基金公司请丹尼尔·斯皮维开发一种算法交易策略，搜索芝加哥股指期货与其纽约标的股票证券的微小价格差。如果芝加哥交易市场的期货价格高于纽约交易市场其所有标的股票，交易策略就是卖出芝加哥交易市场的期货，买进纽约交易市场的股票。如果两个市场价格有所调整，斯皮维设计的程序就会平掉头寸、锁定利润。

算法交易不断发展，人们对任何低风险套利交易机会都趋之若鹜，这就是此交易策略对速度有要求的原因。如果能够抢得交易先机，就很容易赚钱。那些即便是落后了 1 毫秒的人都分不到一杯羹。

斯皮维为对冲基金公司构建的算法要能正常执行，需要利用芝加哥与纽约之间最快的光纤传输网络。但不论是斯皮维还是对冲基金公司都无法获取光纤网的带宽，因为它已经满载且很少有空闲。

斯皮维可用的传输方式是亮光纤，但斯皮维需要的是暗光纤。

如果没有暗光纤，他和他的对冲基金公司客户就不能与使用相似交易策略的投资公司和交易员抗衡了，他们的策略以及策略背后的代码也就没什么用了。

斯皮维一头钻进了通信学、土地权和高速电路的晦涩难懂的世界。

他实地考察了自己想象中的线路途经之地。他了解得越多就越相信，在芝加哥和纽约之间建立一个全新的私有的暗光纤管道是可行的。

最重要的是斯皮维设计的线路比现存最短的光纤线路往返距离少 160 千米，虽然光在光纤内行进此段距离只需不到 4 毫秒。“这几乎是自动化交易的天堂，”伊利诺伊理工大学的教授本·范·弗列特说，“这就好比捡地上的金币——动作最快的人才能得到金币。”

斯皮维需要有钱有势又愿意承担风险的人与自己一起运作这个投资项目。他首先想到的人是他的密西西比老乡詹姆士·巴士德。巴士德有着辉煌的履历，他 20 世纪 80 年代任联邦快递首席运营官，也出任过 AT&T 无线的首席执行官，并帮助公司发展成为美国移动运营商老大，而后他被硅谷有史以来孕育的最传奇的网景公司任命为首席执行官。

巴士德同意担任此项目的首席融资人，他的儿子大卫成了首席执行官，斯皮维担任董事长。暗光纤运营商 Spread Networks 诞生了。

Spread 的光纤线路绵延 1320 多千米，是连接长途电话代码 212 区域和代码 312 区域最短的路径。信息以光速通过此线路，往返纽约和芝加哥耗时 13 毫秒以内，比以前路径的最快耗时少了将近 4 毫秒。

作为最低延迟的通信线路运营商，Spread 收取的费用是现行价格的八到十倍，高达 300 万美元一年。一根一寸粗的光缆大约能承载 200 位暗光纤客户的通信量。

Spread Networks 不过是早在 20 世纪 80 年代就打响且不断升级的硬件之战中出现的最新、最先进的武器。算法成为执行复杂策略的媒介，而运行算法的硬件的发展也使算法的覆盖面和影响力不断扩大。

算法下一步将入侵哪个领域由两个简单的变量决定:破坏的可能性和破坏产生的回报。

作者提出,走在最前面的人就占据了赚钱的优势地位,而有法子走在最前面的人已经就是富人了。这并不是最近才出现的资本主义的瑕疵。

科技总是由享有特权的少数人发展起来的,然而创造最多财富的是普及科技的人。

## 第5章系统里的博弈论

作者说到,1989年IBM批准了一批科学家着手构建一台计算机,声称要打败国际象棋世界冠军。这台计算机最终取名为深蓝(Deep Blue)。1997年,深蓝在其鼎盛时期打败了国际象棋大师加里·卡斯帕罗夫。那时它平均每秒可以权衡两亿步,而卡斯帕罗夫每秒只可以想三步。

2011年年初,IBM的最新发明——沃森(Watson)在智力竞赛节目Jeopardy中全胜人类选手。沃森可以运用600多万条逻辑规则得出一个答案。

在华尔街工作的人都对扑克游戏乐此不疲,因为从很多方面来看,扑克就像是他们生活的一个缩影。它包含了华尔街两大最受重视的特质:刀光火影间的虚张声势和电闪雷鸣般的快速分析。

扑克游戏也充斥着真真假假、隐情内幕、见招拆招、敌进我退,也有高明和愚钝之分。而在众多影响因素中,人的非理性是最危险的。

会利用人类心理学是世界级的扑克玩家最重要的技能。对于高手们而言,扑克是人类有史以来发明的最微妙、最具人性的游戏之一。作者谈到,博弈论是应用数学的一个分支,主要用来解决包含未知因素和多个利益方的复杂难题。

不可预测的人的因素使得扑克游戏成为只有最出色的算法创造者才能成功的领域。

人的本性就是在任何时刻都能做出不理智的行为,而在参与人数众多的扑克比赛中,没有任何算法能够替代丰富的经验和敏锐的直觉。

预测人类将何种看似不合理的行为视作最符合其利益之举,这是算法需要跨越的最后一道坎。作者举了《中情局的最佳员工》的例子。桑德霍姆的算法在扑克领域的突破为算法攻克

无数其他比游戏更为重要的生死攸关的难题打开了一扇门。对于美国安全部队来说,追查恐怖分子下落或是制定国防力量部署方案时,未知的幽灵也给他们带来了许多难题。打击恐怖主义往往要付出高昂的代价,但是博弈论算法正在改变这一事实。桑德霍姆的扑克机器人给了南加州大学计算机科学教授米林德·塔姆贝很多灵感,他想到利用算法抵御恐怖分子的威胁。塔姆贝向桑德霍姆的扑克机器人学习了几条策略,为洛杉矶国际机场构建了一套破坏恐怖分子阴谋的最佳巡逻模式算法。这套算法能够检测到恐怖分子自认为在机场制造死亡、破坏和混乱等恐怖事件的“最佳方法”,然后基于对恐怖分子想法的预测做出摧毁攻击性计划的决策。

2007年，洛杉矶国际机场开始启用该算法，其效果深深震撼了联邦政府。联邦政府请求塔姆贝博士为美国交通安全管理局构建一个在全国范围内发布巡逻命令的计算机程序。与之相似的计算机程序于2011年在匹兹堡国际机场进入试运行阶段。美国交通安全管理局开始使用这套程序来确定飞机上空军中将的人员配置和嗅弹警犬部队的巡逻次数。这样的结果掀起了研究算法的浪潮，人们希望有朝一日能通过一套程序在像阿富汗这样的地方做出关键性战场决策，设计出能降低敌人伏击可能性的护航路线、临时基地和用兵战术。

但早在博弈论算法开始保卫机场、巡查班机以前，美国中央情报局就把它当作自己的高级智能分析师了。20世纪90年代早期，苏联解体之后的两年里冷战有所缓和，美国和俄罗斯情报组织的关系友好。美国中央情报局和以前的克格勃就分享信息和情报系统诚心诚意开展了讨论。俄罗斯人对美国人夸夸其谈的一件东西非常感兴趣。那是一套极为复杂的博弈论算法，它准确预测了反对戈尔巴乔夫的军事政变将会触发苏联的轰然倒塌。

这套算法是由纽约大学政治学系教授布鲁斯·麦斯奇塔和斯坦福大学胡佛研究所的一名高级研究员发明的。

在预测紧张的政治局势将会如何演变发展时，需要考虑的因素成千上万。福克斯新闻或CNN的优秀评论员通常会根据所了解到的关于局势变化的信息，并查阅历史相似事件，然后依靠直觉做出预测。中央情报局的分析员也如此工作，只不过他们掌握着更多的信息。但说到底，情报分析员在某种程度上做出的预测依然是基于主观判断的。

麦斯奇塔发明的算法工作原理有所不同。美国中央情报局聘请他对全球不同地缘政治形势做了1700多个政治和军事预测，发现这位使用博弈论算法的教授的正确率比他们自己的专家高出一倍。

“当我们依靠直觉，就会遗漏太多可能性，”麦斯奇塔说，“数学能给我们答案。”

在预测一个国家政治博弈的结局时，麦斯奇塔的算法需要考虑所有与此相关的各方利益。算法假设的唯一前提就是绝大多数相关人都会为自己的最佳利益奔走游说。也就是说，这件事对于每一个相关人的重要性也是起关键作用的一个因素。也许最高领袖对某一事件有自己的倾向，但整件事的结局对他来讲无关紧要，所以与结局更相关的低级别人员就会施展更多的影响力。因此，要想真正运用算法预测地缘政治，使用者仍然需要掌握情报资料，但他们只需要某些特定的情报。他们需要知道哪些政治团体或个人影响着决策，以及对每一个相关人来说何种决策是最符合其利益的。此外，了解每一个相关人的影响力以及每个人对这一特定事件有多上心也是很有必要的。每一相关方都希望将自己的利益最大化，损失最小化——能够达成这一目的的行动正是他们会采取的行动。

此前，情报部门耗费巨资搜集到的大部分信息，只不过是在分析一些人为何做出了某种选择。他们在意的信息包括每一个相关人的历史、私事和他走到当前职位所历经的过程。但是麦斯奇塔教授说这些情报无关紧要，他的算法能够证明这一点。“你走进一间房，看见里面正在对弈，只要看一眼棋

盘，你就能迅速知道每一个棋手接下来会做什么，”麦斯奇塔说，“他们怎样走到那一步不重要——现实世界的事也是如此。”

如果我们知道每个相关人的动机、影响力和兴趣，就能启动算法工作了。这正是麦斯奇塔的工作流程。他每日都在设置变量、相关人信息以及他们在意某一特定结果的原因，在此之后的工作就交给算法了。

至于伊朗制造核武器的可能性有多大，麦斯奇塔在 2009 年设计了一套算法预测这一事件。机器程序考虑了每一个因素：相关人、相关国家、可能的制裁和伊朗科学家的水平。机器断言在 2014 年伊朗会制造出武器级铀，但算法同时预言伊朗不会用铀来制造核弹。算法表示伊朗的重要决策人只要能让人们意识到他们有能力、有资金开发核燃料就心满意足了。面对拥有核武器带来的弊端——国际制裁、来自以色列的军事打击、石油卖不出去——他们会清醒地权衡利弊。麦斯奇塔说算法大概已经意识到，根据核不扩散条约，伊朗可以开展一些核电项目，依靠这些项目伊朗也可以获得类似的诸多好处，例如增强影响力、受人尊敬以及成为地域权威。这一结论现在看来准确无误，比大多数中情局和传统分析员的预测都要好很多。

绝大多数中情局从业人员都是直接从大学招聘的。因为旧式研究方法规定分析员应该对其研究的地域人文有深厚的了解，所以大多数新成员都毕业于文科而非理工科。因此，很多情报员都致力于研究个人历史以及围绕权力人物上演的戏剧性事件。这些事情谈论起来也许很有趣，但要预测一个遥远国度的人最终会做什么事，它们并非是至关重要的因素。真正重要的是分析出怎样做才符合掌权人的最佳利益。然而，我们的情报员却只是在文化差别、宗教冲突，以及种族、国家和民族之间主流意见的差异上狠下功夫。相比之下，算法编写更为便捷。如果重量级相关人和他们的利益能被清楚了解、明确定义，那么即便是分析预测伊朗这样的领导层不稳定的国家，算法也能比人类分析员更准确地预测他们将会怎样做。

在 20 年总结的硬数据面前，再强势的教条也能被推翻。中情局和国防部中已经有人开始推动机器人在华盛顿外交机构担任更重要的角色。

2010 年 5 月 5 日，麦斯奇塔告诉一小部分人他的算法已经预测埃及总统胡斯尼·穆巴拉克会在一年之内倒台。后来，这一预测果然在 2011 年 2 月 11 日成为现实。

作者提到，电脑红娘的算法之所以会失算，原因在于终结一段关系的问题只能在两人相处的过程中暴露。人们在压力下如何交流、如何携手解决难题更能说明一段关系能否长久。

## 第 6 章 呼叫机器人医生

作者写到，一些医院在诊断入院病人时已经开始利用算法来进行初步判断。这样的医院虽然不多，但数量正与日俱增。在临床诊断方面应用算法的好处越来越明显，尤其是那些将算法纳入常规流程的

医院。需要处理的医学并发症减少了，初次诊断的正确率提高了，病人的死亡率降低了，整体费用也随之减少了。

算法挽救了更多人的生命并且促成了更多的器官移植手术。看到这种情况，大多数公众可能都愿意接受算法进行器官分配。但是，当改变波及他们自己的医生和常规医学治疗时，人们就没那么信任算法了。在医疗保健领域，即便是由于虚假消息、个体的贪婪、恐惧或其他情况引起的抱怨，通常都异常尖刻。所以很难想象，如果健康保险公司建议用算法对身体的持续监控取代现实生活中由人类医生做身体检查，将会引起怎样的轩然大波。在某种程度上，算法一定会取代医生，不过这一前景并没有听起来那么可怕。然而只要一谈到健康领域的问题，很多人的态度总是趋向于：改变是糟糕的，不改变是好的。

所以，人们抵触利用计算机算法进行医疗诊断和治疗，这一态度是可以预料到的。

医疗保健业充满了算法能够轻易攻破的领域。首当其冲会被改变的就是我们检查身体的方式。宫颈抹片、X光成像、磁共振成像以及CT扫描等的高频率、高复杂度以及高费用是导致过去20年医疗保健费用激增的原因。如果人们身体不舒服去看医生，通常都不得不经受一系列的身体检查。虽然某一项检查有99.9%的可能是不必要的，但通常都会被强制执行。因为这些检查能源源不断为医院带来滚雪球般的收入。

是时候利用算法改善医疗系统了，而实现这个目标的第一步就是要让算法成为所有身体检查默认的扫描仪、观察者和分析师。曾经需要极高收费的放射科医师或病理学医师参与的身体检查，算法在将来也能担此重任。实际上算法会做得更好。

控制美国医疗机构检查费用的第一步就是将分析检测结果的任务外包给训练有素的印度医师，而第二步——也是最后一步——答案就是最终将所有工作交给算法。经过完善的算法会比印度、美国，或世界上任何地方的医师都做得好。

作者举了《机器人药剂师的优势》的例子。解读扫描报告、检测结果以及患者元数据的工作人员必然会受到威胁，同样受到威胁的还有为患者制定药物剂型与剂量的药剂师。药剂师的工作需要注重细节，压力很大。另外，药剂师的工作内容容易量化，这意味着他们的工作很容易被计算机入侵。2011年，加州大学旧金山分校开办了一家仅由一台机器人管理的药房。

机器人直接接收往返于医生办公室和药房的电子讯息指令，然后用它颇长的手臂从镶嵌在特殊墙面上的成千上万个药箱里取药并包装好。它还会接收到关于患者的所有信息，然后利用身体内部的算法迅速检查新处方是否有药物冲突或引起并发症，以保证患者服用的药物不会产生对人体有害的交互作用。加大旧金山分校的机器人算法能利用制药公司的电子通知直接更新药物信息。而且与人类不同的是，机器从来不会忘记任何一件事。这并不是说它从不犯错——写得再好的程序也会出故障。但是故障很好修复，而且算法能够对分发药物的机器运行并发测试，增强机器人药剂师的安全冗余度。

旧金山的这个机器人已经毫无纰漏地成功配药 200 万余次。药物包装过程不涉及人类接触，从而杜绝了药物污染的可能。

那么人类药剂师在这一领域干得如何呢？现存的研究结果表明配药错误率高得吓人，其幅度为 4% 到 10%。业界可以接受的保守值是 1%。美国药剂师协会对 50 家制药公司进行了一项全国性调查研究，查出配药错误率平均值为 1.7%。美国每年需要配药的处方单子有 37 亿张，照最保守的统计，每年会有 3700 万例配药错误事故。

由于药剂师从业人员在全国范围内都比较紧缺，这又导致了药剂师工作量大，人员配备不足，出错率更高。因此，这一岗位需要算法驱动的机器人来挑大梁已是无可争辩的事实。挑选药物、交叉检查药物交互作用、保证剂型与剂量完全正确，这些都是算法驱动的机器人可以掌握的技能。况且，很少有如此容易被量化的任务。让算法担此重任也有其商业意义。

## 第 7 章 人的分类

作者表示，在真正的专家眼中，每个人都可以像动物园的动物一样被分类、加标签并编成目录。算法通过解析我们的性格、分辨我们的动机，已经对我们中的很多人了解得很清楚。

我们几乎都被机器解读分析过，虽然大多数人当时并不知情。你可能对下面的话非常熟悉：“为了确保服务质量，您的通话可能会被录音，请您谅解。”不管是银行、信用卡公司、航空公司还是保险公司，只要我们拨通其客服热线，听到这熟悉的措辞，我们就默许了算法监听我们的通话。

真正发生的是机器人程序在窃听你的谈话。它听你说话，分析你的性格类型，判断你为什么要打这个电话。在某些情况下，它会将这些信息转给与你交谈的客服代表。最令人难以置信的是它试图读懂你的想法。一旦机器人程序想了解你——它能在 30 秒内读懂你——当你下次再打来电话的时候，它就会把你的电话转给与你具有相同性格特征的客服。呼叫者与客服的错误配对会增加陷入长时间争论的风险。把性情相投的客户与客服连线在一起会使通话时间更短，客户满意度更高，同时能为公司赚取更多的利润。

作者举了《阿波罗 13 号的运气》的例子。1970 年 4 月 11 日，三个人坐上了高 110 米的“土星 5 号”运载火箭。火箭里装满了低温液氧、液氢，这些燃料一燃烧就会推动火箭上升离开地球大气层。然后助推器会再次点燃，推动火箭离开地球运行轨道到达“阿波罗 13 号”的目的地：月球。宇宙飞船在飞行途中进行常规维护时，加热系统的恒温器开关产生电弧放电作用，瞬间点燃了一个液氧贮箱并引起爆炸，半个飞船都被烧毁散落到宇宙深处。

指令舱的氧气很快就所剩无几，三人进入了唯一能够活命的地方——登月舱，它永远也到不了它本来要去的地方了。登月舱是按两位宇航员生存 45 个小时设计的，但现在却不得不为三位宇航员提供 100 个小时的生存保障。为了保存足够的电力资源来活命，三人关掉了登月舱的大部分电源。

休斯顿地面指挥中心指挥宇航员绕过月球的背面，利用月球引力将航天器送上返回地球的轨道。然后宇航员小心翼翼地启动了登月舱的发动机，圆满完成了轨道修正，这的确是一大技术壮举。

这不过是休斯顿指挥中心设计的众多绝妙工程学特技之一。由于多出来的一个人呼出的二氧化碳超出了氢氧化锂过滤器的极限值，登月舱的二氧化碳浓度之高曾一度危及宇航员的生命。要是没有应对方案，宇航员有可能在自己呼出的二氧化碳中窒息而死。被损毁的服务舱配的过滤器因形状不同不能用到登月舱上。指挥中心指示宇航员用船舱里可用的东西——一个塑料袋、飞行计划书的封面、一卷胶带——做了一个临时应急装置。这个连接到过滤器其中一个氧气管的装置神奇地奏效了，登月舱二氧化碳含量下降到非致死浓度。

两天后，三人降落在南太平洋，这证明了休斯顿指挥组团队的能力，也证明了宇航员们自己的能力。同样不可思议的是三人在压力下的忍耐力和凝聚力。整整 100 个小时，三个人待在为两个人设计的船舱内，完成一件又一件生死攸关的任务。每个人在极端环境下的反应不同。哪怕绝境近在眼前，有些人都能控制自己的感情，冷静地开展手头的任务。然而我们中的大多数人，即便是宇航员，在面对可能的死亡时都难免惊恐万分。要么逃避，要么什么也不做，或者表现得毫无理智。

要预测一个人的反应很困难，更难预测的是在如此孤立无援的环境下，人们如何协作共处。在“阿波罗 13 号”这一意外事件上，美国航空航天局小有运气。斯威格特是在飞船发射前一天加入此次飞行任务的，他接替了飞行前一周患了麻疹的肯·马丁利。这个三人小组还没在一起协作多长时间就被推射到了外太空，但是美国航空航天局对这三人将怎样协作略有所知。评估宇航员性格相容性的太空项目已有将近十年的历史了，这个心理评估项目对美国在太空竞赛中领先苏联起到了至关重要的作用。

美国航空航天局非常清楚，不合适的宇航员组合不利于任务的执行，甚至会危及人员的性命。这就是为什么早期的宇航员要直接从军队选拔，候选人在那里经历过战斗，承受过高压，也面临过死亡。

对候选人在枪林弹雨中表现出来的品质做同行审查和背景调查是选拔流程的重中之重。

## 第 8 章 华尔街与硅谷的较量

作者指出，2000 年，通过计算机程序交易的比率不足美国股市交易量的 10%，人们控制的是交易厅而非电脑。纽约证券交易所的所有交易单都要经过专家经纪人之手，专家经纪人享有查看某一股的绝大部分交易单的特权。交易者在交易厅的下单速度受限于专家及其下属处理交易单的速度，整个交易系统完全是由人操作的。那时统治华尔街的是交易厅，而不是电脑主板或光纤的光脉冲。彼得菲发明的交易系统和模仿彼得菲的山寨系统占领了纳斯达克，但是在纽约证券交易所，扯着嗓门喊单仍然胜过向电脑输入交易指令，更别提任何全自动交易系统了。

然而这一切正在改变，电子交易网络的兴起使得如彼得菲一样的交易者和程序员能大举进攻以前被小集团少数人统治的市场。一个有交易经验的程序员凭借稳定的网络连接、强大的台式电脑和漂亮的代码，就能迅速进入股市交易。

除了更加便捷的网络连接，华尔街所谓的十进制报价方式也让电脑黑客们获得了赚钱的巨大良机。在实行十进制报价方式以前，美国股市的报价单位是以美元的几分之几来计的。在 1997 年以前，最小报价单位是 1/8 美元，也就是 12.5 美分。后来又调整到 1/16 美元，即 6.25 美分。这一价差仍然能让有地位的经纪人和做市商大量敛财，他们所要做的仅是根据投资者的报价来买入和卖出。分数制报价方式以其他所有人尤其是小投资者的利益为代价保证了经纪人的利益。

2001 年，美国证券交易委员会要求所有市场采用十进制报价制度，很多交易量大的股票价差变为 1 美分。现在做市商仅仅靠客户很难赚大钱了，而需要极大的交易量。只有一种方法可以获得这一切：科技和知道如何运用科技的人。

2008 年上半年，自动化电子交易占了全美股市交易量的 60%。金融行业花了七年时间追逐每一位聪明的工程学毕业生、物理学家和博学通才，只要这些人对高起薪和能够买得起两套房子的巨额奖金稍感兴趣，它也会不遗余力吸引他们投身自己的怀抱。渐渐地，对数学、工程学和理科毕业生来说，华尔街成了比半导体、制药和通信行业更吸引人的地方。

为了保证自己处于行业的顶端，金融机构需要两件东西：华盛顿的朋友和用钱能够买到的最聪明的数量统计方面的人才。他们用钱打通了这两个环节。

作者说到，争取最出色的算法人才的战争扩展到了华尔街的公司之间。能写出最漂亮代码或想出最巧妙算法的人能够自由穿梭服务于不同的竞争对手公司。在这一类人里还有一类超级宽客玩家，他们有着独特而宝贵的技能，能够赚到百万美元以上的年薪。

作者举了《电话销售的金额增长了 86 倍》的例子。算法机器对恼人的电话销售行业进行了改革。对待不同性格的人需要使用不同的销售技巧。对思考型的人不要甜言蜜语，而要把好处、事实和节约的金额说到他们的心上。对行动支配型的人，可能这样就能激起他们的购物兴趣：“现在买的话，就能免费得到赠品！”对于不带个人感情的电话销售来说，诀窍就是知道使用什么样的说辞。指导电话销售人员在合适的时候使用正确的说辞是很有价值的。

沃达丰公司的经历就很能说明问题。它愿意利用美国航空航天局研发并经康威的电子忠诚团队完善的方法来进行销售。沃达丰的营销部有自己原来的一套方法帮助话务员推销升级类产品，其他很多公司采用的方法亦是如此，但大多数销售电话均以失败告终。沃达丰采用了康威团队的方法，根据潜在客户的性格来改变其销售策略。对待情感支配型的人要用甜言蜜语和能让他们与亲朋好友保持紧密联系的产品来迎合他们，对待其他性格类型的客户也有专门的计划。令人吃惊的是，销售额最后竟然增长了 86 倍。



作者谈到，让一套计算机算法调查自己的邮件，这种做法不会是大多数员工所乐意接受的。但全球最大的电子邮件平台谷歌邮箱在用关键词搜索我们的邮件并向我们发布针对性广告时，也是在做同样的事情。不仅如此，法律允许雇主利用其员工的公司邮件账号做任何他们想做的事——阅读、删除、转发、公布。在这个世界上，灵活巧妙地利用算法的人积累了越来越多的权力。未来我们将任由计算机程序来评判我们的性格癖好、决定我们的电话连线对象以及衡量我们的能力与价值。

康威的技术理论上会限制我们所交谈的人的类别。可能我们只被允许与有相同性格类型的人交流。这也许能为公司带来更好的盈利，但康威的机器所具有的能力再加上现有科技给我们带来的强行分类却使问题变得更加尖锐。

从表面上看，上述情况可能还不错，但它会导致狭隘、极端及立法僵局。科技为我们构建了一个属于我们自己的舒适世界。我们可以躺在相似的思想、灵魂和观点的安乐椅上。如果一个思想支配型的员工只能与具有同样思维模式的人接触，那他要怎样面对一个情感型的新同事呢？一个员工在他的岗位上可能是全国数一数二的人才，但如果他不能与不同性格的人有效交流，又怎么称得上是一位专业人才呢？和不同类型的人合作共事已然是我们日常生活中要面临的挑战之一，但它同样也会让我们受益匪浅。康威的任务并非是要破坏人际关系，但随着机器逐渐掌控人类交流的不同途径，我们必须承认这可能会给我们的世界带来奇特的影响。

## 第9章 华尔街的损失让我们大家获益

作者举了《Facebook 的粘性》的例子。马克·扎克伯格的故事广为人知。使用 Facebook 的不是科技宅和金融控，而是非常普通的人。他们中的很多人都沉迷于 Facebook 上不断更新的动态、新闻和聊天主题。

2006 年春天，扎克伯格通过哈佛一个朋友的介绍找到了数学能力极强的哈默巴赫尔。一周后，搬到加利福尼亚的哈默巴赫尔成为 Facebook 最初的 100 位员工之一。扎克伯格给了他一个漂亮的头衔：研究科学家。哈默巴赫尔的工作就是用数学和算法分析不同年龄、性别、地理位置和收入的人使用 Facebook 的方式，以及为什么 Facebook 在某些地方发展得很好，而在另一些地方却遭到惨败。

扎克伯格选择哈默巴赫尔是因为公司急需着手处理快要超过其存储容量上限的成堆数据。这个刚刚起步的公司明白数据的价值，但却不知道如何处理这些数据。而华尔街近 20 年早就知道如何处理海量数据了：存储、分类和分割，寻找可以用来牟利的模式、异常和趋势。

但华尔街任何一个数据处理高手都从未遇到过哈默巴赫尔面对的庞大数据。Facebook 拥有近十亿用户，针对其中的每一位都要存储 1000 页数据，包括电脑型号、政治观点、恋爱关系、宗教信仰、最新

地址、信用卡信息、工作情况、常用链接、家庭成员以及最占空间的照片。这还不包括我们的浏览习惯或 Facebook 上大量的广告版块。

哈默巴赫尔把在华尔街所学到的一切都带进了 Facebook 的硅谷办公室。他组建了一支人才济济的队伍，这支队伍由像他一样经验丰富、技能高超的人组成。有些人是从摩根士丹利、骑士交易集团和高盛等有金字招牌的华尔街公司挖过来的，他们厌倦了华尔街。Facebook 需要知道怎样最佳化页面排版，怎样把人们联系在一起，最重要的是怎样让人们的在线时间更长。

哈默巴赫尔构建了工具和算法来监控每小时源源不断涌入 Facebook 的难以想象的海量数据。Facebook 对上网的人来说之所以有那么强的粘性、那么的不可抗拒，部分原因是哈默巴赫尔构建了跟踪用户鼠标点击的系统。系统记录下用户的光标会在什么地方停下，什么样的页面排版吸引的人最多、占用的时间最长。所有的鼠标点击、页面停留和光标游走信息都会受到过滤筛选和仔细调查。最吸引人的东西会被保留下来，其他的被扔进数字垃圾箱。

Facebook 称它的一半用户每天至少登录一次。除了电子邮件平台没有其他网站敢如此宣称。事实证明这家社交网站是如此让人上瘾，以至于有一种以充满蓝色液体的透明注射器装饰上白色 Facebook 商标为特征的文化基因开始在网上传播。有了将近十亿人次的被俘虏受众，就能更便利、更有针对性的根据受众群体的性别、收入、地理位置及更多因素来打广告，获利自然更加丰厚。

Facebook 能把用户的癖好、关系网、愿望以及兴趣爱好提供给它潜在客户。精准的广告投放已经成为了 Facebook 的业务发动机。从另一面来看，这可能也是导致像哈默巴赫尔这样的华尔街数量分析专家离开的原因。哈默巴赫尔全然投入连接全球的 Facebook 理念。但在他眼中，即便是这项工作也正在变得越来越不重要。因为很多最聪明的科技界同仁将大部分时间都花在了设法让人去点击互联网广告上，哈默巴赫尔为此感到痛惜。

作者说到，华尔街的存在就是为了做一件事：尽可能地介入每一项金融交易。

在 Facebook 时，哈默巴赫尔已经积累了一定的财富和远见。他退回来审视科技界，思考到底什么才能真正地改变世界。他意识到，答案就是一种挖掘由数字化和计算机化的世界产生的海量数据的通用铁铲。哈默巴赫尔想要帮助网站、公共事业公司、生物学家和医疗服务人员分类整理并从容解决棘手的大数据问题。领悟到这个答案后不久，他就建立了 Cloudera。

作者谈到，世界上最重要的发明通常并非来自通用电气和微软公司，而是来自专注于某一领域的企业家。这种专注程度是大公司无法具备的。绝大多数大公司认为提高已有产品和既定流程的效率是最值得花时间做的事。这是因为，如果利润额为 1000 亿美元的公司效率提高了 5%，那么净收益也就轻轻松松增加了 50 亿美元，然而几乎没有一件新产品能达到同样的效果。因此，大公司易落入管理过度、创新不足的窠臼。这也就是为什么初创公司如此重要的原因。

## 第 10 章 未来属于算法和它们的创造者

作者说到，朋友之间总是有等级排序的。等级最高的是那些能左右其他人观点、具有影响力的人，等级较低的则是追随者。在工作或家庭环境中亦是如此。有人发号施令，有人听从命令。每个人在某种程度上都懂这个道理，而且绝大多数人都对自己适合的角色有一定的认识，但这种认识很可能与事实不符。你也许感觉自己是个领导者，但实际上你是个听差的。你的朋友们知道这一点，而你并没有意识到。有些人凭借他们的地位、金钱、长相或是仅仅凭其知名度飘到了等级秩序的最上层，这些人是社会上真正有影响力的人。有时候你甚至一眼就能看出谁是这样的人，但通常没那么明显。

一个人的说话方式能透露出他所处的权势等级。影响者很少改变他们的言语模式，而被影响的人经常调整其说话方式以靠近影响者的风格。电子邮件等书面交流也同此理。言语暴露了一个人有无影响力或影响力的大小。决定风格的人处于等级秩序的最上层，而按照他人的风格打造自己语言的人等级则较低。通过观察一个人如何使用冠词、助动词、连词以及高频副词就可以判断出他所在的等级。不管是演讲还是聊天，语法词汇的使用都能彰显我们的风格。

作者举了《用算法有针对性地开展营销活动》的案例。康奈尔大学的计算机科学系教授乔恩·克莱因伯格编写了一套算法，这套算法在我们的只言片语间寻找线索，分辨谁是真正的影响者，谁具有左右舆论、引领潮流和引人注目的本领。这些人能决定选举结果，也能推动新产品面市。算法监听人们的谈话，辨别谁是掌控局面的人以及其他分别处在哪一等级。它也能很好地分析业务往来的电子邮件内容。一旦掌握了谁是真正的影响者，公司就能有针对性地开展营销活动。在网上，针对有影响力的人制定的横幅和文字广告要比普通广告贵很多，而这些高端广告的受众对此并不知情。企图推销产品的公司可以用所谓的浏览器 cookies 跟踪目标受众的浏览行为，密切关注其动向，从而有针对性地进行广告。营销界深谙此道：这些人是通往其他人内心和思想的一扇门。随着大范围自动识别等级秩序中上层人物的算法的出现，在政界、管理界以及市场营销界，更多有针对性的新方法将会应运而生。

社会等级的定量评估无疑已来到我们身边。无论是营销机构、广告商还是随便哪个人，只要取悦了最有影响力的人物就有大把大把的钱可以赚。未来我们还要面对算法带来的各种各样、严谨缜密和毫无人情的计算。社会等级排序仅仅只是一个开始。

作者谈到，如果你留心算法的动态，就会发现算法已经掌控了你的货币市场基金、股票还有退休金账户。它们很快也会掌控你的电话会连线到谁，什么样的音乐能到达你的无线电台，你获得器官移植的救命机会有多大。对数百万人来说，算法还能帮助他们做出人生最重要的决定：选择伴侣。以前，人类是所有重要问题的决策者；而今，算法与人类共同扮演这一角色。在世界转型之际，智力超群的人才的价值也在与日俱增。通过计算机软件和代码，他们的影响力被扩大，其对社会的潜在价值也成倍增加。这些人正在改变人类的生活，把从开车到股票交易再到公司人员配置等一系列重大决策权从人手中移交到算法手中。

作者提到，来自硅谷的最佳产品总是能够使大多数人受益。例如，帮助你与亲朋好友建立更紧密的联系，追溯你的家族起源，教你的女儿学习微积分，或是缓和你与健康保险公司的关系。

医疗业务、检验结果和并发症比率正在被世界上最杰出的工程师构造的算法详细检查，一切的一切在算法坚定的目光下无所遁形。虽然医生这一职业不会消失，但如果想在未来保住饭碗，医生们需要想办法让自己与众不同。因为一般医生所能做的简单重复的工作，计算机程序都能取而代之。

作者又举了《千千万万没找到工作的律师》的案例。律师这一行业很容易受到算法的入侵。任何一宗大型诉讼案件最耗钱的环节就是卷宗分析——收集、研究所有与案件相关的文档，而公司法的法律条文和涉及合同纠纷的诉讼案件通常都有数百万页文档需专人研读分析。虽然阅读这些文档的通常都是律师助理或初级律师，但客户要为这项服务支付的费用仍然高达每人每小时 100~200 美元，而这项工作的本质就是用眼睛服务。眼睛可以用计算机硬件来代替，控制眼睛的大脑则可以用算法来代替。

帕洛阿尔托市的一家叫黑石探索的科技公司正在利用算法取代人类为客户提供法律分析服务。算法的工作效率不亚于人类律师，严密程度更是胜人一筹。而且，算法也更便宜。过去研读分析 120 万页法律文档至少要向客户收取 500 万美元，而今只需 10 万美元。这也许就是为什么 2012 年美国通过律师资格考试的人大概有 54000 人，而市场能提供给他们的工作机会却只有 26000 个。对于这个国家千千万万没找到工作的律师来说这绝非好消息。算法不需要健康保险，也不需要事务所承诺它某天可以成为合伙人。

当下这个经济时代，最不可缺少的就是能构建、维护、完善代码和算法的人才，未来亦是如此。

我们应该寻找那些能够在两个领域里来去自如的特殊人才——既能游刃有余地编写算法代码，又能很好地将高中生培养成下一代人才。虽然找到此类能让学生对复杂的新兴学科感兴趣的技术人才并非易事，但每每找到一个这样的人进入教育界，对社会和学生都会有巨大好处。

很多人有成为量化人才的潜力，只是没有机会尝试。我们不缺聪明人，缺的是有数量分析教育背景的聪明人。美国每所中学都应要求学生至少学习一门编程课。虽然大部分学生都会止步于此，但只要 5% 的学生从此迷上了构造自己的程序和算法，就会为我们的教育和经济提供新的发展动力。很多学生从来没想到过进入编程或数量分析领域。数学对他们来说只是死记硬背、生搬硬套从而应付考试的机械技能，他们从来没有看到数学的另外一面——改变世界。或者，当他们终于意识到这一点时，却已然处在前往另一领域的人生轨道上，而这很可能发生在大学期间。编程和计算机科学课程不应该专属于一小群特定的学生——这是 21 世纪最重要的一项技能，所有的学生都应该有机会学习这项技能。

对于少数敢冒风险、开拓进取的人，凭着编写创新算法的本领，更能打开创业与创新之门。这道门为任何有志之士敞开。

能在数学和科学类考试中取得高分的人，不一定是技术奇才。成为技术奇才在于实践，在于花时间研究流程。

未来，会编写代码的人有许多事情可以做。如果你还能构思并构造出复杂精妙的算法且没有机器人抢在你的前头，那你很有可能统治世界。

## 参考文献

[1] (美)克里斯托弗·斯坦纳 (Christopher Steiner), 著. 李筱莹, 译. 算法帝国[M]. 中国北京: 人民邮电出版社, 2014 年 06 月.

陈阳 杨熙坤